

Identification of Structural Vector Autoregressions by Stochastic Volatility ^{*}

Dominik Bertsche^a Robin Braun^b

First draft: August 14, 2017

This version: December 21, 2017

Abstract

In Structural Vector Autoregressive (SVAR) models, heteroskedasticity can be exploited to identify structural parameters statistically. In this paper, we propose to capture time variation in the second moment of structural shocks by a stochastic volatility (SV) model, assuming that their log variances follow latent AR(1) processes. Estimation is performed by Gaussian Maximum Likelihood and an efficient Expectation Maximization algorithm is developed for that purpose. Since the smoothing distributions required in the algorithm are intractable, we propose to approximate them either by Gaussian distributions or with the help of Markov Chain Monte Carlo (MCMC) methods. We provide simulation evidence that the SV-SVAR model works well in estimating the structural parameters also under model misspecification. We use the proposed model to study the interdependence between monetary policy and the stock market. Based on monthly US data, we find that the SV specification provides the best fit and is favored by conventional information criteria if compared to other models of heteroskedasticity, including GARCH, Markov Switching, and Smooth Transition models. Since the structural shocks identified by heteroskedasticity have no economic interpretation, we 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

^{*}We thank the participants of the Doctoral Workshop on Applied Econometrics at the University of Strasbourg and the Econometrics Colloquium at the University of Konstanz for useful comments on earlier versions of this paper. Financial support by the Graduate School of Decision Science (GSDS) is gratefully acknowledged.

^aDominik Bertsche: University of Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: dominik.bertsche@uni-konstanz.de

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

1 Introduction

Since the seminal contribution of Sims (1980), structural vector autoregressive (SVAR) models have been used extensively in applied macroeconomic research. Based on a reduced form model capturing the common dynamics of time series vectors, identifying restrictions are introduced to back out structural shocks and estimate their dynamic causal effects on the endogenous variables. Popular identifying restrictions 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 Nicoló; 2002; Uhlig; 2005) and identification via external instruments, also known as Proxy SVARs (Mertens & Ravn; 2013; Stock & Watson; 2012).¹

Another approach for identification of SVAR models that enjoys increasing interest is based on statistical identification and assumes heterogeneity in the second moment of structural shocks. Various specifications have been proposed to model heteroskedasticity within SVARs, starting with a simple exogenous breakpoint model (Rigobon; 2003), Markov Switching mechanisms (Lanne, Lütkepohl & Maciejowska; 2010), a GARCH- (Normandin & Phaneuf; 2004) and a Smooth Transition model (Lütkepohl & Netsunajev; 2014). In this paper, we contribute to the literature by proposing a stochastic volatility (SV) model to identify the SVAR parameters. Specifically, we assume that the log variances of structural shocks are latent, each following independent AR(1) processes. 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 macroeconomic literature. For example, within DSGE models fitted to US data, Justiniano & Primiceri (2008) and Fernández-Villaverde & Rubio-Ramírez (2007) use SV models and find substantial time variation in the second moments of their structural shocks. On the empirical side, SV is typically used in time varying parameter VAR models and has been found to describe the volatility of macroeconomic data very well, including the slowly decaying variance of U.S. macroeconomic aggregates known as the Great Moderation (Primiceri; 2005; Koop & Korobilis; 2010). Given this context, it seems natural to exploit the model also for identification purposes within SVAR analysis. Second, the SV model is known to be very flexible, particularly in comparison to ARCH type of models with deterministic variance rules. As pointed out in Kim, Shephard & Chib (1998), this additional flexibility typically translates into superior fit than 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 and therefore, is favored by any conventional information criterion (IC). This is an important aspect, given that recent evidence of Lütkepohl & Schlaak (2017) suggests to choose the heteroskedasticity model of SVARs by information criteria. With the help of a simulation study, the authors find that choosing the model by IC can “reduce the mean squared error of impulse response estimates relative to a model that is chosen arbitrarily based on the personal preferences”. Finally,

¹For an extensive review of identification methods for SVAR models we refer to Kilian & Lütkepohl (2017).

we find that the SV-SVAR model works very well in identifying structural parameters also under model misspecification, and is able to capture volatility patterns generated by different data generating processes (DGPs). In a stylized simulation study involving four different DGPs, the SV model performs well in terms of mean squared error of impulse response functions identified via heteroskedasticity, particularly if compared to other models designed to capture variation in the second moments of the structural shocks.

Unfortunately, estimation of the proposed model is relatively difficult in comparison to alternative specifications for the variance. The main obstacle is that SV models have a non-linear state space representation and hence, standard linear filtering algorithms cannot be applied for evaluate the likelihood function. 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) and Chan & Grant (2016) in evaluating the likelihood function by importance sampling (IS) in a computationally efficient way. To maximize the likelihood function we develop two versions of an Expectation Maximization (EM) algorithm. The first is based on a Laplace approximation for the intractable E-step and relies on sparse matrix algorithms developed for Gaussian Markov random fields (Rue, Martino & Chopin; 2009; Chan; 2017). Therefore, the algorithm is 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 is normal and linear, allowing to compute the expectations necessary in the E-step by standard Kalman smoothing recursions. Thereby, the Laplace approximation can be avoided at the cost of higher computational effort. However, after fitting the model to various simulated and real datasets, our experience is that only negligible gains in the likelihood can be achieved by using the Monte Carlo based algorithm. Therefore, we recommend the usage of the computational more efficient Laplace approximation.

In an empirical application, we use the proposed model to identify the structural parameters of a VAR specified by 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 if compared to other heteroskedastic models typically used to identify the SVAR parameters statistically, the SV model provides superior fit and is favored by all conventional information criteria. Since the structural shocks identified by heteroskedasticity cannot be interpreted without further economic narrative, we follow Lütkepohl & Netšunajev (2017) 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 Gaussian Maximum Likelihood estimation and presents an efficient EM algorithm. In section 4 we go through a testing procedure which allows to assess whether there is enough heteroskedasticity in the data to identify all structural parameters. In section 5, we present simulation evidence while in section 6 we apply the proposed model to study the interdependence between US monetary policy and stock markets. Section 7 concludes.

2 Identification of SVAR via Stochastic Volatility

In the following section, we introduce the SVAR model subject to stochastic volatility in the variances and discuss the conditions under which the structural parameters are identified via heterogeneity in the second moments. Let y_t be a $K \times 1$ vector of endogenous variables. The most general SV-SVAR model reads:

$$y_t = \nu + \sum_{i=1}^p A_i y_{t-i} + u_t, \quad (2.1)$$

$$u_t = BV_t^{\frac{1}{2}} \eta_t, \quad (2.2)$$

where $\eta_t \sim (0, I_n)$ 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_i for $i = 1, \dots, p$ are $K \times K$ matrices of autoregressive coefficients and ν a $K \times 1$ vector of intercepts. Equation (2.2) models the structural part and is set up as a B -model in the terminology of Lütkepohl (2005). The correlated 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 $K \times K$ contemporaneous impact matrix and $V_t^{\frac{1}{2}}$ a stochastic diagonal matrix with strictly positive elements capturing conditional heteroskedasticity in the structural shocks ε_t . The specification yields a time-varying covariance matrix of the reduced form errors u_t given as $\Sigma_t = E(u_t u_t') = BV_t B'$. In this paper, we specify a basic SV model for the first $r \leq K$ diagonal elements of V_t corresponding to the variances of the first r structural shocks:

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

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

We assume that $\omega_{it} \sim \mathcal{N}(0, 1)$ and $E(\varepsilon_t' \omega_t) = 0$ for $\omega_t = [\omega_{1t}, \dots, \omega_{rt}]'$. In words, the first r log variances of ε_t contained in the diagonal elements of V_t are assumed to be latent independent Gaussian AR(1) processes. Their unconditional first and second moments are given by $E(h_{it}) = \mu_i$ and $\text{Var}(h_{it}) = s_i / (1 - \phi_i^2)$. Note that the proposed model for equation (2.2) is very similar to the Generalized Orthogonal GARCH (GO-GARCH) model from Van der

Weide (2002) and Lanne & Saikkonen (2007), with the major difference in the specification of V_t . While for the GO-GARCH the first r diagonal components are modeled by observable univariate GARCH(1,1) processes, we model their logarithms as latent AR(1)'s.

In order to conduct structural analysis with the proposed model, the contemporaneous impact matrix B must be identified. Following Sentana & Fiorentini (2001) and Lanne et al. (2010), we show that identification via heteroskedasticity depends crucially on the number of heteroskedastic shocks r . For that purpose, let $B = [B_1, B_2]$ with $B_1 \in \mathbb{R}^{K \times r}$ and $B_2 \in \mathbb{R}^{K \times (K-r)}$. In Proposition 1 of Appendix A, we show that B_1 is identified up to sign changes, given that the r heteroskedastic shocks have some heterogeneity in their variance which allows to discriminate amongst them.² This means that for all $i \in \{1, \dots, r\}$ and $j \neq i \in \{1, \dots, K\}$ there must be a $t \in \{1, \dots, T\}$ such that $h_{it} \neq h_{jt}$, where $h_{jt} = 1$ for $j \in \{r+1, \dots, K\}$. Orthogonality constraints of the structural shocks also yield that B_2 is identified if $r = K - 1$ (Corollary 1). This means that for full identification of B , all but one structural shock must be heteroskedastic. In case that $r < K - 1$, B is only partially identified and further identifying restrictions are necessary.³

Note that some normalizing constraints are needed to ensure that the scale of the elements in B and h_i is unique. Similar to the GO-GARCH, we normalize the expected unconditional variance of the structural shocks to unity, that is $E(\varepsilon_{it}^2) = 1$. Note that from the properties of a log-normal distribution, $E(\exp(h_{it})) = \exp\left(\mu_i + \frac{s_i}{2(1-\phi_i^2)}\right)$. Therefore, we simply set $\mu_i = -\frac{s_i}{2(1-\phi_i^2)}$ and impose the linear constraint on the first sample moment:

$$A_h h_i = \mu_i, \quad (2.5)$$

where $A_h = \mathbf{1}_T' / T$ and $h_i = [h_{i1}, \dots, h_{iT}]'$. To initialize the latent variables, we assume that at $t = 1$, $h_{i1} \sim \mathcal{N}(\mu_i, s_i / (1 - \phi_i^2))$ which corresponds to the unconditional distribution of h_{it} . Note that an alternative normalization constraint would be to set $E(h_{i1}) = \text{Var}(h_{i1}) = 0$ which implies $E(u_1 u_1') = BB'$ as imposed e.g. by Markov Switching SVAR models (Lanne et al.; 2010; Herwartz & Lütkepohl; 2014). However, the latter would require additional r free parameters to capture nonzero means in the log variances. Furthermore, we find the linear constraint of equation (2.5) to yield numerically stable results at trivial computational extra costs.

3 Maximum Likelihood Estimation

Let $\phi = [\phi_1, \dots, \phi_r]'$ and $s = [s_1, \dots, s_r]'$. In order to estimate the parameter vector $\theta = [\text{vec}([\nu, A_1, \dots, A_p])', \text{vec}(B)', \phi', s']'$, we recur to Gaussian maximum likelihood estimation. Assuming normality of η_t , the likelihood function based on the prediction error decomposition is given as follows:

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

²Also column permutations are allowed if B and V_t are permuted jointly.

³We discuss possibilities to estimate the model under partial identification in section 4.

where $u_t = y_t - \nu - \sum_{i=1}^p A_i y_{t-i}$ and $V_{t|t-1} = E[V_t | \mathcal{F}_{t-1}]$ are one-step ahead predicted variances conditional on information at time $t - 1$. Since the SV model implies a nonlinear state space model, the predictive distributions $p(h_{it} | y_{t-1}, \theta)$ necessary to compute $V_{t|t-1}$ are not available in closed form. Therefore, the likelihood is intractable and standard Kalman based estimation algorithms cannot be applied. Fortunately, 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 et al.; 1994; Ruiz; 1994), Simulated Likelihood (Danielsson & Richard; 1993) and Bayesian methods (Kim et al.; 1998) based on Markov Chain Monte Carlo (MCMC) simulation.⁴ We follow Durbin & Koopman (1997) and Chan & Grant (2016) in evaluating the likelihood function by importance sampling in a computational efficient way. In order to reach a maximum, we develop an Expectation Maximization algorithm that leads to fast and reliable results.

3.1 Evaluation of the Likelihood

To evaluate the likelihood by importance sampling, we further simplify the likelihood function:

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

where $\varepsilon_t = B^{-1}u_t$. 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 ε_i . Note that for the last $(K - r)$ shocks these densities are trivial to compute since $v_{it|t-1} = 1$. However, $\log p(\varepsilon_i | \theta)$ for $i \leq r$ is not tractable. We follow Chan & Grant (2016) and use importance sampling to evaluate these densities. Note that the likelihood evaluation of a heteroskedastic shock reduces to the problem of evaluating the high-dimensional integral:

$$p(\varepsilon_i | \theta) = \int p(\varepsilon_i | \theta, h_i) p(h_i | \theta) dh_i \quad (3.1)$$

for $i = 1, \dots, r$. Let $q(h_i)$ be a proposal distribution from which independent random numbers $h_i^{(1)}, \dots, h_i^{(R)}$ 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 then:

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

Note that in order to be able to assess the precision of the likelihood estimator given in equation (3.2) and assure \sqrt{R} -convergence, the variance of the importance weights must exist. For likelihood evaluation of stochastic volatility models with their high-dimensional integrals, this is not clear a-priori. Based on extreme value theory, Koopman, Shephard &

⁴For an extensive review we recommend the paper of Broto & Ruiz (2004).

Creal (2009) develop formal tests to check the existence of the variance. We recommend their usage and in addition, to assess the Monte Carlo error by re-estimating the likelihood several times and reporting a range of possible values.

In the following, we discuss in detail the choice of the importance density $q(h_i)$ critical to the success of the IS estimator in equation (3.2). Note that the 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, the normalizing constant is unknown which is why we need the IS estimator after all. We follow Durbin & Koopman (1997, 2000) and rely on a Gaussian importance density $\pi_G(h_i|\varepsilon_i, \theta)$ centered at the mode 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 $\pi_G(h_i|\varepsilon_i, \theta)$, note that the normal prior for h_i implies the following explicit form of the zero variance IS density:

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

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

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

and $\Sigma_{h_i} = \text{diag}([\frac{s_i^2}{1-\phi_i}, s_i^2, \dots, s_i^2]')$. Furthermore, it is $\delta_i = H_i^{-1}\tilde{\delta}_i$ with $\tilde{\delta}_i = [\mu_i, (1 - \phi_i)\mu_i, \dots, (1 - \phi_i)\mu_i]'$ (Chan & Grant; 2016). The Gaussian approximation is based on a second order Taylor expansion of the nonlinear density $\log p(\varepsilon_i|h_i, \theta)$ around $\tilde{h}_i^{(0)}$:

$$\log p(\varepsilon_{it}|h_{it}, \theta) \approx \log p(\varepsilon_{it}|\tilde{h}_{it}^{(0)}, \theta) + b_{it}h_{it} - \frac{1}{2}c_{it}h_{it}^2, \quad (3.3)$$

where b_{it} and c_{it} depend on $\tilde{h}_{it}^{(0)}$. Based on the linearized kernel, the approximate smoothing distribution takes the form of a Normal distribution $\pi_G(h_i|\varepsilon_i, \theta)$ 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 $C_i = \text{diag}([c_{i1}, \dots, c_{iT}]')$ and $b_i = [b_{i1}, \dots, b_{iT}]'$. The T -dimensional density has a tridiagonal precision matrix which allows for very fast generation of random samples and likelihood evaluation. The approximation is fitted around the mode of $\log p(\varepsilon_i|h_{it}, \theta)$ obtained by a Newton Raphson method and typically converges in few iterations. Details on the Newton Raphson and on explicit expressions for b_{it} and c_{it} are given in Appendix B. Finally, to account for the linear restriction $A_h h_i = \mu_i$, the mean is corrected by:

$$\bar{\delta}_i^c = \bar{\delta}_i - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}(A_h\bar{\delta}_i - \mu_i), \quad (3.4)$$

which is known as conditioning by kriging (Rue et al.; 2009) and yields the correct expected value of $\pi_G(h_i|\varepsilon_i, \theta)$ under the linear constraint. Note that the corrected covariance

$\text{Cov}(h_i|\varepsilon_i, \theta, A_h h_i = \mu_i) = \bar{Q}_i^{-1} - \bar{Q}_i^{-1} A'_h (A_h \bar{Q}_i^{-1} A'_h)^{-1} A_h \bar{Q}_i^{-1}$ is a full matrix of rank $T - 1$ so that sparse algorithms cannot be exploited anymore for direct sampling and density evaluation. Following Rue & Martino (2007), sampling and evaluation of the importance density under linear constraints can still be implemented at trivial extra costs. Specifically, first a random sample $\tilde{h}_i^{(j)}$ is generated from the unconstrained distribution $\mathcal{N}(\bar{\delta}_i, \bar{Q}_i^{-1})$, 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 restricted density can be achieved efficiently by applying Bayes Theorem:

$$\pi_G^c(h_i|\varepsilon_i, \theta) = \frac{\pi_G(h_i|\varepsilon_i, \theta)\pi(A_h h_i|h_i)}{\pi(A_h h_i)}, \quad (3.5)$$

where $\pi_G(h_i|\varepsilon_i, \theta) \sim \mathcal{N}(\bar{\delta}_i^c, \bar{Q}_i^{-1})$, $\log \pi(A_h h_i|h_i) = -\frac{1}{2} \log |A_h A'_h|$ as well as $\pi(A_h h_i) \sim \mathcal{N}(A_h \delta_i, A_h \bar{Q}_i^{-1} A'_h)$.

3.2 EM Algorithm

To reach an optimum of the likelihood function, we exploit the derivative free 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. Let $h = [h_1, \dots, h_r]$ denote the hidden variables, then the goal is to maximize:

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

Following the exposition of 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:

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

where the inequality arises by Jensen's inequality. The EM algorithm starts with some initial guess $\theta^{(0)}$, and proceeds by iteratively computing:

$$\text{E-step: } \quad \tilde{p}^{(l)} = \arg \max_{\tilde{p}} F(\tilde{p}, \theta^{(l-1)}), \quad (3.6)$$

$$\text{M-step: } \quad \theta^{(l)} = \arg \max_{\theta} F(\tilde{p}^{(l)}, \theta). \quad (3.7)$$

Under mild regularity conditions the EM algorithm converges reliably towards a local optimum.⁵ It is easy to show that the maximum of the E-step in (3.6) is given by the smoothing

⁵For details on convergence, we refer to the textbook treatment in McLachlan & Krishnan (2007).

distribution $p(h|\theta^{(l-1)}, y)$ since then $F(\tilde{p}, \theta)$ equals $\mathcal{L}(\theta)$. The M-step in equation (3.7) reduces to maximizing the criterion function:

$$Q(\theta; \theta^{(l)}) = E_{\theta^{(l-1)}} (\log p(y|h, \theta)p(h|\theta)),$$

where the expectation is taken with respect to $\tilde{p}^{(l)}(h)$. Maximization of the complete data likelihood $\mathcal{L}_c(\theta) = \log p(y|h, \theta)p(h|\theta)$ is easy in the heteroskedastic SVAR model. Unfortunately, a straightforward application of the EM principle is not possible for the SV model since the smoothing density $p(h|\theta^{(l-1)}, y)$ necessary in the E-step is not tractable. We propose two approaches to approximate this density, one based on an analytical approximation and the other based on Monte Carlo integration. Our analytical approximation uses:

$$\tilde{p}^{(l)}(h) = \pi_G^c(h_i|\varepsilon_i, \theta^{(l-1)}), \quad (3.8)$$

which is the Gaussian approximation of the smoothing distribution that we already introduced as importance density. As highlighted by Neal & Hinton (1998), it is not necessary to work with the exact smoothing distribution to get monotonic increases in the likelihood function $\mathcal{L}(\theta)$. Neal & Hinton (1998) show that in fact, $F(\tilde{p}, \theta) = \mathcal{L}(\theta) - D_{KL}(\tilde{p}(h)||p(h|y, \theta))$ 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 $F(\tilde{p}, \theta)$ will yield 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 for more details we refer to Appendix C.1.

The second approach is based on Monte Carlo integration in the E-step, drawing on results of Kim et al. (1998).⁶ It is based on the linearized state equation of the r heteroskedastic structural shocks:

$$\log(\varepsilon_{it}^2) = h_{it} + \log(\eta_t^2), \quad (3.9)$$

where $\eta_t \sim N(0, 1)$, $E(\log(\eta_t^2)) = -1.2704$ and $\text{Var}(\log(\eta_t^2)) = \frac{\pi^2}{2}$. Kim et al. (1998) propose to approximate the $\log \chi^2$ -distribution of the linearized state equation by a mixture of seven normal distributions. The mixture is specified as:

$$p(\log(\eta_t^2)|z_{it} = k) \approx \mathcal{N}(\log(\varepsilon_{it}^2); m_k, v_k^2), \quad (3.10)$$

$$p(z_{it} = k) = p_k, \quad (3.11)$$

with mixture parameters p_k, m_k, v_k^2 tabulated in Table 5 of Appendix C.2. In the following, this mixture representation is exploited to get a Monte Carlo approximation of the E-step. Therefore, let $z_t = [z_{1t}, \dots, z_{rt}]'$ and $z = [z_1, \dots, z_T]'$ be the collection of mixture indicators. Given R random samples of $z^{(j)}$, a Monte Carlo smoothing distribution is given as:

$$p(h_i|\theta, y) \approx \frac{1}{R} \sum_{j=1}^R p(h_i|\theta, y, z_i^{(j)}), \quad (3.12)$$

⁶See also Mahieu & Schotman (1998) for a similar Monte Carlo EM algorithm to estimate a univariate SV model.

where $p(h_i|\theta, y, z_i^{(j)})$ is Gaussian with tractable mean and variance. The random samples of z are generated efficiently by MCMC, involving iteratively sampling between $p(h_i|z_i, \theta^{(l-1)})$ and $p(z_i|h_i, \theta^{(l-1)})$. For computational reasons, we rely on the precision sampler of Chan & Jeliazkov (2009) which exploits the sparsity in the precision matrix and allows for straightforward extension to implement the linear normalizing constraint on h_i . The M-step for the Monte Carlo EM algorithm reduces to maximizing the criterion function:

$$\tilde{Q}(\theta; \theta^{(l)}) = \frac{1}{R} \sum_{j=1}^R \mathbb{E}_{z^{(j)}, \theta^{(l-1)}} \mathcal{L}_c(\theta),$$

where expectation is taken with respect to $p(h_i|\theta^{(l-1)}, y, z_i^{(j)})$. In the remainder, we call the Monte Carlo based algorithm EM-2. For details on the MCMC algorithm and the M-steps, we refer to Appendix C.2.

3.3 Standard Errors

We compute standard errors for the model parameters θ based on the estimated observed information matrix. For algorithm EM-1, we evaluate the likelihood in closed form based on the Gaussian approximation used in the E-step. Based on Bayes Theorem:

$$\log p(\varepsilon_i|\theta) \approx \log p(\varepsilon_i|h_i, \theta) + \log p(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i), \quad (3.13)$$

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 (3.13) the complete log likelihood is approximately given as:

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

which is numerically differentiated twice with respect to the parameter vector θ to obtain the estimated observed information matrix $I(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}$. Based on standard asymptotic theory for ML estimation, $\text{Cov}(\hat{\theta}) \approx I(\hat{\theta})^{-1}$.

For the Monte Carlo based algorithm EM-2, the computation of standard errors is more involved. We use Louis Identity (Louis; 1982) for the observed information matrix:

$$I(\hat{\theta}) = \mathbb{E} \left[I_c(\hat{\theta}) | y \right] - \text{Cov}(S_c(\theta) | y) \Big|_{\theta=\hat{\theta}}, \quad (3.14)$$

where $I_c(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}$, $S_c(\theta) = \frac{\partial \mathcal{L}_c(\theta)}{\partial \theta}$.

Note that the integrals in the expected value and variance of (3.14) are with respect to the smoothing distribution at the ML estimator $p(h|y, \hat{\theta})$ which is intractable for the SV model. However, based on simulated values of the mixture indicators $z^{(j)} (j = 1, \dots, R)$, Monte Carlo integration is feasible with:

$$\mathbb{E} \left[I_c(\hat{\theta}) | y \right] \approx \frac{1}{R} \sum_{j=1}^R -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \theta \partial \theta'} \Big| z^{(j)} \right]_{\theta=\hat{\theta}},$$

$$\text{Cov}(S_c(\theta) | y) \Big|_{\theta=\hat{\theta}} \approx \frac{1}{R} \sum_{j=1}^R \mathbb{E} \left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \theta} \frac{\partial \mathcal{L}_c(\theta)}{\partial \theta'} \Big| z^{(j)} \right]_{\theta=\hat{\theta}},$$

where the second approximation holds since $E(S_c(\theta)|y)|_{\theta=\hat{\theta}} = 0$. The integrals required to compute the expected values are with respect to $p(h|\hat{\theta}, z^{(j)})$ which are a collection of Gaussian distributions. Therefore, their moments can be computed by standard Kalman-smoother recursions. The derivatives necessary to compute the Louis Method are available in Appendix C.3.

4 Testing for Identification

As discussed in section 2, the structural impact matrix B is fully identified by heteroskedasticity if at most one structural shock is homoskedastic, that is if $r \geq K - 1$. If the goal is to identify the structural parameters by the proposed SV-SVAR, one has to ensure that $r \geq K - 1$ for the dataset at hand. In the following, we discuss a testing strategy in order to determine whether there is enough heteroskedasticity in the data for full identification of B . We closely follow Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016) who discuss tests for identification in GARCH-SVAR models. Specifically, they consider the following sequence of hypotheses:

$$H_0 : r = r_0 \quad \text{vs} \quad H_1 : r > r_0, \quad (4.1)$$

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

The testing problem given in (4.1) is nonstandard since parts of the B matrix are identified only under the alternative. Therefore, we follow Lanne & Lütkepohl (2008) who suggest a testing procedure which requires estimation only under H_0 . In particular, suppose that r_0 is the true value, and divide the structural shocks $\varepsilon_t = B^{-1}u_t = [\varepsilon'_{1t}, \varepsilon'_{2t}]'$ where $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Then, under the null hypothesis the first r_0 structural shocks ε_{1t} are allowed to be heteroskedastic while all other shocks ε_{2t} are assumed to be white noise.

Lanne & Saikkonen (2007) propose the following test statistics to check for remaining heteroskedasticity in ε_{2t} . The first statistic is denoted by Q_1 and tests the autocovariances up to a prespecified horizon H of the following time series vector:

$$\xi_t = \varepsilon'_{2t}\varepsilon_{2t} - T^{-1} \sum_{t=1}^T \varepsilon'_{2t}\varepsilon_{2t}.$$

Their corresponding test statistic is given as:

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

where $\tilde{\gamma}(h) = T^{-1} \sum_{t=h+1}^T \xi_t \xi_{t-h}$. Lanne & Saikkonen (2007) show that $Q_1(H)$ is asymptotically $\chi^2(H)$ -distributed if the null is true. The second test statistic denoted as Q_2 is based on the autocovariance matrices of the time series vector

$$\vartheta_t = \text{vech}(\varepsilon_{2t}\varepsilon'_{2t}) - T^{-1} \sum_{t=1}^T \text{vech}(\varepsilon_{2t}\varepsilon'_{2t}),$$

where $\text{vech}(\cdot)$ is the half-vectorization operator as defined e.g. in Lütkepohl (2005). The test statistic is given as:

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

where $\tilde{\Gamma}(h) = T^{-1} \sum_{t=h+1}^T \vartheta_t \vartheta_{t-h}'$. According to Lanne & Saikkonen (2007), Q_2 is asymptotically $\chi^2 \left(\frac{1}{4} H (K - r_0)^2 (K - r_0 + 1)^2 \right)$ -distributed under the null hypothesis.

Note that in practice, ε_{2t} is replaced by the estimated residual from the model under H_0 , where $\hat{\varepsilon}_t = \hat{B}^{-1} \hat{u}_t$. Recall that given $B = \begin{bmatrix} B_1 & B_2 \\ K \times r & K \times (K-r) \end{bmatrix}$, B_2 is not identified for $r < K - 1$. Therefore, to uniquely identify B under the sequence of H_0 's where $r_0 < K - 1$, we impose zero constraints on B_2 . In particular, let $B_2 = \begin{pmatrix} B_{21} \\ B_{22} \end{pmatrix}$, we estimate the model under H_0 imposing a lower triangular structure on the $(K - r) \times (K - r)$ dimensional block B_{22} . As we show in Corollary 2 of Appendix A, this suffices to uniquely identify B if $r < K - 1$.

5 Monte Carlo Comparison

In the following, we provide simulation evidence that the proposed SV-SVAR model works very well in estimating the structural parameters also under model misspecification. Therefore, we set up a stylized comparative Monte Carlo experiment where we simulate datasets from four different heteroskedastic DGPs. For each dataset, we estimate the structural parameters by the SV-SVAR model as well as by the main alternative models: a Breakpoint SVAR (BP-SVAR), a Markov Switching model (MS-SVAR) and a GARCH model (GARCH-SVAR).⁷ To compare their performance, we study cumulated Mean Squared Error (MSE) of the structural impulse response functions which are probably the most widely used tool in SVAR analysis.

In our Monte Carlo design we closely follow Lütkepohl & Schlaak (2017) who study model selection of heteroskedastic SVARs by information criteria. Specifically, we simulate time series of length $T \in \{200, 500\}$ generated by the bivariate VAR(1) process:

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

with $E(u_t u_t') = \Sigma_t = B \Lambda_t B'$ for $t = 1, \dots, T$ and

$$A_1 = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ -1 & 10 \end{pmatrix}.$$

For the diagonal matrix Λ_t , we specify four different heteroskedastic processes listed below.

1. **Breakpoint DGP:** In the first DGP, we model a one time shift in the variance. In particular, $\Lambda_t = I_K$ for $t = 1, \dots, T/2$ and $\Lambda_t = \text{diag}([2, 7]')$ for $t = T/2 + 1, \dots, T$.

⁷For computational reasons, the SV-SVAR model is estimated by EM-1. The breakpoint is estimated by maximizing the likelihood function over a grid of possible breakdates. The GARCH model is based on univariate GARCH(1,1) processes and is estimated as discussed in Lanne & Saikkonen (2007).

2. **Markov Switching DGP:** In the second DGP, we simulate a MS model with two regimes and transition probability matrix:

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

Based on simulated states $s_1, \dots, s_T \in \{1, 2\}$, $\Lambda_{s_t=1} = I_K$ and $\Lambda_{s_t=2} = \text{diag}([2, 7]')$.⁸ The shifts in variances are therefore of equal magnitude than in the Breakpoint DGP.

3. **GARCH DGP:** For the GARCH DGP, we set $\Lambda_t = \text{diag}([\lambda_{1t}, \lambda_{2t}]')$. Each diagonal entry follows a univariate GARCH(1,1) process, that is:

$$\lambda_{it} = c_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \lambda_{i,t-1}, \quad i \in \{1, 2\},$$

where $\varepsilon_{it} = \sqrt{\lambda_{it}} \eta_{it}$ is the i -th structural shock at time t and $\eta_{it} \sim \mathcal{N}(0, 1)$. To ensure that the structural shocks have an unconditional variance of one, we set the intercept to $c_i = (1 - \alpha_i - \beta_i)$. For the remaining GARCH parameters, we set $\alpha_i = 0.05$ and $\beta_i = 0.94$ ($i = 1, 2$) which correspond to values typically estimated for empirical data.⁹

4. **SV DGP:** In the last 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 $\mu_i = 0.5s_i/(1 - \phi_i^2)$ to ensure that $E(\varepsilon_{it}^2) = 1$, $\phi_i = 0.95$ and $s_i = 0.04$ ($i = 1, 2$).¹⁰

The number of simulated series is set to $M = 2000$. For each of the models we estimate A_1 and B based on the m -th simulated series. Based on these estimates, we compute impulse response functions $\hat{\Theta}_i(m) = \hat{A}_1(m)^i \hat{B}(m)$ for $i \geq 0$. Consequently, following the notation of Lütkepohl (2005), the elements of $\hat{\Theta}_i(m)$, $\hat{\theta}_{jk,i}(m)$ ($j, k \in \{1, 2\}$), denote the estimated impulse response functions in variable j caused by a structural innovation k after i periods based on simulation m . The true impulse response functions $\theta_{jk,i}$ are calculated in the same fashion using the true parameter values. Based on these quantities, we compute the cumulated MSE of response functions up to horizon h :

$$\text{MSE}(\theta_{jk})_h = \frac{1}{M} \sum_{m=1}^M \left(\sum_{i=0}^h \left(\hat{\theta}_{jk,i}(m) - \theta_{jk,i} \right)^2 \right). \quad (5.1)$$

Due to unequal normalizing constraints among the volatility models, we rescale the impulse responses functions always considering unit shocks.

⁸To ensure identification through heteroskedasticity, we only accept simulation draws where $\sum_{t=1}^T \mathbb{1}_{\{s_t=i\}} \geq 0.25T$ for $i = 1, 2$.

⁹Note that the choice of GARCH parameters corresponds to an unconditional kurtosis of ε_{it} equal to four. To ensure sufficiently heteroskedastic structural shocks we only accept draws with a sample kurtosis of ε_i of at least 3.5.

¹⁰This choice corresponds to a fairly persistent time series with an expected kurtosis of around 4.5. To ensure sufficiently heteroskedastic structural shocks we only accept draws with a sample kurtosis of at least 3.5.

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	1.17	1.17	2.95	2.18	1.03	1.03	2.39	1.19
	GARCH	1.11	1.13	4.48	2.24	1.09	1.11	1.73	1.14
	SV	0.98	0.95	3.39	1.52	1.01	1.02	3.09	1.22
MS-DGP	BP	1.37	1.49	5.37	3.43	2.30	2.71	15.95	7.20
	MS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	GARCH	1.27	1.38	7.68	4.29	1.74	1.95	7.79	4.25
	SV	0.91	0.92	1.32	0.98	1.02	1.01	1.49	1.21
GARCH-DGP	BP	1.05	1.13	1.51	1.13	1.25	1.40	2.43	1.62
	MS	1.22	1.24	1.58	1.06	1.32	1.38	2.42	2.18
	GARCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SV	0.88	0.87	0.39	0.52	1.02	1.03	1.08	1.09
SV-DGP	BP	1.19	1.27	3.53	1.78	1.67	1.92	15.05	3.37
	MS	1.31	1.32	2.52	1.70	1.45	1.66	10.48	2.72
	GARCH	1.04	1.06	1.56	1.30	1.07	1.10	1.66	1.17
	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 (5.1) and displayed relative to true model MSEs.

In Table 1, the results of the simulation study are provided for a horizon of $h = 5$. For improved readability, we report relative MSEs in comparison to the correctly specified model. We find that the SV-SVAR model performs pretty well regardless of the true DGP or the sample size. The relative MSEs of this model are close to or even below the correctly specified model for almost all θ_{jk} 's, time series lengths and DGPs. One exception, however, are the relative MSEs for θ_{21} with data simulated from a Breakpoint DGP.

If the SV-SVAR is compared to other misspecified models for a certain DGP, we find it to perform better in two out of three DGPs. For a MS DGP as well as a GARCH DGP, all impulse responses estimated by a SV-SVAR have lower cumulative MSE than if they were estimated by the other misspecified models.¹¹ Only in the BP DGP there is no clear advantage with respect to the other misspecified models.

Summing up, our small simulation study yields promising results indicating that choosing a SV-SVAR model can be a good idea if the heteroskedasticity pattern is smoothly changing. However, one might want to be more careful if there is data with a sudden shift in the variance. For those cases, one should also think about a simpler breakpoint model. Alternatively, allowing for a unit root in the log variances could help as well. We leave this

¹¹For the GARCH-DGP with $T = 200$, it even performs better than the correctly specified model. This is likely be driven by the difficulty to estimate the GARCH parameters with few observations.

specification for future research.

6 Interdependence between Monetary Policy and Stock Markets

Since the seminal contribution of Sims (1980), structural VAR models represent a frequently applied instrument to investigate the dynamic causal of monetary policy.¹² To identify the corresponding structural shock, the most simple way involves a Cholesky decomposition of the reduced form VAR covariance matrix with the policy variable ordered last in the VAR model (Christiano, Eichenbaum & Evans; 1999; Bernanke, Boivin & Elias; 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 ordered above react with at least one lag to the 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 part of 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 & Leitimo (2009) propose a way to identify a monetary policy shock in the presence of stock market returns in their model. More specifically, they identify a monetary policy shock with a combination of short- and long-run restrictions on the effects of the policy shock. Besides zero impact restrictions on real variables, a monetary policy shock is not allowed to have a long-term impact on stock market returns in their model. This additional restriction allows to disentangle the monetary policy from financial shocks.

Another promising way to address identification in presence of fast moving variables are Proxy SVARs based on external instruments (Mertens & Ravn; 2013; Stock & Watson; 2012). 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) or Coibion (2012)) or changes in high frequency future prices in a narrow window around FOMC meetings (e.g. Faust, Swanson & Wright (2004); Nakamura & Steinsson (2013); Gertler & Karadi (2015)).¹³

Finally, heteroskedasticity can be exploited to identify the interdependence between monetary policy and financial variables. For example, Rigobon & Sack (2003) combine 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 higher variance

¹²See e.g. Ramey (2016) for an extensive overview of the literature.

¹³Yet another branch of the literature relies on sign restrictions of the impulse response functions (Faust; 1998; Canova & De Nicoló; 2002; Uhlig; 2005) or on a combination of sign restrictions and information in proxy variables (Braun & Brüggemann; 2017).

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 (2017) review various SVAR models subject to conditional heteroskedasticity and use them to test the combination of exclusion restrictions employed by Bjørnland & Leitemo (2009).¹⁴

In the remainder of this section, we follow Lütkepohl & Netšunajev (2017) and revisit the analysis of Bjørnland & Leitemo (2009) using the proposed 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.

6.1 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 are 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 E. We follow Lütkepohl & Netšunajev (2017) in using 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):

$$B = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \quad \text{and} \quad \Xi_\infty = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{bmatrix}, \quad (6.1)$$

where $\Xi_\infty = (I_K - A_1 - \dots - A_p)^{-1}B$ is the long-run impact matrix of the structural shocks on y_t . Note that an asterisk means that the corresponding entries in B and Ξ_∞ matrix are left unrestricted. The last column of B corresponds to the reaction of y_t to a monetary policy shock. Economic activity, consumer and commodity prices are only allowed to respond with one lag to a monetary policy shock, while the stock markets are allowed to move contemporaneously. However, in the long run, a monetary policy shock is assumed to have a neutral 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 adjust the interest rates within the same period. The remaining shocks do not have an economic interpretation. To uniquely identify the model, Bjørnland & Leitemo (2009) disentangle these shocks by a simple recursivity assumption.

¹⁴See also Lütkepohl & Netsunajev (2014) for a similar analysis based on a Smooth Transition SVAR model only.

To be in line with Lütkepohl & Netšunajev (2017), the following restrictions are tested since they are overidentifying in a heteroskedastic model:

R1: Both, B and Ξ_∞ restricted as in (6.1).

R2: Only the last two columns of B and Ξ_∞ are restricted as in (6.1).

R3: Only B is restricted as in (6.1).

R4: Only Ξ_∞ is restricted as in (6.1).

In addition to test R1 against R3 as in Lütkepohl & Netšunajev (2017), we add R4 as a more natural way to test the reasonability of the long-run restriction. We further contribute to the literature by testing Proxy SVAR restrictions that arise if an external instrument z_t 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 linear restrictions on β_1 , denoting the first column of B :

$$\beta_{21} = (\Sigma_{zu'_1}^{-1} \Sigma_{zu'_2})' \beta_{11}. \quad (6.2)$$

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.¹⁵ 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 if the first column of B is identified via either RR's or GK's instrument:

R5rr: IV moment restrictions (6.2) based on the RR shock.

R5gk: IV moment restrictions (6.2) 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 baseline series which is constructed based on the 3 months ahead monthly fed funds futures.¹⁶ Time series plots of both series are available in Appendix E.

6.2 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

¹⁵In 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 $N_z = \sum_{t=1}^T D_t$.

¹⁶We repeat our analysis for the other instruments available in Gertler & Karadi (2015). The results do not change qualitatively.

Table 2: Model Selection by Information Criteria

	Linear	SV EM1	SV EM2	GARCH	STVAR	MS2	MS3
$\ln L$	-3159.34	-2692.24	-2692.34	-2762.57	-2878.25	-2827.39	-2775.23
AIC	6508.69	5614.48	5614.68	5755.14	5980.51	5878.79	5792.46
BIC	6898.43	6086.27	6086.52	6226.94	6439.99	6338.27	6288.87

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 importance sampling gives a range of $[-2692.29, -2692.21]$ and $[-2692.38, -2692.28]$ for $\ln L$, respectively.

residual plots, we compare a SV, a GARCH, a Markov Switching and a Smooth Transition (ST) model. Since we use exactly the same data set as Lütkepohl & Netšunajev (2017) we can directly compare the results except for the SV-SVAR model.

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 it does not matter for the likelihood value of the SV model whether we use the deterministic approximation (EM-1) or a Monte Carlo based E-step (EM-2). Both algorithms yield almost identical likelihood values. To assess the Monte Carlo error of the estimates, we also report a range of values that arise by re-estimating the likelihood 20 times based on $R = 30\,000$ draws of the importance density.¹⁷ Comparing the different models, our results suggest that including time-variation in the second moment is strongly supported by all information criteria. Moreover, among the heteroskedastic models we find that particularly models also used in finance are favored, that is the GARCH and SV model. This might not be surprising given that stock market returns are included in the system. Among these two, the SV model performs slightly better in terms of information criteria. Our results deviate from those of Lütkepohl & Netšunajev (2017) who find that a MS(3) model provides the best description for this dataset. The difference is likely due to the maximization procedures used for the challenging GARCH likelihood. While Lütkepohl & Netšunajev (2017) rely on a sequential estimation procedure we take this to provide starting values and further attempt to compute a local maximum (see Lanne & Saikkonen (2007) for details). Overall, model selection by IC suggests that the SV-SVAR model provides the best description of the data.

In accordance with Lütkepohl & Netšunajev (2017), we also look at standardized residuals as an additional model checking device. Figure 1 provides a plot for the reduced form residuals \hat{u}_t from the linear model, as well as standardized residuals from all models computed as $\hat{u}_{it}/\hat{\sigma}_{ii,t}$ where $\hat{\sigma}_{ii,t}^2$ is the i -th diagonal entry of the estimated covariance matrices $\hat{\Sigma}_t$. These plots clearly suggest that none of the other methods is fully satisfactory in yielding standardized residuals that seem homoskedastic and approximately normally distributed.

¹⁷A 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.

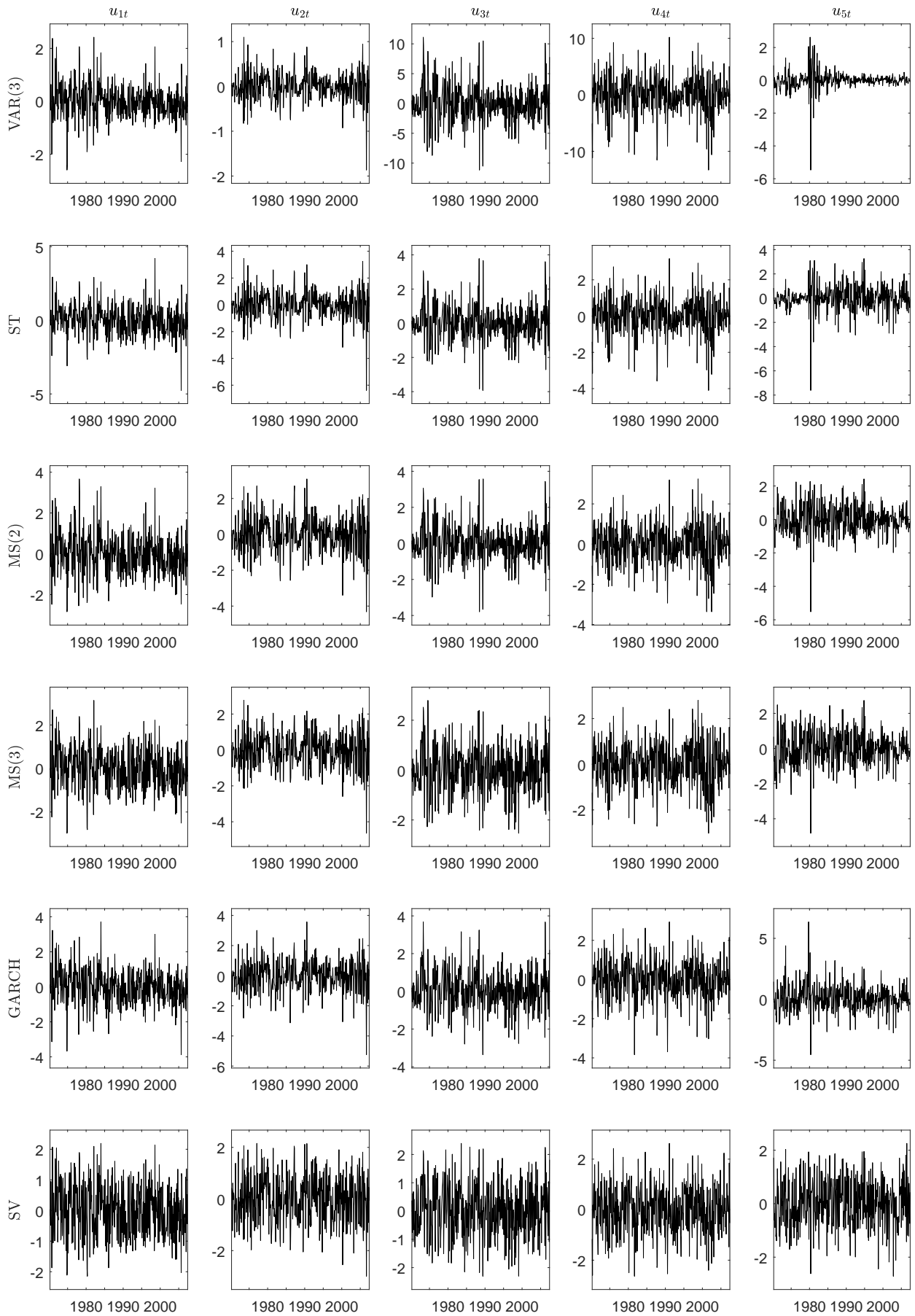


Figure 1: Reduced form residuals for linear VAR model and standardized Residuals of ST-, MS(2)-, MS(3)-, GARCH- and SV-SVAR model.

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
$r_0 = 1$	23.82	1	0.00	250.03	100	0.00
$r_0 = 2$	29.40	1	0.00	140.62	36	0.00
$r_0 = 3$	18.31	1	0.00	43.79	9	0.00
$r_0 = 4$	17.27	1	0.00	17.27	1	0.00
	$Q_1(3)$	dof	p -value	$Q_2(3)$	dof	p -value
$r_0 = 0$	52.34	3	0.00	1433.73	675	0.00
$r_0 = 1$	39.67	3	0.00	528.79	300	0.00
$r_0 = 2$	32.70	3	0.00	221.40	108	0.00
$r_0 = 3$	20.21	3	0.00	60.93	27	0.00
$r_0 = 4$	19.83	3	0.00	19.83	3	0.00

Note: Sequence of tests to check the number of heteroskedastic shocks in the system as introduced in section 4 (Lanne & Saikkonen; 2007).

However, for the SV-SVAR model, standardized residuals seem well behaved with no apparent heteroskedasticity, most of the residuals located between -2 and 2 and virtually no outliers.¹⁸ We conclude that the proposed SV model seems to be the most suitable for our application and continue our analysis based on this model.

In order to be able to test restrictions R1-R5 as overidentifying, it is necessary to have enough heteroskedasticity in the data for full identification of B . Recall that for this to hold, we need at least $r = K - 1$ structural shocks with time-varying variances. As described in section 4, we apply a testing strategy based on a sequence of tests with $H_0 : r = r_0$ against $H_1 : r > r_0$ for $r_0 = 0, 1, \dots, K - 1$. The results of the tests are reported in Table 3. We find that both tests indicate that there is enough heteroskedasticity in the data, and given that each null hypothesis is rejected there is substantial evidence that r equals K in our analysis.

Because of the strong statistical evidence for full identification through heteroskedasticity, we continue our analysis and test the economically motivated restrictions R1-R5 as overidentifying. In Table 4 we provide Likelihood Ratio (LR) test statistics for the restrictions introduced previously.¹⁹ 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. Since we estimate the likelihood values with the help of importance sampling, we assess the Monte Carlo error by re-estimating the likelihoods 20 times and report a range of corresponding p -values.

In line with the findings of Lütkepohl & Netšunajev (2017), our results suggest that R1, the restrictions of Bjørnland & Leitemo (2009) are rejected by the data. To make sure that

¹⁸Formal Jarque-Bera tests on the standardized residuals indeed provide no evidence against normality.

¹⁹The likelihood ratio test statistic is given as $LR = 2(\ln L_{uc} - \ln L_c)$ where $\ln L_c$ is the log likelihood value under the restrictions (H_0), and $\ln L_{uc}$ is the unconstrained log likelihood under the alternative (H_1).

Table 4: Test for Overidentifying Restrictions

H_0	H_1	LR	dof	p -value	p_{\min}	p_{\max}
R1	UC	25.854	10	0.005	0.005	0.006
R2	UC	22.982	7	0.002	0.002	0.002
R3	UC	24.245	9	0.004	0.004	0.004
R1	R3	1.609	1	0.205	0.189	0.240
R4	UC	0.701	1	0.402	0.350	0.460
R5rr	UC	6.398	4	0.171	0.164	0.183
R5gk	UC	256.505	4	0.000	0.000	0.000

Note: For details about overidentifying restrictions see subsection 6.1. Likelihood ratio test statistics are computed as $2(\ln L_{H_1} - \ln L_{H_0})$ and are χ^2 -distributed under H_0 .

this result does not come from the lower triangular block corresponding to the non-identified shocks, Lütkepohl & Netšunajev (2017) 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 (2017), 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 these in the short run. To confirm this result, we also test R4 which corresponds to the long-run restriction on its own. Again, it cannot be rejected which confirms the previous finding.

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 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_t with $\hat{\varepsilon}_t$, the estimated orthogonal shocks of the unconstrained SV-SVAR model. For GK, we find $\text{Corr}(z_t^{GK}, \hat{\varepsilon}_t) = (0.039, -0.066, 0.048, -0.242, 0.430)$, while for RR, $\text{Corr}(z_t^{RR}, \hat{\varepsilon}_t) = (0.042, 0.004, 0.028, -0.017, 0.453)$. 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 more shock. This clearly violates the exogeneity condition of the instrument. Thereby, our results support the argument of Ramey (2016) who questions the exogeneity of 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 $\text{Corr}(z_t^{RR}, \hat{\varepsilon}_{5t}) = 0.45$. In Figure 2 we plot impulse response functions (IRFs) up to 72 months (6 years) of

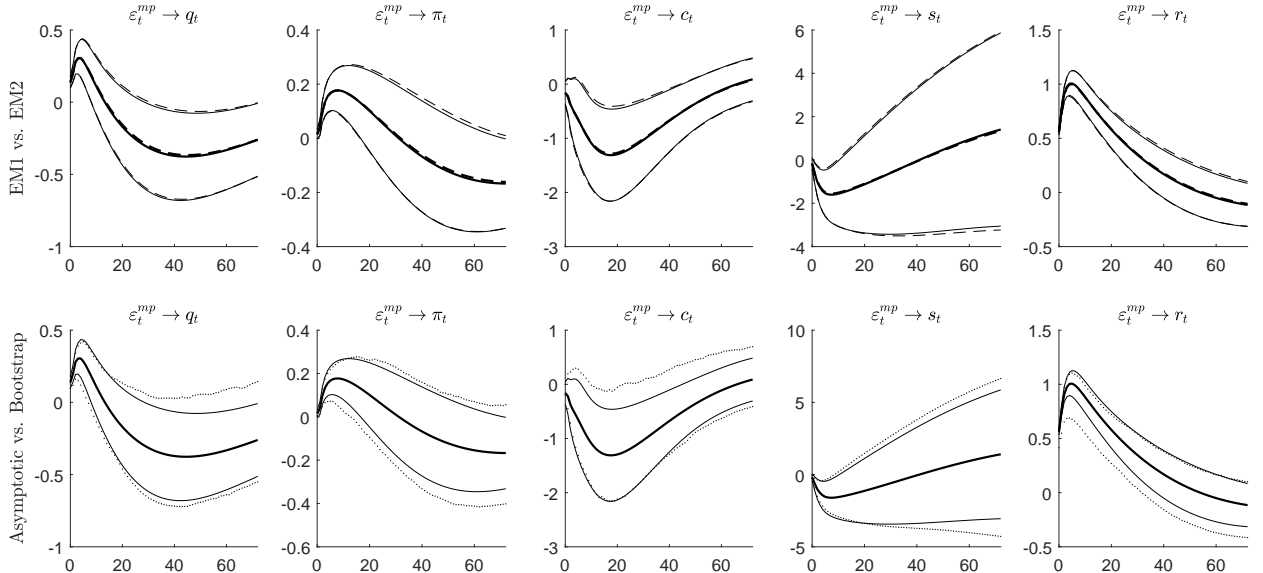


Figure 2: IRFs up to a horizon of 72 months of a monetary policy shock with 68% confidence bounds. The first row plots estimates based on EM-1 (solid line) and EM-2 (dashed line) with corresponding asymptotic confidence intervals. The second row compares asymptotic (solid line) and wild bootstrap (dotted line) confidence bounds both computed based on EM-1.

the system variables in response to a monetary policy shock. Besides mean estimates, we provide 68% asymptotic confidence intervals as well as bounds based on a fixed design wild bootstrap that preserves the second moment properties of the residuals. For details on their computation we refer to Appendix D. Again, we note that there is virtually no difference in using EM-1 or EM-2 to compute the estimates and corresponding standard errors. The IRFs and their asymptotic confidence bounds coincide for all variables at all horizons. In line with the IRFs computed by Lütkepohl & Netšunajev (2017) 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, which seems reasonable given that one would expect a shift in demand towards risk free assets.

7 Conclusion

In this paper, we propose to use a stochastic volatility model to identify parameters of SVAR models by heteroskedasticity. In particular, we assume that the log variance of each structural shock is random and evolves according to an AR(1). Conditions for full and partial identification of the SV-SVAR model are discussed and in order to check whether they are satisfied for a given dataset, a formal testing procedure is provided. With respect to estimation, we develop two EM algorithms for Maximum Likelihood inference. The first algorithm is based on a Laplace approximation of the intractable E-step, while the second is based on Monte Carlo integration. While we leave the choice of algorithm to individual

preferences, we experience that in practice little can be gained by using the Monte Carlo based method. For computational reasons, we therefore recommend the usage of the former. In a small Monte Carlo study, we compare cumulative MSEs of impulse response functions estimated by the proposed model with those obtained by other possible specifications for the variance. The results are promising, and we find that the SV model is very flexible and works comparatively well in identifying the structural parameters also under model misspecification.

In an empirical application, we revisit the model of Bjørnland & Leitemo (2009) who rely on a combination of short- and long-run restrictions to identify monetary policy and stock market shocks. For their dataset, formal model selection supports a SV specification in the variance if compared to other heteroskedastic SVARs. We use the SV-SVAR model to test the identifying restrictions of Bjørnland & Leitemo (2009) as overidentifying. In line with findings of Lütkepohl & Netšunajev (2017) who test the same restrictions based on various existing heteroskedastic SVAR models, all types of short-run restrictions considered are rejected. However, in contrast to Lütkepohl & Netšunajev (2017), we do not reject the long-run restriction. Besides these exclusion restrictions, we also test the idea of using external instruments to identify the monetary policy shock. We find that identification by the instrument of Gertler & Karadi (2015) is rejected in our model. In turn, there is no evidence against identification via the narrative series of Romer & Romer (2004).

References

- Bernanke, B. S., Boivin, J. & Eliasziw, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**(1): 387–422.
- Bernanke, B. S. & Mihov, I. (1998). Measuring monetary policy, *The 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, *The American Economic Review* **79**(4): 655–673.
- Braun, R. & Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Technical Report 16*, Department of Economics, University of Konstanz.
- Broto, C. & Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey, *Journal of Economic Surveys* **18**(5): 613–649.
- 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 Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **49**: 1131–1159.
- Chan, J. 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. & 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. & 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. (1999). Monetary policy shocks: What have we learned and to what end?, in J. Taylor & M. Woodford (eds), *The Handbook of Macroeconomics*, Amsterdam: Elsevier Science Publication, chapter 2, 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, Elsevier, pp. 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, *The 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.
- Gonçalves, S. & Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**(1): 89–120.
- Harvey, A., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models, *The 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.
- Justiniano, A. & Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *The 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, *The 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. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**(6): 1131–1149.
- Lanne, M., Lütkepohl, H. & Maciejowska, K. (2010). Structural vector autoregressions with markov switching, *Journal of Economic Dynamics and Control* **34**(2): 121–131.
- Lanne, M. & Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**(1): 61–75.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **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. & Netsunajev, A. (2014). Structural vector autoregressions with smooth transition in variances: The interaction between us monetary policy and the stock market, *Technical Report 1388*, DIW Berlin.
- Lütkepohl, H. & Netšunajev, A. (2017). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**(Supplement C): 2 – 18.
- Lütkepohl, H. & Schlaak, T. (2017). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Technical Report 1672*, DIW Berlin.
- Mahieu, R. J. & Schotman, P. C. (1998). An empirical application of stochastic volatility models, *Journal of Applied Econometrics* **13**(4): 333–360.
- 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.
- Nakamura, E. & Steinsson, J. (2013). High frequency identification of monetary non-neutrality: The information effect, *Technical Report 19260*, National Bureau of Economic Research.

- Neal, R. M. & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants, *Learning in graphical models*, 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, *The Review of Economic Studies* **72**(3): 821–852.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation, in J. B. Taylor & H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 8, pp. 71–162.
- Rigobon, R. (2003). Identification through heteroskedasticity, *The Review of Economics and Statistics* **85**(4): 777–792.
- Rigobon, R. & Sack, B. (2003). Measuring the reaction of monetary policy to the stock market, *The Quarterly Journal of Economics* **118**(2): 639–669.
- 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. (2007). Approximate bayesian inference for hierarchical gaussian markov random field models, *Journal of Statistical Planning and Inference* **137**(10): 3177 – 3192. Special Issue: Bayesian Inference for Stochastic Processes.
- 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**: 1–48.
- 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, *Technical Report 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.

A Identification

To ensure identification of B in (2.2) we show that under sufficient heterogeneity in the second moments, there is no B^* different from B except for column permutations and sign changes which yields an observationally equivalent model with the same second moment in u_t for all $t = 1, \dots, T$.

Proposition 1. *Let $\Sigma_1 = BB'$ and $\Sigma_t = BV_t^*B'$ ($t = 2, \dots, T$), where $B = [B_1, B_2]$ with $B_1 \in \mathbb{R}^{K \times r}$, $B_2 \in \mathbb{R}^{K \times (K-r)}$ and $V_t^* = \text{diag}(v_{1t}, \dots, v_{rt}, 1_{K-r})$ be nonsingular $K \times K$ covariance matrices. If for $r < K$*

$$\forall i \in \{1, \dots, r\} : \forall j \neq i \in \{1, \dots, K\} : \exists t \in \{2, \dots, T\} : v_{it} \neq v_{jt} \quad (\text{A.1})$$

holds, matrix B_1 is unique up to multiplication of its columns by -1 .

Proof. Suppose $Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{pmatrix}$, where $Q_1 \in \mathbb{R}^{r \times r}$, $Q_2, Q_3 \in \mathbb{R}^{(K-r) \times r}$ and $Q_4 \in \mathbb{R}^{(K-r) \times (K-r)}$

$$\text{satisfies} \quad \Sigma_1 = BB' = BQQ'B \quad \text{and} \quad (\text{A.2})$$

$$\Sigma_t = BV_t^*B' = BQV_t^*Q'B' \quad (t = 2, \dots, T). \quad (\text{A.3})$$

From (A.2) directly follows that Q is an orthogonal matrix, i.e. $QQ' = I_K$ what implies

$$Q_1Q_1' + Q_3Q_3' = I_r, \quad (\text{A.4})$$

$$Q_2Q_1' + Q_4Q_3' = 0, \quad (\text{A.5})$$

$$Q_2Q_2' + Q_4Q_4' = I_{K-r}. \quad (\text{A.6})$$

Furthermore, as $V_t^* = \begin{pmatrix} \Lambda_t & 0 \\ 0 & I_{K-r} \end{pmatrix}$ with $\Lambda_t = \text{diag}(v_{1t}, \dots, v_{rt})$, (A.3) yields

$$Q_1\Lambda_tQ_1' + Q_3Q_3' = \Lambda_t \quad \xrightarrow{(\text{A.4})} \quad Q_1 \underbrace{(I_r - \Lambda_t)}_{=: \Lambda_t^*} Q_1' = \Lambda_t^*, \quad (\text{A.7})$$

$$Q_2\Lambda_tQ_1' + Q_4Q_3' = 0 \quad \xrightarrow{(\text{A.5})} \quad Q_2\Lambda_tQ_1' = Q_2Q_1'. \quad (\text{A.8})$$

Let q_{1i} ($i = 1, \dots, r$) be the rows of Q_1 . Due to (A.7), $q_{1i}\Lambda_t^*q_{1i}' = 1 - v_{it}$ has to hold for all i and t . Because of (A.1) for all i there exists a $t \in \{2, \dots, T\}$ with $v_{it} \neq 1$, so $q_{1i} \neq 0$ has to hold for all $i = 1, \dots, r$. Moreover, because $q_{1i}\Lambda_t^*q_{1j}' = 0$ holds for all $i \neq j$ and t due to (A.7), $q_{1i} \neq c \cdot q_{1j}$ has to hold for all $c \neq 0$. Therefore, the rows of Q_1 are linearly independent so that Q_1 has full rank and is thus invertible.

With (A.8) and the invertibility of Q'_1 it follows $Q_2\Lambda_t = Q_2$ for all t why Q_2 equals the zero matrix because for any i there exists a t such that $v_{it} \neq 1$ due to (A.1). Using $Q_2 = 0$ and (A.6) directly yields $Q_4Q'_4 = I_{K-r}$, so Q_4 is an orthogonal matrix and therefore invertible. In addition, because of (A.5), $Q_2 = 0$ and the invertibility of Q_4 , Q_3 has to be the zero matrix. Following to that, (A.4) delivers $Q_1Q'_1 = I_r$, i.e. Q_1 is an orthogonal matrix. Consequently, (A.7) reduces to $Q_1\Lambda_tQ'_1 = \Lambda_t$ for all $t \in \{2, \dots, T\}$. Using assumption (A.1) one can show equivalent to Proposition 1 in Lanne et al. (2010) that Q_1 is a diagonal matrix with ± 1 entries on the diagonal. This proves the uniqueness of B_1 apart from sign reversal of its columns. \square

Using Proposition 1 with $V_t^* = V_1^{-1}V_t$ (cf. (2.3)) for $t = 1, \dots, T$ such that $V_1^* = I_K$ shows that an observationally equivalent model with the same second moment properties can be obtained by $B^* = BQ$ with $Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 \in \mathbb{R}^{r \times r}$ a diagonal matrix with ± 1 entries on the diagonal 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 . Furthermore, joint column permutations of B_1 and V_t^* for all $t = 1, \dots, T$ obviously keep the second moment properties.

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 .*

Proof. For $r = K - 1$ matrix Q_4 is a scalar with $Q_4^2 = 1 \Rightarrow Q_4 = \pm 1$. So, full Q is a diagonal matrix with ± 1 entries on the diagonal. This proves the uniqueness of the full matrix B apart from sign reversal of its columns. \square

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

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 is diagonal with ± 1 entries on the diagonal. This proves the uniqueness of B apart from sign reversal of its columns. \square

B Importance Density

To derive the Gaussian approximation $\pi_G(h_i|\varepsilon_i, \theta)$ for $i = 1, \dots, r$ of the importance density we start with Bayes theorem for the true importance density (Chan & Grant; 2016):

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

The prior for the log variances $h_i \sim \mathcal{N}(\delta_i, Q_i^{-1})$ gives the kernel

$$\log p(h_i) \propto -\frac{1}{2} (h_i - \delta_i)' Q_i (h_i - \delta_i). \quad (\text{B.2})$$

Moreover, the conditional distribution of the structural shocks $\varepsilon_{it}|h_{it}, \theta \sim \mathcal{N}(0, \exp(h_{it}))$ yields $\log p(\varepsilon_{it}|h_{it}, \theta) \propto -\frac{1}{2} (h_{it} + \varepsilon_{it}^2 e^{-h_{it}})$, so the partial derivatives are

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

Using that, the nonlinear density $\log p(\varepsilon_i|h_i, \theta)$ can be approximated by a second order Taylor approximation around $\tilde{h}_i^{(0)}$:

$$\begin{aligned} \log p(\varepsilon_i|h_i, \theta) &\approx \log p(\varepsilon_i|\tilde{h}_i^{(0)}, \theta) + (h_i - \tilde{h}_i^{(0)})' f_i - \frac{1}{2} (h_i - \tilde{h}_i^{(0)})' C_i (h_i - \tilde{h}_i^{(0)}) \\ &= -\frac{1}{2} \left(h_i' C_i h_i - 2h_i' \underbrace{(f_i + C_i \tilde{h}_i^{(0)})}_{=: b_i} \right) + \text{constant}. \end{aligned} \quad (\text{B.3})$$

Combining (B.1), (B.2) and (B.3) provides the normal kernel

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

for the smoothing density which gives the its Gaussian approximation:

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

Newton-Raphson method:

The Newton-Raphson method to evaluate $\pi_G^c(h_i|\varepsilon_i, \theta)$ at its mode 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

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

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

C EM Algorithm

To fix notation, define the following quantities:

$$\begin{aligned}
Y^0 &:= (y_1, \dots, y_T) && K \times T, \\
A &:= (\nu, A_1, \dots, A_p) && K \times Kp + 1, \\
Y_t^0 &:= \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} && Kp \times 1, \\
x_t &:= \begin{pmatrix} 1 \\ Y_t^0 \end{pmatrix} && Kp + 1 \times 1, \\
X &:= (x_1, \dots, x_T) && Kp + 1 \times T, \\
y^0 &:= \text{vec}(Y^0) && KT \times 1, \\
\alpha &:= \text{vec}(A) && K(Kp + 1) \times 1, \\
U &:= (u_1, \dots, u_T) && K \times T, \\
u &:= \text{vec}(U) && KT \times 1, \\
V^{-1} &:= (\exp(-h_1), \dots, \exp(-h_T)) && K \times T.
\end{aligned}$$

Then, the VAR can be compactly written as:

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

with $Z = (X' \otimes I_K)$, $E(uu') = \tilde{\Sigma}_u$. 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 $\Sigma_e^{-1} = \text{diag}(\text{vec}(V^{-1})')$.

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

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

Starting values are set in the same way for both algorithms. That is

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

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

$$\begin{aligned}
\hat{\phi} &= [0.9, \dots, 0.9]', \\
\hat{s} &= [0.05, \dots, 0.05]',
\end{aligned}$$

which correspond to persistent heteroskedasticity with initial kurtosis of about four for the structural shocks $\varepsilon_i, i = 1, \dots, r$.

C.1 EM-1

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

1. E-step:

For $i = 1, \dots, r$, compute the approximate smoothing densities $\pi_G^c(h_i|\theta^{(l-1)}, \varepsilon_i)$ as described in Appendix B with mean $\bar{\delta}_i^c$ and variance $\bar{Q}_i^{-1} - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}A_h\bar{Q}_i^{-1}$. Note that directly inverting \bar{Q}_i^{-1} is unnecessary costly since we only need the marginal variances $\text{Var}(h_{it}|\theta^{(l-1)}, \varepsilon_i)$ and the first off-diagonal corresponding to $\text{Cov}(h_{it}, h_{i,t-1}|\theta^{(l-1)}, \varepsilon_i)$. Similar to the Kalman Smoother recursions, they can be obtained without computing the whole inverse using sparse matrix routines (Rue & Martino; 2007).

2. M-step: Taking expectation with respect to the approximations of $p(h_i|\theta^{(l-1)}, \varepsilon_i)$ for $i = 1, \dots, r$ and maximizing yields:

(a) Update ϕ_i and s_i for $i = 1, \dots, r$:

$$\begin{aligned}\phi_i^{(l)} &= \frac{S_{xy}^i}{S_{xx}^i}, \\ s_i^{(l)} &= (T-1)^{-1} \left(S_{yy}^i - 2\phi_i^{(l)}S_{xy}^i + \left(\phi_i^{(l)}\right)^2 S_{xx}^i \right),\end{aligned}$$

with:

$$\begin{aligned}S_{xx}^i &= \sum_{t=1}^{T-1} \left[\text{Var}(h_{it}|\theta^{(l-1)}, \varepsilon_i) + \left(\text{E}(h_{it}|\theta^{(l-1)}, \varepsilon_i) - \mu_i \right)^2 \right], \\ S_{yy}^i &= \sum_{t=2}^T \left[\text{Var}(h_{it}|\theta^{(l-1)}, \varepsilon_i) + \left(\text{E}(h_{it}|\theta^{(l-1)}, \varepsilon_i) - \mu_i \right)^2 \right], \\ S_{xy}^i &= \sum_{t=2}^T \left[\text{Cov}(h_{it}, h_{i,t-1}|\theta^{(l-1)}, \varepsilon_i) \right. \\ &\quad \left. + \left(\text{E}(h_{it}|\theta^{(l-1)}, \varepsilon_i) - \mu_i \right) \left(\text{E}(h_{i,t-1}|\theta^{(l-1)}, \varepsilon_i) - \mu_i \right) \right].\end{aligned}$$

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

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

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

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

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

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

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

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

We iterate between steps 1.-3. until the relative change in the expected complete data likelihood becomes negligible. Note that strictly spoken, the algorithm is a Generalized EM algorithm since the M-step of $\text{vec}(B)$ depends on α .

C.2 EM-2

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

1. E-Step: in order to compute $E(h|\theta^{(l-1)}, \varepsilon)$, we recur to Monte Carlo integration. In particular, for $i = 1, \dots, r$ we simulate random draws of the mixture indicators $z_i^{(j)}$ for $j = 1, \dots, R$ by MCMC using the methodology of Kim et al. (1998) implemented based on the precision sampler of Chan & Jeliazkov (2009). Note that to break the autocorrelation of the chain, we only keep every 50th draw. Based on these draws, a Monte Carlo E-step is given as:

$$E(h_i|\theta^{(l-1)}, \varepsilon_i) = \frac{1}{R} \sum_{j=1}^R E(h_i|\theta^{(l-1)}, \varepsilon_i, z_i^{(j)}),$$

where the expectation is taken with respect to the Gaussian distribution $p(h_i|\theta^{(l-1)}, \varepsilon_i, z_i^{(j)})$ with tractable mean and variance given as:

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

where $y_i^* = (\ln(\varepsilon_{i1}^2), \dots, \ln(\varepsilon_{iT}^2))'$, and

$$\begin{aligned} G_{ij} &= \text{diag} \left(v^2(z_{i1}^{(j)}), \dots, v^2(z_{iT}^{(j)}) \right)^{-1}, \\ m_{ij} &= \text{diag} \left(m(z_{i1}^{(j)}), \dots, m(z_{iT}^{(j)}) \right). \end{aligned}$$

For the mixture distribution values see Table 5. Note that again, the moments are corrected for the linear constraint, that is $\bar{\delta}_{ij}^c = \bar{\delta}_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} (A_h \bar{\delta}_{ij} - \mu_i)$ and $\text{Cov}(h_i|\varepsilon_i, \theta^{(l-1)}, z_i^{(j)}, A_h h_i = \mu_i) = \Sigma_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} A_h \Sigma_{ij}$.

2. M-steps:

(a) Update ϕ_i and s_i for $i = 1, \dots, r$:

$$\phi_i^{(l)} = \frac{\tilde{S}_{xy}^i}{\tilde{S}_{xx}^i},$$

$$s_i^{(l)} = (T-1)^{-1} \left(\tilde{S}_{yy}^i - 2\phi_i^{(l)} \tilde{S}_{xy}^i + \left(\phi_i^{(l)}\right)^2 \tilde{S}_{xx}^i \right),$$

with:

$$\tilde{S}_{xx}^i = R^{-1} \sum_{j=1}^R \sum_{t=1}^{T-1} \left[\text{Var} \left(h_{it} | \theta^{(l-1)}, \varepsilon_{it}, z_{it}^{(j)} \right) + \left(\mathbb{E} \left(h_{it} | \theta^{(l-1)}, \varepsilon_{it}, z_{it}^{(j)} \right) - \mu_i \right)^2 \right],$$

$$\tilde{S}_{yy}^i = R^{-1} \sum_{j=1}^R \sum_{t=2}^T \left[\text{Var} \left(h_{it} | \theta^{(l-1)}, \varepsilon_{it}, z_{it}^{(j)} \right) + \left(\mathbb{E} \left(h_{it} | \theta^{(l-1)}, \varepsilon_{it}, z_{it}^{(j)} \right) - \mu_i \right)^2 \right],$$

$$\tilde{S}_{xy}^i = R^{-1} \sum_{j=1}^R \sum_{t=2}^T \left[\text{Cov} \left(h_{it}, h_{i,t-1} | \theta^{(l-1)}, \varepsilon_{it}, \varepsilon_{i,t-1}, z_{it}^{(j)}, z_{i,t-1}^{(j)} \right) \right. \\ \left. + \left(\mathbb{E} \left(h_{it} | \theta^{(l-1)}, \varepsilon_{it}, z_{it}^{(j)} \right) - \mu_i \right) \left(\mathbb{E} \left(h_{i,t-1} | \theta^{(l-1)}, \varepsilon_{i,t-1}, z_{i,t-1}^{(j)} \right) - \mu_i \right) \right].$$

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

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

where everything is as in EM-1 but

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

(c) Update B as in EM-1.

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

We recommend to set the starting values based on the results of EM-1, which are quickly available. We set R adaptively as proposed by Caffo, Jank & Jones (2005). In particular, if it becomes unlikely that the expected complete data likelihood is increased by an additional cycle of M-steps, R is increased by 100. Furthermore, the algorithm is stopped if a very large $R > R^*$ is needed to get a significant increase in the expected complete data likelihood. We set $R^* = 3000$.

Table 5: Mixture Components

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 $\log(\chi_{(1)}^2)$ distribution adjusted by its mean -1.2704 .

C.3 Derivatives complete data likelihood

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

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

Furthermore, let $\Sigma_t = BV_t B'$, $\beta = \text{vec}(B)$, $\alpha = \text{vec}(A)$, $\tilde{X}_t = (x_t' \otimes I_K)$, 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 (C.1) with respect to α and β are given as:

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

Note that the other 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 (C.1).

D Impulse Response Functions

Structural impulse responses of an SVAR are contained in the structural vector moving-average (SV-MA) representation:

$$y_t = \sum_{j=0}^{\infty} \underbrace{\Phi_j B}_{\Theta_j} \underbrace{V_t^{\frac{1}{2}} \eta_t}_{\varepsilon_t},$$

where Φ_j $K \times K$ is a sequence of exponentially decaying matrices given as:

$$\Phi_j = \sum_{i=1}^j \Phi_{j-i} A_i, \quad \text{for } j \geq 1,$$

and $\Phi_0 = I_K$.

D.1 Asymptotic Distribution

For the ease of exposition, assume that there is no intercept included into the SV-SVAR model. For confidence intervals of IRFs based on the asymptotic distribution, let $\theta = [\text{vec}(A)', \text{vec}(B)', \phi', s']'$. From Maximum Likelihood theory:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_\theta),$$

with $\Sigma_\theta = \mathcal{I}(\theta)^{-1}$ where $\mathcal{I}(\theta) = \text{E} \left[-\frac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta' \partial \theta} \right]$. Let $\alpha = \text{vec}(A)$, $\beta = \text{vec}(B)$, then

$$\Sigma_\theta = \begin{pmatrix} \Sigma_\alpha & & & \\ \Sigma_{\alpha,\beta} & \Sigma_\beta & & \\ \Sigma_{\alpha,\phi} & \Sigma_{\beta,\phi} & \Sigma_\phi & \\ \Sigma_{\alpha,s} & \Sigma_{\beta,s} & \Sigma_{\phi,s} & \Sigma_s \end{pmatrix}.$$

An application of the delta method yields (Brüggemann, Jentsch & Trenkler; 2016) the asymptotic distribution of the structural impulse responses:

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

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 $G_i = \frac{\partial \text{vec}(\Phi_i)}{\partial \alpha'} = \sum_{m=0}^{i-1} [J(\mathbf{A}')^{i-1-m}] \otimes \Phi_m$ for $i \geq 1$, $J = (I_K, 0, \dots, 0)$ be a $K \times Kp$ matrix and \mathbf{A} the companion matrix. Finally, $C_{i,\beta} = \frac{\partial \text{vec}(\Theta_i)}{\partial \beta'} = (I_K \otimes \Phi_i)$ for $i \geq 0$.

D.2 Fixed Design Wild Bootstrap

The fixed-design wild bootstrap (Gonçalves & Kilian; 2004) relies on the set of residuals:

$$\hat{u}_t = y_t - \hat{v} - \hat{A}_1 y_{t-1} - \dots - \hat{A}_p y_{t-p}, \quad t = 1, \dots, T.$$

Let ξ_t be i.i.d. random variables with $\text{E}(\xi_t) = 0$, $\text{E}(\xi_t^2) = 1$ and $\text{E}(\xi_t^4) < \infty$. Then, bootstrap time series y_t^* are constructed as:

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1} + \dots + \hat{A}_p y_{t-p} + \xi_t \hat{u}_t, \quad t = 1, \dots, T. \quad (\text{D.1})$$

We choose ξ_t to be from the Rademacher distribution with outcomes -1 or 1 and probabilities $p(\xi_t = 1) = \frac{1}{2} = p(\xi_t = -1)$. For N_{bs} time series generated as in equation (D.1), we fit the model and compute $\hat{\theta}^* = \text{vec}(\hat{\Theta}^*)$. We compute IRFs based on Hall's percentile intervals given as:

$$\left[\hat{\theta} - c_{(1-\alpha/2)}^*, \hat{\theta} - c_{\alpha/2}^* \right],$$

where c_γ^* denotes the γ quantile of $[\hat{\theta}^* - \hat{\theta}]$.

E Data

The time series data used in section 6 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 Lütkepohl & Netsunajev (2014) and Lütkepohl & Netšunajev (2017), we use the sample period 1970M1-2007M6. Except 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).

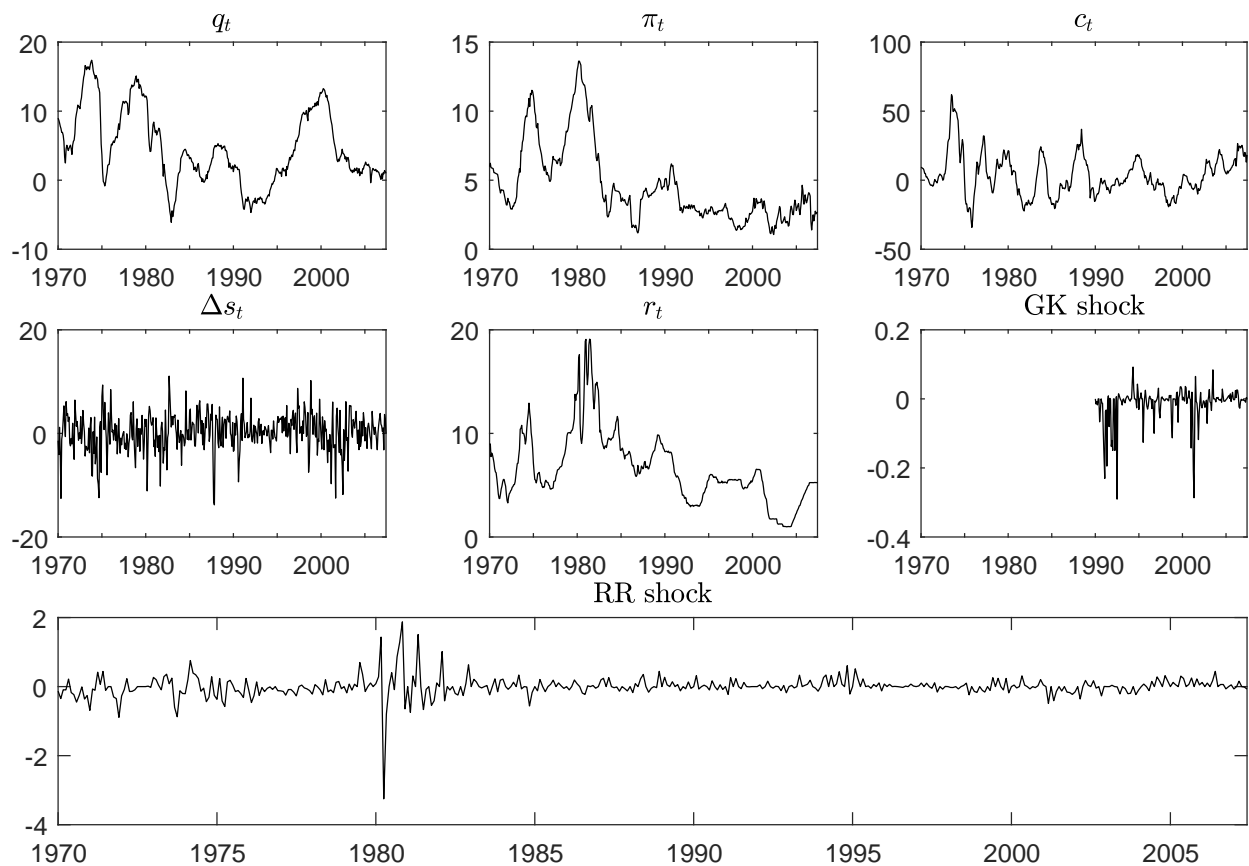


Figure 3: Time Series Data