BANK OF ENGLAND

Staff Working Paper No. 869

Identification of structural vector autoregressions by stochastic volatility

Dominik Bertsche and Robin Braun

June 2020

BANK OF ENGLAND

Staff Working Paper No. 869
Identification of structural vector autoregressions by stochastic volatility

Dominik Bertsche[1] and Robin Braun[2]

Abstract

We propose to exploit stochastic volatility for statistical identification of structural vector autoregressive models (SV-SVAR). We discuss full and partial identification of the model and develop efficient Expectation Maximization algorithms for Maximum Likelihood inference. Simulation evidence suggests that, compared to alternative models, the SV-SVAR works well in identifying structural parameters also under misspecification of the variance process. We apply the model to study the importance of oil supply shocks for driving oil prices. Since shocks identified by heteroskedasticity may not be economically meaningful, we exploit the framework to test instrumental variable restrictions which are overidentifying in the heteroskedastic model. Our findings suggest that conventional supply shocks are negligible drivers of oil prices, while news shocks about future supply account for almost all the variation.

Key words: Structural vector autoregression (SVAR), identification via heteroskedasticity, stochastic volatility, external instruments.

JEL classification: C32, Q43.

(1) University of Konstanz. Email: dominik.bertsche@uni-konstanz.de
(2) Bank of England. Email: robin.braun@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. We are grateful for the excellent comments we received from Ralf Brüggemann, Helmut Lütkepohl, Daniel Lewis, Tomasz Wozniak, Thoore Schlaak, Christian Hansen, anonymous referees and participants in various seminars and conferences for their comments on different stages of this project. Financial support by the Graduate School of Decision Science (GSDS) and the German Science Foundation (grant number: BR 2941/3-1) is gratefully acknowledged.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Structural vector autoregressive (SVAR) models are extensively used in empirical macroeconomics. Various strategies have been proposed to identify the model, including the introduction of exclusion- and sign restrictions on the effects of structural shocks (Bernanke & Mihov 1998, Blanchard & Quah 1989, Uhlig 2005), identification via the use of external instruments (Stock & Watson 2012, Mertens & Ravn 2013) and exploiting statistical properties of structural shocks (Rigobon 2003, Lanne et al. 2017, Gourieroux et al. 2017).

In this paper, we discuss identification and estimation of SVARs by a stochastic volatility (SV) model. Specifically, we assume that the log-variances of structural shocks are latent, each following independent AR(1) processes, i.e. autoregressive processes of order one. Drawing on recent advances in Lewis (2019), we show that in conjunction with a fixed impact matrix, our model yields additional restrictions that allow to pin down a unique set of orthogonal shocks. Besides identification, we extensively discuss Maximum Likelihood inference and provide fast algorithms for this purpose.

A stochastic volatility model for the variance of structural shocks can be an attractive specification for many reasons. First, SV models are very popular to capture volatility in theoretical and empirical macroeconomics. For example, Justiniano & Primiceri (2008) and Fernández-Villaverde & Rubio-Ramírez (2007) allow for SV within fitted DSGE models, finding substantial time variation in the second moments of their structural shocks. Furthermore, SV models are often used to complement time-varying parameter VARs and have been found to provide a good description of volatility patterns in macroeconomic data (Primiceri 2005, Koop & Korobilis 2010). Finally, the literature has documented great forecasting performance of VAR models with stochastic volatility (see e.g. Clark (2011), D'Agostino et al. (2013) and Clark & Ravazzolo (2015)). Given this context, it seems natural to exploit the model also for identification purposes of SVARs. Second, a stochastic volatility specification is known to be more flexible than models with deterministic variance processes. This is because the SV model, in contrast to alternative specifications, includes shocks in the volatility equation that do not depend on the innovations in the VAR equation. As pointed out in Kim et al. (1998), this additional flexibility typically translates into superior fit in comparison to equally parameterized models from the GARCH family. This

is an important aspect, given that recent evidence of Lütkepohl & Schlaak (2018) suggests that choosing the heteroskedasticity model in SVARs by information criteria translates into more precise structural parameter estimates.

The additional flexibility of the stochastic volatility model comes at the cost of higher complexity for likelihood inference. The SV specification implies a nonlinear state space model, and therefore, standard filtering algorithms cannot be applied to evaluate the likelihood function. To cope with this issue, a large part of this paper focuses on the development of reliable Expectation Maximization (EM) algorithms. We propose two versions which trade off computational complexity and accuracy. The first is based on a second order Taylor approximation of the intractable smoothing distribution necessary to compute the E-step (Durbin & Koopman 1997), and exploits recent advances in sparse matrix algorithms developed for Gaussian Markov random fields (Rue et al. 2009, Chan 2017). Therefore, the algorithm is fast and converges within seconds. The second EM algorithm corrects for the approximation error in the smoothing distribution by Importance Sampling. Hence, at the cost of an additional computational burden, a greater level of accuracy may be achieved.

Our paper fits into the literature of identifying structural shocks in SVARs by heteroskedasticity. A variety of other models have been proposed in this literature, starting with a simple breakpoint model (Rigobon 2003), a Markov Switching model (Lanne et al. 2010), a GARCH model (Normandin & Phaneuf 2004) and a Smooth Transition model (Lütkepohl & Netšunajev 2017*b*). Furthermore, Lewis (2019) discusses identification and estimation of heteroskedastic SVARs in a general GMM framework. We complement this literature by discussing the SV specification in detail. Within a simulation exercise we provide evidence that, in comparison to the alternative models, the SV-SVAR works particularly well in estimating the structural parameters under misspecification of the variance process, proofing itself capable to capture volatility patterns generated by very different data generating processes. By simulating data from SVAR models subject to four distinct variance specifications, we find that the SV model compares favourably in terms of the mean squared error of estimated impulse response functions.

Also related to our paper is the work of Carriero et al. (2019) who exploit a SV model to

identify uncertainty shocks within a SVAR. However, while we focus on the identification of the traditional SVAR model and implement classical inference, the model of Carriero et al. (2019) is more structurally tailored to their application, and is based on Bayesian inference.

We illustrate the usefulness of our methodology contributing to a recent debate on the importance of oil supply shocks for driving oil prices (Kilian & Murphy 2014, Baumeister & Hamilton 2019, Herrera & Rangaraju 2019). We find that for this application, the SV model provides superior fit and is favoured by all conventional information criteria, if compared to other specifications for the variance. Since the structural shocks identified by heteroskedasticity are not guaranteed to be economically meaningful, we test instrumental variable (IV) restrictions used to identify oil supply shocks as overidentifying. We find no evidence against IV restrictions implied by the supply shock of Kilian (2009) as well as the news shock of Känzig (2019). Our main results suggest that conventional supply shocks are negligible drivers of oil prices, while the news components account for almost all the variation. A related analysis was conducted in Lütkepohl & Netšunajev (2014), who also exploit heteroskedasticity to disentangle supply from demand shocks in the crude oil market. However, in comparison to their study, the model we consider in our application is closer to state of the art specifications in terms of variables included, lag length of the VAR, and identifying constraints considered.

The paper is structured as follows. Section 2 introduces the SVAR model with stochastic volatility and discusses under which conditions the structural parameters are identified. Section 3 considers Maximum Likelihood estimation, reviews procedures to test for identification and discusses tests for overidentifying restrictions. In Section 4, we present simulation evidence while in Section 5 we apply the proposed model to study oil supply shocks. Section 6 concludes.

2 Identification of SVARs by stochastic volatility

Let y_t be a $K \times 1$ vector of endogenous variables. We consider the heteroskedastic SVAR(p) model reading:

$$y_t = \nu + \sum_{j=1}^{p} A_j y_{t-j} + u_t, \tag{2.1}$$

$$u_t = B V_t^{\frac{1}{2}} \eta_t, \tag{2.2}$$

where $\eta_t \sim (0, I_K)$ is assumed to be a white noise error term. Equation (2.1) corresponds to a standard reduced form VAR(p). Here, A_j's are $K \times K$ autoregressive coefficient matrices and ν is a $K \times 1$ vector of intercepts. Since we only consider stable time series throughout the paper, we assume:

$$\det A(z) = \det(I_K - A_1 z - \ldots - A_p z^p) \neq 0, \qquad \text{for } |z| \leq 1.$$

Equation (2.2) models the structural part and is set up as a B-model in the terminology of Lütkepohl (2005). The reduced form errors u_t are linked to the structural shocks $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$ through a $K \times K$ invertible contemporaneous impact matrix B. The structural shocks are potentially heteroskedastic and/or non-normal, captured by $V_t^{\frac{1}{2}}$, a stochastic diagonal matrix with strictly positive elements. To be more specific, we assume that there are $r \leq K$ heteroskedastic shocks which are ordered such that they appear first in vector ε_t. To model the time-varying second moments of these shocks, we specify independent Gaussian AR(1)'s for the first r components in ε_t:

$$V_t = \begin{pmatrix} \operatorname{diag}(\exp[h_{1t}, \ldots, h_{rt}]) & 0 \\ 0 & I_{K-r} \end{pmatrix}, \tag{2.3}$$

$$h_{it} = \phi_i h_{i,t-1} + \sqrt{s_i}\omega_{it}, \qquad \text{for } i = 1, \ldots, r, \tag{2.4}$$

where $\omega_{it} \sim \mathcal{N}(0,1)$, $\mathrm{E}(\varepsilon'_{1:r,t}\omega_t) = 0$ for $\omega_t = (\omega_{1t}, \ldots, \omega_{rt})'$ and $|\phi_i| < 1$. Note that omitting an intercept in equation (2.4) means to set $\mathrm{E}(h_{it}) = 0$ for $i = 1, \ldots, r$ and $t = 2, \ldots, T$ due to the stability condition. Furthermore, the initial states are assumed to be initialized

from the unconditional distribution $h_{i1} \sim \mathcal{N}(0, s_i/(1-\phi_i^2))$. This specification yields an unconditional covariance matrix of the reduced form errors u_t given as:

$$\Sigma = \mathrm{E}(u_t u_t') = BE(V_t)B' = BVB', \tag{2.5}$$

with $V = \mathrm{diag}\left[\exp(\sigma_{h_1}^2/2), \ldots, \exp(\sigma_{h_r}^2/2), 1_{K-r}\right]$ and $\sigma_{h_i}^2 = s_i/(1-\phi_i^2)$ being the unconditional variance of the underlying log-variance process ($i = 1, \ldots, r$).

The proposed model for equation (2.2) is similar to the Generalized Orthogonal GARCH (GO-GARCH) model of Van der Weide (2002) and Lanne & Saikkonen (2007), with the major difference in the specification (2.3)-(2.4) of V_t. While for the GO-GARCH the first r diagonal components are modeled by deterministic GARCH(1,1) processes, we specify AR(1)'s for their logarithms. In addition to their stability, we will also assume that their variances are finite, i.e. $0 < s_i < \infty$. This directly implies that ε_t is a strictly stationary stochastic process with finite second moment, which will aid in the identification analysis. In particular, the following basic properties of the model can be derived in a straightforward manner (see e.g. Jacquier et al. (1994)) for $i = 1, \ldots, r$ and $\tau > 0$:

$$\gamma_i(\tau) = \mathrm{Cov}(\varepsilon_{it}^2, \varepsilon_{i,t+\tau}^2) = \exp(\sigma_{h_i}^2)(\exp(\sigma_{h_i}^2\phi_i^\tau) - 1), \tag{2.6}$$

$$\kappa_i = \frac{\mathrm{E}(\varepsilon_{it}^4)}{\mathrm{E}(\varepsilon_{it}^2)^2} = \mathrm{E}(\eta_{it}^4)\exp\left(\sigma_{h_i}^2\right). \tag{2.7}$$

The model is able to capture two features that are often observed in structural shocks. First, heteroskedasticity can be captured when $\phi_i > 0$. The respective autocovariance function in the second moment of ε_{it} is given by equation (2.6), displaying an exponential decay in ϕ_i. This autocovariance function is found to be very flexible allowing to capture a large variety of heteroskedasticity patterns, an argument that we can confirm by simulation evidence. Second, the model can capture heavy tailed errors and the respective kurtosis function κ_i can be decomposed into a part that is due to the kurtosis of the standardized structural shocks η_{it} and a component which inflates the value depending on the underlying SV parameters. That is, given a conditional Gaussian error distribution in ε_{it}, excess kurtosis kicks in as soon as the SV process is nontrivial, that is $s_i > 0$. This means that even if a shock is homoskedastic ($\phi_i = 0$), the model is still able to capture heavy tails

under conditional Gaussianity. We argue that this is one key advantage with respect to a model from the GARCH family, which are generally unable to generate homoskedastic shocks featuring excess kurtosis given the assumption of conditional Gaussianity.

In the following, we will use equations (2.6) and (2.7) to discuss identification in detail. Due to the symmetry of the covariance matrix, identification in the SV-SVAR model cannot be discussed based on equation (2.5) solely. For that purpose, we rely on Lewis (2019) who treats identification by time-varying volatility in a more general context requiring no specific functional form. In particular, identification can be analyzed based on the lag $\tau > 0$ autocovariance in the squared reduced form residuals $\xi_t = \text{vech}(u_t u_t')$. This function takes the following form (Lewis 2019):

$$\text{Cov}(\xi_t, \xi_{t+\tau}) = L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)' L_K', \tag{2.8}$$

where L_K is an elimination matrix such that $\text{vech}(A) = L_K \,\text{vec}(A)$, G_K is a selection matrix such that $\text{vec}(D) = G_K d$ for $D = \text{diag}(d)$ and $M_\tau = \text{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$. In the Supplementary Appendix A.1, we derive this function for the SV-SVAR model. Note that one autocovariance has $\sum_{i=1}^{5} \binom{i+K-3}{K-2}$ unique elements ($K \geq 2$), while the structural model contains K^2 entries in B and r autocovariances in $\gamma_i(\tau)$'s, implicitly parameterized nonlinearly by the underlying SV processes. Lewis (2019) proves general identification of the elements in M_τ and B under the restriction that the diagonal of B is fixed at unity. In order to account for the alternative normalization implied by (2.3)-(2.5), Proposition 1 summarizes identification of B for any $r \leq K$.

Proposition 1. *Let V_t be modeled by equations* (2.3) *and* (2.4) *with $|\phi_i| < 1$, $\phi_i \neq 0$ and $0 < s_i < \infty$ for $i = 1, \ldots, r$. Let $B = (B_1, B_2)$ with $B_1 \in \mathbb{R}^{K \times r}$ and $B_2 \in \mathbb{R}^{K \times (K-r)}$. Since SV-SVAR model* (2.1)-(2.4) *implies uncorrelated structural shocks ε_{it} with independent variance processes for $i = 1, \ldots, r$, these conditions impose that matrix B_1 is identified up to permutation and sign switches.*

Proof. See Supplementary Appendix A.2.

Although $r = K$ ensures identification of the full B-matrix, this is not a necessary condition. The orthogonality constraints implied by equation (2.5) and the scaling of V_t

imposed by equation (2.4) yield enough structure to identify the model in case of $r = K-1$, which is summarized in Corollary 1.

Corollary 1. *Assume the setting from Proposition 1 for the special case $r = K-1$. Then, the entire matrix $B \in \mathbb{R}^{K\times K}$ is unique up to multiplication of its columns by -1 and permutation of its first $K-1$ columns.*

Proof. See Supplementary Appendix A.3.

The presented results are broadly in line with those provided by Lewis (2019). However, our results deviate in the sense that identification is given also under $r = K-1$ heteroskedastic shocks. Furthermore, the simple structure assumed for the SV-SVAR allows for a much simpler proof.

In case of a lower number of heteroskedastic shocks ($r < K-1$), without further exclusion restrictions the impact matrix is only partially identified. To disentangle the remaining structural shocks ε_{it} for $i = r+1, \ldots, K$ in that case, it is sufficient to impose a lower triangular structure on the lower right $(K-r)\times(K-r)$ block of B:

Corollary 2. *Assume the setting from Proposition 1 for $r \leq K-2$. Moreover, separate $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, $B_{11} \in \mathbb{R}^{r\times r}$, $B_{21} \in \mathbb{R}^{(K-r)\times r}$, $B_{12} \in \mathbb{R}^{r\times(K-r)}$ and $B_{22} \in \mathbb{R}^{(K-r)\times(K-r)}$. Let B_{22} be restricted to a lower triangular matrix. Then, the full matrix B is unique up to multiplication of its columns by -1 and permutation of its first r columns.*

Proof. See Supplementary Appendix A.4.

Some authors (see e.g. Arias et al. (2018, 2019) or Baumeister & Hamilton (2019)) prefer to work with what is known as A-model in the terminology of Lütkepohl (2005). An A-model is a re-parameterization of the SVAR in equation (2.1) and (2.2), obtained by premultiplying both sides with $\mathrm{A} = B^{-1}$. Then, the coefficients in A provide the marginal effects among the variables and accordingly, give the model implied causal structure. For $r \geq K-1$, full identification of B up to permutation and sign switches obviously implies a fully identified $\mathrm{A} = B^{-1}$. In case of partial identification ($r \leq K-2$), the identification of the first r columns in B is equivalent to the identification of the first r rows in A, and hence also the first r rows of $\mathrm{A}A_j$ for $j = 1, \ldots, p$ (see Supplementary Appendix A.5).

To obtain the identification results, we have only assumed uncorrelated shocks with independent variance processes. Note that if one is willing to assume mutual independence of structural errors ε_{it}'s, we do not need the assumption of heteroskedasticity ($\phi_i \neq 0$). In that case, the identification of the SV-SVAR model can be directly obtained from results in Lanne et al. (2017), requiring that (besides mutual independence) $(K-1)$ components in ε_t are non-Gaussian. Within the SV-SVAR model, this condition is met as soon as the necessary amount of SV processes are non-trivial ($s_i > 0$). Equivalent results for partial identification under mutual independence of ε_{it}'s are available in Maxand (2018). Note that, instead of independence, one could assume a set of co-kurtosis constraints and still obtain equivalent identification results (Lanne & Luoto 2019). In particular, it is easy to show that the proof of Proposition 1 in Supplementary Appendix A.2 goes through for $\phi_i = 0$, if one assumes $\mathrm{E}[\varepsilon_{it}^3 \varepsilon_{jt}] = 0$ and $\mathrm{Cov}[\varepsilon_{it}^2, \varepsilon_{jt}^2] = 0$ for $i \neq j$. In this case, the proof would rely on the additional information from the covariance matrix of ξ_t, instead of the autocovariance matrix (i.e. equation (2.8) for $\tau = 0$).

In order to draw on standard asymptotic results during estimation, it is necessary to restrict our analysis to a uniquely identified SV-SVAR model for any $r \in \{1, \ldots, K\}$. Hence, we need to fix a configuration for sign and permutation of the shocks, and impose the additional restrictions in case of $r \leq K-2$. We can obtain a globally identified SV-SVAR by considering only impact matrices B that are contained in the set:

$$\mathcal{B}_{r,K}^{(1)} := \{B \in \mathbb{R}^{K \times K} : \text{ lower right } (K-r) \times (K-r) \text{ block of } B \text{ is lower triangular}\}.$$

Note that the elements of this set are subject to $(K-r)(K-r-1)/2$ restrictions considered in Corollary 2, yielding the necessary restrictions to identify the model also if $r \leq K-2$. We note that an alternative approach, pursued in e.g. in Lanne & Saikkonen (2007) or Lütkepohl & Milunovich (2016), would be to re-parameterize B exploiting a polar decomposition of the unconditional variance. Implementing their approach is equivalent to ours in terms of free parameters ($K^2 - (K-r)(K-r-1)/2$), but would require an additional step of the Delta method to map standard errors back from the polar decomposition to B. Hence, for $r \leq K-2$ we prefer to work with a restricted B, and in the following we denote the free parameters of B by $\beta = S_B \operatorname{vec}(B)$ where S_B is the corresponding

$(K^2 - 0.5(K-r)(K-r-1)) \times K^2$ selection matrix.

To obtain a globally identified model, we further fix an unique permutation and sign configuration (Lanne & Saikkonen 2007):

Identification scheme. *Let $B \in \mathcal{B}_{r,K}^{(1)}$ and transform it to $\bar{B} = M(B) = BPD$ as follows:*

(i) $P = diag(P_1, I_{K-r})$ with P_1 a $r \times r$ permutation matrix such that $G = B_1 P_1$ satisfies $|g_{ii}| > |g_{ij}|$ for all $i = 1, \ldots, r$ and $j = i+1, \ldots, r$.

(ii) D is a diagonal matrix with ± 1 entries such that all diagonal elements of BPD are positive.

Define $\mathcal{B}_{r,K}^{(2)} = \{\bar{B} \in \mathcal{B}_{r,K}^{(1)} : \exists B \in \mathcal{B}_{r,K}^{(1)} : M(B) = \bar{B}\}$.

Then, Proposition 1 and Corollaries 1 and 2 ensure unique identification of model (2.1)-(2.4) within $\mathcal{B}_{r,K}^{(2)}$.

Before we continue with inference, we discuss an additional constraint that we impose on the log-variances. Note that we identify the scale of the structural shocks by setting $\mathrm{E}(h_{it}) = 0$. However, this constraint holds only in expectation, and for very persistent SV processes, the sample moment can be very uninformative about the scale. Therefore, throughout this paper we additionally fix the sample mean of the log-variances, i.e. we set $1/T \sum_{t=1}^{T} h_{it} = 0$ for $i = 1, \ldots, r$. Note that this constraint leads to a rank reduction of the covariance matrix implied by the Gaussian AR(1) model. This is similar in spirit to imposing the alternative normalizing constraint that $\mathrm{E}(h_{i1}) = \mathrm{Var}(h_{i1}) = 0$, implying that $\mathrm{E}(u_1 u_1') = BB'$ which is typically used to identify the scaling in Markov Switching SVAR models (Lanne et al. 2010, Herwartz & Lütkepohl 2014).

3 Maximum likelihood estimation

In order to estimate the model we propose a full Maximum Likelihood approach. In the following, denote by $\theta = [\mathrm{vec}(\nu, A_1, \ldots, A_p)', \beta', \phi', s']'$ the full vector of parameters in the SV-SVAR model where $\phi = [\phi_1, \ldots, \phi_r]'$ and $s = [s_1, \ldots, s_r]'$. Assuming normality of the standardized structural shocks η_t, the log-likelihood function based on the prediction error

decomposition is given as follows:

$$\begin{aligned}\mathcal{L}(\theta) =& -T\log|B| + \sum_{i=1}^{K}\log p(\varepsilon_i|\theta),\\ =& -T\log|B| + \sum_{i=1}^{K}\sum_{t=1}^{T}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\mathrm{E}[h_{it}|\mathcal{F}_{t-1}] - \frac{1}{2}\varepsilon_{it}^2\mathrm{E}[\exp(-h_{it})|\mathcal{F}_{t-1}]\right],\end{aligned}$$

where $\varepsilon_t = B^{-1}u_t$ with $u_t = y_t - \nu - \sum_{j=1}^{p} A_j y_{t-j}$, and $\varepsilon_i = [\varepsilon_{i1},\ldots,\varepsilon_{iT}]'$. Since the SV model implies a nonlinear state space model, the predictive distributions $p(h_t|\theta,\mathcal{F}_{t-1})$ that are necessary to compute the expected values in the log-likelihood are not available in closed form. To overcome this difficulty, we follow Durbin & Koopman (1997) in evaluating the likelihood function by Importance Sampling. Furthermore, for maximization we provide two versions of an Expectation Maximization algorithm that approximate the E-steps with different levels of accuracy. Both algorithms rely on fast sparse matrix implementations put forward by Chan & Jeliazkov (2009) and applied to SV models in Chan & Grant (2016).

3.1 Evaluation of the likelihood

To evaluate the likelihood function, it is convenient to split the log-likelihood into the following components:

$$\mathcal{L}(\theta) = -T\log|B| + \log p(\varepsilon_{1:r}|\theta) + \log p(\varepsilon_{r+1:K}|\theta),$$

where $\varepsilon_{i:j} = [\varepsilon'_{i:j,1},\ldots,\varepsilon'_{i:j,T}]'$ with $\varepsilon_{i:j,t} = [\varepsilon_{it},\ldots,\varepsilon_{jt}]'$. Note that for $i = r+1,\ldots,K$, $\mathrm{E}[h_{it}|\mathcal{F}_{t-1}] = 0$ and $\mathrm{E}[\exp(-h_{it})|\mathcal{F}_{t-1}] = 1$. Hence, the second term is simply given by $\log p(\varepsilon_{r+1:K}|\theta) = -\frac{T(K-r)}{2}\log(2\pi) - \frac{1}{2}\sum_{i=r+1}^{K}\sum_{t=1}^{T}\varepsilon_{it}^2$. To estimate the intractable part $\log p(\varepsilon_{1:r}|\theta)$, we use an Importance Sampling (IS) approach. This involves integrating out the latent log-variances:

$$p(\varepsilon_{1:r}|\theta) = \int p(\varepsilon_{1:r}|\theta,h_{1:r})p^c(h_{1:r}|\theta)dh_{1:r}, \tag{3.1}$$

where $p(\varepsilon_{1:r}|\theta,h_{1:r}) = \prod_{i=1}^{r}\prod_{t=1}^{T}(2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}h_{it})\exp(-\frac{1}{2}\varepsilon_{it}^2\exp(-h_{it}))$ and $p^c(h_{1:r}|\theta)$ is a Gaussian distribution of dimension Tr governed by the underlying AR(1) structure. Note

that the corresponding covariance matrix is of reduced rank $(T-1)r$ as a direct consequence of the r zero mean constraints that we impose, denoted in the following by $A_h h_{1:r} = 0_{r\times 1}$ with $A_h = (1_{1\times T} \otimes I_r)/T$. Exploiting Bayes theorem (Rue 2001), the constrained density can be conveniently written in terms of the unconstrained density $p(h_{1:r}|\theta)$ and correction terms:

$$p^c(h_{1:r}|\theta) = \frac{p(h_{1:r}|\theta)\pi_1(A_h h_{1:r}|h_{1:r})}{\pi_2(A_h h_{1:r}|\theta)}, \tag{3.2}$$

where $\pi_1(A_h h_{1:r}|h_{1:r}) = |A_h A_h'|^{-\frac{1}{2}} = T^{r/2}$ if $A_h h_{1:r} = 0_{r\times 1}$ holds, and zero else. Furthermore, $p(h_{1:r}|\theta) \sim \mathcal{N}(\delta, Q^{-1})$ and $\pi_2(A_h h_{1:r}|\theta) \sim \mathcal{N}(0, A_h Q^{-1} A_h')$, where $\delta = 0_{Tr\times 1}$ and Q is a (sparse) precision matrix implied by the Gaussian AR(1) model (2.4) that we assume for the log-variances. In particular, it is given by $Q = H'\Sigma_h^{-1}H$ where H a $(Tr \times Tr)$ (sparse) matrix with unit diagonal, and elements on the r-th diagonal below the main diagonal given by $-\left(1_{(T-1)\times 1} \otimes \phi\right)$. Likewise, $\Sigma_h = \text{diag}\left(\sigma_{h_1}^2, \ldots, \sigma_{h_r}^2, \left[1_{(T-1)\times 1} \otimes s\right]'\right)$ is a $(Tr \times Tr)$ matrix.

A simulation consistent importance density estimator for the intractable integral in equation (3.1) is then given by:

$$\widehat{p(\varepsilon_{1:r}|\theta)} = \frac{1}{R}\sum_{j=1}^{R} \frac{p(\varepsilon_{1:r}|\theta, h_{1:r}^{(j)})p^c(h_{1:r}^{(j)}|\theta)}{q(h_{1:r}^{(j)})}, \tag{3.3}$$

where $h_{1:r}^{(j)}$ for $j = 1, \ldots, R$ are independent draws from an importance density $q(h_{1:r})$.

The accuracy of the IS estimator crucially depends on the choice for the importance density $q(h_{1:r})$ which we discuss in the following. First, note that the zero variance importance density is given by the intractable smoothing distribution $p^c(h_{1:r}|\theta, \varepsilon_{1:r}) \propto p(\varepsilon_{1:r}|\theta, h_{1:r})p^c(h_{1:r}|\theta)$. We follow Durbin & Koopman (1997) and use a Gaussian importance density denoted by $\pi_G^c(h_{1:r}|\theta, \varepsilon_{1:r})$, which is centered at the mode of $p^c(h_{1:r}|\theta, \varepsilon_{1:r})$ with precision equal to the curvature at this point. Here, we rely on the fast algorithms that exploit the sparsity of the precision matrix used e.g. in Rue (2001) for general Gaussian Markov random fields and Chan & Grant (2016) for stochastic volatility models in particular.

To derive $\pi_G^c(h_{1:r}|\theta, \varepsilon_{1:r})$, we follow the exposition of Chan & Grant (2016). For ease of exposition, assume there is no linear constraint on the log-variances. Then, the zero

variance importance density is proportional to:

$$p(h_{1:r}|\theta,\varepsilon_{1:r}) \propto \exp\left(-\frac{1}{2}h'_{1:r}Qh_{1:r} + \log p(\varepsilon_{1:r}|\theta,h_{1:r})\right), \tag{3.4}$$

A Gaussian importance density can be obtained exploiting a second order Taylor expansion of $\log p(\varepsilon_{1:r}|\theta,h_{1:r})$ around some properly chosen $\tilde{h}^{(0)}_{1:r}$:

$$\log p(\varepsilon_{1:r}|\theta,h_{1:r}) \approx \log p(\varepsilon_{1:r}|\theta,\tilde{h}^{(0)}_{1:r}) + [h_{1:r}-\tilde{h}^{(0)}_{1:r}]'f - \frac{1}{2}[h_{1:r}-\tilde{h}^{(0)}_{1:r}]'C[h_{1:r}-\tilde{h}^{(0)}_{1:r}], \tag{3.5}$$

where $f = \frac{\partial \log p(\varepsilon_{1:r}|\theta,h_{1:r})}{\partial h_{1:r}}\Big|_{h_{1:r}=\tilde{h}^{(0)}_{1:r}}$ and $C = -\frac{\partial^2 \log p(\varepsilon_{1:r}|\theta,h_{1:r})}{\partial h_{1:r}\partial h'_{1:r}}\Big|_{h_{1:r}=\tilde{h}^{(0)}_{1:r}}$. Plugging the linearized kernel into equation (3.4), an approximate smoothing distribution $\pi_G(h_{1:r}|\theta,\varepsilon_{1:r})$ takes the form of a normal distribution for $h_{1:r}$ with precision matrix $\bar{Q} = Q + C$ and mean $\bar{\delta} = \bar{Q}^{-1}b$, where $b = f + C\tilde{h}^{(0)}_{1:r}$. The Tr-dimensional density has a tridiagonal precision matrix which allows for fast generation of random samples and likelihood evaluation. The approximation is evaluated at the mode of the smoothing distribution obtained by a Newton-Raphson method that typically converges in few iterations. Details on the Newton-Raphson method and on explicit expressions for f and C are given in Supplementary Appendix B.

In order to take into account the linear constraint on the average log-variances, the importance density requires a slight modification. Applying Bayes' theorem yields a constraint density $\pi^c_G(h_{1:r}|\theta,\varepsilon_{1:r})$ which is also Gaussian but has mean and covariance:

$$\mathrm{E}(h_{1:r}|\theta,\varepsilon_{1:r},A_hh_{1:r}{=}0) = \bar{\delta} - \bar{Q}^{-1}A'_h(A_h\bar{Q}^{-1}A'_h)^{-1}A_h\bar{\delta}, \tag{3.6}$$

$$\mathrm{Cov}(h_{1:r}|\theta,\varepsilon_{1:r},A_hh_{1:r}{=}0) = \bar{Q}^{-1} - \bar{Q}^{-1}A'_h(A_h\bar{Q}^{-1}A'_h)^{-1}A_h\bar{Q}^{-1}. \tag{3.7}$$

This allows for an easy adjustment of the Newton-Raphson by including the update given in equation (3.6) at each iteration. However, imposing the linear restriction yields a non-sparse precision and a reduced rank covariance which impedes efficient sampling and density evaluation. Following Rue (2001), sampling and evaluation of $\pi^c_G(h_{1:r}|\theta,\varepsilon_{1:r})$ can still be implemented at trivial extra costs by what is known as 'conditioning by kriging'. Specifically, a random sample $\tilde{h}^{(j)}_{1:r}$ is first generated from $\pi_G(h_{1:r}|\theta,\varepsilon_{1:r})$, exploiting the spar-

sity in $\bar{Q}^{-1}$. In a second step, the draw is corrected for the linear constraint by setting $h_{1:r}^{(j)} = \tilde{h}_{1:r}^{(j)} - \bar{Q}^{-1}A_h'(A_h\bar{Q}^{-1}A_h')^{-1}A_h\tilde{h}_{1:r}^{(j)}$. As with the prior distribution, evaluation of the adjusted IS density can be achieved efficiently evaluating Bayes' theorem:

$$\pi_G^c(h_{1:r}|\theta, \varepsilon_{1:r}) = \frac{\pi_G(h_{1:r}|\theta, \varepsilon_i)\pi_1(A_h h_{1:r}|h_{1:r})}{\pi_2(A_h h_{1:r})}, \tag{3.8}$$

where $\pi_1(A_h h_{1:r}|h_{1:r}) = T^{\frac{r}{2}}$ and $\pi_2(A_h h_{1:r}) \sim \mathcal{N}(0, A_h\bar{Q}^{-1}A_h')$.

The importance estimator is found to yield a reliable performance at very little computational expense. However, we still recommend to assess the quality of estimator (3.3) by reporting its standard error which can be computed e.g. by the batch means method. Furthermore, for the validity of the standard error and $\sqrt{R}$-convergence of the IS estimator, the variance of the importance weights has to exist. Since for the high-dimensional integral (3.1) this is not clear a priori, we advise to test for the existence of the variance using the test of Koopman et al. (2009). For sample sizes typically used in macroeconomics we do not expect this to be a serious issue.

3.2 EM algorithm

In order to optimize the likelihood function, we exploit the Expectation Maximization algorithm introduced by Dempster et al. (1977). The EM procedure is particularly suitable for maximization problems under the presence of hidden variables. In our setting, these hidden variables are the set of r log-variances which for simplicity, are denoted by $h = h_{1:r}$ in the following. Our goal is to maximize:

$$\mathcal{L}(\theta) = \log p(y|\theta) = \log \int p(y|\theta, h)p^c(h|\theta)dh.$$

Following Neal & Hinton (1998) and Roweis & Ghahramani (2001), let $\tilde{p}(h)$ be any distribution of the hidden variables, possibly depending on θ and y. Then, a lower bound

on $\mathcal{L}(\theta) = \log \int p(y|\theta, h)p(h|\theta)dh$ can be obtained by an application of Jensen's inequality:

$$\mathcal{L}(\theta) = \log \int \frac{p(y|\theta, h)p^c(h|\theta)}{\tilde{p}(h)} \tilde{p}(h)dh \tag{3.9}$$

$$\geq \int \log \left(\frac{p(y|\theta, h)p^c(h|\theta)}{\tilde{p}(h)} \right) \tilde{p}(h)dh \tag{3.10}$$

$$= \int \log \left(p(y|\theta, h)p^c(h|\theta) \right) \tilde{p}(h)dh - \int \log \left(\tilde{p}(h) \right) \tilde{p}(h)dh \tag{3.11}$$

$$=: F(\tilde{p}, \theta). \tag{3.12}$$

The EM algorithm starts with some initial parameter vector $\theta^{(0)}$ and proceeds by iteratively maximizing:

$$\text{E-step:} \qquad \tilde{p}^{(l)} = \arg\max_{\tilde{p}} \; F(\tilde{p}, \theta^{(l-1)}), \tag{3.13}$$

$$\text{M-step:} \qquad \theta^{(l)} = \arg\max_{\theta} \; F(\tilde{p}^{(l)}, \theta). \tag{3.14}$$

Under mild regularity conditions the EM algorithm converges towards a local optimum (McLachlan & Krishnan 2007). It is easy to show that the E-step in (3.13) is given by setting $\tilde{p}^{(l)}$ equal to the smoothing distribution $p(h|\theta^{(l-1)}, y)$. This can be seen by noting that for this choice, equation (3.10) holds with equality which means that the lower bound $F(\tilde{p}, \theta)$ exactly equals the log-likelihood $\mathcal{L}(\theta)$. Furthermore, the M-step in equation (3.14) is given by maximizing the criterion function:

$$Q(\theta; \theta^{(l-1)}) = \int \log \left(p(y|\theta, h)p^c(h|\theta) \right) \tilde{p}^{(l)}(h)dh = \mathrm{E}_{\theta^{(l-1)}} \left(\mathcal{L}_c(\theta) \right), \tag{3.15}$$

where the expectation is taken with respect to $\tilde{p}^{(l)}(h)$ and $\mathcal{L}_c(\theta) = \log \left(p(y|\theta, h)p^c(h|\theta) \right)$ is the complete data log-likelihood.

For the SV-SVAR model, the complete data log-likelihood is rather simple, implying that for some tractable $\tilde{p}^{(l)}$, computing the expectations and maximizing it with respect to θ is straightforward. However, since the smoothing distribution is not tractable, we cannot simply set $\tilde{p}^{(l)} = p(h|\theta^{(l-1)}, y)$. Instead, we develop two algorithms which approximate this density to a different extent. The first algorithm relies on an E-step which is based on the same analytical approximation of the smoothing distribution used as importance density.

The second algorithm corrects for the approximation error by Importance Sampling.

3.2.1 Analytical approximation

Our analytical approximation is based on the following E-step:

$$\tilde{p}^{(l)}(h) = \pi_G^c\left(h|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right), \tag{3.16}$$

which is the Gaussian approximation of the smoothing distribution that we already introduced as importance density. This E-step corresponds to maximizing $F(\tilde{p}, \theta^{(l-1)})$ with respect to $\tilde{p}$ considering only the family of Gaussian distributions. To motivate this approach, we follow the arguments of Neal & Hinton (1998) who argue that it is not necessary to work with the exact smoothing distribution in the EM algorithm to get monotonic increases in the log-likelihood function $\mathcal{L}(\theta)$. In fact, it can be shown that $F(\tilde{p}, \theta) = \mathcal{L}(\theta) - D_{KL}\left(\tilde{p}(h)||p^c(h|y, \theta)\right)$ where $D_{KL}(\cdot||\cdot)$ is the Kullback - Leibler (KL) divergence measure. Therefore, if the Gaussian approximation is close to the smoothing density in a KL sense, iteratively optimizing $F(\tilde{p}, \theta)$ yields convergence to a point very close to the corresponding local maximum of $\mathcal{L}(\theta)$. We found the resulting algorithm to perform very well both in our simulation studies and applications, and in the following we refer to it as EM-1. Details on the corresponding M-steps are given in Supplementary Appendix B.3.

3.2.2 Monte Carlo approximation

Recall that the optimal E-step is given by setting $\tilde{p}^{(l)} = p^c(h|\theta^{(l-1)}, y)$, which we know only up to normalizing constant. Hence, it is possible to correct for the error of EM-1 by replacing the integral in the E-step of equation (3.15) by a simulation consistent Importance Sampling analogue:

$$\hat{Q}(\theta; \theta^{(l-1)}) = \frac{1}{R}\sum_{j=1}^{R} w^{(j)}(\theta^{(l-1)}) \log\left(p(y|\theta, h^{(j)})p^c(h^{(j)}|\theta)\right), \tag{3.17}$$

where the importance weights are given by:

$$w^{(j)}(\theta^{(l-1)}) \propto \frac{p^c(h^{(j)}|\theta^{(l-1)}, y)}{\pi_G^c(h^{(j)}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)})},$$

and the draws $h^{(j)}, j = 1, \ldots, R$ are generated from the IS density $\pi_G^c(h^{(j)}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)})$. In the remainder, we call the Monte Carlo based algorithm EM-2 and for details on the M-steps, we refer to Supplementary Appendix B.3.

Note that compared to EM-1, this algorithm can correct for the approximation error for $R \to \infty$, however, at the cost of a higher computational burden. Therefore, we recommend to first run EM-1, before starting to iterate based on EM-2. In our experience, both algorithms EM-1 and EM-2 result in estimates of the SVAR parameter that are almost indistinguishable. However, the parameter estimates of the SV processes (ϕ_i, s_i) could slightly differ sometimes. Therefore, while EM-1 is likely to suffice for fast exploratory structural analysis, we recommend to run the more accurate EM-2 to present any final results.

3.3 Properties of the estimator

Because the SV-SVAR model is a special case of a Hidden Markov Model, the asymptotic properties of the Maximum Likelihood estimator can be inferred from Cappé et al. (2005). Let $\hat{\theta}$ denote the ML estimator, under appropriate regularity conditions and the global identifying constraints considered in Section 2, $\hat{\theta}$ is consistent and asymptotically normally distributed:

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1}), \tag{3.18}$$

where $\mathcal{I}(\theta) = -\mathrm{E}\left(\frac{\partial^2 \log p(y|\theta)}{\partial\theta\partial\theta'}\right)$ is the information matrix. Furthermore, a strongly consistent estimator for the asymptotic variance is given by $\widehat{\mathcal{I}(\theta)} = T^{-1}\mathcal{J}(\hat{\theta})$, where $\mathcal{J}(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta'}\big|_{\theta=\hat{\theta}}$ is the observed information matrix evaluated at the ML estimator. We rely on Oakes identity (Oakes 1999) to evaluate the Hessian, given by:

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \left[\frac{\partial^2 Q(\theta;\theta^{(l)})}{\partial\theta\partial\theta'} + \frac{\partial^2 Q(\theta;\theta^{(l)})}{\partial\theta\partial\theta^{(l)'}}\right]_{\theta^{(l)}=\theta}. \tag{3.19}$$

In Supplementary Appendix B.4, we provide closed form expressions for the expected gradient $\partial Q(\theta;\theta^{(l)})/\partial\theta$ and Hessian $\partial^2 Q(\theta;\theta^{(l)})/(\partial\theta\partial\theta')$ of the complete data log-likelihood. Finally, the Jacobian of the expected gradient with respect to $\theta^{(l)}$ is computed using finite differences. We use the same underlying uniform random variables to re-compute the expectations when we compute the Jacobian for estimates of EM-2.

Identification of the SVAR model is ultimately useful to conduct structural analysis. Since Impulse Response Functions (IRFs) and Forecast Error Variance Decompositions (FEVDs) are likely to be the most widely used tool for that purpose, we describe in Supplementary Appendix B.5 how to conduct inference on these quantities within our model. In particular, we describe a Delta Method approach to quantify uncertainty of the identified IRFs and FEVDs, following work of Lütkepohl (1990) and Brüggemann et al. (2016).

3.4 Testing for identification

To test how many columns of B can be identified by heteroskedasticity, we recommend to follow well established procedures proposed by Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016) to test identification in SVAR-GARCH models. The idea is to conduct the following sequence of tests:

$$H_0 : r = r_0 \qquad \text{vs} \qquad H_1 : r > r_0, \tag{3.20}$$

for $r_0 = 0, \ldots, K-1$. If all null hypotheses up to $r_0 = K-2$ can be rejected, there is evidence that all parameters in B can be identified by heteroskedasticity. The testing problem given in (3.20) is nonstandard since parts of the parameter space differ between null and alternative hypothesis. Therefore, Lanne & Saikkonen (2007) suggest test statistics which require estimation under H_0 only. If r_0 is the true number of heteroskedastic errors, one can separate the structural shocks $\varepsilon_t = B^{-1}u_t = (\varepsilon_{1t}', \varepsilon_{2t}')'$ into a heteroskedastic part $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and homoskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Under the null ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is homoskedastic white noise, which can be tested for remaining heteroskedasticity. A detailed description of these test are given in Supplementary Appendix C. We highlight that according to simulation evidence by Lütkepohl & Milunovich (2016), these tests display

a substantial lack of power in sample sizes typically available in macroeconomics. Hence, if the null hypothesis can be rejected for all r_0's up to $K-2$, this can be interpreted as strong evidence in favor of model identification.

Alternatively, a new testing procedure is proposed in Lewis (2019). In particular, he first establishes a minimum rank of the covariance matrix $\text{Cov}(\xi_t, \xi_{t+\tau})$, such that a SVAR can be identified by heteroskedasticity. In a second step, he proposes to use the test statistic of Cragg & Donald (1996) to determine the rank of the sample counterpart. While this testing strategy seems promising, the small sample properties are yet to be explored.

3.5 Testing overidentifying restrictions

Identification by heteroskedasticity is a statistical procedure and therefore, needs to be combined with economic theory before the results can be interpreted in a meaningful way. One way to introduce economic theory into the SV-SVAR is to statistically test economically motivated restrictions, as they become overidentifying in the heteroskedastic model. If there is no evidence against a certain set of (economic) restrictions, the corresponding shock(s) can be interpreted in the usual way. See for example, Normandin & Phaneuf (2004) and Lütkepohl & Netšunajev (2017*a*) who test conventional short- and long run restrictions imposed to identify monetary policy shocks. Within the Likelihood framework considered in this paper, these restrictions can be easily tested e.g. via Likelihood ratio- or Wald tests.

Avoiding exclusion restrictions, an alternative identification strategy involves the use of instrumental variables (IV) (Mertens & Ravn 2013, Stock & Watson 2012). The identifying assumptions are that an external instrument z_t is correlated with the structural shock it is designed for (relevance) and uncorrelated with all remaining shocks (exogeneity). Without loss of generality, assume that the first shock is identified by the instrument. Then, Mertens & Ravn (2013) show that these assumptions can be translated into the following set of linear restrictions on $b_{\bullet 1}$, denoting the first column of B:

$$b_{2:K,1} = (\Sigma_{zu_1'}^{-1} \Sigma_{zu_2'})' b_{1,1}. \tag{3.21}$$

where $b_{\bullet 1} = (b_{1,1}, b'_{2:K,1})'$ with $b_{1,1}$ scalar and $b_{2:K,1} \in \mathbb{R}^{K-1}$. Furthermore, $\text{Cov}(z_t, u'_t) = \Sigma_{zu'} = [\Sigma_{zu'_1}, \Sigma_{zu'_2}]$ with $\Sigma_{zu'_1}$ scalar and $\Sigma'_{zu'_2} \in \mathbb{R}^{K-1}$. In a previous version of this paper, we proposed to replace theoretical moments with sample moments and simply use Likelihood ratio tests to test the IV restrictions given in equation (3.21). However, given the stochastic nature of the constraint, the test is likely to reject too often using critical values from the corresponding $\chi^2(K-1)$ distribution. This problem was noted in Podstawski et al. (2018) who instead, propose to augment a heteroskedastic SVAR model with an equation that relates the instrument to the structural shock. Within the augmented model, Likelihood ratio tests are valid as they take into account the uncertainty of the moment constraint.

In the following, we describe a simpler two-step procedure to test the identifying constraints from an IV approach. Given a set of estimated structural shocks from the SV-SVAR model $\hat{\varepsilon}_t$, we conduct an auxiliary regression as a simple device for testing IV conditions:

$$z_t = \psi \hat{\varepsilon}_t + \sigma_z u_t^\star, \tag{3.22}$$

where $u_t^\star \sim (0, 1)$ is a white noise error term. First, consider the relevance constraint which implies $\text{E}[\varepsilon_{1t} z_t] \neq 0$. This can be tested through the null hypothesis $H_0 : \psi_1 = 0$ vs. $H_1 : \psi_1 \neq 0$. If the null is rejected, this can be interpreted as evidence in favor of instrument relevance. In turn, exogeneity requires that $\text{E}[\varepsilon_{jt} z_t] = 0, j > 1$. Likewise, one can set up a null hypothesis $H_0 : (\psi_2, \dots, \psi_K) = 0$ vs. $H_1 : \exists j \in \{2, \dots, K\} : \psi_j \neq 0$, and rejecting the null constitutes statistical evidence against instrument exogeneity. Given asymptotic normality of the auxiliary regression parameters, $T^{1/2}(\hat{\psi} - \psi) \xrightarrow{d} \mathcal{N}(0, V^\star)$, simple Student t- and Wald tests can be conducted.

Note that $V^\star$ needs to account for the fact that $\hat{\varepsilon}_t$'s are constructed based on estimates for θ, the parameters of the SV-SVAR. Let $\mathcal{L}_2(\psi, \sigma_z | \theta)$ be the log-likelihood function of auxiliary regression (3.22) under normality of $u_t^\star$. Then, following Greene (2000) and Murphy & Topel (2002), a valid asymptotic covariance matrix for the second step is given by:

$$V^\star = V_2 + V_2[FV_1F' - RV_1F' - FV_1R']V_2,$$

where V_1 is the asymptotic covariance matrix of the SV-SVAR parameters θ, V_2 is the

asympotic covariance matrix of $\hat{\psi}$ ignoring the fact that $\hat{\varepsilon}_t$ is estimated, $F = \mathrm{E}\left[\frac{\partial \mathcal{L}_2}{\partial \psi} \cdot \frac{\partial \mathcal{L}_2}{\partial \theta'}\right]$ and $R = \mathrm{E}\left[\frac{\partial \mathcal{L}_2}{\partial \psi} \cdot \frac{\partial \mathcal{L}}{\partial \theta'}\right]$. In practice, we replace F and R with sample averages, while for V_2 we use the standard asymptotic covariance matrix $V_2 = \hat{\sigma}_z^2(\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'/T)^{-1}$.

4 Monte Carlo study

An important question for practitioners is how a heteroskedastic SVAR model performs in estimating structural parameters under misspecification of the variance process. To shed some light on this question, we conduct a comparative Monte Carlo (MC) study. Specifically, we compare the estimation performance of the SV-SVAR model under misspecification to that of alternative heteroskedastic SVARs, namely a simple Breakpoint model (BP-SVAR), Markov Switching models (MS-SVAR) and a GARCH model (GARCH-SVAR). Furthermore, we also compare the performance to a GMM estimator which does not make any assumptions about the volatility process.

Our analysis involves generating a large number of data sets from the four stated heteroskedastic SVARs. Then, we estimate each model and compare the relative estimation performance of the misspecified to the correctly specified model. We focus on estimation of structural IRFs which are probably the most widely used tool in SVAR analysis. Furthermore, they are nonlinear functions of both, the structural impact matrix and reduced form autoregressive parameters. Thus, they are particularly suited to summarize the overall estimation performance of a SVAR model. As a metric of comparison, we use cumulated Mean Squared Errors (MSEs) of the IRF estimates.

The following data generating processes (DGPs) are specified to simulate the data sets, closely resembling the MC design of Lütkepohl & Schlaak (2018). Time series of lengths $T \in \{200, 500\}$ are generated by the following bivariate VAR(1) process:

$$y_t = A_1 y_{t-1} + u_t,$$

with $u_t \sim \mathcal{N}(0, B\Lambda_t B')$ for $t = 1, \ldots, T$ and:

$$A_1 = \begin{pmatrix} 0.6 & 0.35 \\ -0.1 & 0.7 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0.5 & 2 \end{pmatrix}.$$

For the diagonal matrix Λ_t, the following DGPs are specified:

1. **BP-SVAR**: The BP-SVAR is subject to a one time change in the variance. We set $\Lambda_t = I_2$ for $t = 1, \ldots T/2$ and $\Lambda_t = \text{diag}(2, 7)$ for $t = T/2 + 1, \ldots T$.

2. **MS(2)-SVAR**: The specified MS-SVAR involves a switching variance with the same regimes than the BP-SVAR. We specify the transition probability matrix:

$$P = \begin{pmatrix} .95 & .05 \\ .1 & .9 \end{pmatrix}.$$

   Based on simulated states $s_1, \ldots, s_T \in \{1, 2\}$, $\Lambda_{s_t=1} = I_2$ and $\Lambda_{s_t=2} = \text{diag}(2, 7)$.

3. **GARCH-SVAR**: For this specification, the diagonal elements of $\Lambda_t = \text{diag}(\lambda_{1t}, \lambda_{2t})$ follow univariate GARCH(1,1) processes with unit unconditional variance:

$$\lambda_{it} = (1 - \alpha_i - \beta_i) + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \lambda_{i,t-1}, \qquad i \in \{1, 2\},$$

   where $\varepsilon_t = B^{-1} u_t$ is the vector of structural shocks at time t. We set $\alpha_i = 0.15$ and $\beta_i = 0.8$ $(i = 1, 2)$.

4. **SV-SVAR**: For this DGP, $\Lambda_t = \text{diag}(\exp(h_{1t}), \exp(h_{2t}))$ with:

$$h_{it} = \phi_i h_{i,t-1} + \sqrt{s_i}\omega_{it},$$

   where $\omega_{it} \sim \mathcal{N}(0, 1)$. We set $\phi_i = 0.95$ and $s_i = 0.04$ $(i = 1, 2)$ what corresponds to fairly persistent processes in the variance often observed in macroeconomic and financial data.

To avoid that our results are influenced by weak identification, we only accept data sets in the MS(2)-SVAR DGP if at least 25% of the observations are associated with either of

the regimes. Likewise, for the GARCH and SV DGPs, only data sets with an empirical kurtosis of the simulated structural shocks of at least 3.6 are accepted.

A total of M=1000 data sets are simulated for each variance specification. In the following, let $\hat{\Theta}_{jk,i}(m)$ for $j,k \in \{1,2\}$ denote the estimated impulse response function in variable j caused by structural shock k after i periods based on estimates for the m-th data set. Our metric of comparison is then given as:

$$\text{MSE}\left(\Theta_{jk}\right)_h = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{i=0}^{h}\left(\hat{\Theta}_{jk,i}(m) - \Theta_{jk,i}\right)^2\right). \tag{4.1}$$

We choose horizon h=5 as in Lütkepohl & Schlaak (2018). To compute parameter estimates, we use algorithm EM-1 for the SV-SVAR model. For the BP-SVAR we maximize a Gaussian likelihood over a grid of possible break-dates, while for the MS-SVARs we use the EM algorithm outlined in Herwartz & Lütkepohl (2014). Finally, for the GARCH-SVAR we compute ML estimates based on the procedure of Lanne & Saikkonen (2007). The GMM estimator is implemented with a simple two-step procedure, where the first step is based on the unconditional variance $\text{E}[u_t u_t']$ and the second step based on the first order autocovariance of $\xi_t = \text{vech}(u_t u_t')$, hereby following Lewis (2019). Note that the estimated models rely on different normalizing constraints for the structural shocks which is why we rescale all impulse response functions to unit shock size.

The results of the simulation study are provided in Table 1. For improved readability, we report relative MSEs in comparison to the correctly specified model. Overall, we find that the SV-SVAR model performs very well regardless of the true DGP or the sample size for each of the impulse responses Θ_{jk}. In fact, the largest deterioration that we document in terms of MSE is 75% in Θ_{21} of the Markov Switching DGP. This contrasts all other models included into the Monte Carlo study which are subject to a very heterogeneous performance. Whenever they are inherently misspecified, we find relative MSEs of much higher orders of magnitude. For example, with detoriations of up to 24 times, estimates based on a MS(2)-SVAR seem completely unreliable for data generated by the SV and GARCH DGPs. Admittedly, the complexity of a MS model can be increased by adding additional states. Therefore, we also report estimates based on a MS(3) for the SV and GARCH DGPs.

Table 1: Cumulated MSEs at horizon $h = 5$

		T=200				T=500			
		Θ_{11}	Θ_{12}	Θ_{21}	Θ_{22}	Θ_{11}	Θ_{12}	Θ_{21}	Θ_{22}
BP-DGP	BP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MS(2)	1.00	1.01	1.01	1.00	1.00	1.00	1.00	1.00
	GARCH	1.60	1.79	1.58	1.14	1.20	1.24	1.19	1.04
	GMM	6.74	8.37	7.23	1.57	9.20	11.77	10.40	1.65
	SV	1.27	1.39	1.31	1.07	1.10	1.12	1.10	1.03
MS-DGP	BP	3.23	3.72	4.71	1.37	7.98	9.75	12.01	1.79
	MS(2)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	GARCH	3.89	4.43	3.45	1.26	3.52	4.14	3.90	1.28
	GMM	4.95	5.61	5.05	1.40	8.48	10.30	9.30	1.58
	SV	1.57	1.72	1.75	1.09	1.29	1.38	1.33	1.08
GARCH-DGP	BP	3.88	4.23	2.56	1.26	11.58	12.67	4.99	1.47
	MS(2)	8.37	9.16	3.62	1.27	21.40	24.15	7.24	1.38
	MS(3)	4.02	4.30	2.00	1.13	5.58	6.01	2.33	1.20
	GARCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	GMM	5.94	6.45	3.04	1.52	7.41	8.24	3.37	1.65
	SV	1.14	1.14	1.04	1.01	1.09	1.09	1.06	1.04
SV-DGP	BP	3.59	3.86	2.36	1.19	8.60	9.64	4.37	1.36
	MS(2)	6.15	6.81	3.50	1.19	14.34	16.03	5.64	1.31
	MS(3)	4.90	5.39	2.11	1.14	3.15	3.38	1.72	1.15
	GARCH	2.60	2.87	1.85	1.16	1.51	1.56	1.23	1.08
	GMM	7.01	7.88	3.95	1.36	12.21	13.81	5.05	1.45
	SV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: MSEs of impulse response functions calculated as in (4.1) and displayed relative to true model MSEs.

While indeed this yields substantial improvements, we still register deteriorations in MSE up to 500%. The GMM implementation also struggles with precise estimation and yields relatively poor performance in comparison to likelihood methods. Our findings are in line with the simulation conducted in Lewis (2019), who explains this by the inherent difficulty to estimate higher order moments precisely.

If we compare the IRF estimates of the SV-SVAR to all other misspecified models in a certain DGP, we find it to perform strictly better in two out of three DGPs. Specifically, for residuals generated by a MS(2) and a GARCH model, all impulse responses estimated by the SV-SVAR have lower cumulative MSEs than the other misspecified models. Only

if the structural errors are simulated with a one time shift in the variance there is no clear advantage of the SV model over the MS model. However, this is not surprising given that the latter is perfectly able to capture such sudden shifts in the variance.

Finally, we find that the SV-SVAR model also compares favorable if its performance is directly matched to the most related model, the GARCH-SVAR. In particular, the SV-SVAR model always performs better when both models are misspecified. Furthermore, while there is almost no deterioration in the MSE of the SV-SVAR estimates in a GARCH-DGP, the other way around we record substantially higher relative MSEs.

Summing up, our small simulation study yields promising results indicating that the SV-SVAR may be a good choice to identify structural shocks for different types of heteroskedasticity patterns and to estimate the corresponding impulse response functions.

5 The importance of oil supply shocks for oil prices

Since Kilian (2009), a large amount of research has been devoted to understand the drivers of oil prices in terms of supply and demand shocks. Using variance decompositions of structural VAR models, a wide range of estimates has been found for the relative importance of oil supply shocks, ranging between close to 0% to more than 40% depending on the underlying identification strategy. In particular, papers that have imposed very small short term supply elasticities for identification, arrive at estimates close to the lower bound (Kilian & Murphy 2012, 2014, Herrera & Rangaraju 2019). On the other hand, when larger supply elasticities are imposed for identification, one may find supply shocks to be equally important than demand shocks (Baumeister & Hamilton 2019, Caldara et al. 2019). Finally, recent evidence by Känzig (2019) suggests that rather than traditional supply shocks, news about future oil supply are the main drivers of oil prices. In the remainder of this section, we use the methodology developed in this paper to reassess the importance of oil supply shocks for driving oil prices.

5.1 Model and identifying constraints

We follow the convention by Kilian & Murphy (2014) and Baumeister & Hamilton (2019) (BH19 henceforth) and study structural disturbances within a four-dimensional SVAR model based on:

$$y_t = [100\Delta q_t, 100\Delta wip_t, 100p_t, 100\Delta inv_t],$$

where q_t is the log of global crude oil production (in million barrels per day), wip_t is the log of an index for world industrial production (see BH19), p_t is the log of real WTI (West Texas Intermediate) spot price, and Δinv_t is the change in global inventories expressed as fraction of last period's oil production. As in BH19, we set $p = 12$ and use monthly data covering January 1974 until December 2016, thereby excluding earlier data when oil prices were regulated.

To identify different oil supply shocks, we primarily rely on identification by heteroskedasticity. In a second step, we use the two-step procedure outlined in Section 3.5 to test IV restrictions implied by using a set of instruments for oil supply shocks. While this ensures that the statistical shocks can be interpreted economically, it also allows to discriminate between a competing set of instruments. For conventional oil supply shocks, we will consider the shocks series of Kilian (2009) (K09), BH19, Caldara et al. (2019) (CCI19) and Kilian (2008) (K08). For a supply news shock, we take the instrument constructed by Känzig (2019) (DK19). K09 and BH19 are structural shocks from identified SVAR models, while CCI19 and K08 are constructed with a narrative approach. Finally, the instrument of DK19 exploits high frequency variation in oil price futures around quota announcements from the Organization of the Petroleum Exporting Countries (OPEC).

5.2 Statistical and economic analysis

We start our analysis with formal model selection for the variance specification following the suggestion of Lütkepohl & Schlaak (2018). By means of information criteria, we compare the SV model to a set of alternatives available in the literature: a GARCH(1,1), a Smooth Transition (ST) and different specifications of a Markov Switching model. Table 2 reports corresponding log-likelihood values, Akaike information criteria (AIC) and Bayesian

Table 2: Model selection

	linear	MS(2)	MS(3)	STVAR	GARCH(1, 1)	SV (EM-2)
$\ln L$	-3649.96	-3515.60	-3424.76	-3514.01	-3447.38	-3417.20
AIC	7735.92	7467.21	7301.51	7464.03	7334.76	7274.41
BIC	8661.57	8387.73	8255.81	8384.55	8263.73	8203.38

Note: $\ln L$ - log-likelihood, AIC=$-2\ln L + 2 \times n_p$ and BIC=$-2\ln L + \ln(T) \times n_p$ with n_p the number of free parameters. For the likelihood of the SV model, application of the batch means method yields a standard error of 0.04. The test of Koopman et al. (2009) indicates that the variance of the importance weights is finite.

Table 3: Tests of identification

	$Q_1(1)$	dof	p-value	$Q_2(1)$	dof	p-value	$LM(1)$	df	p-val
$r_0 = 0$	25.36	1.00	0.00	232.74	100.00	0.00	252.31	100.00	0.00
$r_0 = 1$	31.66	1.00	0.00	122.14	36.00	0.00	121.93	36.00	0.00
$r_0 = 2$	26.83	1.00	0.00	68.71	9.00	0.00	67.32	9.00	0.00
$r_0 = 3$	4.40	1.00	0.04	4.40	1.00	0.04	3.78	1.00	0.05

Note: Sequence of tests for the number of heteroskedastic shocks in the system. See C for more details about the test statistics.

information criteria (BIC) for a linear VAR and the heteroskedastic models. Our results suggest that including time-variation in the second moment is strongly supported by both information criteria. Furthermore, we find that the SV model performs best in terms of information criteria. Therefore, from a statistical point of view, the SV model is the most suitable to continue our analysis.

Before we can test the IV restrictions as overidentifying, we make sure that there is enough heteroskedasticity in the VAR residuals to identify the structural shocks. As described in Section 3.4, we apply a sequence of tests with $H_0 : r = r_0$ against $H_1 : r > r_0$ for $r_0 = 0, 1, \ldots K - 1$. The results are reported in Table 3. We find strong evidence that $r \geq 3$ in our model, implying that the model can be fully identified by heteroskedasticity. For the remaining analysis, we continue with $r = 4$, given that all $p-$values are smaller or equal to 5% when testing $r_0 = 3$ against $r_0 > 3$.

Given strong evidence in favor of identification by heteroskedasticity, we can proceed with testing the IV restrictions implied by the external instruments included in our analysis. As outlined in Section 3.5, we do so regressing each external instrument on the estimated structural shocks. In Table 4, we report simple pairwise correlations as well as

Table 4: Correlations, regression coefficients and tests for instrument relevance/exogeneity

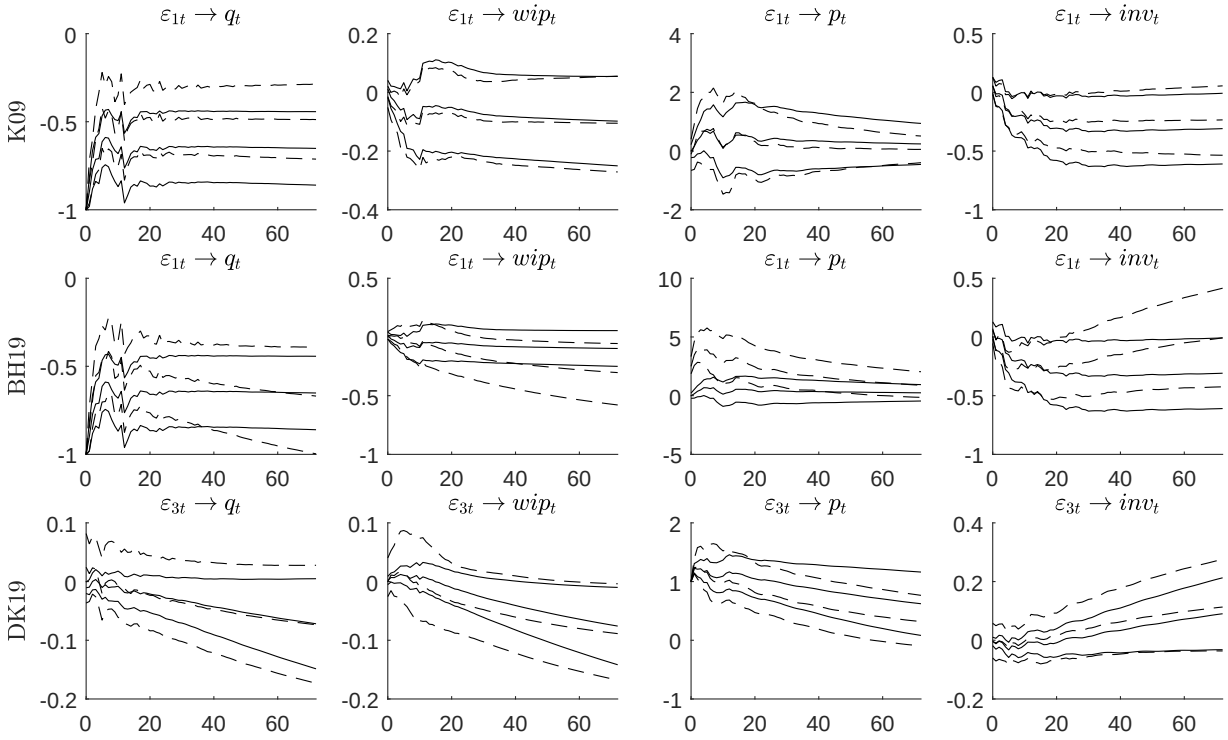
	K09	BH19	K08	CCI19	DK19
$\text{Corr}(\hat{\varepsilon}_{1t}, z_t)$	**0.89**	**0.85**	**0.24**	**0.81**	0.09
$\text{Corr}(\hat{\varepsilon}_{2t}, z_t)$	0.07	0.00	0.02	0.10	0.02
$\text{Corr}(\hat{\varepsilon}_{3t}, z_t)$	0.01	−0.42	−0.07	−0.58	**0.38**
$\text{Corr}(\hat{\varepsilon}_{4t}, z_t)$	0.00	0.05	0.12	0.25	0.01
$\hat{\psi}_1$ (s.e.)	0.55 (0.06)	0.97 (0.06)	0.16 (0.10)	0.90 (0.12)	0.20 (0.14)
$\hat{\psi}_2$ (s.e.)	0.07 (0.05)	0.06 (0.08)	0.02 (0.05)	0.29 (0.09)	−0.16 (0.13)
$\hat{\psi}_3$ (s.e.)	−0.02 (0.03)	−0.55 (0.07)	−0.04 (0.05)	−0.04 (0.09)	0.70 (0.15)
$\hat{\psi}_4$ (s.e.)	−0.01 (0.04)	0.06 (0.07)	0.08 (0.04)	0.42 (0.11)	0.06 (0.12)
$\text{E}[z_t\varepsilon_{jt}] \neq 0$	j=1	j=1	j=1	j=1	j=3
relevance (p-val)	0	0	0.11	0	0
exogeneity (p-val)	0.5	0	0.23	0	0.40

Note: Sample correlations, regression coefficients $\hat{\psi}_i, i = 1, \ldots, K$ and standard errors (s.e.) from the instrumental variables tests discussed in 3.5. Instruments included: Kilian (2009) (K09), Baumeister & Hamilton (2019) (BH19), Kilian (2008) (K08), Caldara et al. (2019) (CCI19) and Känzig (2019) (DK19).

the estimated coefficients with their standard errors. We find that the first structural shock $\hat{\varepsilon}_{1t}$ is strongly correlated with the instruments for conventional supply shocks (K09, BH19, K08 and CCI19), displaying correlations from 25% up to 89%. In turn, the instrument for the news shock (DK19) is highly correlated (38%) with the third shock of the model. Therefore, we consider $\hat{\varepsilon}_{1t}$ as our candidate for the conventional supply shock, while $\hat{\varepsilon}_{3t}$ will be our candidate for the news shock.

Moreover, Table 4 also provides p-values that we obtain testing both instruments relevance and exogeneity separately (see Section 3.5). For these tests, we assume that K09, BH19, K08 and CCI19 are targeting the conventional supply shock ($\hat{\varepsilon}_{1t}$), while DK19 aims to identify the news shock ($\hat{\varepsilon}_{3t}$). First, we find evidence that with the exception of the K08 series, all the external instruments are relevant for the shock they are targeting. Only for K08, a p-value of 11% means that one cannot reject the null hypothesis of non-relevance at any conventional significance level. These findings confirm evidence in Montiel Olea et al. (2018) that the K08 might be a weak instrument. With respect to exogeneity, we find strong evidence against using BH19 and CCI19 as an instrument for a supply shock. On the contrary, we cannot reject the null hypothesis of exogeneity for the remaining instruments (K09, K08 and DK19). Based on our test results, we conclude that there is no

Figure 1: Impulse responses for oil supply shocks



IRFs up to a horizon of 72 months with 90% confidence intervals (CIs). Solid lines give IRFs obtained by identification through heteroskedasticity, while dashed lines give IRFs based on identification by SVAR-IV. CIs for the latter are computed using the methodology of Montiel Olea et al. (2018).

evidence against identifying $\hat{\varepsilon}_{1t}$ by the shock series of K09. Moreover, there is no evidence against imposing the IV restrictions implied by the DK19 instrument. While these results are valuable on their own, they also allow us to continue with structural analysis labeling $\hat{\varepsilon}_{1t}$ as a conventional supply shock and $\hat{\varepsilon}_{3t}$ as a supply news shock.

In Figure 1, we plot IRFs with 90% confidence intervals for the effects of conventional supply (ε_{1t}) and supply news shocks (ε_{3t}). For comparison, we also draw IRFs obtained from (homoskedastic) models identified by K09, BH19 and DK19 as respective external instruments. First, note that the IRFs from IV-SVARs identified by K09 and DK19 are very similar to those obtained by the SV-SVAR. This is in line with our testing results, suggesting that K09 and DK19 are valid instruments in this setting. In turn, IRFs based on an IV-SVAR identified by BH19 differ substantially, particularly in the response of oil prices. Ultimately, this explains why the corresponding IV restrictions are rejected in the heteroskedastic model.

Table 5: Forecast error variance decomposition of real oil prices

	h=12	h=24	h=48	h=72
ε_{1t}	0.49	0.50	0.49	0.49
(s.e.)	(0.82)	(1.13)	(1.30)	(1.35)
ε_{3t}	95.75	95.97	96.17	96.22
(s.e.)	(11.44)	(15.15)	(20.86)	(27.18)

Note: Expressed as share (%) of total variation. Values in brackets indicate estimated standard errors.

Our findings suggest that a conventional supply shock, normalized to decrease global oil production by 1%, is associated with a slight decrease in world industrial production, an increase in price of about 0.5%, and a drop in inventories. However, the 90% confidence intervals suggest that there is substantial uncertainty associated with these estimates. A supply news shock, standardized to increase real oil prices by 1%, is associated with a slowly decreasing oil production and a drop in world economic activity. On the contrary, inventories seem to slowly build up over time. The full set of IRFs including a comparison with results from other heteroskedasticity models is provided in Supplementary Appendix D.

Finally, to shed light on the drivers of oil prices, Table 5 provides estimates for the contributions of supply shocks to the corresponding forecast error variance. With values around 0.5%, we find that for all horizons between 1-6 years, supply shocks account for very little variation in oil prices. On the contrary, for the same horizons, oil supply shocks are estimated to account for as much as 95% of the variation. These results support the findings in Kilian (2009), Kilian & Murphy (2014) as well as Känzig (2019).

6 Conclusion

In this paper, we have considered stochastic volatility to identify structural parameters of SVAR models. The resulting model (SV-SVAR) can generate patterns of heteroskedasticity which are very typical in VAR analysis and therefore, we expect it to be useful in a wide range of applications.

We discuss conditions for full and partial identification and propose to estimate the model by full Maximum Likelihood. For this purpose, we develop two efficient EM algorithms which approximate the intractable E-step to a different extent. One algorithm is based on a Laplace approximation while the other corrects for the approximation error using Monte Carlo integration. Besides discussing estimation algorithms, we state the main properties of the estimator and present tools to approximate the asymptotic covariance matrix. Tests considered by Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016) can be used to determine the number of heteroskedastic shocks and to test for identification. To label the shocks, one possibility is to test economic restrictions as overidentifying. Here, we provide a novel two-step approach to test identification by instrumental variables.

To demonstrate the flexibility of the SV-SVAR model, we conduct a Monte Carlo study investigating how precise Impulse Response Functions are estimated under misspecification of the variance process. In contrast to alternative heteroskedastic SVARs, we find that the proposed model performs very well regardless of the DGP specified for the variance.

In an empirical application, we revisit a current debate on the importance of supply shocks for oil prices. Formal model selection strongly supports a SV specification in the variance if compared to other heteroskedastic SVARs. We use the SV-SVAR to test IV restrictions as overidentifying. We find no evidence against identification of supply shocks using the shock series of Kilian (2009) and Känzig (2019). We find that supply news shocks explain for almost all the variation in oil prices, while conventional supply shocks are negligible in that context.

References

Arias, J. E., Rubio-Ramírez, J. F. & Waggoner, D. F. (2018), 'Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications', *Econometrica* **86**(2), 685–720.

Arias, J. E., Rubio-Ramírez, J. F. & Waggoner, D. F. (2019), 'Inference in Bayesian Proxy-SVARs', *FRB Atlanta Working Paper* .

Baumeister, C. & Hamilton, J. D. (2019), 'Structural interpretation of vector autoregres-

sions with incomplete identification: Revisiting the role of oil supply and demand shocks', *American Economic Review* **109**(5), 1873–1910.

Bernanke, B. S. & Mihov, I. (1998), 'Measuring monetary policy', *Quarterly Journal of Economics* **113**(3), 869–902.

Blanchard, O. J. & Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances', *The American Economic Review* **79**(4), 655–673.

Brüggemann, R., Jentsch, C. & Trenkler, C. (2016), 'Inference in VARs with conditional heteroskedasticity of unknown form', *Journal of Econometrics* **191**(1), 69–85.

Caldara, D., Cavallo, M. & Iacoviello, M. (2019), 'Oil price elasticities and oil price fluctuations', *Journal of Monetary Economics* **103**, 1–20.

Cappé, O., Moulines, E. & Ryden, T. (2005), *Inference in Hidden Markov Models*, Springer.

Carriero, A., Clark, T. E. & Marcellino, M. (2019), 'The identifying information in vector autoregressions with time-varying volatilities: An application to endogenous uncertainty', *Working Paper* .

Chan, J. C. C. (2017), 'The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling', *Journal of Business & Economic Statistics* **35**(1), 17–28.

Chan, J. C. C. & Grant, A. L. (2016), 'On the observed-data deviance information criterion for volatility modeling', *Journal of Financial Econometrics* **14**(4), 772–802.

Chan, J. C. C. & Jeliazkov, I. (2009), 'Efficient simulation and integrated likelihood estimation in state space models', *International Journal of Mathematical Modelling and Numerical Optimisation* **1**(1-2), 101–120.

Clark, T. E. (2011), 'Real-time density forecasts from bayesian vector autoregressions with stochastic volatility', *Journal of Business & Economic Statistics* **29**(3), 327–341.

Clark, T. E. & Ravazzolo, F. (2015), 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics* **30**(4), 551–575.

Cragg, J. G. & Donald, S. G. (1996), 'On the asymptotic properties of ldu-based tests of the rank of a matrix', *Journal of the American Statistical Association* **91**(435), 1301–1309.

D'Agostino, A., Gambetti, L. & Giannone, D. (2013), 'Macroeconomic forecasting and structural change', *Journal of applied econometrics* **28**(1), 82–101.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **39**, 1–38.

Durbin, J. & Koopman, S. J. (1997), 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika* **84**(3), 669–684.

Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2007), 'Estimating macroeconomic models: A likelihood approach', *Review of Economic Studies* **74**(4), 1059–1087.

Gourieroux, C., Monfort, A. & Renne, J. P. (2017), 'Statistical inference for independent component analysis: Application to structural VAR models', *Journal of Econometrics* **196**(1), 111–126.

Greene, W. H. (2000), 'Econometric analysis 4th edition', *International edition, New Jersey: Prentice Hall* pp. 201–215.

Herrera, A. M. & Rangaraju, S. K. (2019), 'The effect of oil supply shocks on us economic activity: What have we learned?', *Journal of Applied Econometrics* .

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

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business & Economic Statistics* **12**(4), 371–389.

Justiniano, A. & Primiceri, G. E. (2008), 'The time-varying volatility of macroeconomic fluctuations', *American Economic Review* **98**(3), 604–641.

Känzig, D. R. (2019), 'The macroeconomic effects of oil supply shocks: New evidence from OPEC announcements', *Working Paper* .

Kastner, G., Frühwirth-Schnatter, S. & Lopes, H. F. (2017), 'Efficient bayesian inference for multivariate factor stochastic volatility models', *Journal of Computational and Graphical Statistics* **26**(4), 905–917.

Kilian, L. (2008), 'Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?', *The Review of Economics and Statistics* **90**(2), 216–240.

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

Kilian, L. & Murphy, D. P. (2012), 'Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models', *Journal of the European Economic Association* **10**(5), 1166–1188.

Kilian, L. & Murphy, D. P. (2014), 'The role of inventories and speculative trading in the global market for crude oil', *Journal of Applied Econometrics* **29**(3), 454–478.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**(3), 361–393.

Koop, G. & Korobilis, D. (2010), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**(4), 267–358.

Koopman, S. J., Shephard, N. & Creal, D. (2009), 'Testing the assumptions behind importance sampling', *Journal of Econometrics* **149**(1), 2–11.

Lanne, M. & Luoto, J. (2019), 'Gmm estimation of non-gaussian structural vector autoregression', *Journal of Business & Economic Statistics* **0**(0), 1–13.

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

Lanne, M., Meitz, M. & Saikkonen, P. (2017), 'Identification and estimation of non-Gaussian structural vector autoregressions', *Journal of Econometrics* **196**(2), 288–304.

Lanne, M. & Saikkonen, P. (2007), 'A multivariate generalized orthogonal factor GARCH model', *Journal of Business & Economic Statistics* **25**(1), 61–75.

Lewis, D. J. (2019), 'Identifying shocks via time-varying volatility', *working paper* .

Lütkepohl, H. (1990), 'Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models', *The Review of Economics and Statistics* **72**(1), 116–125.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

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

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

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

Lütkepohl, H. & Schlaak, T. (2018), 'Choosing between different time-varying volatility models for structural vector autoregressive analysis', *Oxford Bulletin of Economics and Statistics* (forthcoming).

Magnus, J. R. & Neudecker, H. (2019), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Maxand, S. (2018), 'Identification of independent structural shocks in the presence of multiple gaussian components', *Econometrics and Statistics* .

McLachlan, G. & Krishnan, T. (2007), *The EM Algorithm and Extensions*, Vol. 382, John Wiley & Sons.

Mertens, K. & Ravn, M. O. (2013), 'The dynamic effects of personal and corporate income tax changes in the United States', *American Economic Review* **103**(4), 1212–1247.

Montiel Olea, J. L., Stock, J. H. & Watson, M. W. (2018), Inference in SVARs identified with external instruments, Technical report, Working Paper, version dated November.

Murphy, K. M. & Topel, R. H. (2002), 'Estimation and inference in two-step econometric models', *Journal of Business & Economic Statistics* **20**(1), 88–97.

Neal, R. M. & Hinton, G. E. (1998), *A view of the EM algorithm that justifies incremental, sparse, and other variants*, Springer, pp. 355–368.

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

Oakes, D. (1999), 'Direct calculation of the information matrix via the EM', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(2), 479–482.

Podstawski, M., Schlaak, T. & Rieth, M. (2018), 'Monetary policy, external instruments and heteroskedasticity'.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

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

Roweis, S. & Ghahramani, Z. (2001), 'Learning nonlinear dynamical systems using the Expectation–Maximization algorithm', *Kalman filtering and neural networks* **6**, 175–220.

Rue, H. (2001), 'Fast sampling of gaussian markov random fields', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2), 325–338.

Rue, H., Martino, S. & Chopin, N. (2009), 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations', *Journal of the Royal Statistical Society Series B-Statistical Methodology* **71**(2), 319–392.

Stock, J. H. & Watson, M. W. (2012), 'Disentangling the channels of the 2007-09 recession', *Brookings Papers on Economic Activity* **43**(1), 81–156.

Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**, 381–419.

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

Appendix A Derivations and proofs

To ensure identification of impact matrix B in model (2.1)-(2.4) we show that under sufficient heterogeneity in the second moments of the structural shocks, i.e. $r \geq K-1$, there is no $\tilde{B}$ different from B except for column permutations and sign changes which yields an observationally equivalent model with the same time-varying second moment properties in reduced form errors u_t for all $t = 1, \ldots, T$. Furthermore, for $r < K-1$, we show which parameters in impact matrix B are identified and which are not. This also yields an identification scheme for this scenario. We start with the derivation of the autocovariance function of the second moments of reduced form residuals u_t.

A.1 Autocovariance function of the second moments

The autocovariance function of the second moments of the structural shocks for $\tau > 0$ is:

$$\mathrm{Cov}\left(\mathrm{vec}\left(\varepsilon_t\varepsilon_t'\right), \mathrm{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) = \left[E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right) - E\left(\varepsilon_{it}\varepsilon_{jt}\right)E\left(\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right)\right]_{ijkl}.$$

Since the structural shocks are uncorrelated and have independent variance processes, the law of iterated expectations yields that the entries of this expression are only non-zero if both $i = j = k = l$ and $i \leq r$ hold for $i, j, k, l \in \{1, \ldots, K\}$. Thus, it is:

$$\operatorname{Cov}\left(\operatorname{vec}\left(\varepsilon_t \varepsilon_t'\right), \operatorname{vec}\left(\varepsilon_{t+\tau} \varepsilon_{t+\tau}'\right)\right) = G_K M_\tau G_K',$$

with G_K being a selection matrix such that $\operatorname{vec}(D) = G_K d$ for a diagonal matrix $D = \operatorname{diag}(d)$ and $M_\tau = \operatorname{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$ with $\gamma_i(\tau) = \exp(\sigma_{h_i}^2)(\exp(\sigma_{h_i}^2 \phi_i^\tau) - 1)$ and $\sigma_{h_i}^2 = s_i/(1 - \phi_i^2)$. Briefly recall that we define $\xi_t = \operatorname{vech}\left(u_t u_t'\right) = L_K \operatorname{vec}\left(u_t u_t'\right)$ (Lewis 2019). Consequently, the autocovariance function in ξ_t reads:

$$\begin{aligned}
\operatorname{Cov}\left(\xi_t, \xi_{t+\tau}\right) &= L_K \operatorname{Cov}\left(\operatorname{vec}\left(u_t u_t'\right), \operatorname{vec}\left(u_{t+\tau} u_{t+\tau}'\right)\right) L_K' \\
&= L_K \left(B \otimes B\right) \operatorname{Cov}\left(\operatorname{vec}\left(\varepsilon_t \varepsilon_t'\right), \operatorname{vec}\left(\varepsilon_{t+\tau} \varepsilon_{t+\tau}'\right)\right) \left(B \otimes B\right)' L_K' \\
&= L_K \left(B \otimes B\right) G_K M_\tau G_K' \left(B \otimes B\right)' L_K'.
\end{aligned}$$

A.2 Proof of Proposition 1

Proof. Suppose $\tilde{B} = BQ$ and $\tilde{\varepsilon}_t = Q^{-1} \varepsilon_t$ with $Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{pmatrix}$, where $Q_1 \in \mathbb{R}^{r \times r}$, $Q_2, Q_3' \in \mathbb{R}^{(K-r) \times r}$ and $Q_4 \in \mathbb{R}^{(K-r) \times (K-r)}$ define an observationally equivalent model. Hence, the log-variances $\tilde{h}_i$ of $\tilde{\varepsilon}_i$ for $i = 1, \ldots, r$ are modeled by AR(1) processes (2.4) with parameters $|\tilde{\phi}_i| < 1$, $\tilde{\phi}_i \neq 0$ and $0 < \tilde{s}_i < \infty$. Consequently, restriction (2.5) implies:

$$\operatorname{E}(u_t u_t') = BQ\tilde{V}Q'B' = BVB', \tag{A.1}$$

where $V = \operatorname{E}(V_t) = \operatorname{diag}(V_1, I_{K-r})$, $V_1 = \operatorname{diag}(\exp(\sigma_{h_1}^2/2), \ldots, \exp(\sigma_{h_r}^2/2))$, $\sigma_{h_i}^2 = s_i/(1-\phi_i^2)$ and $\tilde{V}$ analogue. Since $s_i, \tilde{s}_i > 0$ for $i = 1, \ldots, r$, the diagonal elements of $V_1, \tilde{V}_1$ are nonzero why they are of full rank. The diagonality of $Q\tilde{V}Q' = V$ due to (A.1) implies:

$$Q_2 \tilde{V}_1 Q_1' + Q_4 Q_3' = 0, \tag{A.2}$$

$$Q_2 \tilde{V}_1 Q_2' + Q_4 Q_4' = I_{K-r}. \tag{A.3}$$

Furthermore, the autocovariance function with lag $\tau > 0$ in the second moment of the reduced form errors $\xi_t = \text{vech}(u_t u_t')$ defined in (2.8) imposes:

$$\begin{aligned}\text{Cov}(\xi_t, \xi_{t+\tau}) &= L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)' L_K' \\ &= L_K(\tilde{B} \otimes \tilde{B})G_K \tilde{M}_\tau G_K'(\tilde{B} \otimes \tilde{B})' L_K' \\ &= L_K(B \otimes B)(Q \otimes Q)G_K \tilde{M}_\tau G_K'(Q \otimes Q)'(B \otimes B)' L_K',\end{aligned} \tag{A.4}$$

where $M_\tau = \text{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$ with elements $\gamma_i(\tau) = \exp(\sigma_{h_i}^2)(\exp(\sigma_{h_i}^2 \phi_i^\tau) - 1)$ and $\tilde{M}_\tau$ analogue for the autocovariance in $\text{vec}(\tilde{\varepsilon}_t \tilde{\varepsilon}_t')$. As $s_i, \tilde{s}_i > 0$ and $\phi_i, \tilde{\phi}_i \neq 0$, it is $\gamma_i(\tau), \tilde{\gamma}_i(\tau) \neq 0$ for $i = 1, \ldots, r$. Furthermore, (A.4) implies $(Q \otimes Q)G_K \tilde{M}_\tau G_K'(Q \otimes Q)' = G_K M_\tau G_K'$ what yields the following conditions:

$$\forall i = 1, \ldots, r : \sum_{l=1}^{r} q_{il}^4 \tilde{\gamma}_l(\tau) = \gamma_i(\tau) \neq 0, \tag{A.5}$$

$$\forall a_j \in \{0, 1, 2, 3\} : \sum_{j=1}^{K} a_j = 4 : \sum_{l=1}^{r} \left(\prod_{j=1}^{K} q_{jl}^{a_j} \right) \tilde{\gamma}_l(\tau) = 0. \tag{A.6}$$

Because of (A.6), it is $\sum_{l=1}^{r} q_{\bullet l} \underbrace{q_{il}^2 q_{jl} \tilde{\gamma}_l(\tau)}_{=:\lambda_{ijl}} = 0$ for all $i, j \in \{1, \ldots, K\}$ with $i \neq j$. As Q is a full rank matrix, its column vectors $q_{\bullet l}$ are linearly independent such that $\lambda_{ijl} = 0$ for all $l \in \{1, \ldots, r\}, i, j \in \{1, \ldots, K\} : i \neq j$. As in addition $\tilde{\gamma}_l(\tau) \neq 0$, considering the first r columns of Q, i.e. matrix $(Q_1', Q_2')'$, only one element per column can be different from zero.

Because of (A.5), in each row of $r \times r$ matrix Q_1 at least one element has to be non-zero. Following the previous argument, these r non-zero entries correspond to the r non-zero entries in $(Q_1', Q_2')'$. This directly implies that Q_2 is a zero matrix and Q_1 has exactly one element different from zero per row and column.

The fact that $Q_2 = 0$ and (A.3) directly imply $Q_4 Q_4' = I_{K-r}$, i.e. Q_4 is an orthogonal matrix. Then, (A.2) yields that $Q_3 = 0$.

Since Q_1 has exactly one non-zero entry per row and column, it can be decomposed into $Q_1 = D_1 P_1 S_1$ where D_1 is diagonal with ± 1 entries, P_1 is a permutation matrix and S_1 is any diagonal matrix that rescales the columns of B. Regardless of sign switches and

permutations, think of rescaled structural shocks $\tilde{\varepsilon}_{jt} = c_j \varepsilon_{jt}$. For the reduced form errors this means:

$$\begin{aligned} u_{it} &= \sum_{j=1}^{K} \tilde{b}_{ij} \tilde{\varepsilon}_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} c_j \varepsilon_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} c_j \exp\left(\frac{h_{jt}}{2}\right) \eta_{jt} \\ &= \sum_{j=1}^{K} \tilde{b}_{ij} \exp\left(\frac{h_{jt} + 2\log(c_j)}{2}\right) \eta_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} \exp\left(\frac{\tilde{h}_{jt}}{2}\right) \eta_{jt}. \end{aligned}$$

Since the log-variance process $\tilde{h}_{jt}$ is restricted to have zero mean, it is:

$$\mathrm{E}(\tilde{h}_{jt}) = \underbrace{\mathrm{E}(h_{jt})}_{=0} + 2\log(c_j) = 0 \quad \Leftrightarrow \quad c_j = 1.$$

Hence, the restriction of the log-variance process to mean zero fixes the scaling of B (Kastner et al. 2017), i.e. $S_1 = I_r$. Therefore, it is $Q_1 = D_1 P_1$ and thus an orthogonal matrix why also full matrix Q is orthogonal. Moreover, it is shown that block B_1 is identified up to permutation and sign switches.

□

A.3 Proof of Corollary 1

Using Proposition 1 shows that an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors can be obtained by $\tilde{B} = BQ$ if and only if Q has the structure $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 = D_1 P_1$ with D_1 a diagonal matrix with ± 1 entries on the diagonal, P_1 a permutation matrix and $Q_4 \in \mathbb{R}^{(K-r)\times(K-r)}$ any orthogonal matrix. Thus, the decomposition $B = (B_1, B_2)$ with $B_1 \in \mathbb{R}^{K\times r}$ and $B_2 \in \mathbb{R}^{K\times(K-r)}$ yields uniqueness of B_1 apart from multiplication of its columns by -1 and permutation. Moreover, in case that $r = K - 1$, column vector B_2 is also unique up to multiplication with -1:

Proof. For $r = K - 1$, matrix Q_4 is a scalar with $Q_4^2 = 1 \Rightarrow Q_4 = \pm 1$. So, full matrix Q can be decomposed in a diagonal matrix with ± 1 entries and a permutation matrix that has an entry of one in the very right bottom corner. This proves the uniqueness of the full

matrix B apart from sign reversal of its columns and permutation of its first $r = K - 1$ columns. □

A.4 Proof of Corollary 2

Proof. Let $Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$ be a $K \times K$ matrix such that $BQ = \begin{pmatrix} B_{11}Q_1 & B_{12}Q_4 \\ B_{21}Q_1 & B_{22}Q_4 \end{pmatrix}$ has the same structure as B, i.e. $B_{22}Q_4$ is still a lower triangular matrix. Thereby, it directly follows that Q_4 is a lower triangular matrix itself. Moreover, because Q_4 is orthogonal, it is also normal and therefore diagonal. Any diagonal and orthogonal matrix has ± 1 entries on the diagonal. So, full matrix Q can be decomposed in a diagonal matrix D having ± 1 entries and a permutation matrix P having an identity block in the lower right $(K-r) \times (K-r)$ block. Thus, matrix B is unique up to multiplication of its columns with -1 and permutation of its first r columns. □

A.5 Partial identification of A-model

Let B be unrestricted and partitioned as in Corollary 2. If BQ should imply an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors, it is $Q = \text{diag}(Q_1, Q_4)$ with $Q_1 = D_1 P_1$ with D_1 a diagonal matrix with ± 1 entries on the diagonal and P_1 a permutation matrix. For the corresponding A-model this implies:

$$A = (BQ)^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

with $A_{11} = Q_1' B_{11}^{-1} + Q_1' B_{11}^{-1} B_{12} (B_{22} - B_{21} B_{11}^{-1} B_{12})^{-1} B_{21} B_{11}^{-1}$ and $A_{12} = -Q_1' B_{11}^{-1} B_{12} (B_{22} - B_{21} B_{11}^{-1} B_{12})^{-1}$ (Magnus & Neudecker 2019). Hence, the first r rows of A do not depend on Q_4 but only on $Q_1 = D_1 P_1$. Consequently, they are identified up to permutation and sign switches. The structural model in A-model form reads:

$$\mathrm{A} y_t = \tilde{\nu} + \sum_{j=1}^{p} \tilde{A}_j y_{t-j} + \varepsilon_t,$$

with $\tilde{\nu} = \mathrm{A}\nu$ and $\tilde{A}_j = \mathrm{A}A_j$ $(j = 1, \ldots, p)$. Consequently, the first r equations of the A-model are identified up to permutation and sign switches.

Appendix B Estimation

B.1 Importance density

To derive the Gaussian approximation of the (unrestricted) IS density $\pi_G(h_{1:r}|\theta, \varepsilon_{1:r})$, we closely follow the exposition of Chan & Grant (2016). We start with an application of Bayes' theorem which gives the zero variance importance density:

$$\log p(h_{1:r}|\theta, \varepsilon_i) \propto \log p(\varepsilon_{1:r}|\theta, h_{1:r}) + \log p(h_{1:r}). \tag{B.1}$$

The assumption of normality in both the transition and measurement equation gives:

$$\log p(h_{1:r}) \propto -\frac{1}{2} h'_{1:r} Q h_{1:r}, \tag{B.2}$$

$$\log p(\varepsilon_{1:r}|\theta, h_{1:r}) \propto \sum_{i=1}^{r} \sum_{t=1}^{T} -\frac{1}{2} \left(h_{it} + \varepsilon_{it}^2 e^{-h_{it}} \right). \tag{B.3}$$

Since the measurement equation is nonlinear in h_i, the normalizing constant of the smoothing distribution in equation (B.1) is not known. An approximate distribution, however, can be obtained by a second order Taylor approximation of the measurement equation (B.3). The corresponding partial derivatives are given as:

$$\begin{aligned} \frac{\partial \log p(\varepsilon_{it}|\theta, h_{it})}{\partial h_{it}} &= -\frac{1}{2} + \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: f_{it} \quad \Rightarrow \quad f = (f_{1:r,1}, \ldots, f_{1:r,T})', \\ -\frac{\partial^2 \log p(\varepsilon_{it}|\theta, h_{it})}{\partial h_{it}^2} &= \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: c_{it} \quad \Rightarrow \quad C = \operatorname{diag}(c_{1:r,1}, \ldots, c_{1:r,T}), \end{aligned}$$

with $f_{1:r,t} = [f_{1t}, \ldots, f_{rt}]$ and $c_{1:r,t} = [c_{1t}, \ldots, c_{rt}]$. A second order Taylor approximation around $\tilde{h}_{1:r}^{(0)}$ then yields:

$$\begin{aligned} \log p(\varepsilon_{1:r}|\theta, h_{1:r}) &\approx \log p(\varepsilon_{1:r}|\theta, \tilde{h}_{1:r}^{(0)}) + (h_{1:r} - \tilde{h}_{1:r}^{(0)})' f - \frac{1}{2}(h_{1:r} - \tilde{h}_{1:r}^{(0)})' C (h_{1:r} - \tilde{h}_{1:r}^{(0)}) \\ &\propto -\frac{1}{2}[h_{1:r}' C h_{1:r} - 2h_{1:r}' \underbrace{(f + C\tilde{h}_{1:r}^{(0)})}_{=:b}]. \end{aligned} \tag{B.4}$$

Combining (B.1), (B.2) and (B.4) provides an approximation of the smoothing distribution which takes the form of a normal kernel:

$$\log p(h_i|\theta, \varepsilon_i) \propto -\frac{1}{2}[h_{1:r}' \underbrace{(C + Q)}_{=:\bar{Q}} h_{1:r} - 2h_{1:r}' b].$$

Consequently, the approximate smoothing density is:

$$\pi_G\left(h_{1:r}|\theta, \varepsilon_{1:r}\right) \sim \mathcal{N}\left(\bar{\delta}, \bar{Q}^{-1}\right), \qquad \text{with} \quad \bar{\delta} = \bar{Q}^{-1} b.$$

The restricted density $\pi_G^c\left(h_{1:r}|\theta, \varepsilon_{1:r}\right)$ is constructed as outlined in Section 3. Note that $\pi_G^c\left(h_{1:r}|\theta, \varepsilon_{1:r}\right)$ yields a good approximation only if $\tilde{h}_{1:r}^{(0)}$ is chosen appropriately. In the following, we sketch how the Newton-Raphson method is used to evaluate the IS density at the mode of the smoothing distribution (B.1).

B.2 Newton-Raphson method

The Newton-Raphson method is implemented as follows: $h_{1:r}$ is initialized by some vector $h_{1:r}^{(0)}$ satisfying the linear constraint, i.e. $A_h h_{1:r}^{(0)} = 0_{r\times 1}$. Then, $h_{1:r}^{(l)}$ is used to evaluate $\bar{Q}$, $\bar{\delta}$ and to iterate:

$$\begin{aligned} \tilde{h}_{1:r}^{(l+1)} &= h_{1:r}^{(l)} + \bar{Q}^{-1}\left(-\bar{Q} h_{1:r}^{(l)} + b\right) = \bar{Q}^{-1} b = \bar{\delta}, \\ h_{1:r}^{(l+1)} &= \tilde{h}_{1:r}^{(l+1)} - \bar{Q}^{-1} A_h' \left(A_h \bar{Q}^{-1} A_h'\right)^{-1} A_h \tilde{h}_{1:r}^{(l+1)}, \end{aligned}$$

for $l \geq 0$ until convergence, i.e. until $\left\| h_{1:r}^{(l+1)} - h_{1:r}^{(l)} \right\| < \varepsilon$ holds for a specified tolerance level ε.

B.3 EM algorithms

To fix notation, define the following quantities:

$$\begin{aligned}
Y^0 &:= (y_1, \ldots, y_T) && K \times T, \\
A &:= (\nu, A_1, \ldots, A_p) && K \times (Kp+1), \\
Y_t^0 &:= \left(y'_{t-1}, \ldots, y'_{t-p}\right)' && Kp \times 1, \\
x_t &:= \left(1, Y_t^{0'}\right)' && (Kp+1) \times 1, \\
X &:= (x_1, \ldots, x_T) && (Kp+1) \times T, \\
y^0 &:= \mathrm{vec}(Y^0) && KT \times 1, \\
\alpha &:= \mathrm{vec}(A) && [K(Kp+1)] \times 1, \\
U &:= (u_1, \ldots, u_T) && K \times T, \\
u &:= \mathrm{vec}(U) && KT \times 1, \\
V_{(-1)} &:= (\exp(-h_1), \ldots, \exp(-h_T)) && K \times T.
\end{aligned}$$

Using this, VAR equation (2.1) can be compactly written as:

$$y^0 = Z\alpha + u,$$

with $Z = (X' \otimes I_K)$.

This yields the following compact representation of the complete data log-likelihood:

$$\begin{aligned}
\mathcal{L}_c(\theta) \propto & -T \ln |B| - \frac{1}{2}\left(y^0 - Z\alpha\right)' \left(I_T \otimes B^{-1}\right)' \Sigma_e^{-1} \left(I_T \otimes B^{-1}\right)\left(y^0 - Z\alpha\right) \\
& + \sum_{i=1}^{r} \left\{ -\frac{T}{2}\ln(s_i) + \frac{1}{2}\ln\left(1-\phi_i^2\right) - \frac{1}{2s_i}\left(\left[1-\phi_i^2\right]h_{i1}^2 + \sum_{t=2}^{T}\left(h_{it} - \phi_i h_{i,t-1}\right)^2\right)\right. \\
& \left. + \frac{1}{2}\ln\left(\frac{s_i}{\left(1-\phi_i^2\right)}\frac{T\left(1-\phi_i^2\right) - 2\phi_i\left(1-\phi_i^T\right)}{T^2\left(1-\phi_i\right)^2}\right)\right\},
\end{aligned} \tag{B.5}$$

where $\Sigma_e^{-1} = \mathrm{diag}(\mathrm{vec}(V_{(-1)}))$ and the last term origins from the constraint imposed on the prior (see equation (3.2)). In particular, it is obtained when multiplying out $\frac{1}{2}\ln(A_h Q^{-1} A_h')$, the normalizing constant of $\pi_2(A_h h_{1:r}|\theta)$.

The EM algorithm requires starting values, which we simply set:

$$\hat{\alpha}^{(0)} = \left(\left[(XX')^{-1} X \right] \otimes I_K \right) y^0,$$
$$\hat{\beta}^{(0)} = S_B \operatorname{vec}\left(\hat{B}^{(0)} \right), \quad \text{with } \hat{B}^{(0)} = (T^{-1}\hat{U}\hat{U}')^{\frac{1}{2}} Q, \text{ and } \hat{U} = Y^0 - \hat{A}X,$$

where Q is a $K \times K$ orthogonal matrix uniformly drawn from the space of K-dimensional orthogonal matrices. In case that $r < K - 1$, we postmultiply Q with a fixed orthogonal matrix Q_2 that rotates $\hat{B}^{(0)}$ such that the lower right $(K-r) \times (K-r)$ block of $\hat{B}^{(0)} Q_2$ is lower triangular. Furthermore, we set the $r \times 1$ vectors:

$$\hat{\phi}^{(0)} = [0.95, \ldots, 0.95]',$$
$$\hat{s}^{(0)} = [0.02, \ldots, 0.02]',$$

which correspond to persistent heteroskedasticity with initial kurtosis of about 3.7 for the estimated structural shocks $\hat{\varepsilon}_i, i = 1, \ldots, r$.

Based on starting values $\theta^{(0)} = \left[\hat{\alpha}^{(0)\prime}, \hat{\beta}^{(0)\prime}, \hat{\phi}^{(0)\prime}, \hat{s}^{(0)\prime} \right]'$, the EM algorithm iteratively cycles through the following steps for $l \geq 1$:

E-Steps

Recall that the E-step is computing the expected value of the complete data log-likelihood:

$$Q(\theta; \theta^{(l-1)}) = \mathrm{E}_{\theta^{(l-1)}} \left[\mathcal{L}_c(\theta) \right],$$

where the expectations are built with respect to the smoothing distribution $p(h_{1:r}|\theta^{(l-1)}, y)$. The expected complete data log-likelihood is given by:

$$
\begin{aligned}
E_{\theta^{(l-1)}}[\mathcal{L}_c(\theta)] \propto & -T\ln|B| - \frac{1}{2}\left(y^0 - Z\alpha\right)'\left(I_T \otimes B^{-1}\right)'\mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]\left(I_T \otimes B^{-1}\right)\left(y^0 - Z\alpha\right) \\
& + \sum_{i=1}^{r}\left\{-\frac{T}{2}\ln(s_i) + \frac{1}{2}\ln\left(1-\phi_i^2\right) + \frac{1}{2}\ln\left(\frac{T\left(1-\phi_i^2\right) - 2\phi_i\left(1-\phi_i^T\right)}{T^2\left(1-\phi_i\right)^2}\right.\right. \\
& \left.\times\frac{s_i}{(1-\phi_i^2)}\right) - \frac{1}{2s_i}\left(\left[1-\phi_i^2\right]\mathrm{E}_{\theta^{(l-1)}}[h_{i1}^2] + \sum_{t=2}^{T}\left(\mathrm{E}_{\theta^{(l-1)}}[h_{it}^2]\right.\right. \\
& \left.\left.\left.-2\phi_i\mathrm{E}_{\theta^{(l-1)}}[h_{it}h_{i,t-1}] + \phi_i^2\mathrm{E}_{\theta^{(l-1)}}[h_{i,t-1}^2]\right)\right)\right\}.
\end{aligned}
$$

Therefore, we require computing the expectations of $\mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]$, with elements $\mathrm{E}_{\theta^{(l-1)}}[\exp(-h_{it})]$, $\mathrm{E}_{\theta^{(l-1)}}[h_{it}^2] = \mathrm{Var}_{\theta^{(l-1)}}[h_{it}] + (\mathrm{E}_{\theta^{(l-1)}}[h_{it}])^2$, and $\mathrm{E}_{\theta^{(l-1)}}[h_{it}h_{i,t-1}] = \mathrm{Cov}_{\theta^{(l-1)}}[h_{it}, h_{i,t-1}] + \mathrm{E}_{\theta^{(l-1)}}[h_{it}]\mathrm{E}_{\theta^{(l-1)}}[h_{i,t-1}]$.

1. EM-1: Here, we compute the moments based on the Gaussian approximation of the smoothing density $\pi_G^c\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)$. The first two moments are given by:

$$
\mathrm{E}\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}, A_h h_{1:r}{=}0\right) = \bar{\delta} - \bar{Q}^{-1}A_h'(A_h\bar{Q}^{-1}A_h')^{-1}A_h\bar{\delta}, \tag{B.6}
$$

$$
\mathrm{Cov}\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}, A_h h_{1:r}{=}0\right) = \bar{Q}^{-1} - \bar{Q}^{-1}A_h'(A_h\bar{Q}^{-1}A_h')^{-1}A_h\bar{Q}^{-1}. \tag{B.7}
$$

   Computation of both variances $\mathrm{Var}_{\theta^{(l-1)}}[h_{it}]$ and first order autocovariances $\mathrm{Cov}_{\theta^{(l-1)}}[h_{it}, h_{i,t+1}]$ can be obtained without computing the whole inverse of $\bar{Q}$ using sparse matrix routines (Rue et al. 2009). An efficient implementation in Matlab is available at the MathWorks File Exchange (see *sparseinv* by Tim Davis). Finally, we compute:

$$
\mathrm{E}_{\theta^{(l-1)}}[\exp(-h_{it})] = \exp\left(-\mathrm{E}_{\theta^{(l-1)}}[h_{it}] + 0.5\mathrm{Var}_{\theta^{(l-1)}}[h_{it}]\right), \tag{B.8}
$$

   which follows from (approximate) normality of $h_{1:r}$.

2. EM-2: Here, we compute the corresponding moments by Importance Sampling. In

particular, we approximate the moments by a Monte Carlo integral:

$$\mathrm{E}_{\theta^{(l-1)}}[g(h_{1:r})] = \int g(h_{1:r}) p\left(h_{1:r}|\varepsilon_{1:r}^{(l-1)}, \theta^{(l-1)}\right) dh_{1:r} \approx R^{-1} \sum_{j=1}^{R} w_{(j)} g(h_{1:r}^{(j)}), \quad \text{(B.9)}$$

where $h_{1:r}^{(j)}$ is drawn from $\pi_G^c\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)$ and $w_{(j)} \propto \frac{p\left(h_{1:r}^{(j)}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)}{\pi_G^c\left(h_{1:r}^{(j)}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)}$. Note that in practice, it is computationally more efficient to repeat the IS estimators of equation (B.9) for each $h_i, i = 1, \ldots, r$ separately. Furthermore, we recommend to start iterating with EM-2 only after EM-1 has converged. To facilitate convergence analysis, we compute the expectation always with the same underlying uniform random numbers until convergence. We find that for sample sizes typically used in macroeconomics ($T \approx 500$), one should choose $R >> 10\,000$ in order to guarantee a sufficient level of accuracy. In our application, we set $R = 50\,000$.

M-Steps

Conditional on the approximate smoothing density of log-variances h_i $(i = 1, \ldots, r)$, we update parameters of both state and measurement equation of the SV-SVAR model.

1. Update ϕ_i and s_i for $i = 1, \ldots, r$:

   Conditional on the moments of the approximate smoothing density we maximize the expected value of the complete data log-likelihood (B.5) with respect to the state equation parameters. Therefore, define:

$$\nabla G(\hat{\phi}, \hat{s}) = \mathrm{E}\left[\frac{\partial \mathcal{L}_c}{\partial \phi'}, \frac{\partial \mathcal{L}_c}{\partial s'}\right]'_{\phi=\hat{\phi}, s=\hat{s}}, \qquad H(\hat{\phi}, \hat{s}) = \mathrm{E}\begin{pmatrix} \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s'} \\ \frac{\partial^2 \mathcal{L}_c}{\partial s \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial s \partial s'} \end{pmatrix}_{\phi=\hat{\phi}, s=\hat{s}}.$$

   The detailed expressions for first and second derivatives of the expected complete data log-likelihood are printed in Section B.4. Then, set $\hat{\phi}_{(k)} = \hat{\phi}^{(l-1)}$ and $\hat{s}_{(k)} = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set:

$$\begin{pmatrix} \hat{\phi}_{(k+1)} \\ \hat{s}_{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_{(k)} \\ \hat{s}_{(k)} \end{pmatrix} - \left(H\left(\hat{\phi}_{(k)}, \hat{s}_{(k)}\right)\right)^{-1} \nabla G\left(\hat{\phi}_{(k)}, \hat{s}_{(k)}\right),$$

until $\left\| \begin{pmatrix} \hat{\phi}_{(k+1)} \\ \hat{s}_{(k+1)} \end{pmatrix} - \begin{pmatrix} \hat{\phi}_{(k)} \\ \hat{s}_{(k)} \end{pmatrix} \right\|$ is smaller than a specified threshold, e.g. 0.001. Then, set $\hat{\phi}^{(l)} = \hat{\phi}_{(k+1)}$ and $\hat{s}^{(l)} = \hat{s}_{(k+1)}$.

2. Update α. Let $Z = (X' \otimes I_K)$, then:

$$\hat{\alpha}^{(l)} = (Z'\tilde{\Sigma}_u^{-1}Z)^{-1}(Z'\tilde{\Sigma}_u^{-1}y^0),$$

with $\tilde{\Sigma}_u^{-1} = \left(I_T \otimes \hat{B}^{(l-1)'}\right)^{-1} \hat{\Sigma}_e^{-1} \left(I_T \otimes \hat{B}^{(l-1)}\right)^{-1}$ and $\hat{\Sigma}_e^{-1} = \mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]$.

3. Update β. Recall $\beta = S_B \operatorname{vec}(B)$, define $\hat{U} = Y^0 - \hat{A}^{(l)}X$ and set:

$$\begin{aligned} \hat{\beta}^{(l)} =& \arg\max_{\beta} \mathrm{E}\left[\mathcal{L}_c(\beta) \middle| \hat{A}^{(l)}, \hat{\phi}^{(l)}, \hat{s}^{(l)}, y\right] \\ \propto& -T\ln|B| - \frac{1}{2}\operatorname{vec}(B^{-1}\hat{U})'\hat{\Sigma}_e^{-1}\operatorname{vec}(B^{-1}\hat{U}). \end{aligned}$$

Finally, set $\theta^{(l)} = \left[\hat{\alpha}^{(l)'}, \hat{\beta}^{(l)'}, \hat{\phi}^{(l)'}, \hat{s}^{(l)'}\right]'$, $l = l + 1$ and return to the E-step. We iterate between E-step and M-steps until the relative change in the expected complete data log-likelihood becomes negligible. To be more precise, the algorithm is a Generalized EM algorithm since the M-step of impact matrix B depends on VAR coefficients α.

B.4 Derivatives expected complete data log-likelihood

The respective derivatives of the expected complete data log-likelihood (B.5) are given in the following. If a certain cross derivative is not stated explicitly, it is zero. Define $S_{xx,i} = \sum_{t=1}^{T-1} \mathrm{E}[h_{it}]^2 + \mathrm{Var}(h_{it})$, $S_{yy,i} = \sum_{t=2}^{T} \mathrm{E}[h_{it}]^2 + \mathrm{Var}(h_{it})$ and $S_{xy,i} = \sum_{t=2}^{T} \mathrm{E}[h_{it}]\mathrm{E}[h_{i,t-1}] + \mathrm{Cov}(h_{it}, h_{i,t-1})$, and:

$$c_i = \frac{s_i}{(1-\phi_i^2)} \frac{T\left(1-\phi_i^2\right) - 2\phi_i\left(1-\phi_i^T\right)}{T^2\left(1-\phi_i\right)^2} = \frac{s_i}{T(1-\phi_i^2)} + \frac{2s_i}{T^2(1-\phi_i^2)} \sum_{j=1}^{T-1}(T-j)\phi_i^j.$$

Then, the complete set of first and second derivatives are given by:

$$
\begin{aligned}
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial s_i}\right] &= -\frac{T}{2s_i} + \frac{1-\phi_i^2}{2s_i^2}\mathrm{E}[h_{i1}^2] + \frac{1}{2s_i^2}(S_{yy,i} - 2\phi_i S_{xy,i} + \phi_i^2 S_{xx,i}) + \frac{\partial \frac{1}{2}\ln c_i}{\partial s_i}, \\
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \phi_i}\right] &= -\frac{\phi_i}{1-\phi_i^2} + \frac{\phi_i}{s_i}\mathrm{E}[h_{i1}^2] + \frac{1}{s_i}S_{xy,i} - \frac{1}{s_i}S_{xx,i}\phi_i + \frac{\partial \frac{1}{2}\ln c_i}{\partial \phi_i}, \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial s_i}\right] &= -\frac{\phi_i}{s_i^2}\mathrm{E}[h_{i1}^2] - \frac{1}{s_i^2}S_{xy,i} + \frac{1}{s_i^2}S_{xx,i}\phi_i + \frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i \partial s_i}, \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i^2}\right] &= \frac{T}{2s_i^2} - \frac{1}{s_i^3}\left(S_{yy,i} - 2S_{xy,i}\phi_i + \phi_i^2 S_{xx,i} + (1-\phi_i^2)\mathrm{E}[h_{i1}^2]\right) + \frac{\partial^2 \frac{1}{2}\ln c_i}{\partial s_i^2}, \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i^2}\right] &= -\frac{S_{xx,i}}{s_i} - \frac{1+\phi_i^2}{(1-\phi_i^2)^2} + \frac{\mathrm{E}[h_{i1}^2]}{s_i} + \frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i^2},
\end{aligned}
$$

where the derivatives with respect to $\frac{1}{2}\ln c_i$ are given by:

$$
\begin{aligned}
\frac{\partial \frac{1}{2}\ln c_i}{\partial s_i} &= \frac{1}{2c_i}\frac{\partial c_i}{\partial s_i} = \frac{1}{2c_i}\left(\frac{1}{T(1-\phi_i^2)} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\frac{\phi_i^j}{(1-\phi_i^2)}\right), \\
\frac{\partial \frac{1}{2}\ln c_i}{\partial \phi_i} &= \frac{1}{2c_i}\frac{\partial c_i}{\partial \phi_i} = \frac{1}{2c_i}\left(\frac{2\phi_i}{T(1-\phi_i^2)^2} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{j\phi_i^{j-1}}{1-\phi_i^2} + 2\frac{\phi_i^{j+1}}{(1-\phi_i^2)^2}\right)\right)s_i, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i \partial s_i} &= \frac{1}{2c_i}\left(\frac{2\phi_i}{T(1-\phi_i^2)^2} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{j\phi_i^{j-1}}{(1-\phi_i^2)} + \frac{2\phi_i^{j+1}}{(1-\phi_i^2)^2}\right)\right) - \frac{1}{2c_i^2}\frac{\partial c_i}{\partial s_i}\frac{\partial c_i}{\partial \phi_i}, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial s_i^2} &= -\frac{1}{2c_i^2}\left(\frac{\partial c_i}{\partial s_i}\right)^2, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i^2} &= \frac{1}{2c_i}\frac{\partial^2 c_i}{\partial \phi_i^2} - \frac{1}{2c_i^2}\left(\frac{\partial c_i}{\partial \phi_i}\right)^2, \\
\frac{\partial^2 c_i}{\partial \phi_i^2} &= s_i\left(\frac{2(1+3\phi_i^2)}{T(1-\phi_i^2)^3}\right. \\
&\quad \left. + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{(j-1)j\phi_i^{j-2}}{1-\phi_i^2} + \frac{2j\phi_i^j}{(1-\phi_i^2)^2} + \frac{2(j+1)\phi_i^j}{(1-\phi_i^2)^2} + \frac{8\phi_i^{j+2}}{(1-\phi_i^2)^3}\right)\right).
\end{aligned}
$$

Furthermore, let $\mathrm{E}[\Sigma_t^{-1}] = B^{-1'}\mathrm{E}[V_t^{-1}]B^{-1}$ whereas $\mathrm{E}[V_t^{-1}] = \mathrm{diag}\left(\mathrm{E}[V_{1t}^{-1}], I_{K-r}\right)$ and $\mathrm{E}[V_{1t}^{-1}] = \mathrm{diag}\left(\mathrm{E}[\exp(-h_{1t})], \ldots, \mathrm{E}[\exp(-h_{rt})]\right)$. Thereby, $\mathrm{E}[\exp(-h_{it})]$ is obtained using (B.8) for EM-1 and (B.9) for EM-2, respectively. Moreover, let $\beta = S_B\,\mathrm{vec}(B)$, $\alpha = \mathrm{vec}(A)$, $\tilde{X}_t = (x_t' \otimes I_K)$, such that $\mathrm{vec}(Ax_t) = \tilde{X}_t\alpha$ and $K^{(K,K)}$ be the $K^2 \times K^2$ commutation matrix.

Then, the first and second derivatives of (B.5) with respect to α and β are given as:

$$
\begin{aligned}
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \alpha'}\right] &= \left(\sum_{t=1}^{T} y_t' \mathrm{E}[\Sigma_t^{-1}] \tilde{X}_t\right) - \alpha' \left(\sum_{t=1}^{T} \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}] \tilde{X}_t\right), \\
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \beta'}\right] &= \left[-T \operatorname{vec}\left(\left[B^{-1}\right]'\right)' + \operatorname{vec}\left(\sum_{t=1}^{T} \mathrm{E}[\Sigma_t^{-1}] u_t \varepsilon_t'\right)'\right] S_B', \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha' \partial \beta}\right] &= -\sum_{t=1}^{T} \left[\left(\varepsilon_t' \otimes \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}]\right) + \left(\tilde{X}_t' \left[B^{-1}\right]' \otimes u_t' \mathrm{E}[\Sigma_t^{-1}]\right)\right] S_B', \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha \partial \alpha'}\right] &= -\left(\sum_{t=1}^{T} \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}] \tilde{X}_t\right), \\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \beta \partial \beta'}\right] &= S_B \left[T\left(B^{-1} \otimes \left[B^{-1}\right]'\right) K^{(K,K)} - \sum_{t=1}^{T} \left(\varepsilon_t \varepsilon_t' \otimes \mathrm{E}[\Sigma_t^{-1}]\right)\right. \\
&\qquad \left. - \left(\sum_{t=1}^{T} \left(B^{-1} \otimes \mathrm{E}[\Sigma_t^{-1}] u_t \varepsilon_t'\right) + \left(\varepsilon_t u_t' \mathrm{E}[\Sigma_t^{1}] \otimes \left[B^{-1}\right]'\right)\right) K^{(K,K)}\right] S_B'.
\end{aligned}
$$

B.5 Inference on impulse response functions and variance decompositions

Following Lütkepohl (2005), the IRFs are elements of the coefficient matrices $\Theta_i = \Phi_i B$ in the Vector Moving Average (VMA) representation of the model:

$$y_t = \mu_y + \sum_{i=0}^{\infty} \Phi_i B \varepsilon_t,$$

where $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$ are the structural shocks, $\mu_y = (I_K - A_1 - \ldots - A_p)^{-1} \nu$ is the unconditional mean of y_t and $\Phi_i \in \mathbb{R}^{K \times K}$ $(i = 0, 1, \ldots)$ is a sequence of exponentially decaying matrices

given as: $\Phi_i = J\mathbf{A}^i J'$ with $J = [I_K, 0, \ldots, 0]$ and:

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \ldots & I_K & 0 \end{pmatrix}.$$

The elements of Θ_i, $\Theta_{jk,i}$'s are the impulse response functions in variable j to a structural innovation k after i periods.

We conduct inference on the estimated quantities $\hat{\Theta}_i$ based on their asymptotic distribution. Given that the IRFs are nonlinear functions of the model parameters, the distribution can be inferred based on the result that $T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1})$. Let $\alpha = \text{vec}(A)$ with $A = (A_1, \ldots, A_p)$, i.e. the intercept ν is not included in α only in this subsection, $\beta = S_B \,\text{vec}(B)$ and partition the asymptotic covariance matrix of $\hat{\theta}$ into:

$$\mathcal{I}(\theta)^{-1} = \Sigma_\theta = \begin{pmatrix} \Sigma_\nu & & & & \\ \Sigma_{\nu,\alpha} & \Sigma_\alpha & & & \\ \Sigma_{\nu,\beta} & \Sigma_{\alpha,\beta} & \Sigma_\beta & & \\ \Sigma_{\nu,\phi} & \Sigma_{\alpha,\phi} & \Sigma_{\beta,\phi} & \Sigma_\phi & \\ \Sigma_{\nu,s} & \Sigma_{\alpha,s} & \Sigma_{\beta,s} & \Sigma_{\phi,s} & \Sigma_s \end{pmatrix}.$$

As in Brüggemann et al. (2016), an application of the Delta method yields the asymptotic distribution of the structural impulse responses:

$$\sqrt{T}(\hat{\Theta}_i - \Theta_i) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\Theta}_i}), \qquad i = 0, 1, 2, \ldots,$$

where:

$$\Sigma_{\hat{\Theta}_i} = C_{i,\alpha}\Sigma_\alpha C_{i,\alpha}' + C_{i,\beta}\Sigma_\beta C_{i,\beta}' + C_{i,\alpha}\Sigma_{\alpha,\beta}' C_{i,\beta}' + C_{i,\beta}\Sigma_{\alpha,\beta} C_{i,\alpha}',$$

with $C_{0,\alpha} = 0$, $C_{i,\alpha} = \frac{\partial \,\text{vec}(\Theta_i)}{\partial \alpha'} = (B' \otimes I_K)G_i$ and $G_i = \frac{\partial \,\text{vec}(\Phi_i)}{\partial \alpha'} = \sum_{j=0}^{i-1} [J(\mathbf{A}')^{i-1-j}] \otimes \Phi_j$ for $i \geq 1$. Finally, $C_{i,\beta} = \frac{\partial \,\text{vec}(\Theta_i)}{\partial \beta'} = (I_K \otimes \Phi_i)S_B'$ for $i \geq 0$. Similarly, for the accumulated

structural impulse responses $\Xi_n = \sum_{i=0}^{n} \Theta_i$, we get:

$$\sqrt{T}\left(\hat{\Xi}_n - \Xi_n\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\Xi}_n}), \qquad n = 0, 1, 2, \ldots,$$

where:

$$\Sigma_{\hat{\Xi}_n} = P_n \Sigma_\alpha P_n' + \bar{P}_n \Sigma_\beta \bar{P}_n' + P_n \Sigma_{\alpha,\beta}' \bar{P}_n' + \bar{P}_n \Sigma_{\alpha,\beta} P_n',$$

with $P_n = (B' \otimes I_K) F_n$, $F_0 = 0$, $F_n = G_1 + \cdots + G_n$, $\bar{P}_n = (I_K \otimes \Psi_n) S_B'$ and $\Psi_n = \sum_{i=0}^{n} \Phi_i$.

Since we also consider structural impulse responses that are normalized to unity, note that the Delta method yields for $\Theta_i^* = \Phi_i B^*$ $(i = 0, 1, 2, \ldots)$ and $B^* = B\tilde{B}$ with $\tilde{B} = \mathrm{diag}(1_{K\times 1} \oslash b)$, $b = \mathrm{diag}(B)$ and $\oslash$ being the elementwise division:

$$\sqrt{T}\left(\hat{\Theta}_i^* - \Theta_i^*\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\Theta}_i^*}), \qquad i = 0, 1, 2, \ldots,$$

where:

$$\Sigma_{\hat{\Theta}_i^*} = C_{i,\alpha^*} \Sigma_\alpha C_{i,\alpha^*}' + C_{i,\beta^*} \Sigma_\beta C_{i,\beta^*}' + C_{i,\alpha^*} \Sigma_{\alpha,\beta}' C_{i,\beta^*}' + C_{i,\beta^*} \Sigma_{\alpha,\beta} C_{i,\alpha^*}',$$

with $C_{i,\alpha^*} = \frac{\partial \,\mathrm{vec}(\Theta_i^*)}{\partial \alpha'} = (B^* \otimes I_K) G_i$, $C_{i,\beta^*} = \frac{\partial \,\mathrm{vec}(\Theta_i^*)}{\partial \beta'} = [(\tilde{B} \otimes \Phi_i) - (I_K \otimes \Theta_i) G_K \mathrm{diag}(1_{K\times 1} \oslash [b \odot b]) G_K'] S_B'$, G_K a selection matrix such that $\mathrm{vec}(D) = G_K d$ for a diagonal matrix $D = \mathrm{diag}(d)$ and $\odot$ being the elementwise multiplication.

The proportion of the h-step ahead forecast error variance in variable k that is accounted for by innovations in variable j is denoted by:

$$\zeta_{kj,h} = \sum_{i=0}^{h-1} (e_k' \Theta_i e_j)^2 / \mathrm{MSE}_k(h),$$

where e_j is the j-th column of I_K and $\mathrm{MSE}_k(h) = \sum_{i=0}^{h-1} e_k' \Phi_i B B' \Phi_i' e_k$. Corresponding standard errors can be obtained by a slight modification of the Delta method used in Lütkepohl

(1990) for $k, j = 1, \ldots, K$ and $h \geq 1$:

$$\begin{aligned}
&\sqrt{T}\left(\hat{\zeta}_{kj,h} - \zeta_{kj,h}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\hat{\zeta}_{kj,h}}\right), \quad \text{with} \\
\Sigma_{\hat{\zeta}_{kj,h}} &= d_{kj,h}\Sigma_{\alpha} d'_{kj,h} + \bar{d}_{kj,h}\Sigma_{\beta}\bar{d}'_{kj,h} + \bar{d}_{kj,h}\Sigma_{\alpha,\beta} d'_{kj,h} + d_{kj,h}\Sigma'_{\alpha,\beta}\bar{d}'_{kj,h}, \\
d_{kj,h} &= 2\sum_{i=1}^{h-1}\left[\mathrm{MSE}_k(h)(e'_k\Theta_i e_j)(e'_j B' \otimes e'_k)G_i - (e'_k\Theta_i e_j)^2 \sum_{m=1}^{h-1}(e'_k\Phi_m BB' \otimes e'_k)G_m\right] \\
&\quad \Big/ \mathrm{MSE}_k^2(h)\,, \\
\bar{d}_{kj,h} &= \sum_{i=0}^{h-1}\left[2\mathrm{MSE}_k(h)(e'_k\Theta_i e_j)(e'_j \otimes e'_k\Phi_i) - (e'_k\Theta_i e_j)^2\left(\sum_{m=0}^{h-1} e'_k\Phi_m \otimes e'_k\Phi_m\right)\right. \\
&\quad \left.\left(K^{(K,K)} + I_{K^2}\right)(B \otimes I_K)\right] S'_B \Big/ \mathrm{MSE}_k^2(h)\,,
\end{aligned}$$

and $K^{(K,K)}$ being the $K^2 \times K^2$ commutation matrix.

Appendix C Tests for identification

In this part, we quickly describe the tests for identification via heteroskedasticity proposed by Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016). Recall the following sequence of tests:

$$H_0 : r = r_0 \qquad \text{vs} \qquad H_1 : r > r_0. \tag{C.1}$$

Suppose that r_0 is the true number of heteroskedastic errors, and separate the structural shocks $\varepsilon_t = B^{-1}u_t = (\varepsilon'_{1t}, \varepsilon'_{2t})'$ into a heteroskedastic part $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and homoskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Under the null ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is homoskedastic white noise. To test for remaining heteroskedasticity in ε_{2t}, Lanne & Saikkonen (2007) propose to use Portmanteau types of test statistics on the second moment of ε_{2t}. In particular, they construct the following time series:

$$\rho_t = \varepsilon'_{2t}\varepsilon_{2t} - T^{-1}\sum_{t=1}^{T}\varepsilon'_{2t}\varepsilon_{2t}, \tag{C.2}$$

$$\vartheta_t = \mathrm{vech}(\varepsilon_{2t}\varepsilon'_{2t}) - T^{-1}\sum_{t=1}^{T}\mathrm{vech}(\varepsilon_{2t}\varepsilon'_{2t}), \tag{C.3}$$

with vech$(\cdot)$ being the half-vectorization operator as defined e.g. in Lütkepohl (2005). Based on these time series, autocovariances up to a prespecified horizon H are tested considering the following statistics:

$$Q_1(H) = T \sum_{h=1}^{H} \left(\frac{\tilde{\gamma}(h)}{\tilde{\gamma}(0)} \right)^2, \tag{C.4}$$

$$Q_2(H) = T \sum_{h=1}^{H} \operatorname{tr} \left[\tilde{\Gamma}(h)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(h) \tilde{\Gamma}(0)^{-1} \right], \tag{C.5}$$

where $\tilde{\gamma}(h) = T^{-1} \sum_{t=h+1}^{T} \rho_t \rho_{t-h}$ and $\tilde{\Gamma}(h) = T^{-1} \sum_{t=h+1}^{H} \vartheta_t \vartheta_{t-h}'$. It is shown that under the null, $Q_1(H) \xrightarrow{d} \chi^2(H)$ and $Q_2(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$. Alternatively, Lütkepohl & Milunovich (2016) propose a test based on the auxiliary model:

$$\tilde{\vartheta}_t = \delta_0 + D_1 \tilde{\vartheta}_{t-1} + \ldots + D_H \tilde{\vartheta}_{t-h} + \tilde{\zeta}_t, \tag{C.6}$$

where $\tilde{\vartheta}_t = \text{vech}(\varepsilon_{2t} \varepsilon_{2t}')$. Under the null hypothesis, $D_1 = \ldots = D_h = 0$, and a standard LM statistic is given by:

$$LM(H) = \frac{1}{2} T (K-r_0)(K-r_0+1) - T \operatorname{tr}[\hat{\Sigma}_{\tilde{\zeta}} \tilde{\Gamma}(0)^{-1}],$$

where $\hat{\Sigma}_{\tilde{\zeta}}$ is the estimated residual covariance matrix from auxiliary model (C.6). Under the null, the test statistic converges to $LM(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$.

Appendix D Complementary results

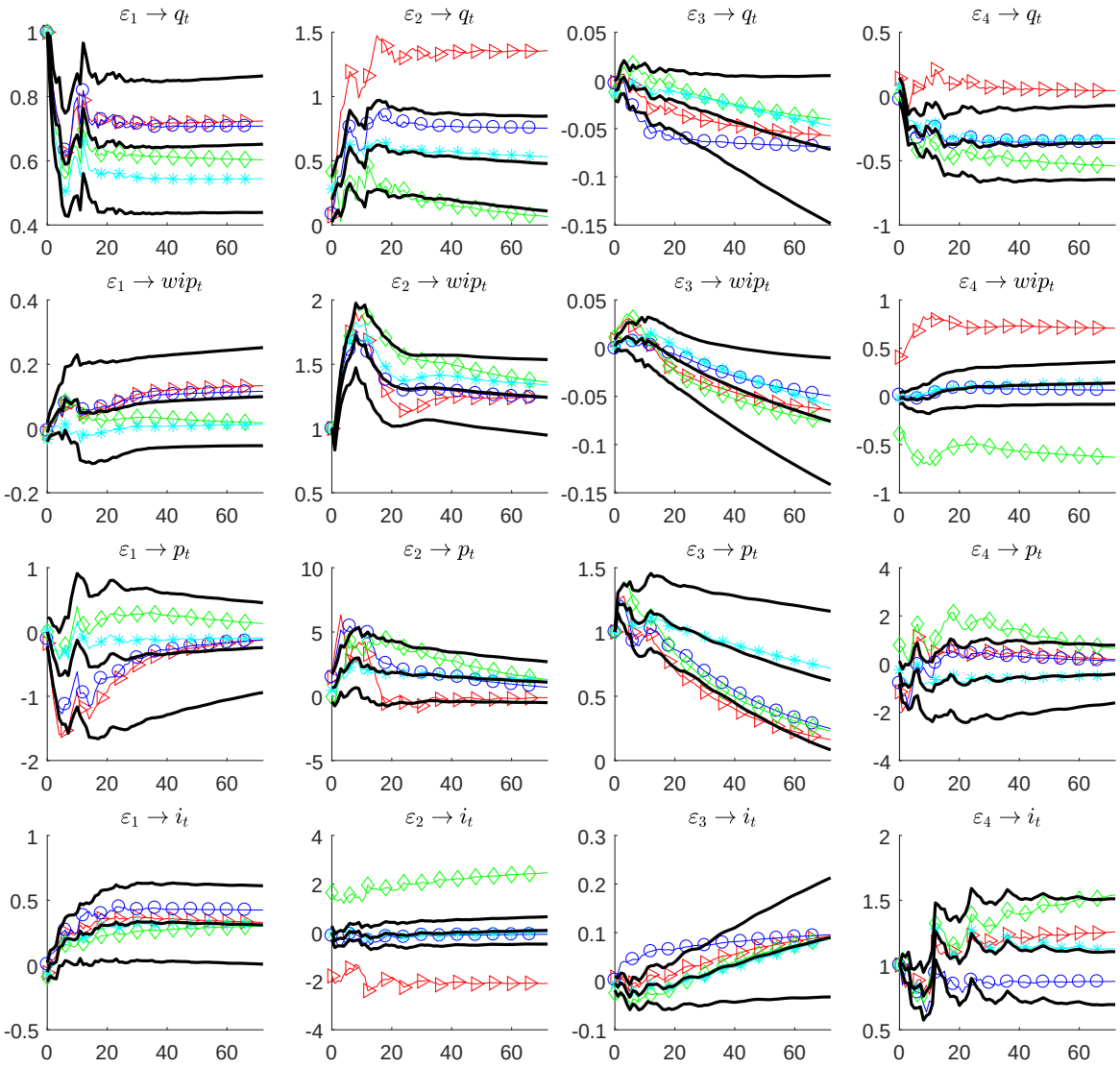


Figure 2: Complete set of standardized IRFs with 90% confidence intervals (solid lines). For comparison, we also provide estimates identified by alternative volatility estimators: MS(2) (triangle), MS(3) (circles), STVAR (diamonds) and GARCH (stars).