Identification of Structural Vector Autoregressions by Stochastic Volatility*

Dominik Bertsche[a] Robin Braun[b]

first draft: December 13, 2017
this version: September 11, 2018

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 EM algorithms for Maximum Likelihood inference. Simulation evidence suggests that the SV-SVAR works well in identifying structural parameters also under misspecification of the variance process, particularly if compared to alternative heteroskedastic SVARs. We apply the model to study the interdependence between monetary policy and stock markets. Since shocks identified by heteroskedasticity may not be economically meaningful, we exploit the framework to test conventional exclusion restrictions as well as Proxy SVAR restrictions which are overidentifying in the heteroskedastic model.

Keywords: Structural Vector Autoregression (SVAR), Identification via heteroskedasticity, Stochastic Volatility, Proxy SVAR

JEL classification: C32

[a]Corresponding Author, University of Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: dominik.bertsche@uni-konstanz.de

[b]University of Konstanz, Graduate School of Decision Science, Department of Economics, Box 129, 78457 Konstanz, Germany, email: robin.braun@uni-konstanz.de

*We thank Ralf Brüggemann, Helmut Lütkepohl, participants of the 1st Vienna Workshop on Economic Forecasting, the Doctoral Workshop on Applied Econometrics at the University of Strasbourg, the IAAE 2018, ESEM 2018, the 2018 Annual Congress of the Verein für Socialpolitik and the Econometrics Colloquium at the University of Konstanz for useful comments on this paper. Financial support by the Graduate School of Decision Science (GSDS) is gratefully acknowledged.

1 Introduction

Following Sims (1980), structural vector autoregressive (SVAR) models have been used extensively in empirical macroeconomics. Based on a reduced form VAR, identifying restrictions are imposed to back out a unique set of structural shocks and estimate their dynamic effects on the endogenous variables. Popular approaches for identification include short- and long-run restrictions on the effects of structural shocks (Sims; 1980; Bernanke & Mihov; 1998; Blanchard & Quah; 1989), sign restrictions (Faust; 1998; Canova & De Nicolo; 2002; Uhlig; 2005) and identification via external instruments, also known as Proxy SVARs (Stock & Watson; 2012; Mertens & Ravn; 2013; Montiel-Olea, Stock & Watson; 2016). Furthermore, a growing body of literature exploits statistical properties of the data to identify SVAR models, assuming non-Gaussianity (Lanne, Meitz & Saikkonen; 2017; Gourieroux, Monfort & Renne; 2017) or heteroskedasticity of the structural shocks (see Lütkepohl & Netšunajev (2017a) for a review).[1]

To model and identify heteroskedastic shocks within SVAR models, a variety of variance models have been used in the literature. These include a simple breakpoint model (Rigobon; 2003), a Markov Switching model (Lanne, Lütkepohl & Maciejowska; 2010), a GARCH model (Normandin & Phaneuf; 2004) and a Smooth Transition model (Lütkepohl & Netšunajev; 2017b). More recently, Lewis (2018) discusses identification and estimation of heteroskedastic SVARs in a GMM framework without specifying any functional forms for the variance.

In this paper, we contribute to this literature in discussing identification and efficient estimation of SVARs with a stochastic volatility (SV) model. Specifically, we assume that the log variances of structural shocks are latent, each following independent AR(1) processes. Following the methodology Lewis (2018), we show that in conjunction with a fixed impact matrix, this yields additional restrictions that allow to pin down a unique set of orthogonal shocks. To the best of our knowledge, this model has not yet been used for identification in the SVAR literature.

A stochastic volatility model for the variance of structural shocks is an attractive specification for various reasons. First, SV models enjoy increasing popularity 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). Given this context, it seems natural to exploit the model also for identification purposes of SVAR models. Second, the SV model is known to be more flexible than models with deterministic variance processes. As pointed out in Kim, Shephard & Chib (1998), this additional flexibility typically translates into superior fit in comparison to equally parameterized models from the GARCH family. We find this to be confirmed in our empirical example where a simple SV model provides the best model fit with a relatively small amount of parameters

[1] For a textbook treatment of identification in SVARs we refer to Kilian & Lütkepohl (2017).

and therefore, is favored by any conventional information criterion (IC). This is an important aspect, given that recent evidence of Lütkepohl & Netšunajev (2017a) suggests to choose the heteroskedasticity model of SVARs by information criteria. Finally we provide evidence that, in comparison to alternative heteroskedastic SVARs, the SV-SVAR model works 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 (DGPs). More specifically, by simulating data from SVAR models subject to four distinct variance specifications we find that the SV model performs superior in terms of mean squared error of estimated impulse response functions.

Since the SV specification implies a nonlinear state space model, standard linear filtering algorithms cannot be applied to evaluate the likelihood function which makes estimation of the SV-SVAR model relatively challenging. However, many estimation methods have been proposed in the literature to overcome this difficulty starting with Generalized Methods of Moments (Melino & Turnbull; 1990), Quasi Maximum Likelihood (Harvey, Ruiz & Shephard; 1994; Ruiz; 1994), Simulated Likelihood (Danielsson & Richard; 1993) and Bayesian methods (Kim et al.; 1998) based on Markov Chain Monte Carlo (MCMC) simulation. In this paper, we follow Durbin & Koopman (1997) in evaluating the likelihood function by importance sampling. To maximize the likelihood function we develop two versions of an Expectation Maximization (EM) algorithm. The first is based on a second order Taylor approximation of the intractable smoothing distribution necessary in the E-step and relies on sparse matrix algorithms developed for Gaussian Markov random fields (Rue, Martino & Chopin; 2009; Chan; 2017). Therefore, the algorithm is very fast and typically converges within seconds. Our second EM algorithm approximates the E-step by Monte Carlo integration, exploiting that the error term of a log-linearized state equation can be accurately approximated by a mixture of normal distributions (Kim et al.; 1998). Conditional on simulated mixture indicators, the model has a normal linear state space representation allowing to compute the expectations necessary in the E-step by standard Kalman smoothing recursions. Thereby, the second order approximation can be avoided at the cost of higher computational effort.

In an empirical application, we use the proposed model to identify the structural parameters of a VAR specified in Bjørnland & Leitemo (2009). Within conventional SVAR analysis, they study the interdependence between monetary policy and the stock market based on short- and long-run restrictions. We find that compared to other heteroskedastic SVAR models, the SV specification provides superior fit and is favored by all conventional information criteria. Since structural shocks identified by heteroskedasticity are not guaranteed to be economically meaningful, we follow Lütkepohl & Netšunajev (2017a) and test the exclusion restrictions used by Bjørnland & Leitemo (2009). In addition, we also test Proxy SVAR restrictions which arise if the narrative series of Romer & Romer (2004) and Gertler & Karadi (2015) are used as external instruments to identify a monetary policy shock. Our results indicate that the short-run restrictions of Bjørnland & Leitemo (2009) and Proxy SVAR restrictions based on the shock of Gertler & Karadi (2015) are rejected by the data. However, we do neither find

evidence against imposing the long-run restriction of Bjørnland & Leitemo (2009) nor against identifying a monetary policy shock by the Romer & Romer (2004) series.

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 and reviews a procedure to test for identification. In section 4, we present simulation evidence while in section 5 we apply the proposed model to study the interdependence between US monetary policy and stock markets. Section 6 concludes.

2 Identification of SVAR via Stochastic Volatility

Let y_t be a $K \times 1$ vector of endogenous variables. We consider the heteroskedastic SVAR 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) model for y_t, capturing common dynamics across the time series data by a linear specification. Here, A_j for $j = 1, \ldots, p$ are $K \times K$ matrices of autoregressive coefficients 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 error terms u_t are decomposed into a linear function of K structural shocks $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$, with B a $K \times K$ invertible contemporaneous impact matrix and $V_t^{\frac{1}{2}}$ a stochastic diagonal matrix with strictly positive elements capturing potential heteroskedasticity and/or non-normality in each structural shock. This specification yields a time-varying covariance matrix of the reduced form errors u_t given as $\Sigma_t = \mathrm{E}(u_t u_t') = B V_t B'$. Throughout the paper, we assume that there are $r \leq K$ heteroskedastic shocks which are ordered such that they appear first in the vector ε_t. To model the time varying second moment of these shocks, we specify an independent Gaussian AR(1) log stochastic volatility model for each of the r heteroskedastic components:

$$V_t = \begin{bmatrix} \mathrm{diag}(\exp([h_{1t}, \ldots, h_{rt}]')) & 0 \\ 0 & I_{K-r} \end{bmatrix}, \tag{2.3}$$

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

where $\omega_{it} \sim \mathcal{N}(0,1)$ and $\mathrm{E}(\varepsilon_t'\omega_t) = 0$ for $\omega_t = [\omega_{1t}, \ldots, \omega_{rt}]'$. Furthermore, the initial states are assumed to be initialized from the unconditional distribution $h_{i1} \sim \mathcal{N}(\mu_i, s_i/(1-\phi_i^2))$. Note that the proposed model for equation (2.2) is very 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 model their logarithms as latent AR(1)'s. We will assume that the underlying AR(1) processes of the log-volatilities are stable with finite variance implying that for $i = 1, \ldots, r$, $|\phi_i| < 1$ and $0 < s_i < \infty$. It immediately follows that ε_t is assumed to be a strictly stationary stochastic process with finite second moment, which will aid in the identification analysis. In particular, the following basic properties can be derived for the model in a straightforward manner (see e.g. Jacquier, Polson & Rossi (1994)): for $i = 1, \ldots, r$,

$$\gamma_i(\tau) = \mathrm{Cov}(\varepsilon_{it}^2, \varepsilon_{i,t+\tau}^2) = \exp(2\mu_i + \sigma_{h_i}^2)(\exp(\sigma_{h_i}^2\phi_i^\tau) - 1) \tag{2.5}$$

$$\kappa_i = \frac{E(\varepsilon_{it}^4)}{(\varepsilon_{it}^2)^2} = \mathrm{E}(\eta_{it}^4)\exp\left(\sigma_{h_i}^2\right), \tag{2.6}$$

$$\mathrm{E}(\varepsilon_{it}^2) = \mathrm{E}(\exp(h_{it})\eta_{it}^2) = \mathrm{E}(\exp(h_{it}))\mathrm{E}(\eta_{it}^2) = \mathrm{E}(\exp(h_{it})) = \exp\left(\mu_i + \frac{1}{2}\sigma_{h_i}^2\right), \tag{2.7}$$

where $\sigma_{h_i}^2 = s_i/(1-\phi_i^2)$ is the unconditional variance of the underlying log-volatility process.

The model is able to capture the main stylized facts of structural shocks that are typically encountered in empirical SVAR analysis. First, heteroskedasticity can be modeled by setting $\phi_i > 0$. The respective autocovariance function in the second moment of ε_{it} is given by equation (2.5), displaying an exponential decay ϕ_i. This autocovariance function has been found to be very flexible enabling to capture a large variety of heteroskedasticity patterns, an argument that we can confirm based on our 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 the shock is homoskedastic ($\phi_i = 0$), the model is still able to capture heavy tails under conditional Gaussianity. In this particular case, the structural error would be independent and identically distributed following a mixture of log-normal and Gaussian errors.[2] We argue that this is a 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 conditionally Gaussianity. Finally, equation (2.7) gives the unconditional scale of the structural shocks as a function of the underlying SV parameters.

In the following, we will use equations (2.5)-(2.7) to discuss identification in detail. First, note that the structural shocks are latent variables and a unique scaling must be obtained.

[2]Note the similarity to a t-distribution, which can be represented as a product of an independent Gamma and Gaussian random variable.

For this purpose, we follow the widely used normalization of setting the scale to $\mathrm{E}(\varepsilon_t\varepsilon_t') = I_K$. Using equation (2.7), this can be achieved by restricting the mean of the AR(1) processes to $\mu_i = -0.5s_i/(1-\phi_i^2)$. It follows that the structural parameters in B are related to the unconditional reduced form covariance matrix by:

$$\mathrm{E}(u_t u_t') = \Sigma_u = BB'. \tag{2.8}$$

Given this normalization, a standard interpretation applies in that the jth column of B corresponds to the average contemporaneous response of the endogenous variables y_t to shock ε_{jt} of size "one standard deviation".

Due to the symmetry of the covariance matrix, identification in the SV-SVAR model cannot be discussed based on equation (2.8) solely. For that purpose, we follow the approach of Lewis (2018) who treats identification by time varying volatility in a more general context requiring no specific functional forms. In particular, identification can be analyzed based on the lag τ autocovariance in the squared reduced form residuals $\xi_t = \mathrm{vech}(u_t u_t')$. In Appendix A.1, we derive this function for the SV-SVAR model taking the following very compact form:

$$\mathrm{Cov}(\xi_t, \xi_{t+\tau}) = L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)' L_K' \tag{2.9}$$

where L_K is the elimination matrix such that $\mathrm{vech}(A) = L_K \,\mathrm{vec}(A)$, G_K is a selection matrix with zeros and ones such that $\mathrm{vec}(D) = G_K d$ for $D = \mathrm{diag}(d)$ and $M_\tau = \mathrm{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$. 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 in B and r autocovariances in $\gamma_i(\tau)$, implicitly parameterized nonlinearly by the underlying SV processes. Proposition 1 summarizes which set of structural parameters of the model in B are (locally) identified for any $r \leq K$.

Proposition 1. *Let $B = [B_1, B_2]$ with $B_1 \in \mathbb{R}^{K\times r}$, $B_2 \in \mathbb{R}^{K\times(K-r)}$ and. Assume the stable SV-SVAR model presented above with $|\phi_i| < 1$, $\phi_i \neq 0$ and $0 < s_i < \infty$ for $i = 1, \ldots, r$, implying that equations* (2.8) *and* (2.9) *hold. Then, matrix B_1 is unique up to permutation and sign switches.*

Proof. See Appendix A.2.

In fact, it is not necessary that $r = K$ shocks are heteroskedastic in order that the model is fully identified. The orthogonality constraints implied by equation (2.8) yield enough structure to fully 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 Appendix A.3.

Our results are in line with those derived for SVAR and factor models identified with heteroskedasticity based on a GARCH model (see Sentana & Fiorentini (2001); Milunovich & Yang

(2013)). Furthermore, they coincide with the general results of identification provided by Lewis (2018) which also cover our model. However, our analysis is slightly different in that we assume a different normalization of the shocks, which gives us full identification under $r = K - 1$ heteroskedastic shocks based on orthogonality constraints. Furthermore, it allows for a simplified proof adapted to the parametric framework we consider.

At this point we highlight that identification of the model can also be discussed based on non-Gaussianity implied by the SV model. If one is willing to assume mutual independence in ε_{it}, the SV-SVAR model as discussed in this paper is covered by the general framework of Lanne et al. (2017). Specifically, the structural parameters in B are identified up to permutation and sign if the structural shocks are strictly stationary with finite second moments, mutually independent and with at most one Gaussian component. For the SV-SVAR model, this means that in order to achieve strict stationarity and finite second moments, we need $s_i < \infty$ and $|\phi_i| < 1 \ \forall i$ as discussed above. Furthermore, under conditionally Gaussian errors, at most one structural shock can display a degenerate SV process with $s_i = 0$, implying a Gaussian marginal. Analogous results regarding to partial identification are available in Maxand (2017). As in proposition A.2, the structural parameters associated with the non-Gaussian shocks are locally identified up to permutation and sign-changes.

Before we continue with estimation of the model, we discuss an additional constraint that we impose. Note that we identify the scale of the structural shocks by setting $\mu_i = -\frac{s_i}{2(1-\phi_i^2)}$, implying that $\mathrm{E}(\varepsilon_t \varepsilon_t') = I_K$. However, this constraint holds only in expectation and for very persistent heteroskedasticity patterns, the sample moment can be very distinct in finite samples. For that purpose, we additionally impose the sample constraint:

$$A_h h_i = \mu_i, \tag{2.10}$$

where $A_h = \mathbf{1_T}'/T$ and $h_i = [h_{i1}, \ldots, h_{iT}]'$. We find this constraint given in (2.10) to yield satisfactory results in finite sample sizes, yielding a scalings of the latent shocks very close to unit (sample) variance. Note that this constraint leads to a rank reduction of the covariance matrix implied for h_i by the Gaussian AR(1) model, which has to be accounted for during estimation. Note that 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). However, this would require that we leave μ_i unrestricted implying an additional parameter to estimate, which is why we stick with restriction (2.10).

3 Maximum Likelihood Estimation

In order to estimate the model, we propose a full Maximum Likelihood approach. Let $\phi = [\phi_1, \ldots, \phi_r]'$, $s = [s_1, \ldots, s_r]'$ and the parameter vector $\theta = [\mathrm{vec}([\nu, A_1, \ldots, A_p])', \mathrm{vec}(B)', \phi', s']'$. Assuming normality of the standardized structural shocks η_t, the log-likelihood function based

on the prediction error decomposition is given as follows:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \left[-\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |BV_{t|t-1}B'| - \frac{1}{2} u_t' (BV_{t|t-1}B')^{-1} u_t \right],$$

where $u_t = y_t - \nu - \sum_{j=1}^{p} A_j y_{t-j}$ and $V_{t|t-1} = \mathrm{E}[V_t|\mathcal{F}_{t-1}]$ are one-step ahead predicted variances conditional on the information set at time $t-1$. Since the SV model implies a nonlinear state space model, the predictive distributions $p(h_t|\theta, y_{t-1})$ necessary to compute $V_{t|t-1}$ are not available in closed form. That is, the likelihood is intractable and standard Kalman filter algorithms cannot be applied. To overcome this difficulty, we follow Durbin & Koopman (1997) and Chan (2013) in evaluating the likelihood function by importance sampling in a computationally efficient way. Furthermore, to maximize the likelihood, we develop two versions of an Expectation Maximization algorithm which lead to fast and reliable results.

3.1 Evaluation of the Likelihood

To show how the likelihood can be evaluated by importance sampling, we slightly manipulate the log-likelihood function. For that purpose, let $\varepsilon_t = B^{-1}u_t$ and $v_{i,t|t-1}$ the i-th diagonal element of $V_{t|t-1}$, then:

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

where we have used that $\log |BV_{t|t-1}B'| = 2 \log |B| + \sum_{i=1}^{K} \log \left(v_{i,t|t-1} \right)$. Therefore, given autoregressive coefficients and contemporaneous impact matrix, likelihood evaluation of the SV-SVAR model reduces to the evaluation of K univariate densities for each structural shock. For $i = r+1, \ldots, K$ these densities are trivial to compute since $v_{i,t|t-1} = 1$. However, the densities $\log p(\varepsilon_i|\theta)$ for $i \leq r$ are not tractable. Their evaluation equals computing the following high-dimensional integral for $i = 1, \ldots, r$:

$$p(\varepsilon_i|\theta) = \int p(\varepsilon_i|\theta, h_i) p^c(h_i|\theta) dh_i. \tag{3.1}$$

where $p(\varepsilon_i|\theta, h_i)$ is a Gaussian distribution and $p^c(h_i|\theta)$ the prior density implied by the Gaussian AR(1) model subject to the constraint $A_h h_i = \mu_i$.

To evaluate this integral, we use an importance sampling estimator. Therefore, let $q(h_i)$ be a proposal distribution from which independent random draws $h_i^{(1)}, \ldots, h_i^{(R)}$ can be generated, and further let $q(h_i)$ dominate $p(\varepsilon_i|\theta, h_i)p(h_i|\theta)$. An unbiased importance sampling estimator

of the integral in equation (3.1) is:

$$\widehat{p(\varepsilon_i|\theta)} = \frac{1}{R}\sum_{j=1}^{R}\frac{p(\varepsilon_i|\theta, h_i^{(j)})p^c(h_i^{(j)}|\theta)}{q(h_i^{(j)})}. \tag{3.2}$$

Plugging (3.2) into the SV-SVAR log-likelihood yields an IS estimator of the SV-SVAR log-likelihood function:

$$\widehat{\mathcal{L}(\theta)} = -T\log|B| + \sum_{i=1}^{r}\log\widehat{p(\varepsilon_i|\theta)} + \sum_{i=r+1}^{K}\log p(\varepsilon_i|\theta). \tag{3.3}$$

The accuracy of the IS estimator crucially depends on our choice for the importance densities $q(h_i)$ which we discuss in the following. First, note that the optimal (zero variance) importance density is given by the smoothing distribution $p(h_i|\theta, \varepsilon_i) \propto p(\varepsilon_i|\theta, h_i)p(h_i|\theta)$. However, since the likelihood of the measurement equation is nonlinear in h_i, the normalizing constant is unkown which is why we rely on IS in the first place. We follow Durbin & Koopman (1997, 2000) and use a Gaussian importance density denoted by $\pi_G(h_i|\theta, \varepsilon_i)$, which is centered at the mode of $p(h_i|\theta, \varepsilon_i)$ with precision equal to the curvature at this point. For computational reasons, we rely on fast algorithms that exploit the sparse precision matrices of Gaussian Markov random fields as used e.g. in Rue et al. (2009) for a broad class of models and Chan & Grant (2016) for stochastic volatility models in particular.

To derive $\pi_G(h_i|\theta, \varepsilon_i)$, we follow the exposition of Chan & Grant (2016). For a moment, assume that there was no linear constraint on h_i. Then, normality implies the following explicit form of the zero variance IS density:

$$p(h_i|\theta, \varepsilon_i) \propto \exp\left(-\frac{1}{2}(h_i - \delta_i)'Q_i(h_i - \delta_i) + \log p(\varepsilon_i|\theta, h_i)\right),$$

where $Q_i = H_i'\Sigma_{h_i}^{-1}H_i$ with

$$H_i = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_i & 1 & 0 & \dots & 0 \\ 0 & -\phi_i & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\phi_i & 1 \end{pmatrix},$$

and $\Sigma_{h_i} = \text{diag}([\frac{s_i}{1-\phi_i}, s_i, \dots, s_i]')$. Furthermore, $\delta_i = H_i^{-1}\tilde{\delta}_i$ with $\tilde{\delta}_i = [\mu_i, (1-\phi_i)\mu_i, \dots, (1-\phi_i)\mu_i]'$. The Gaussian approximation is based on a second order Taylor expansion of the non-linear density $\log p(\varepsilon_i|\theta, h_i)$ around some properly chosen $\tilde{h}_i^{(0)}$:

$$\log p(\varepsilon_{it}|\theta, h_{it}) \approx \log p(\varepsilon_{it}|\theta, \tilde{h}_{it}^{(0)}) + b_{it}h_{it} - \frac{1}{2}c_{it}h_{it}^2, \tag{3.4}$$

where b_{it} and c_{it} depend on $\tilde{h}_{it}^{(0)}$. Based on the linearized kernel, an approximate smoothing distribution $\pi_G(h_i|\theta,\varepsilon_i)$ takes the form of a Normal distribution with precision matrix $\bar{Q}_i = Q_i + C_i$ and mean $\bar{\delta}_i = \bar{Q}_i^{-1}(b_i + Q_i\delta_i)$, where $C_i = \text{diag}([c_{i1},\ldots,c_{iT}]')$ and $b_i = [b_{i1},\ldots,b_{iT}]'$. The T-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 b_{it} and c_{it} are given in Appendix B.1.

As discussed in section 2, our prior density for h_i is subject to the normalizing constraint $A_h h_i = \mu_i$. Therefore, the IS density $\pi_G(h_i|\theta,\varepsilon_i)$ needs a slight modification to account for this linear constraint. In particular, an application of Bayes' theorem yields a constraint density $\pi_G^c(h_i|\theta,\varepsilon_i)$ which is also Gaussian but has mean and covariance:

$$\bar{\delta}_i^c = \bar{\delta}_i - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}(A_h\bar{\delta}_i - \mu_i), \tag{3.5}$$

$$\text{Cov}(h_i|\theta,\varepsilon_i, A_h h_i{=}\mu_i) = \bar{Q}_i^{-1} - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}A_h\bar{Q}_i^{-1}. \tag{3.6}$$

Note that imposing the linear restriction yields a non-sparse precision and a reduced rank covariance which impedes direct efficient sampling and density evaluation. Following Rue et al. (2009), sampling and evaluation of $\pi_G^c(h_i|\theta,\varepsilon_i)$ can still be implemented at trivial extra costs by what is known as "conditioning by kriging". Specifically, a random sample $\tilde{h}_i^{(j)}$ is first generated from $\pi_G(h_i|\theta,\varepsilon_i)$, exploiting the sparse precision $\bar{Q}_i^{-1}$. In a second step, the draw is corrected for the linear constraint by setting $h_i^{(j)} = \tilde{h}_i^{(j)} - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}(A_h\tilde{h}_i^{(j)} - \mu_i)$. Also evaluation of the adjusted IS density can be achieved efficiently by applying Bayes' Theorem:

$$\pi_G^c(h_i|\theta,\varepsilon_i) = \frac{\pi_G(h_i|\theta,\varepsilon_i)\pi(A_h h_i|h_i)}{\pi_1(A_h h_i)}, \tag{3.7}$$

where $\log\pi(A_h h_i|h_i) = -\frac{1}{2}\log|A_h A_h'|$ and $\pi_1(A_h h_i) \sim \mathcal{N}(A_h\bar{\delta}_i, A_h\bar{Q}_i^{-1}A_h')$. Note that the same routine can be used to evaluate the prior density $p^c(h_i|\theta)$ which displays the same constraint. That is, the constraint prior density is evaluated as follows:

$$p^c(h_i|\theta) = \frac{p(h_i|\theta)\pi(A_h h_i|h_i)}{\pi_2(A_h h_i)} \tag{3.8}$$

where $p(h_i|\theta) \sim \mathcal{N}(\delta_i, Q_i)$, $\pi_2(A_h h_i) \sim \mathcal{N}(A_h\delta_i, A_h Q_i^{-1}A_h')$ and $\pi(A_h h_i|h_i)$ is as above.

Finally, we recommend to assess the quality of the 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) that has to be estimated this is not clear a-priori, we advise to test for the existence of the variance using e.g. the test of Koopman, Shephard & Creal (2009). However, 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 first introduced by Dempster, Laird & Rubin (1977). The EM procedure is particularly suitable for maximization problems under the presence of hidden variables. In our setting, the hidden variables are the set of r log variances denoted by $h = [h_1, \ldots, h_r]$. Our goal is to maximize:

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

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 $\mathcal{L}(\theta)$ can be obtained by an application of Jensen's inequality:

$$\begin{aligned}
\mathcal{L}(\theta) &= \log \int p(y|\theta, h)p(h|\theta)dh && (3.9)\\
&= \log \int \frac{p(y|\theta, h)p(h|\theta)}{\tilde{p}(h)}\tilde{p}(h)dh && (3.10)\\
&\geq \int \log\left(\frac{p(y|\theta, h)p(h|\theta)}{\tilde{p}(h)}\right)\tilde{p}(h)dh && (3.11)\\
&= \int \log\left(p(y|\theta, h)p(h|\theta)\right)\tilde{p}(h)dh - \int \log\left(\tilde{p}(h)\right)\tilde{p}(h)dh && (3.12)\\
&=: F(\tilde{p}, \theta). && (3.13)
\end{aligned}$$

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

$$\begin{aligned}
&\text{E-step:} \quad \tilde{p}^{(l)} = \arg\max_{\tilde{p}} F(\tilde{p}, \theta^{(l-1)}), && (3.14)\\
&\text{M-step:} \quad \theta^{(l)} = \arg\max_{\theta} F(\tilde{p}^{(l)}, \theta). && (3.15)
\end{aligned}$$

Under mild regularity conditions the EM algorithm converges reliably towards a local optimum.[3] It is easy to show that the E-step in (3.14) 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 (3.11) 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 (3.15) is given by maximizing the criterion function:

$$\begin{aligned}
Q(\theta; \theta^{(l-1)}) &= \int \log\left(p(y|\theta, h)p(h|\theta)\right)\tilde{p}^{(l)}(h)dh && (3.16)\\
&= \mathrm{E}_{\theta^{(l-1)}}\left(\mathcal{L}_c(\theta)\right), && (3.17)
\end{aligned}$$

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

[3] For details on convergence, we refer to the textbook treatment in McLachlan & Krishnan (2007).

For the SV-SVAR model, the complete data log-likelihood is rather simple and we refer to Appendix B.3 for an explicit expression. It follows that for a given choice of $\tilde{p}^{(l)}$, computing the M-Step is straightforward. However, since the smoothing distribution in SV models is generally 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, one based on an analytical approximation and the other based on Monte Carlo integration. In the following, we use that independence among the structural errors implies that the smoothing distribution can be factored as: $p(h|\theta^{(l-1)}, y) = \prod_{i=1}^{r} p(h_i|\theta^{(l-1)}, y)$.

3.2.1 Analytical Approximation

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

$$\tilde{p}^{(l)}(h) = \prod_{i=1}^{r} \pi_G^c(h_i|\theta^{(l-1)}, \varepsilon_i), \tag{3.18}$$

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 distributions in the EM algorithm to get monotonic increases in the log-likelihood function $\mathcal{L}(\theta)$. In fact, it can be shown that $F(\tilde{p}, \theta) = \mathcal{L}(\theta) - D_{KL}\left(\tilde{p}(h)||p(h|y, \theta)\right)$ 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 $\mathcal{L}(\theta)$. In the following, we refer to this algorithm as EM-1 and provide details in Appendix B.3.

3.2.2 Monte Carlo Approximation

The second approach is based on Markov Chain Monte Carlo (MCMC) integration and draws on the results of Kim et al. (1998).[4] The idea is to consider the linearized state space representation of the r independent SV equations:

$$\log(\varepsilon_{it}^2) = h_{it} + \log(\eta_{it}^2), \tag{3.19}$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sqrt{s_i}\omega_{it}, \tag{3.20}$$

[4]See also Mahieu & Schotman (1998) for a similar Monte Carlo EM algorithm to estimate a univariate SV model.

where $\eta_{it} \sim N(0,1)$ and $\omega_{it} \sim N(0,1)$. Kim et al. (1998) propose to closely approximate the log-χ^2 error distribution in (3.19) by a mixture of seven normals. In particular, they specify:

$$p(\log(\eta_t^2)|z_{it} = k) \sim \mathcal{N}(\log(\varepsilon_{it}^2); m_k, v_k^2), \tag{3.21}$$

$$p(z_{it} = k) = p_k, \tag{3.22}$$

with mixture parameters p_k, m_k, v_k^2 for $k = 1, \ldots, 7$ tabulated in Appendix B.3. The advantage of representing the transformed measurement error with a normal mixture is that conditional on a realization of the indicators $z_i = [z_{i1}, \ldots, z_{iT}]'$, the state space model is both, linear and Gaussian which allows for closed form computations of $p(h_{it}|\theta, z_{it}, y)$ by Kalman smoothing recursions.

We exploit this property in our Monte Carlo EM algorithm in the following way. First, consider the mixture representation of the intractable smoothing distribution:

$$p(h|\theta^{(l-1)}, y) \approx \int p(h|\theta^{(l-1)}, z, y) p(z|\theta^{(l-1)}, y) dz.$$

Using this distribution in the EM algorithm yields the following objective function in the M-step:

$$Q(\theta; \theta^{(l-1)}) \approx \int \int \log\left[p(y|\theta, h) p(h|\theta)\right] p(h|\theta^{(l-1)}, z, y) p(z|\theta^{(l-1)}, y) dz dh.$$

To approximatively solve this high-dimensional integral, we simulate a large number of mixture indicators z from $p(z|\theta^{(l-1)}, y)$ by MCMC methods and consider the Monte Carlo counterpart:

$$Q(\theta, \theta^{(l-1)}) \approx \frac{1}{R} \sum_{j=1}^{R} \mathrm{E}_{\theta^{(l-1)}}^{(j)}[\mathcal{L}(\theta)],$$

where the expectation is now taken with respect to the tractable Gaussian distribution $p(h|\theta^{(l-1)}, z^{(j)}, y)$ which can be computed by Kalman smoothing recursions.

In order to generate random draws of the mixture indicators we follow the MCMC scheme of Kim et al. (1998) which involves iteratively drawing from the conditional distributions $p(h_i|\theta^{(l-1)}, z_i, y)$ and $p(z_i|\theta^{(l-1)}, h_i, y)$. For computational reasons we rely on the precision sampler of Chan & Jeliazkov (2009) which exploits the sparsity in the precision matrix. Furthermore, it allows for a straightforward extension to implement the linear normalizing constraint on h_i. In the remainder, we call the Monte Carlo based algorithm EM-2 and for details on the MCMC algorithm and respective M-steps, we refer to Appendix B.3.

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, Moulines & Ryden

(2005). Let $\hat{\theta}$ denote the ML estimator, under appropriate regularity conditions, $\hat{\theta}$ is consistent and asymptotically normally distributed:

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1}), \tag{3.23}$$

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 as:

$$\widehat{\mathcal{I}(\theta)} = T^{-1}\mathcal{J}(\hat{\theta}) \tag{3.24}$$

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.

To compute estimator (3.24) in algorithm EM-1, note that we can evaluate an approximate log-likelihood in closed form based on the Gaussian approximation which we rely on in the E-step. In particular, based on Bayes' Theorem:

$$\log p(\varepsilon_i|\theta) \approx \log p(\varepsilon_i|\theta, h_i) + \log p(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i), \tag{3.25}$$

which can be evaluated for any h_i. For convenience, the r likelihoods for the heteroskedastic structural shocks are evaluated at the mean $h_i = \bar{\delta}_i^c$, such that the exponential term in $\pi_G^c(h_i|\theta, \varepsilon_i)$ drops out. Therefore, based on (3.25) an approximate complete log-likelihood is given as:

$$\mathcal{L}_a(\theta) = -T \log |B| + \sum_{i=1}^{r} \left[\log p(\varepsilon_i|\theta, h_i) + \log p(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i)\right] + \sum_{i=r+1}^{K} \log p(\varepsilon_i|\theta).$$

We take the second derivative of this approximation with respect to the parameter vector θ using numerical differentiation to obtain an approximation of the observed information matrix $\mathcal{J}_1(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}_a(\theta)}{\partial \theta' \partial \theta}\big|_{\theta=\hat{\theta}}$.

For the Monte Carlo based algorithm EM-2, no closed form approximation of the likelihood is available which makes the computation of the information matrix estimator more involved. We apply Louis Identity (Louis; 1982) to the observed information matrix:

$$\mathcal{J}_2(\hat{\theta}) = \mathrm{E}\left[\mathcal{J}_c(\hat{\theta})|y\right] - \mathrm{Cov}(S_c(\hat{\theta})|y), \tag{3.26}$$

where $\mathcal{J}_c(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \theta \partial \theta'}\big|_{\theta=\hat{\theta}}$, $S_c(\hat{\theta}) = \frac{\partial \mathcal{L}_c(\theta)}{\partial \theta}\big|_{\theta=\hat{\theta}}$ are the observed information matrix and score of the complete data log-likelihood $\mathcal{L}_c$. The integrals necessary to compute expected value and variance are with respect to the smoothing distribution at the ML estimator $p(h|\hat{\theta}, y)$ which is intractable for the SV model. However, based on simulated values of the mixture indicators

$z^{(j)}(j = 1, \ldots, R)$, Monte Carlo integration is feasible with:

$$\begin{aligned}
\mathrm{E}\left[\mathcal{J}_c(\hat{\theta})|y\right] \approx & \frac{1}{R}\sum_{j=1}^{R} -\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial\theta\partial\theta'} \;\middle|\; z^{(j)}, y\right]_{\theta=\hat{\theta}}, \\
\mathrm{Cov}(S_c(\hat{\theta})) \approx & \frac{1}{R}\sum_{j=1}^{R} \mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial\theta}\frac{\partial \mathcal{L}_c(\theta)}{\partial\theta'} \;\middle|\; z^{(j)}, y\right]_{\theta=\hat{\theta}},
\end{aligned}$$

where the second approximation holds since $\mathrm{E}(S_c(\hat{\theta})|y) = 0$. The integrals required to compute the expected values are with respect to the tractable Gaussian distributions $p(h|\hat{\theta}, z^{(j)}, y)$. The derivatives necessary to apply the Louis Method are available in closed form and given in Appendix B.4.

3.4 Inference on Structural Impulse Response Functions

Identification of the SVAR model is ultimately useful to conduct structural analysis. Since Impulse Response Functions (IRFs) are likely to be the most widely used tool for that purpose, we quickly outline how to conduct inference on these quantities with our model.

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

$$y_t = \mu_y + \sum_{j=0}^{\infty} \Phi_j B \varepsilon_t,$$

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 $\Phi_j \in \mathbb{R}^{K\times K}$ $(j = 0, 1, \ldots)$ is a sequence of exponentially decaying matrices given as: $\Phi_j = J\mathbf{A}^j J'$ 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 $\alpha = \mathrm{vec}(A)$ with $A =$

$[A_1, \ldots, A_p]$, $\beta = \text{vec}(B)$ 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, Jentsch & Trenkler (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 $G_i = \frac{\partial \text{vec}(\Phi_i)}{\partial \alpha'} = \sum_{mj=0}^{i-1} [J(\mathbf{A}')^{i-1-j}] \otimes \Phi_j$ for $i \geq 1$. Finally, $C_{i,\beta} = \frac{\partial \text{vec}(\Theta_i)}{\partial \beta'} = (I_K \otimes \Phi_i)$ 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 $P_n = (B' \otimes I_K)F_n$, $F_0 = 0$, $F_n = G_1 + \cdots + G_n$, $\bar{P}_n = (I_K \otimes \Psi_n)$ and $\Psi_n = \sum_{i=0}^{n} \Phi_i$.

3.5 Testing for Identification

For valid likelihood inference on the structural parameters including the impact matrix B, the model must be identified. As highlighted in section 2, at most one component in ε is allowed to be homoskedastic if the model is to be identified solely by heteroskedasticity. To determine the number of heteroskedastic shocks in a given application, we recommend to follow a procedure considered by Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016) within SVAR-GARCH models. The idea is to conduct the following sequence of tests:

$$H_0 : r = r_0 \qquad \text{vs} \qquad H_1 : r > r_0, \tag{3.27}$$

for $r_0 = 0, \ldots, K-1$. If all null hypotheses up to $r_0 = K-2$ can be rejected, there is evidence for sufficient heteroskedasticity in the data to fully identify B.

The testing problem given in (3.27) 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. In particular, 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 homeskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Note that if the null is true ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is white noise. To test for remaining heteroskedasticity in ε_{2t}, Lanne & Saikkonen (2007) propose to use Portmanteau types of statistics on the second moment of ε_{2t}. In particular, they construct the following time series:

$$\xi_t = \varepsilon_{2t}'\varepsilon_{2t} - T^{-1}\sum_{t=1}^{T}\varepsilon_{2t}'\varepsilon_{2t}, \tag{3.28}$$

$$\vartheta_t = \operatorname{vech}(\varepsilon_{2t}\varepsilon_{2t}') - T^{-1}\sum_{t=1}^{T}\operatorname{vech}(\varepsilon_{2t}\varepsilon_{2t}'), \tag{3.29}$$

with $\operatorname{vech}(\cdot)$ 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:

$$Q_1(H) = T\sum_{h=1}^{H}\left(\frac{\tilde{\gamma}(h)}{\tilde{\gamma}(0)}\right)^2, \tag{3.30}$$

$$Q_2(H) = T\sum_{h=1}^{H}\operatorname{tr}\left[\tilde{\Gamma}(h)'\tilde{\Gamma}(0)^{-1}\tilde{\Gamma}(h)\tilde{\Gamma}(0)^{-1}\right], \tag{3.31}$$

where $\tilde{\gamma}(h) = T^{-1}\sum_{t=h+1}^{T}\xi_t\xi_{t-h}$ 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 $Q_2(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$.

To apply these tests, we must be able to estimate the model under H_0 which requires additional restrictions on B if $r_0 < K-1$. To uniquely disentangle the shocks in ε_{2t}, it turns out that 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 $B = \begin{pmatrix} B_{11} & B_{21} \\ B_{12} & B_{22} \end{pmatrix}$, $B_{11} \in \mathbb{R}^{r\times r}$, $B_{12} \in \mathbb{R}^{(K-r)\times r}$, $B_{21} \in \mathbb{R}^{r\times(K-r)}$ and $B_{22} \in \mathbb{R}^{(K-r)\times(K-r)}$. Let B_{22} be restricted to be 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 Appendix A.4.

We conclude with a remark regarding the small sample properties of the tests. Based on extensive simulation studies, Lütkepohl & Milunovich (2016) find a substantial lack in power

for 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.

4 Monte Carlo Study

An important question for practitioners is how a heteroskedastic SVAR model performs in estimating structural parameters under inherent misspecification of the variance process. To shed some light on this question, we conduct a small scale 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).

Our analysis involves generating a large number of datasets 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 datasets, closely resembling the MC design of Lütkepohl & Schlaak (2018).[5] 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 $t = 1, \dots, T$ 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 $t = 1, \dots T/2$ and $\Lambda_t = \text{diag}([2, 7]')$ for $t = T/2 + 1, \dots T$.

2. **MS(2)-SVAR**: The specified MS-SVAR involves a switching variance with the same

[5] Some difference to their design comes from our choice of the impact matrix. In particular, we use what we think are more realistic values of the impact matrix in a sense that they lead to less dramatic changes in the VAR error variance.

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 $\beta_i = 0.8$ $(i = 1, 2)$ which correspond to values typically estimated for empirical data.

4. **SV-SVAR**: For this DGP, $\Lambda_t = \text{diag}([\exp(h_{1t}), \exp(h_{2t})]')$ with:

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

where $\omega_{it} \sim \mathcal{N}(0, 1)$. We set $\mu_i = -0.5 s_i / (1 - \phi_i^2)$ such that $\text{E}(\varepsilon_{it}^2) = 1$. Furthermore, 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 driven by issues regarding to weak identification, we only accept datasets 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 datasets with an empirical kurtosis of the simulated structural shocks of at least 3.6 are accepted.

A total of M=1000 datasets 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 dataset. 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. Furthermore, 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). 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

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	1.61	1.79	1.58	1.14	1.20	1.24	1.19	1.04
	SV	1.22	1.32	1.21	1.06	1.09	1.11	1.09	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
	SV	1.74	1.94	1.54	1.08	1.23	1.30	1.29	1.08
GARCH-DGP	BP	3.88	4.23	2.56	1.26	11.58	12.67	4.99	1.47
	MS(2)	8.18	9.01	3.67	1.29	21.71	24.52	7.14	1.38
	MS(3)	3.95	4.23	1.98	1.13	5.19	5.60	2.22	1.19
	GARCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SV	1.15	1.16	1.04	1.03	1.10	1.10	1.06	1.04
SV-DGP	BP	3.35	3.53	2.26	1.18	8.52	9.52	4.36	1.35
	MS(2)	5.62	6.10	3.28	1.19	13.60	15.22	5.72	1.30
	MS(3)	4.20	4.58	2.02	1.12	3.12	3.34	1.74	1.14
	GARCH	2.41	2.60	1.77	1.15	1.50	1.54	1.23	1.07
	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.

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 register in terms of MSE is found to be 94% in θ_{12} 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 MSE 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. Admittably, 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. While indeed this yields substantial improvements, we still register detoriations in MSE up to 460%.

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 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 safe choice to identify structural shocks for different types of heteroskedasticity patterns and to estimate the corresponding impulse response functions.

5 Interdependence between Monetary Policy and Stock Markets

SVAR models are a widely used tool to investigate the dynamic effects of monetary policy, see e.g. Ramey (2016) for an extensive overview of the literature. To identify the structural shocks, the most simple way uses a Cholesky decomposition of the covariance matrix in a reduced form VAR with the policy variable ordered last (Christiano, Eichenbaum & Evans; 1999; Bernanke, Boivin & Eliasz; 2005). In accordance with theoretical economic models featuring nominal rigidities (Christiano, Eichenbaum & Evans; 2005), this implies that only the central bank is allowed to respond to all movements in the economy on impact, while all variables in the system ordered above react with at least one lag to a monetary policy shock. While this seems reasonable for slowly moving real macroeconomic aggregates, such a recursivity assumption becomes unrealistic once fast moving financial variables are included into the SVAR analysis.

Over the last years, many other identification schemes have been developed to study the effects of monetary policy shocks avoiding the use of a recursiveness assumption. Bjørnland & Leitemo (2009) propose to identify a monetary policy shock under the presence of stock market returns by a combination of short- and long-run restrictions. Besides zero impact restrictions on real variables, a monetary policy shock is furthermore restricted to have a zero long-term impact on stock markets. This additional restriction allows the authors to disentangle monetary policy innovations from financial shocks.

Another promising way to address identification in presence of fast moving variables are Proxy SVARs based on external instruments. If there is an external time series that is correlated with the structural shock to be identified and uncorrelated with all other shocks in the system, no exclusion restrictions are necessary at all. Recently, many narrative measures have been proposed to identify monetary policy shocks. Widely used are proxies constructed based on either readings of Federal Open Market Committee (FOMC) minutes (e.g. Romer & Romer (2004); Coibion (2012)) or changes in high frequency future prices in a narrow window around FOMC meetings (e.g Faust, Swanson & Wright (2004); Nakamura & Steinsson (2018); Gertler & Karadi (2015)).[6]

[6]Yet another branch of the literature relies on sign restrictions of the impulse response functions (Faust; 1998; Canova & De Nicolo; 2002; Uhlig; 2005) or on a combination of sign restrictions and information in proxy variables (Braun & Brüggemann; 2017).

Finally, heteroskedasticity can be exploited to identify the interdependence between monetary policy and financial variables. For example, Rigobon (2003) combines identification via heteroskedasticity and economic narratives to estimate the reaction of monetary policy to stock market returns. Also Wright (2012) links economic and statistical identification within a daily SVAR, assuming that monetary policy shocks have a higher variance around FOMC meetings. Even if no economic narrative is available for the statistically identified structural parameters, the heteroskedastic SVAR model can be used to formally test conventional identifying restrictions. For example, Lütkepohl & Netšunajev (2017a) review various heteroskedastic SVAR models and use them to test the combination of exclusion restrictions employed by Bjørnland & Leitemo (2009).[7] Their analysis includes a GARCH-SVAR, two specifications of a MS-SVAR and a SVAR featuring a Smooth Transition model for the variance (STVAR).

To illustrate the use of our methods, we repeat the analysis of Lütkepohl & Netšunajev (2017a) complemented by the SV-SVAR model. Besides testing the short- and long-run restrictions used by Bjørnland & Leitemo (2009), we additionally test Proxy SVAR restrictions that arise if the narrative series of Romer & Romer (2004) and Gertler & Karadi (2015) are used as instruments for a monetary policy shock.

5.1 Model and Identifying Constraints

The VAR model of Bjørnland & Leitemo (2009) is based on the following variables: $y_t = (q_t, \pi_t, c_t, \Delta s_t, r_t)'$, where q_t is a linearly detrended index of log industrial production, π_t the annualized inflation rate based on consumer prices, c_t the annualized change in log commodity prices as measured by the World Bank, Δs_t S&P500 real stock returns and r_t the federal funds rate. For detailed description of the data sources, transformations and time series plots see Appendix C. As in Lütkepohl & Netšunajev (2017a), we use an extended sample period including data from 1970M1 until 2007M6, summing up to a total of 450 observations. To make our results comparable, we also choose $p = 3$ lags which is supported by the AIC applied within a linear VAR model.

In our analysis, we test the following set of short- and long-run constraints used by Bjørnland & Leitemo (2009):

$$B = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \quad \text{and} \quad \Xi_\infty = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{bmatrix}, \tag{5.1}$$

where $\Xi_\infty = (I_K - A_1 - \ldots - A_p)^{-1}B$ is the long-run impact matrix of the structural shocks on y_t. Note than an asterisk means that the corresponding entry in B and Ξ_∞ is left unrestricted. The

[7] See also Lütkepohl & Netšunajev (2017b) for a similar analysis based on a Smooth Transition SVAR model only.

last columns of B and Ξ_∞ correspond to the reaction of y_t to a monetary policy shock. Economic activity, consumer- and commodity prices are only allowed to respond with a delay of one month to a monetary policy shock, while stock markets are allowed to react contemporaneously. However, in the long run, a monetary policy shock is assumed to have a zero effect on the stock market. The fourth column of B corresponds to a stock price shock which is constrained to have no contemporaneous impact on activity and prices while the central bank is allowed to adjust the interest rates within the same period. The remaining shocks do not have an economic interpretation. To identify the model, Bjørnland & Leitemo (2009) simply disentangle these shocks by imposing a recursivity assumption. As outlined before, restrictions (5.1) are overidentifying in heteroskedastic SVAR models and can be tested against the data. In line with Lütkepohl & Netšunajev (2017a), the following set of restrictions is tested:

R1: Both, B and Ξ_∞ restricted as in (5.1).

R2: Only the last two columns of B and Ξ_∞ are restricted as in (5.1).

R3: Only B is restricted as in (5.1).

We further contribute to the literature by testing Proxy SVAR restrictions that arise if an external instrument z is used for identification of a structural shock. The identifying assumptions are that the instrument 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 the relevance and exogeneity assumption can be translated into the following set of linear restrictions on β_1, denoting the first column of B:

$$\beta_{21} = (\Sigma_{zu_1'}^{-1}\Sigma_{zu_2'})'\beta_{11}. \tag{5.2}$$

where $\beta_1 = [\beta_{11}, \beta_{21}']'$ with β_{11} scalar and $\beta_{21} \in \mathbb{R}^{K-1}$. Furthermore, $\Sigma_{zu'} = \text{Cov}(z, u') = [\Sigma_{zu_1'}, \Sigma_{zu_2'}]$ with $\Sigma_{zu_1'}$ scalar and $\Sigma_{zu_2'}' \in \mathbb{R}^{K-1}$. In practice, elements of $\Sigma_{zu'}$ are estimated by the corresponding sample moments.[8] To identify a monetary policy shock, we use the narrative series constructed by Romer & Romer (2004) (RR henceforth) and Gertler & Karadi (2015) (GK henceforth). We test the following Proxy SVAR restrictions that arise when the first column of B is identified via either RR's or GK's instrument:

R4rr: IV moment restrictions (5.2) based on the RR shock.

R4gk: IV moment restrictions (5.2) based on the GK shock.

We use the RR series extended by Wieland & Yang (2016) which is available for the whole sample. The GK shock is only available for a subsample starting in 1990M1. We use their

[8] In particular, at each M-step we compute $\hat{\Sigma}_{zu'} = N_z^{-1}\sum_{t=1}^{T} D_t\hat{u}_t z_t'$ where D_t is a dummy indicating whether the instrument is available at time t and $N_z = \sum_{t=1}^{T} D_t$.

Table 2: Model Selection by Information Criteria

	Linear	SV-EM1	SV-EM2	GARCH	STVAR	MS(2)	MS(3)
$\ln L$	-3159.3	-2680.4	-2677.9	-2763.6	-2878.3	-2827.4	-2775.3
AIC	6508.7	5590.9	5585.8	5757.2	5980.5	5878.8	5792.6
BIC	6898.4	6062.6	6057.6	6229.0	6440.0	6338.3	6289.0

Note: $\ln L$ - log-likelihood function, 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 SV-EM1 and SV-EM2, application of the batch means method yields approximate 95%-confidence intervals of [-2680.48,-2680.33] and [-2678.11,-2677.68], respectively.

baseline series which is constructed based on the three months ahead monthly fed funds futures.[9] Time series plots of both series are available in Appendix C.

5.2 Statistical Analysis

Before we start testing the aforementioned restrictions, we conduct formal model selection for the variance specification of the structural shocks. By means of information criteria and residual plots, we compare the SV model to those models included in Lütkepohl & Netšunajev (2017a): a GARCH, a Smooth Transition (ST) and different specifications of a Markov Switching model. This allows us to directly compare our results.

Table 2 reports log-likelihood values, Akaike information criteria (AIC) and Bayesian information criteria (BIC) for a linear VAR and all heteroskedastic models. First of all, we highlight that there is only a small gain in terms of likelihood value of the SV model using the Monte Carlo based algorithm (EM-2) compared to the deterministic approximation (EM-1). To assess the Monte Carlo error of the estimates, we also report approximate 95%-confidence intervals based on an application of the batch means method and R =100,000 draws of the importance density.[10] Comparing the different models, our results suggest that including time-variation in the second moment is strongly supported by both information criteria. Moreover, among the heteroskedastic models we find that particularly models designed for financial variables are favored, that is the GARCH model and the SV model. This may be not surprising given that stock market returns are included in the system.

Among all models considered, we find that the SV model performs best in terms of information criteria. In this regard, our results deviate from those of Lütkepohl & Netšunajev (2017a) who find that the MS(3) model provides the best description for this dataset.[11]

[9] We repeat our analysis for the other instruments available in Gertler & Karadi (2015). The results do not change qualitatively.

[10] A formal test of Koopman et al. (2009) indicates that the variance of the importance weights is finite which further supports the validity of our likelihood estimates.

[11] We also find a better ranking for the GARCH model compared to MS(3). Most likely, this is caused by a different estimation procedure. Specifically, Lütkepohl & Netšunajev (2017a) do only approximatively maximize the likelihood by a sequential estimation procedure.

Table 3: Tests of Identification in SV-SVAR Model

	$Q_1(1)$	dof	p-value	$Q_2(1)$	dof	p-value
$r_0 = 0$	15.02	1	0.00	596.60	225	0.00
$r_0 = 1$	23.89	1	0.00	250.09	100	0.00
$r_0 = 2$	29.53	1	0.00	141.07	36	0.00
$r_0 = 3$	18.43	1	0.00	43.70	9	0.00
$r_0 = 4$	17.30	1	0.00	17.30	1	0.00
	$Q_1(3)$	dof	p-value	$Q_2(3)$	dof	p-value
$r_0 = 0$	52.34	3	0.00	1433.70	675	0.00
$r_0 = 1$	39.75	3	0.00	528.72	300	0.00
$r_0 = 2$	32.85	3	0.00	222.01	108	0.00
$r_0 = 3$	20.36	3	0.00	60.93	27	0.00
$r_0 = 4$	19.86	3	0.00	19.86	3	0.00

Note: Sequence of tests to check the number of heteroskedastic shocks in the system as introduced in section 3.5 (Lanne & Saikkonen; 2007).

In accordance with Lütkepohl & Netšunajev (2017a), we also consider standardized residuals as an additional model checking device. Figure 1 provides a plot for the standardized residuals of all models computed as $\hat{u}_{it}/\hat{\sigma}_{ii,t}$ where $\hat{\sigma}^2_{ii,t}$ is the i-th diagonal entry of the estimated VAR covariance matrix $\hat{\Sigma}_t$. These plots clearly suggest that none of the other methods is fully satisfactory in yielding standardized residuals that seem to be homoskedastic and approximately normally distributed. However, for the SV-SVAR model, standardized residuals seem well behaved with no apparent heteroskedasticity and virtually no outliers. To confirm this impression, we provide complementary test results in Appendix C.1 concerned with remaining heteroskedasticity and non-normality in standardized structural shocks. We find that only for the shocks of the SV-SVAR model, there is no evidence against both normality and homoskedasticity. To conclude, statistical analysis suggests that the proposed SV-SVAR is the most adequate for this application and we continue our analysis based on this model.

In order to test restrictions R1-R4 as overidentifying, it is necessary to count with enough heteroskedastic shocks ($r \geq K-1$) to fully identify the impact matrix B. As described in section 3.5, 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 $r = K$ in our model, implying that the model can be fully identified by heteroskedasticity.

We continue our analysis and test the economically motivated restrictions R1-R4 as overidentifying. In Table 4 we provide Likelihood Ratio (LR) test statistics for the restrictions introduced previously.[12] Note that if B is identified under H_0, they have a standard asymptotic $\chi^2(n_r)$-distribution with n_r being the number of restrictions tested. Since we estimate the likelihood values with the help of importance sampling, we account for the Monte Carlo error

[12]This table is based on parameter estimates provided by EM-1. A corresponding Table based on EM-2 can be found in Appendix C.1 and does not differ qualitatively.

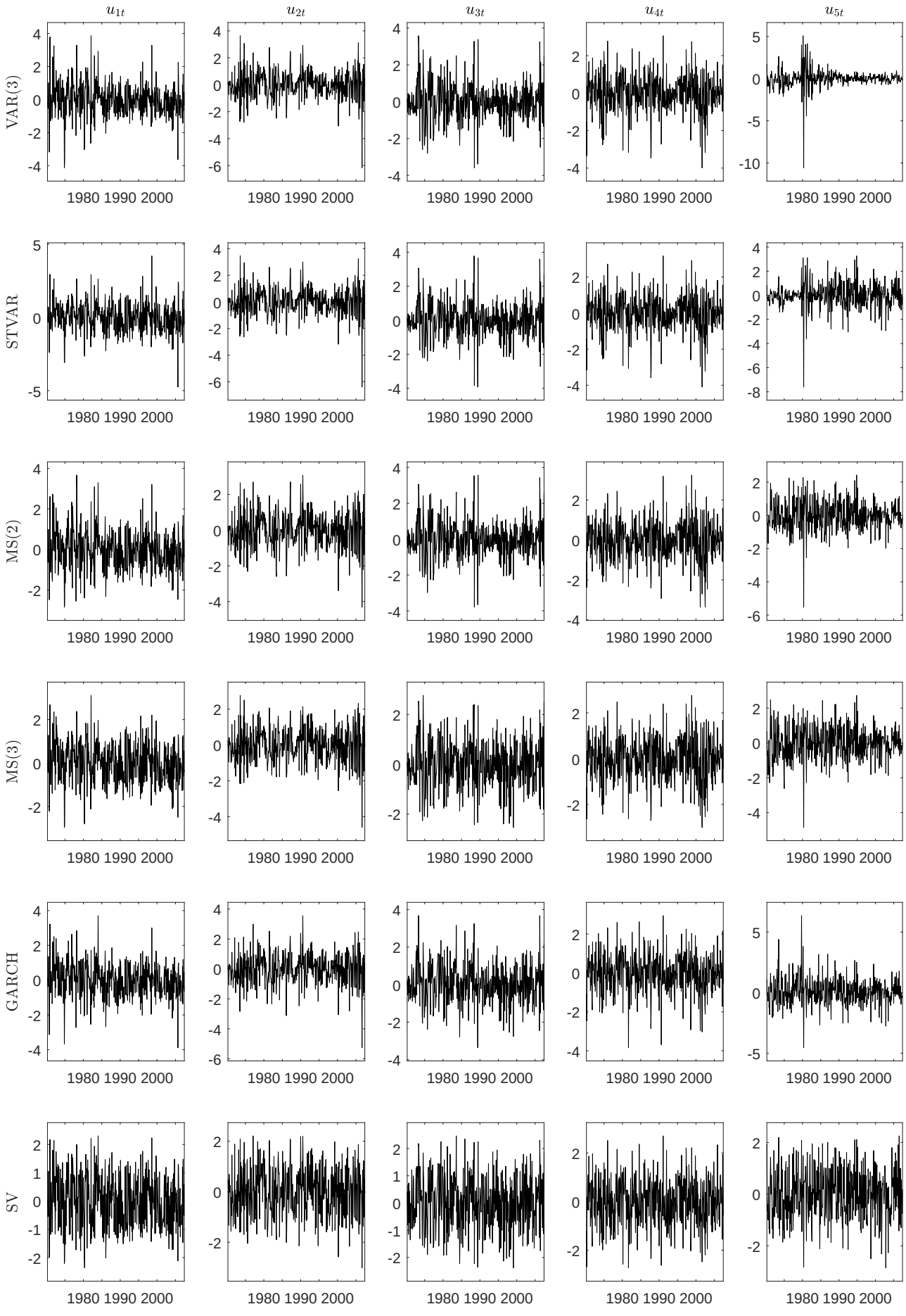

Figure 1: Standardized residuals of linear, ST-, MS(2)-, MS(3)-, GARCH- and SV-SVAR model.

Table 4: Test for Overidentifying Restrictions (EM-1)

H_0	H_1	LR	dof	p-value	$p_{.025}$	$p_{.975}$
R1	UC	25.649	10	0.0042	0.0039	0.0046
R2	UC	22.750	7	0.0019	0.0017	0.0020
R3	UC	24.004	9	0.0043	0.0040	0.0046
R1	R3	1.653	1	0.1986	0.1957	0.2016
R4rr	UC	7.169	4	0.1272	0.0943	0.1705
R4gk	UC	256.480	4	0.0000	0.0000	0.0000

Note: For details about overidentifying restrictions see subsection 5.1. Likelihood ratio test statistics are computed as $2\left(\ln L_{H_1} - \ln L_{H_0}\right)$ and are approximatively χ^2-distributed under H_0. Right columns report approximate 95%-confidence intervals for the p-value resulting from an application of the batch means method to the LR test statistic.

by applying the batch means method and reporting approximate 95%-confidence intervals for the p-values.

In line with the findings of Lütkepohl & Netšunajev (2017a), our results suggest that R1, the restrictions of Bjørnland & Leitemo (2009), are rejected by the data. To make sure that this result does not come from the lower triangular block corresponding to the economically meaningless shocks, Lütkepohl & Netšunajev (2017a) also propose to test R2, which are the restrictions in B corresponding to the impact of monetary policy and stock market shocks. Within the SV model, these restrictions are also rejected. Testing for the zero restrictions in B in isolation (R3) also results in a rejection. However, in contrast to Lütkepohl & Netšunajev (2017a), we find that the long-run restriction is not rejected at any conventional significance level if R1 is tested against R3. This indicates that the long-run restriction is less of a problem, but rather are those in the short run. This key difference in the empirical analysis might arise due to more precisely estimated IRFs by the SV-SVAR model, strongly supported by statistical evidence. The fact that we are able to draw a different empirical conclusion emphasizes the importance of model selection in the context of heteroskedastic SVARs.

With respect to the Proxy SVAR restrictions, we find that identifying a monetary policy shock with the shock series of Gertler & Karadi (2015) is strongly rejected by the data with a likelihood ratio test statistic exceeding 250. In turn, identification via the narrative series of Romer & Romer (2004) cannot be rejected at any conventional significance level. To further understand these results, we compute sample correlations of the instruments z with $\hat{\varepsilon}$, the estimated structural shocks of the unconstrained SV-SVAR model. For GK, we find $\text{Corr}(z^{GK}, \hat{\varepsilon}) = (0.039, -0.067, 0.050, -0.242, 0.419)$, while for RR, $\text{Corr}(z^{RR}, \hat{\varepsilon}) = (0.042, 0.005, 0.031, -0.021, 0.453)$. While both shocks are subject to a strong correlation with one of the statistically identified shocks, the instrument of GK is highly correlated with at least one additional shock. This clearly violates the exogeneity condition on the instrument. Thereby, our results support the argument of Ramey (2016) who questions the exogeneity of

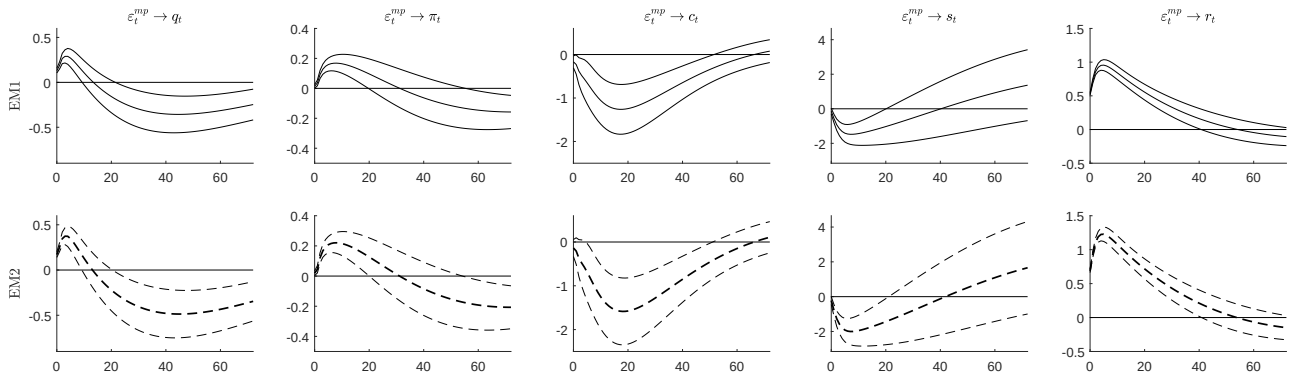

Figure 2: IRFs up to a horizon of 72 months of a monetary policy shock with 68% confidence bounds. Figures compare estimates based on EM-1 (solid line) and EM-2 (dashed line) with corresponding asymptotic confidence intervals.

the GK instrument finding that it is autocorrelated and predictable by Greenbook variables. In turn, for the RR shock we find that there is little correlation with the remaining structural residuals of the SVAR. This clearly explains why identification via the RR shock is not rejected. Since the Proxy SVAR restrictions based on RR cannot be rejected, we can interpret the last shock of the unconstrained model as a monetary policy shock for which $\text{Corr}(z^{RR}, \hat{\varepsilon}_5) = 0.45$. In Figure 2 we plot impulse response functions (IRFs) up to 72 months (6 years) of the system variables in response to a monetary policy shock. Besides point estimates, we provide 68% asymptotic confidence intervals. Again, we note that there is qualitatively no difference in using EM-1 or EM-2 to compute the estimates and corresponding standard errors.[13] The IRFs and their asymptotic confidence intervals coincide for all variables at all horizons. In line with the IRFs computed by Lütkepohl & Netšunajev (2017a) based on other heteroskedastic models, an unexpected tightening in monetary policy is associated with a puzzling short-term increase in activity and prices before they reach negative values on the medium and long term. In turn, commodity prices as well as stock market returns are found to react significantly negative in the short run. This fact seems reasonable given that one would expect a shift in demand towards risk free assets.

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 discussed conditions for full and partial identification and proposed to estimate the model by Gaussian Maximum Likelihood. For this purpose, we developed two EM algorithms which approximate the intractable E-step to a different extent. One algorithm is based on a Laplace

[13]There is only a slight difference in scaling of impulse responses because of a slightly rescaled monetary policy shock in EM-2.

approximation while the other relies on MCMC integration. We leave the choice of algorithm to individual preferences, but find that in pratice little is gained by using the computationally more burdensome Monte Carlo EM. Besides discussing optimization, we stated the main properties of the estimator and present tools to approximate the asymptotic covariance matrix. We also described how inference on Impulse Response Functions can be conducted based on our model. Tests considered by Lanne & Saikkonen (2007) can be used to determine the number of heteroskedastic shocks and to test for identification.

To demonstrate the flexibility of the SV-SVAR model, we conducted 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 have revisited the model of Bjørnland & Leitemo (2009) who rely on a combination of short- and long-run restrictions to disentangle monetary policy from stock market shocks. Formal model selection strongly supports a SV specification in the variance if compared to other heteroskedastic SVARs used by Lütkepohl & Netšunajev (2017a) in this context. The SV-SVAR is used to formally test the exclusion restrictions of Bjørnland & Leitemo (2009) as overidentifying, and additionally test Proxy SVAR restrictions that arise if external instruments are used to identify a monetary policy shock.

Future research in several directions could be pursued. First, a Bootstrap procedure would provide a valuable alternative to summarize estimation uncertainty in the SV-SVAR model. Second, there is a need for more powerful statistical tests of identification in conditional heteroskedastic SVAR models. Furthermore, the impact of weak identification on inference needs to be investigated. Finally, it would be interesting to assess semiparametric methods to identify SVAR models by heteroskedasticity which do not require the specification of a particular variance model.

References

Bernanke, B. S., Boivin, J. & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *Quarterly Journal of Economics* **120**(1): 387–422.

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

Bjørnland, H. C. & Leitemo, K. (2009). Identifying the interdependence between US monetary policy and the stock market, *Journal of Monetary Economics* **56**(2): 275–282.

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

Braun, R. & Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Working Paper 16*, Department of Economics, University of Konstanz.

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

Caffo, B. S., Jank, W. & Jones, G. L. (2005). Ascent-based Monte Carlo expectation– maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2): 235–251.

Canova, F. & De Nicolo, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **49**(6): 1131–1159.

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

Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast, *Journal of Econometrics* **176**(2): 162–172.

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.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1999). Chapter 2: Monetary policy shocks: What have we learned and to what end?, Vol. 1 of *Handbook of Macroeconomics*, Elsevier, pp. 65 – 148.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small?, *American Economic Journal-Macroeconomics* **4**(2): 1–32.

Danielsson, J. & Richard, J. F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* **8**: S153–S173.

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.

Durbin, J. & Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(1): 3–56.

Faust, J. (1998). The robustness of identified VAR conclusions about money, *Carnegie - Rochester Conference Series on Public Policy, Vol 49 - December 1998* **49**: 207–244.

Faust, J., Swanson, E. T. & Wright, J. H. (2004). Identifying VARs based on high frequency futures data, *Journal of Monetary Economics* **51**(6): 1107–1131.

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

Gertler, M. & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal-Macroeconomics* **7**(1): 44–76.

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.

Harvey, A., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**(2): 247–264.

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.

Kilian, L. & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*, Cambridge University Press.

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., 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. (2018). Identifying shocks via time-varying volatility.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society Series B-Methodological* **44**(2): 226–233.

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. (2017a). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**(Supplement C): 2 – 18.

Lütkepohl, H. & Netšunajev, A. (2017b). 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).

Mahieu, R. J. & Schotman, P. C. (1998). An empirical application of stochastic volatility models, *Journal of Applied Econometrics* **13**(4): 333–359.

Maxand, S. (2017). Identification of independent structural shocks in the presence of multiple gaussian components.

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

Melino, A. & Turnbull, S. M. (1990). Pricing foreign-currency options with stochastic volatility, *Journal of Econometrics* **45**(1-2): 239–265.

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.

Milunovich, G. & Yang, M. (2013). On identifying structural var models via arch effects, *Journal of Time Series Econometrics* **5**(2): 117–131.

Montiel-Olea, José, L., Stock, J. H. & Watson, M. W. (2016). Uniform inference in SVARs identified with external instruments, *Harvard Manuscript* .

Nakamura, E. & Steinsson, J. (2018). High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics* (forthcoming).

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.

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

Ramey, V. (2016). Chapter 2 - macroeconomic shocks and their propagation, Vol. 2 of *Handbook of Macroeconomics*, Elsevier, pp. 71 – 162.

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

Romer, C. D. & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.

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

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.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* **63**(1): 289–306.

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

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.

Stock, J. H. & Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 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.

Wieland, J. F. & Yang, M.-J. (2016). Financial dampening, *Working Paper 22141*, National Bureau of Economic Research.

Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound?, *The Economic Journal* **122**(564): F447–F466.

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 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 includes one possible identification scheme for this scenario. We start with the derivation of the autocovariance function of the second moments of reduced form residuals u.

A.1 Autocovariance function in second moment

The autocovariance function of the second moments of the structural shocks is:

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

The entries of this expression are only non-zero if both $i = j = k = l$ and $i \leq r$ hold for $i, j, k, l \in \{1, \ldots, K\}$ due to the structure of the SV-SVAR model (2.1)-(2.4). 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 and M_τ as defined in section 2. Briefly recap that we define (Lewis; 2018):

$$\xi_t = \text{vech}\left(u_t u_t'\right) = L_K \text{vec}\left(u_t u_t'\right).$$

Consequently, the autocovariance function in ξ reads:

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

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 satisfying (2.8) and (2.9). Due to (2.8), it is:

$$\Sigma_u = BB' = \tilde{B}\tilde{B}' = BQQ'B.$$

Hence, Q has to be an orthogonal matrix, i.e. $QQ' = I_K$. To keep the autocovariance function in the second moment of the reduced form errors, it is:

$$\begin{aligned}
\text{Cov}\left(\xi_t, \xi_{t+\tau}\right) &= L_K \left(\tilde{B} \otimes \tilde{B}\right) \text{Cov}\left(\text{vec}\left(\tilde{\varepsilon}_t \tilde{\varepsilon}_t'\right), \text{vec}\left(\tilde{\varepsilon}_{t+\tau} \tilde{\varepsilon}_{t+\tau}'\right)\right) \left(\tilde{B} \otimes \tilde{B}\right)' L_K' \\
&= L_K \left(\tilde{B} \otimes \tilde{B}\right) \left(Q \otimes Q\right)' G_K M_\tau G_K' \left(Q \otimes Q\right) \left(\tilde{B} \otimes \tilde{B}\right)' L_K'.
\end{aligned}$$

As we still have a SV-SVAR model, $\left(Q \otimes Q\right)' G_K M_\tau G_K' \left(Q \otimes Q\right)$ must have the same form as $G_K M_\tau G_K'$, i.e. it is a diagonal matrix with exactly r non-zero entries $\tilde{\gamma}_i(\tau)$ located at elements $(i-1)K + i$ for $i = 1, \ldots, r$ on the diagonal. Thus, it is:

$$G_K \begin{pmatrix} \tilde{\gamma}_1(\tau) & & & \\ & \ddots & & \\ & & \tilde{\gamma}_r(\tau) & \\ & & & 0_{K-r} \end{pmatrix} G_K' = \left(Q \otimes Q\right)' G_K \begin{pmatrix} \gamma_1(\tau) & & & \\ & \ddots & & \\ & & \gamma_r(\tau) & \\ & & & 0_{K-r} \end{pmatrix} G_K' \left(Q \otimes Q\right).$$

This yields the following conditions:

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

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

Because of (A.2), it is $\sum_{l=1}^{r} q'_{l\bullet} \underbrace{q_{li}^2 q_{lj}}_{=:\lambda_{lij}} = 0$ for all $i, j \in \{1, \ldots, K\}$ with $i \neq j$. As it is an orthogonal matrix, row vectors $q_{l\bullet}$ of matrix Q are linearly independent such that $\lambda_{lij} = 0$ for all $l \in \{1, \ldots, r\}, i, j \in \{1, \ldots, K\} : i \neq j$. Consequently, considering the first r rows of Q, i.e. matrix $[Q_1, Q_3]$, only one element per row can be different from zero. Due to the orthogonality, this element has to be ± 1.
Because of (A.1), in each column 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 ± 1 entries in $[Q_1, Q_3]$. This directly implies that Q_3 is a zero matrix and Q_1 has exactly one element different from zero per row and column which is ± 1. Thus, Q_1 can be decomposed in DP where D is a diagonal matrix with ± 1 entries and P is a permutation matrix.
In addition, orthogonality of Q yields that Q_2 has to be a zero matrix. Finally, Q_4 has to be a $(K-r) \times (K-r)$ orthogonal matrix to satisfy $QQ' = I_K$. Therefore, block B_1 is unique up to permutation and sign changes.

□

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 $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 = DP$ with D a diagonal matrix with ± 1 entries on the diagonal, P 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 having a one in the very last element. This proves the uniqueness of the full matrix B apart from sign reversal of its columns and permutation of its first r 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 $BQ = \begin{pmatrix} B_{11}Q_1 & B_{21}Q_4 \\ B_{12}Q_1 & B_{22}Q_4 \end{pmatrix}$ 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. □

Appendix B Estimation

B.1 Importance Density

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

$$\log p(h_i|\theta,\varepsilon_i) \propto \log p(\varepsilon_i|\theta,h_i) + \log p(h_i). \tag{B.1}$$

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

$$\log p(h_i) \propto -\frac{1}{2}\left(h_i-\delta_i\right)' Q_i \left(h_i-\delta_i\right), \tag{B.2}$$

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

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:

$$\begin{aligned}
\frac{\partial \log p(\varepsilon_{it}|\theta,h_{it})}{\partial h_{it}} &= -\frac{1}{2}+\frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: f_{it} \quad \Rightarrow \quad f_i = (f_{i1},\dots,f_{iT})', \\
-\frac{\partial^2 \log p(\varepsilon_{it}|\theta,h_{it})}{\partial h_{it}^2} &= \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: c_{it} \quad \Rightarrow \quad C_i = \operatorname{diag}\left([c_{i1},\dots,c_{iT}]'\right).
\end{aligned}$$

A second order Taylor approximation around $\tilde{h}_i^{(0)}$ then yields:

$$\begin{aligned}
\log p(\varepsilon_i|\theta,h_i) &\approx \log p(\varepsilon_i|\theta,\tilde{h}_i^{(0)}) + \left(h_i-\tilde{h}_i^{(0)}\right)' f_i - \frac{1}{2}\left(h_i-\tilde{h}_i^{(0)}\right)' C_i \left(h_i-\tilde{h}_i^{(0)}\right) \\
&= -\frac{1}{2}\left(h_i' C_i h_i - 2h_i' \underbrace{\left(f_i + C_i\tilde{h}_i^{(0)}\right)}_{=:b_i} \right) + \text{constant}.
\end{aligned} \tag{B.4}$$

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

$$\log p(h_i|\theta,\varepsilon_i) \propto -\frac{1}{2}\left(h_i' \underbrace{\left(C_i+Q_i\right)}_{=:\bar{Q}_i} h_i - 2h_i'\left(b_i+Q_i\delta_i\right) \right).$$

Consequently, the approximate smoothing density is:

$$\pi_G\left(h_i|\theta,\varepsilon_i\right) \sim \mathcal{N}\left(\bar{\delta}_i,\bar{Q}_i^{-1}\right), \qquad \text{with} \quad \bar{\delta}_i = \bar{Q}_i^{-1}\left(b_i+Q_i\delta_i\right).$$

The restricted density $\pi_G^c\left(h_i|\theta,\varepsilon_i\right)$ is constructed as outlined in section 3. Note that $\pi_G^c\left(h_i|\theta,\varepsilon_i\right)$ yields a good approximation only if $\tilde{h}_i^{(0)}$ 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_i is initialized by some vector $h_i^{(0)}$ satisfying the linear constraint, i.e. $A_h h_i^{(0)} = \mu_i$. Then, $h_i^{(l)}$ is used to evaluate $\bar{Q}_i$, $\bar{\delta}_i$ and to iterate:

$$\begin{aligned}
\tilde{h}_i^{(l+1)} &= h_i^{(l)} + \bar{Q}_i^{-1}\left(-\bar{Q}_i h_i^{(l)} + \bar{\delta}_i\right) = \bar{Q}_i^{-1}\bar{\delta}_i, \\
h_i^{(l+1)} &= \tilde{h}_i^{(l+1)} - \bar{Q}_i^{-1}A_h'\left(A_h\bar{Q}_i^{-1}A_h'\right)^{-1}\left(A_h\tilde{h}_i^{(l+1)} - \mu_i\right)
\end{aligned}$$

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

B.3 EM Algorithm

To fix notation, define the following quantities:

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

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

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

with $Z = (X' \otimes I_K)$, $\mathrm{E}(uu') = \tilde{\Sigma}_u$. Note that its inverse is given by $\tilde{\Sigma}_u^{-1} = ([B^{-1}]' \otimes I_T)\Sigma_e^{-1}(B^{-1} \otimes I_T)$ where $\Sigma_e^{-1} = \mathrm{diag}(\mathrm{vec}(V^{-1}))$.

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

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

Both algorithms EM-1 and EM-2 require some starting values. They are set in the same way for both alternatives. That is:

$$\begin{aligned}\hat{\alpha}^{(0)} &= \left(\left[(XX')^{-1}X\right]\otimes I_k\right)y^0, \\ \hat{B}^{(0)} &= (T^{-1}\hat{U}\hat{U}')^{\frac{1}{2}}Q, \quad \text{with } \hat{U} = Y^0 - \hat{A}X,\end{aligned}$$

where Q is a $K\times K$ orthogonal matrix uniformly drawn from the space of K-dimensional orthogonal matrices. Furthermore, we set the $r\times 1$ vectors:

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

which correspond to persistent heteroskedasticity with initial kurtosis of about 3.7 for the estimated structural shocks $\hat{\varepsilon}_i, i=1,\ldots,r$.
Note that in order to satisfy linear restriction (2.10) we set for $i=1,\ldots,r$ and $l\geq 1$:

$$\hat{\mu}_i^{(l-1)} = -\frac{\hat{s}_i^{(l-1)}}{2}\Big/\left(1-\left(\hat{\phi}_i^{(l-1)}\right)^2\right).$$

EM-1

Because of $\hat{\varepsilon}_t^{(l-1)} = \hat{B}^{(l-1)}(y_t - \hat{A}^{(l-1)}x_t)$, it is equivalent to condition the approximate smoothing densities π_G^c and their moments to $\left(\theta^{(l-1)}, \hat{\varepsilon}_i^{(l-1)}\right)$ or $\left(\theta^{(l-1)}, y\right)$, respectively. Based on starting values $\theta^{(0)} = \left[\left(\hat{\alpha}^{(0)}\right)', \mathrm{vec}\left(\hat{B}^{(0)}\right)', \left(\hat{\phi}^{(0)}\right)', \left(\hat{s}^{(0)}\right)'\right]'$, the EM algorithm iteratively cycles through the following steps for $l\geq 1$:

1. E-step: For $i=1,\ldots,r$, evaluate the moments of the approximate smoothing densities, mean $\bar{\delta}_i^c$ and variance $\bar{Q}_i^{-1} - \bar{Q}_i^{-1}A_h'\left(A_h\bar{Q}_i^{-1}A_h'\right)^{-1}A_h\bar{Q}_i^{-1}$, as described in Appendix B.1. Thereby, directly inverting $\bar{Q}_i$ is unnecessary costly since we only need its diagonal elements representing the marginal variances $\mathrm{Var}(h_{it}|\theta^{(l-1)}, y)$ and the entries of the first off-diagonal corresponding to $\mathrm{Cov}(h_{it}, h_{i,t-1}|\theta^{(l-1)}, y)$. Similar to the Kalman smoother recursions, they can be obtained without computing the whole inverse using sparse matrix routines based on Takahashi's equations (Rue et al.; 2009). An efficient implementation in Matlab is available at the MathWorks File Exchange (see *sparseinv* by Tim Davis).

2. M-step: 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.

   (a) 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 $\frac{\partial^{a_1+a_2}\mathcal{L}_c}{\partial^{a_1}\phi\partial^{a_2}s} = \left[\frac{\partial^{a_1+a_2}\mathcal{L}_c}{\partial^{a_1}\phi_1\partial^{a_2}s_1}, \ldots, \frac{\partial^{a_1+a_2}\mathcal{L}_c}{\partial^{a_1}\phi_r\partial^{a_2}s_r}\right]$ for $a_1, a_2 \in \{0, 1, 2\}$ with $a_1 + a_2 \leq 2$, $\nabla G(\phi, s) = \mathrm{E}\left[\frac{\partial\mathcal{L}_c}{\partial\phi}, \frac{\partial\mathcal{L}_c}{\partial s}\right]$ and $H(\phi, s) = \mathrm{E}\begin{pmatrix} \mathrm{diag}\left(\frac{\partial^2\mathcal{L}_c}{\partial\phi^2}\right) & \mathrm{diag}\left(\frac{\partial^2\mathcal{L}_c}{\partial\phi\partial s}\right) \\ \mathrm{diag}\left(\frac{\partial^2\mathcal{L}_c}{\partial\phi\partial s}\right) & \mathrm{diag}\left(\frac{\partial^2\mathcal{L}_c}{\partial s^2}\right) \end{pmatrix}$. The detailed expressions for first and second derivatives of the complete data log-likelihood are printed in 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 $\hat{s}^{(l)} = \hat{s}_{k+1}$.

   (b) 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 $\tilde{\Sigma}_u^{-1} = \left(\left[\left(\hat{B}^{(l-1)}\right)^{-1}\right]' \otimes I_T\right)\hat{\Sigma}_e^{-1}\left(\left(\hat{B}^{(l-1)}\right)^{-1} \otimes I_T\right)$ and $\hat{\Sigma}_e^{-1} = \mathrm{diag}(\mathrm{vec}(\hat{V}^{-1}))$. Furthermore, it is:

$$\begin{aligned} \hat{V}^{-1} =& \mathrm{E}(V^{-1}|\theta^{(l-1)}, y) = (\hat{v}_1^{-1}, \ldots, \hat{v}_T^{-1}) \in \mathbb{R}^{K\times T}, \quad \text{with} \\ \hat{v}_t^{-1} =& \exp\left(-\mathrm{E}(h_t|\theta^{(l-1)}, y) + \frac{1}{2}\mathrm{Var}(h_t|\theta^{(l-1)}, y)\right). \end{aligned}$$

   The latter is based on the properties of a log-normal distribution. Note that for $i = r+1, \ldots, K$, $\hat{v}_{it}^{-1} = 1$.

   (c) Update B. Therefore, define $\hat{U} = Y^0 - \hat{A}^{(l)}X$, then:

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

3. Set $\theta^{(l)} = \left[\left(\hat{\alpha}^{(l)}\right)', \text{vec}\left(\hat{B}^{(l)}\right)', \left(\hat{\phi}^{(l)}\right)', \left(\hat{s}^{(l)}\right)'\right]'$, $l = l + 1$ and return to step 1.

We iterate between steps 1.-3. 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.

EM-2

In EM-2, the expectations in the E-step are approximated by MCMC integration. Based on starting values, $\theta^{(0)}$, the algorithm iterates between the following steps for $l \geq 1$:

1. E-Step: In order to compute the expectations necessary in the EM algorithm, we recur to Monte Carlo integration. In particular, for each of the heteroskedastic shocks ($i = 1, \ldots, r$), we simulate random draws of the mixture indicators $z_i^{(j)}$ for $j = 1, \ldots, R$ and compute:

$$Q(\theta, \theta^{(l-1)}) \approx \frac{1}{R} \sum_{j=1}^{R} \text{E}_{\theta^{(l-1)}}^{(j)}[\mathcal{L}(\theta)], \tag{B.6}$$

   where the expectations are taken with respect to the tractable distribution $p(h|\theta^{(l-1)}, z^{(j)}, y)$. To generate random draws of z, we rely on the methodology of Kim et al. (1998). For each of the heteroskedastic shocks ($i = 1, \ldots, r$), this involves iteratively drawing from the following conditional distributions:

   (a) $z_i^{(j)} \sim p\left(z_i \middle| \theta^{(l-1)}, h_i^{(j-1)}, y\right)$. The mixture indicators are drawn for each $t = 1, \ldots, T$ from the discrete conditional distribution $P\left(z_{it}^{(j)} = k\right) = q_{it,k}$ for $k = 1, \ldots, 7$ where:

$$q_{it,k} = \frac{p_k \phi\left(y_{it}^* - h_{it}; m_k, v_k^2\right)}{\sum_{k=1}^{7} p_k \phi\left(y_{it}^* - h_{it}; m_k, v_k^2\right)},$$

   with $y_{it}^* = \log\left[\left(\hat{\varepsilon}_{it}^{(l-1)}\right)^2\right]$, $\hat{\varepsilon}_t^{(l-1)} = \left(\hat{B}^{(l-1)}\right)^{-1}\left(y_t - \hat{A}^{(l-1)} x_t\right)$ and $\phi(\cdot; m_k, v_k^2)$ indicating the pdf of a normal distribution with mean m_k and variance v_k^2. Mixture parameters p_k's, m_k's and v_k's are tabulated in Table 5.

   (b) $h_i^{(j)} \sim p(h_i|\theta^{(l-1)}, z_i^{(j)}, y)$. To draw the log variances, first a random sample from the unconstrained conditional distribution $\tilde{h}_i^{(j)} \sim \mathcal{N}(\bar{\delta}_{ij}, \Sigma_{ij})$ is generated using the precision sampler of Chan & Jeliazkov (2009). The unconstrained moments are given

as:

$$\begin{aligned}
\Sigma_{ij}^{-1} &= H_i' \Sigma_{h_i}^{-1} H_i + G_{ij}, \\
\bar{\delta}_{ij} &= \Sigma_{ij} \left(H_i' \Sigma_{h_i}^{-1} H_i \delta_i + G_{ij}(y_i^* - m_{ij}) \right),
\end{aligned}$$

and

$$\begin{aligned}
y_i^* &= \left(\log \left[\left(\hat{\varepsilon}_{i1}^{(l-1)} \right)^2 \right], \ldots, \log \left[\left(\hat{\varepsilon}_{iT}^{(l-1)} \right)^2 \right] \right)', \\
G_{ij} &= \operatorname{diag} \left(v^2 \left(z_{i1}^{(j)} \right), \ldots, v^2 \left(z_{iT}^{(j)} \right) \right)^{-1}, \\
m_{ij} &= \operatorname{diag} \left(m \left(z_{i1}^{(j)} \right), \ldots, m \left(z_{iT}^{(j)} \right) \right).
\end{aligned}$$

In a next step, the draw is corrected to account for the linear constraint. That is:

$$h_i^{(j)} = \tilde{h}_i^{(j)} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} \left(A_h \tilde{h}_i^{(j)} - \hat{\mu}_i^{(l-1)} \right),$$

which yields a draw from the correct distribution under the linear constraint. The moments of this distribution are:

$$\begin{aligned}
\bar{\delta}_{ij}^c &= \bar{\delta}_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} \left(A_h \bar{\delta}_{ij} - \hat{\mu}_i^{(l-1)} \right), \\
\operatorname{Cov} &\left(h_i \middle| \theta^{(l-1)}, z_i^{(j)}, y, A_h h_i = \hat{\mu}_i^{(l-1)} \right) = \Sigma_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} A_h \Sigma_{ij}.
\end{aligned}$$

Note that the corrected moments are those used to compute the Monte Carlo expected complete data log-likelihood from equation (B.6). As in EM-1, we only compute the diagonal and first off-diagonal of the covariance matrix Σ_{ij} using the same sparse matrix routines.

2. M-steps: Conditional on the mixture indicators $z_i^{(j)}$ $(i = 1, \ldots, r; j = 1, \ldots, R)$, first and second moments of h_i's are given. Thus, as in EM-1, we maximize the expected complete data log-likelihood using Newton-Raphson updates in state equation parameters, a closed-form update in VAR parameters and numerical optimization in the impact matrix.

   (a) Update ϕ_i and s_i for $i = 1, \ldots, r$: Conditional on the mixture indicators z, the expected value of the complete data log-likelihood (B.5) is maximized. To do so, define $\nabla G_R(\phi, s) = \frac{1}{R} \sum_{j=1}^{R} \mathrm{E} \left(\frac{\partial \mathcal{L}_c}{\partial \phi}, \frac{\partial \mathcal{L}_c}{\partial s} \middle| z^{(j)} \right)$ containing the first and $H_R(\phi, s) = \frac{1}{R} \sum_{j=1}^{R} \mathrm{E} \begin{pmatrix} \operatorname{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi^2} \right) & \operatorname{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) \\ \operatorname{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) & \operatorname{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial s^2} \right) \end{pmatrix} \middle| z^{(j)} \end{pmatrix}$ including the second derivatives. The detailed expressions are printed in subsection B.4. All expectations of functions of the log-variances are uniquely determined by the sampled mixture indicators. 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

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

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 $\hat{s}^{(l)} = \hat{s}_{k+1}$.

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

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

where everything is as in EM-1 but:

$$\hat{v}_t^{-1} = R^{-1}\sum_{j=1}^{R} \exp\left(-\mathrm{E}\left(h_t|\theta^{(l-1)}, z_t^{(j)}, y\right) + \frac{1}{2}\mathrm{Var}\left(h_t|\theta^{(l-1)}, z_t^{(j)}, y\right)\right).$$

(c) Update B as in EM-1.

3. Set $\theta^{(l)} = \left[\left(\hat{\alpha}^{(l)}\right)', \mathrm{vec}\left(\hat{B}^{(l)}\right)', \left(\hat{\phi}^{(l)}\right)', \left(\hat{s}^{(l)}\right)'\right]'$, $l = l + 1$ and return to step 1.

We recommend to set the starting values based on the results of EM-1, which are quickly available. We increase the number of MCMC replications deterministically over the EM iterations. This is necessary since automated strategies as the ascent-based MCEM algorithm (Caffo, Jank & Jones; 2005) fail to converge due to the substantial amount of parameters to be estimated in the VAR equation. That is, we first run a burn-in period of 300 EM steps using $R = 50$ and then proceed with another 100 EM iterations using $R = 500$. Subsequently, we increase R to 50,000 and iterate EM steps until the stopping criterion of Caffo et al. (2005) applies. This usually happens after a small number of additional EM steps using 50,000 MCMC replications.

Table 5: Mixture Components

k	$p_k = \Pr(z_{it} = k)$	m_k	v_k^2
1	0.00730	−10.12999	5.79596
2	0.10556	−3.97281	2.61369
3	0.00002	−8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	−1.08819	1.26261

Note: Seven Normal Mixture components to approximate a $\log\left(\chi^2_{(1)}\right)$ distribution adjusted by its mean −1.2704.

B.4 Derivatives complete data log-likelihood

The respective derivatives of the complete data log-likelihood (B.5) are given in the following. Let $\bar{h}_{it} = h_{it} - \mu_i$ for $i = 1, \dots, r$ and $t = 1, \dots, T$. First and second derivatives with respect to state equation parameters ϕ_i and s_i are given as follows:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_c(\theta)}{\partial s_i} &= -\frac{1}{2s_i}\left(T - \frac{1-\phi_i^2}{s_i}\bar{h}_{i1}^2 + \bar{h}_{i1} + \sum_{t=2}^{T}\left[\frac{\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}}{1+\phi_i} - \frac{\left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right)^2}{s_i}\right]\right), \\
\frac{\partial \mathcal{L}_c(\theta)}{\partial \phi_i} &= -\frac{\phi_i}{1-\phi_i^2}\left(1 + \bar{h}_{i1}\right) + \frac{\phi_i}{s_i}\bar{h}_{i1}^2 - \frac{1}{s_i}\sum_{t=2}^{T}\left[\left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right)\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right], \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial s_i} &= -\frac{\phi_i}{2\left(1-\phi_i^2\right)^2} + \frac{\phi_i\bar{h}_{i1}}{\left(1-\phi_i^2\right)s_i} - \frac{\phi_i\bar{h}_{i1}^2}{s_i^2} + \frac{1}{s_i^2}\sum_{t=2}^{T}\left[\left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right)\right. \\
&\quad \left.\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right] - \frac{1}{s_i}\sum_{t=2}^{T}\left[\frac{1}{2\left(1+\phi_i\right)}\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right. \\
&\quad \left.+ \left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right)\left(\frac{\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \frac{1}{2\left(1-\phi_i^2\right)}\right)\right], \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i^2} &= \frac{1}{s_i}\left(\frac{T}{2s_i} + \frac{\bar{h}_{i1}}{s_i} - \frac{\bar{h}_{i1}^2\left(1-\phi_i^2\right)}{s_i^2} - \frac{1}{4\left(1-\phi_i^2\right)} - \frac{T-1}{4\left(1+\phi_i\right)^2}\right. \\
&\quad \left.+\frac{1}{s_i}\sum_{t=2}^{T}\left[\frac{\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}}{1+\phi_i} - \frac{\left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right)^2}{s_i}\right]\right), \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i^2} &= -\frac{1+\phi_i^2}{\left(1-\phi_i^2\right)^2}\left(1 + \bar{h}_{i1}\right) - \frac{s_i\phi_i^2}{\left(1-\phi_i^2\right)^3} + \frac{\bar{h}_{i1}^2}{s_i} + \frac{2\phi_i^2\bar{h}_{i1}^2}{\left(1-\phi_i^2\right)^2} \\
&\quad - \frac{1}{s_i}\sum_{t=2}^{T}\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)^2 - \left(\frac{1-3\phi_i}{\left(1-\phi_i^2\right)^2} + \frac{4\phi_i^2(1-\phi_i)}{\left(1-\phi_i^2\right)^3}\right)\sum_{t=2}^{T}\left(\bar{h}_{it} - \phi_i\bar{h}_{i,t-1}\right).
\end{aligned}
$$

Furthermore, let $\Sigma_t = BV_tB'$, $\beta = \text{vec}(B)$, $\alpha = \text{vec}(A)$, $\tilde{X}_t = (x_t' \otimes I_K)$, such that $\text{vec}(Ax_t) = \tilde{X}_t\alpha$ 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:

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

Note that the cross derivatives $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial \alpha}$, $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial \beta}$, $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i \partial \alpha}$ and $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i \partial \beta}$ are equal to zero due to the structure of the complete data log likelihood (B.5).

Appendix C Data and complementary Results

The time series data used in section 5 is based on $y_t = (q_t, \pi_t, c_t, \Delta s_t, r_t)'$, where

- q_t is the logarithm of industrial production (linearly detrended),
- π_t is the growth rate of the consumer price index (in %),
- c_t denotes the annualized change in the logarithm of the World Bank commodity price index (in %),
- Δs_t is the first difference of the logarithm of the CPI deflated real S&P500 index,
- r_t is the Federal Funds rate.

As in Lütkepohl & Netšunajev (2017a) and Lütkepohl & Netšunajev (2017b), we use the updated sample period 1970M1-2007M6. Except for c_t, the data can be downloaded from the FRED. The commodity price index is provided by the World Bank. The transformed data set is readily available at `http://sfb649.wiwi.hu-berlin.de/fedc/discussionPapers_formular_content.php`.

The monetary policy instruments of Gertler & Karadi (2015) and Romer & Romer (2004) are obtained from the homepage of Valerie Ramey: `http://econweb.ucsd.edu/~vramey/`

research.html#data. Note that the RR series used in our analysis is the one extended by Wieland & Yang (2016).

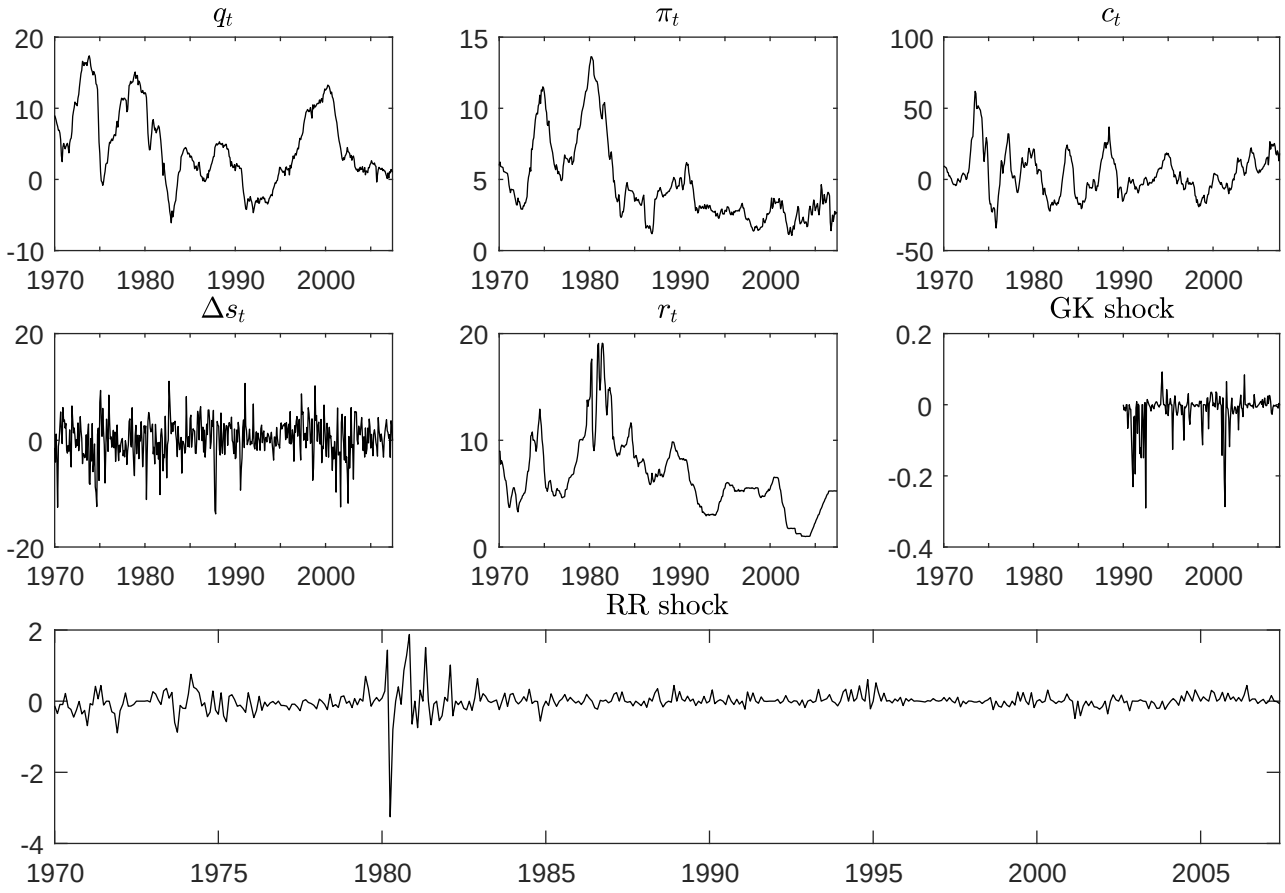

Figure 3: Time Series Data

C.1 Complementary results

Table 6: Tests on standardized structural shocks

	Normality		Heteroskedasticity			
	MJB	p-value	Q_1	p-value	Q_2	p-value
Linear	12,911.0	0.000	52.34	0.000	1433.70	0.000
STVAR	49,785.0	0.000	40.25	0.000	1475.50	0.000
MS(2)	291.8	0.000	13.59	0.004	811.21	0.000
MS(3)	48.6	0.000	9.40	0.024	844.74	0.000
GARCH	555.0	0.000	8.22	0.042	627.45	0.904
SV	16.1	0.096	3.25	0.355	623.88	0.921

Note: Multivariate Jarque-Bera (MJB) test conducted as in (Lütkepohl; 2005, p. 181). Test statistics Q_1 and Q_2 as discussed in section 3.5, applied to estimated standardized structural shocks $\hat{\varepsilon}_t / \exp\left(\hat{h}_t/2\right)$.

Table 7: Test for Overidentifying Restrictions (EM-2)

H_0	H_1	LR	dof	p-value	$p_{.025}$	$p_{.975}$
R1	UC	27.341	10	0.0023	0.0017	0.0032
R2	UC	23.693	7	0.0013	0.0009	0.0018
R3	UC	25.868	9	0.0021	0.0015	0.0030
R1	R3	1.543	1	0.2142	0.1390	0.3438
R4rr	UC	5.779	4	0.2163	0.1388	0.3294
R4gk	UC	256.590	4	0.0000	0.0000	0.0000

Note: For details about overidentifying restrictions see subsection 5.1. Likelihood ratio test statistics are computed as $2\left(\ln L_{H_1} - \ln L_{H_0}\right)$ and are approximatively χ^2-distributed under H_0. Right columns report an approximate 95%-confidence interval for the p-value resulting from an application of the batch means method to the LR test statistic.