Partial Identification of Heteroskedastic Structural VARs: Theory and Bayesian Inference

Helmut Lütkepohl[a], Fei Shang[b], Luis Uzeda[c], Tomasz Woźniak[b]

[a]*Freie Universität Berlin & DIW Berlin*
[b]*University of Melbourne*
[c]*Bank of Canada*

This is a preliminary version. Please, do not distribute.
19 February 2019

Abstract

In this study, we propose structural vector autoregressive models in which the structural parameters are identified via stochastic volatility process for time-varying conditional variances. Our focus is on the question of whether a particular structural shock is identified by heteroskedasticity without the need of imposing any sign or exclusion restrictions. Therefore, we derive a set of parametric restrictions under which the structural matrix is partially or globally unique, and Savage-Dickey density ratios are used to assess the validity of the identification conditions. Also, a novel and efficient Hamiltonian Monte Carlo sampler is proposed for Bayesian estimation of the structural effects. We apply identification through heteroskedasticity to estimate the dynamic output effects of unanticipated changes in tax policy that have been identified in previous studies by zero or sign restrictions as well as by using narrative measures as proxies.

Keywords: Identification Through Heteroskedasticity, Hamiltonian Monte Carlo, Stochastic Volatility
JEL classification: C11, C12, C32,

1. Partial identification in heteroskedastic SVARs

In this section we establish results for partial identification of structural parameters that are applicable within a broad class of heteroskedastic SVAR models. Theoretical results in this paper can be interpreted as general matrix results and thus extendable to different modeling frameworks. Nevertheless, to make notation less cumbersome, Theorem 1, Corollary 1 and Corollary 2 are presented below using SVAR-consistent notation.

1.1. Model

We begin by describing two common representations for SVAR models. Consider first the following structural form:

$$B_0 y_t = \mu + B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t, \tag{1}$$

where y_t is an N-dimensional vector of observable time series variables, B_i, $i = 0, 1, \ldots, p$, are $(N \times N)$ slope coefficient matrices with B_0 having a unit main diagonal, μ is an N-dimensional constant term and w_t is the N-dimensional structural error term which is instantaneously and serially uncorrelated, has mean zero and time-varying unconditional or conditional variances. Depending on the model used, the time-varying covariances, $\mathbb{E}[w_t w_t']$, or conditional covariances, $\mathbb{E}[w_t w_t' | w_{t-1}, w_{t-2}, \ldots]$, of w_t are denoted by $\Lambda_t = \text{diag}\left(\sigma_{1.t}^2, \ldots, \sigma_{N.t}^2\right)$, where the $\sigma_{n.t}^2$ are the unconditional or conditional variances.

We also consider an alternative structural form of the model:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B w_t, \tag{2}$$

where A_i, $i = 1, \ldots, p$, are $(N \times N)$ slope coefficient matrices, ν is a N- dimensional constant term and B is a nonsingular $(N \times N)$ structural matrix which represents the impact effects of the structural errors w_t on the observed variables.

Model (2) differs from model (1) in an important respect. While in model (1) the components of w_t and, hence, the structural shocks, are linked to the equations, they can

be freely permuted in model (2) by simply permuting the columns of B. Although this additional flexibility can be convenient when specific shocks are of interest, it has some drawbacks for inference. This aspect will be further discussed later. We will primarily focus on model (1) to simplify our inference procedures.

Both structural form models have the same reduced from

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{3}$$

where u_t is the N-dimensional reduced-form error term which is $u_t = B_0^{-1} w_t$ if model (1) is considered or $u_t = B w_t$ if model (2) is under consideration. The assumptions for w_t imply that u_t has mean zero, is serially uncorrelated and has time-varying unconditional or conditional covariance matrices. In other words, depending on the model used, the unconditional covariances, $\mathbb{E}\left[u_t u_t'\right]$, or the conditional covariances, $\mathbb{E}\left[u_t u_t' | u_{t-1}, u_{t-2}, \dots\right]$, may be time-varying and are denoted by Σ_t.

1.2. Identification

It is well known that the structural matrices B_0 or B are not identified without additional restrictions. To derive conditions for partial identification of some of the parameters of these matrices we first define I_N as an identity matrix of order N, and then, we state a general matrix result.

Theorem 1. *Let Σ_t, $t = 0, 1, \dots$, be a sequence of positive definite $(N \times N)$ matrices and $\Lambda_t = \operatorname{diag}\left(\sigma_{1.t}^2, \dots, \sigma_{N.t}^2\right)$ a sequence of $(N \times N)$ diagonal matrices with $\Lambda_0 = I_N$. Suppose there exists a a nonsingular $(N \times N)$ matrix B such that*

$$\Sigma_t = B \Lambda_t B', \qquad t = 0, 1, \dots. \tag{4}$$

Let $\boldsymbol{\sigma}_n^2 = (1, \sigma_{n.1}^2, \sigma_{n.2}^2, \dots)$ be a possibly infinite dimensional vector. Then the n^{th} column of B is unique up to sign if $\boldsymbol{\sigma}_n^2 \neq \boldsymbol{\sigma}_i^2 \quad \forall i \in \{1, \dots, N\} \backslash \{n\}$.

Proof. The proof is given in Appendix A.1. It proceeds by showing that, under the

conditions of the theorem, any other nonsingular $(N \times N)$ matrix B_* satisfying Equation (4) has the same n^{th} column as B. □

Theorem 1 has obvious implications for models set up as in (2). It implies that, if the n^{th} component of w_t has a sequence of variances $\boldsymbol{\sigma}_n^2 = (1, \sigma_{n.1}^2, \sigma_{n.2}^2, \dots)$ which is different from the variance sequence of any other component of w_t, the n^{th} column of B will be identified up to sign. It can also be used as in Lütkepohl & Woźniak (2017) to show that in the structural form setup (1), a single row of B_0 is identified if the corresponding error term has a variance vector which is different from the variance vectors of all other equations. More precisely, Theorem 1 implies the following result:

Corollary 1. *Let Σ_t, $t = 0, 1, \dots$, be a sequence of positive definite $(N \times N)$ matrices and $\Lambda_t = \operatorname{diag}\left(\sigma_{1.t}^2, \dots, \sigma_{N.t}^2\right)$ be a sequence of $(N \times N)$ diagonal matrices with positive diagonal elements. Suppose there exists a nonsingular $N \times N$ matrix B_0 with unit main diagonal such that*

$$\Sigma_t = B_0^{-1} \Lambda_t B_0^{-1\prime}, \quad t = 0, 1, \dots. \tag{5}$$

Let $\tilde{\boldsymbol{\sigma}}_n^2 = (\sigma_{n.1}^2/\sigma_{n.0}^2, \sigma_{n.2}^2/\sigma_{n.0}^2, \dots)$ be a possibly infinite dimensional vector. Then the n^{th} row of B_0 is unique if $\tilde{\boldsymbol{\sigma}}_n^2 \neq \tilde{\boldsymbol{\sigma}}_i^2 \; \forall i \in \{1, \dots, N\} \setminus \{n\}$.

Proof. – See Appendix A.2. □

Note that the vectors $\tilde{\boldsymbol{\sigma}}_j^2$ contain variances relative to the variances for $t = 0$ which could be seen as a reparametrization of the model. Of course, one could state the result in terms of the original variances. We are using relative variances in the corollary because it makes it easier to state the result and also leads directly to the verification procedure discussed in Section 4.

The corollary generalizes Theorem 1 of Lütkepohl & Woźniak (2017). It provides a general result on identification of a single equation through heteroskedasticity. It shows that a structural shock and hence the corresponding structural equation is identified if the sequence of variances relative to the corresponding element in Λ_0 is not equal to the relative variance sequences of any of the other shocks. In fact, the theorem and the

corollary generalize a number of results that have been presented in the related literature. It is worth discussing some of these results in detail to show the implications of our partial identification results for impulse response analysis and for different volatility modeling frameworks. We elaborate on these points below.

1.3. Impulse responses

Structural impulse responses are computed from the reduced-form impulse responses Φ_i, $i = 0, 1, \ldots$, which are obtained by the following recursions from the p^{th}-order reduced-form VAR slope coefficients:

$$\Phi_i = \begin{cases} I_N \text{ for } i = 0, \\ \sum_{j=1}^{p} A_j \Phi_{i-j} \text{ for } i = 1, 2, \ldots. \end{cases} \tag{6}$$

The structural impulse responses are the elements of the matrices $\Theta_i = \Phi_i B_0^{-1}$, $i = 0, 1, \ldots$, if model (1) is considered or of $\Theta_i = \Phi_i B$, $i = 0, 1, \ldots$, if model (2) is under consideration. Thus, for computing them, the structural matrices B_0 or B are needed. In particular, if just one shock is identified through heteroskedasticity, the following result formalizes the implications of Corollary 1 for impulse response analysis:

Corollary 2. *If the n^{th} row of B_0 is identified in model (1), then the n^{th} column of B_0^{-1} is unique and the structural impulse responses can be obtained by right-multiplying the matrices Φ_i by the n^{th} column of B_0^{-1}.*

Proof. – See Appendix A.3. □

We present theoretical results for impulse responses based on the SVAR in (1) since this is specification used for the empirical application in this paper. However, it is straightforward to verify that – for the SVAR in (2) – if just one shock is uniquely identified and the corresponding column of the B matrix is given, then all the responses to the n^{th} shock are uniquely obtained up to sign by right-multiplying the Φ_i matrices by the n^{th} column of B.

2. Implications for Volatility Modeling

In this section we will discuss the implications for the previous general identification results for some special volatility models that have been used in the literature on identification through heteroskedasticity (see, e.g., Kilian & Lütkepohl, 2017, Chapter 14).

2.1. Regime changes in volatility

Early work on identification through heteroskedasticity in the SVAR literature as based on models with exogenously changing volatility during the sample period (e.g., Rigobon (2003), Rigobon & Sack (2003), Lanne & Lütkepohl (2008, 2014)). Suppose there is only one change in volatility after some time period T_1 such that $\Sigma_t = \Sigma_1$ for $t = 1, \ldots, T_1$ and $\Sigma_t = \Sigma_2$ for $t = T_1 + 1, \ldots, T$. It is assumed that $\Sigma_1 \neq \Sigma_2$. Then there exists a decomposition

$$\Sigma_1 = BB' \qquad \text{and} \qquad \Sigma_2 = B\Lambda B', \tag{7}$$

where $\Lambda = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. For this case, the matrix of impact effects of the shocks, B, was shown to be identified apart from column ordering and column sign if all diagonal elements of Λ are distinct (Lanne, Lütkepohl & Maciejowska, 2010).

For this case Theorem 1 implies that, for example, the first column of B is identified and, hence, the first shock is identified if $\sigma_1^2 \neq \sigma_i^2$ for $i = 2, \ldots, N$. This condition can of course also be satisfied if $\sigma_2^2 = \cdots = \sigma_N^2$. Thus, Theorem 1 implies that a single shock can also be identified through heteroskedasticity if the other shocks are not identified. In fact, all the other shocks may be homoskedastic if there is a variance change in only one of the shocks.

On the other hand, Theorem 1 implies that a homoskedastic shock can also be identified. If all other shocks are heteroskedastic, the homoskedastic shock with variance sequence $\tilde{\sigma}_n^2 = (1, 1, \ldots)$ would be identified. In fact, it may be the only identified shock if all other shocks have identical variance sequences different from $(1, 1, \ldots)$.

Clearly, if all diagonal elements of Λ are distinct, each of the columns of B is identified up to sign and hence all the shocks are identified. Of course, changing the ordering of the

diagonal elements of Λ changes the columns of the B matrix in the decomposition (7) and, hence, the ordering of the columns of B can change as long as the order of the diagonal elements of Λ is not fixed. The important point of our partial identification result is of course, that identification of a subset of the shocks is also possible even if some of the shocks are not identified through the change in variance. The model with two volatility states is also relevant if a smooth transition in the variances is considered as in Lütkepohl & Netšunajev (2017b).

In the more general setup considered in Lütkepohl (2013), Lütkepohl & Netšunajev (2017a) and others, where more than two volatility states are assumed, the identification conditions become more complicated. Suppose there are M volatility states such that

$$\Sigma_t = \mathbb{E}\left(u_t u_t'\right) = \begin{cases} \Sigma_1 & \text{for} \quad t \in \mathcal{T}_1, \\ \vdots & \\ \Sigma_M & \text{for} \quad t \in \mathcal{T}_M, \end{cases} \tag{8}$$

where $\mathcal{T}_m = \{T_{m-1} + 1, \ldots, T_m\}$ for $m = 1, \ldots, M$ denote M given volatility regimes with $T_0 = 0$ and $T_M = T$. The T_m represent the points of volatility changes. Then, assuming a decomposition of the different covariance matrices in the form of

$$\Sigma_1 = BB' \qquad \Sigma_m = B\Lambda_m B', \qquad m = 2, \ldots, M,$$

with diagonal matrices $\Lambda_m = \operatorname{diag}\left(\sigma_{1.m}^2, \ldots, \sigma_{N.m}^2\right)$ for $m = 2, \ldots, M$, the conditions for (local) identification of B up to column permutation and column sign are that for any two subscripts $n, l \in \{1, \ldots, N\}$, $n \neq i$, there is a $j \in \{2, \ldots, M\}$ such that $\sigma_{n.j}^2 \neq \sigma_{l.j}^2$ (see Lanne et al., 2010). It is easy to see that if the $\sigma_{n.t}^2$ associated with each period t are lined up in our vectors $\boldsymbol{\sigma}_n^2 = (1, \sigma_{n.2}^2, \ldots)$, these vectors would be pairwise distinct if the identification condition of Lanne et al. (2010) is satisfied and vice versa. In other words, for this specific model setup, the earlier identification condition up to column permutation and column sign is equivalent to the condition obtained from Theorem 1.

Again we emphasize, however, that Theorem 1 also implies partial identification

conditions for a subset of the shocks (or columns of B) even if the conditions for full identification of all shocks are not satisfied. Moreover, Theorem 1 allows one to determine precisely which shocks are identified and which ones are not. Since the same set of conditions is also relevant if the heteroskedasticity is driven by a Markov process as in Lanne et al. (2010), Herwartz & Lütkepohl (2014), Lütkepohl & Velinov (2016), Netšunajev (2013), Woźniak & Droumaguet (2015) and others, Theorem 1 also implies conditions for full and partial identification for this case.

2.2. GARCH and stochastic volatiliy

For the typical Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and SV processes that are used in this context, Theorem 1 also applies, of course. GARCH models have been used in this context by Normandin & Phaneuf (2004), Bouakez & Normandin (2010) and Lütkepohl & Milunovich (2016) for example, and SV models have been proposed by Bertsche & Braun (2018). In these models the conditional covariance matrices of the reduced form errors for model (2) are given by

$$\Sigma_t = B\Lambda_t B',$$

where $\Lambda_t = \text{diag}\left(\sigma_{1.t}^2, \ldots, \sigma_{N.t}^2\right)$ is a diagonal matrix. If the $\sigma_{n.t}^2$ vary stochastically, as in GARCH or SV dynamics, they will not be proportional with probability 1 and, hence, satisfy the conditions for identification of Theorem 1. So if any one of the structural errors has changing conditional variances, it will be identified, even if all the other components have constant conditional variance. This discussion also shows that Theorem 1 generalizes results for full identification in Sentana & Fiorentini (2001) to the case of partial identification. Also, in contrast to Sentana & Fiorentini (2001), linear independence of the variance vectors is not explicitly stated. However, the condition in Theorem 1 requires that the first component of the variance vectors is 1. It is easy to see that two vectors which share the same first component and are different, will also be linearly independent. Although our condition for partial identification is, hence, that a specific

variance vector is linearly independent from all other variance vectors, they don't have to be collectively linearly independent if only partial identification is of interest. If all shocks are identified, however, then our condition also implies linear independence of all the variance vectors jointly.

3. SVAR model with stochastic volatility

Our empirical model is the SVAR model from equation (1), that is, the specification with structural matrix B_0 whose diagonal elements are set to ones. The model presented below supplements this specification with the assumption of conditionally normally distributed and heteroskedastic error terms with conditional variances following the SV process. We chose a specific conditional variance process for the structural shocks that facilitates the verification of the conditions from Corollary 1.

To obtain a flexible framework that facilitates the estimation of models with unrestricted or restricted matrix B_0, the approach proposed by Amisano & Giannini (1997), used by Canova & Pérez Forero (2015) and adapted for the case of identification through heteroskedasticity by Lütkepohl & Woźniak (2017) is helpful. Let an $(r \times 1)$ vector b_0 collect all of the unrestricted elements of the matrix B_0 column by column. We impose restrictions on the matrix of contemporaneous effects by setting:

$$\operatorname{vec}(B_0) = S_b b_0 + s_b, \tag{9}$$

where vec denotes a column-wise vectorization of a matrix, S_b and s_b are respectively an $(N^2 \times r)$ matrix and an $(N^2 \times 1)$ vector. Their elements are fixed for a particular model and typically set to zeros and ones.

An additional assumption for the structural shocks w_t from equation (1) is their joint conditional normality given the past observations of vector y_t:

$$w_t | y_{t-1}, y_{t-2}, \cdots \sim \mathcal{N}\left(\mathbf{0}_{(N\times 1)}, \operatorname{diag}\left(\sigma_{1.t}^2, \ldots, \sigma_{N.t}^2\right)\right). \tag{10}$$

Note that the joint normality and a diagonal covariance matrix imply conditional independence between the structural shocks.

In our setting, the latent log-volatilities underlying the SV process are expressed in terms of their non-centered parametrization proposed by Frühwirth-Schnatter & Wagner (2010) and adapted for the SV model by Chan (2016):

$$\sigma_{n.t}^2 = \exp\left\{\omega_n h_{n.t} + \bar{h}_{n.0}\right\}, \tag{11}$$

for $t = 0, 1, \ldots,$ where $h_{n.0} = 0$. The expression $\left(\omega_n h_{n.t} + \bar{h}_{0.n}\right)$ is the logarithm of the conditional variance of the n^{th} shock at time t. The choice of the non-centered parameterization greatly facilitates the verification of the conditions from Corollary 1 as it is discussed below in Section 4.

The equation above implies that the conditional variance of the n^{th} structural shock at time $t = 0$ is given by $\sigma_{n.0}^2 = \exp\left\{\bar{h}_{n.0}\right\}$, where $\bar{h}_{n.0}$ is the initial condition that is estimated for each n. Importantly, a sequence of $\tilde{\sigma}_{n.t}^2 = \exp\{\omega_n h_{n.t}\}$ for $t = 1, \ldots, T$ denotes relative conditional variances of the n^{th} element of w_t, that is, the value of the conditional variance relative to the corresponding initial condition, that is, $\tilde{\sigma}_{n.t}^2 = \sigma_{n.t}^2 / \sigma_{n.0}^2$. Parameter ω_n can be interpreted as the square root of the conditional variance of $\omega_n h_{n.t}$ given $h_{n.t-1}$. Note that both, ω_n and $h_{n.t}$ are identified up to a sign (see e.g. Frühwirth-Schnatter & Wagner, 2010). Moreover, if $\omega_n = 0$ for some n then the n^{th} element of w_t is homoskedastic with the time-invariant conditional variances are equal to $\exp\left\{\bar{h}_{n.0}\right\}$ for all t. Otherwise, this shock is heteroskedastic with the conditional variance process modeled with the SV process. In our empirical example, we use the Savage-Dickey Density Ratio proposed by Chan (2016) to verify a hypothesis given by the restriction $\omega_n = 0$ to verify the conditional heteroskedasticity of the shocks in our model.

Finally, the latent log-volatilities driving the conditional variances follow a state

equation given by:

$$h_{n.t} = h_{n.t-1} + v_{n.t}, \qquad \text{and} \qquad v_{n.t} \sim \mathcal{N}(0,1), \tag{12}$$ [1]

for each $n = 1, \ldots, N$. The notation in equation (12) is simplified and we stress that The choice of a particular form of the SV process is not essential for our purposes and the state equation in (12) could be replaced by an alternative specification. This choice seems not to be crucial for modeling conditional variances of macroeconomic variables as shown by Eisenstat & Strachan (2016).

4. Verifying partial identification

The sequence of the relative conditional variances of $w_{n.t}$, $\tilde{\sigma}^2_{n.t} = \exp\{\omega_n h_{n.t}\}$ for $t = 1, \ldots, T$, is essential from the point of view of verifying partial identification of heteroskedastic SVARs. Corollary 1 can be interpreted such that if for some $i \neq n$ $\tilde{\sigma}^2_{i.t} = \tilde{\sigma}^2_{n.t}$ for all $t = 1, \ldots, T$, then its conditions are violated and the n^{th} structural shock is not identified through heteroskedasticity. This condition for relative variances is in fact equivalent to the condition from Theorem 1 and, thus, the procedure described below applies also to the case of the SVAR model from equation (2) with the unrestricted matrix of contemporaneous effects B.

It is easy to show that an equivalent set of restrictions for the hypothesis that the n^{th} element of w_t is not identified through heteroskedasticity can be presented in terms of the state equation innovations. Then, if for some $i \neq n$ the restrictions:

$$\omega_i = \omega_n, \tag{13a}$$

$$v_{i.t} = v_{n.t}, \tag{13b}$$

[1] Note that the SV process in equation (12) is just an alternative parameterization of the standard SV model. To see this write the noncentered process $h_{n.t}$ in terms of the corresponding logarithm of conditional variance, denoted by $\tilde{h}_{n.t}$, as $h_{n.t} = (\tilde{h}_{n.t} - h_{0.n})/\omega_n$. Then the autoregressive process for $\tilde{h}_{n.t}$ implied by equation (12) is given by: $\tilde{h}_{n.t} = \tilde{h}_{n.t-1} + \omega_n v_{n.t}$.

hold for all $t = 1, \ldots, T$, then the conditions of Corollary 1 are violated. The restriction (13b), given the SV process from equation (12), implies that $h_{i.t}$ and $h_{n.t}$ are equal to each other for all t. This condition, together with the restriction (13a) sets $\tilde{\sigma}^2_{i.t}$ to $\tilde{\sigma}^2_{n.t}$ for all t which results exactly in the condition violating Corollary 1 discussed above. On the contrary, if $\omega_i \neq \omega_n$ and $v_{n.t} \neq v_{i.t}$ for all $i \in \{1, \ldots, N\} \backslash \{n\}$ and for some $t = 1, \ldots, T$, then the n^{th} element of w_t is uniquely identified through heteroskedasticity. This opens the possibility of statistical verification of the partial identification through heteroskedasticity.

To facilitate the statistical verification the restrictions from equation (13) define an $(N \times 1)$ vector $\omega = (\omega, \ldots, \omega_N)'$. Then the restriction from equation (13a) can be presented as:

$$\omega = S_\omega \tilde{\omega}, \tag{14}$$

where an $(N \times 1)$ vector $\tilde{\omega}$ is a vector of unrestricted parameters, and an $(N \times N)$ matrix S_ω is a selection matrix. The matrix S_ω is used to impose restrictions on vector ω by selecting which elements of vector $\tilde{\omega}$ enter the model specification and the likelihood function. In the unrestricted model for which restriction (13a) does not hold for each i the selection matrix is set to an identity matrix, $S_\omega = I_N$.

Consider a specification of matrix S_ω for the verification of partial identification of the n^{th} element of w_t. Define $N - 1$ independent Bernoulli distributed random indicators denoted by κ_i for $i \in \{1, \ldots, N\} \backslash \{n\}$ and specify the selection matrix by:

$$S_\omega = I_N + \sum_{i \in \{1, \ldots, N\} \backslash \{n\}} \kappa_i S_{\omega.i}. \tag{15}$$

The restrictions from equation (13a) are imposed by an appropriate specification of $(N \times N)$ selection matrices $S_{\omega.i}$. In particular, matrix $S_{\omega.i}$ contains zeros everywhere except for the elements in its i^{th} row. In this row, the i^{th} element is set to -1, the n^{th} element is set to 1, and the remaining elements are equal to zero. Note that the model with unrestricted vector ω is obtained by setting $\kappa_i = 0$ for all $i \in \{1, \ldots, N\} \backslash \{n\}$ which results in $S_\omega = I_N$.

To illustrate the workings of the mechanism from equation (15) we consider a 3-variable

Table 1: Examples of Bernoulli indicators κ_i related selection matrices S_ω and resulting restricted vectors ω for verifying partial identification of the second shock ($n = 2$) in a tri-variate ($N = 3$) SVAR model.

	$\kappa_1 = 0$	$\kappa_1 = 1$
$\kappa_3 = 0$	$S_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix}$	$S_\omega = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\omega = \begin{bmatrix} \omega_2 \\ \omega_2 \\ \omega_3 \end{bmatrix}$
$\kappa_3 = 1$	$S_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_2 \end{bmatrix}$	$S_\omega = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\omega = \begin{bmatrix} \omega_2 \\ \omega_2 \\ \omega_2 \end{bmatrix}$

($N = 3$) example in which we are interested in the partial identification of the second shock ($n = 2$). Then, the selection matrices on the right-hand side of equation (15) are given by:

$$S_{\omega.1} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad S_{\omega.3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}. \tag{16}$$

The resulting selection matrix S_ω depends on the values of κ_1 and κ_3. Table 1 presents the resulting selection matrices S_ω and the constrained vectors ω for all of the possible combinations of the values of κ_1 and κ_3.

At the same time, the restrictions on the state-equation error terms need to be imposed as in equation (13b). Define $(NT \times 1)$ vectors $h = (h_{1.1}, \ldots, h_{1.T}, \ldots, h_{N.1}, \ldots, h_{N.T})'$ and $v = (v_{1.1}, \ldots, v_{1.T}, \ldots, v_{N.1}, \ldots, v_{N.T})'$, as well as an $(T \times T)$ matrix H with 1 on the main diagonal, -1 on the first diagonal below the main diagonal, and zeros elsewhere. Then, the log-volatility propagation equation (12) in an extended matrix notation given by:

$$(I_N \otimes H)h = v. \tag{17}$$

We introduce the restrictions by setting:

$$v = S_v \tilde{v}, \tag{18}$$

where S_v is an $(NT \times NT)$ selection matrix and $\tilde{v}$ is an $(NT \times 1)$ vector of the unrestricted residuals following an NT-variate standard normal distribution. We impose restrictions simultaneously on ω and v by setting:

$$S_v = S_\omega \otimes I_T. \tag{19}$$

Therefore, matrix S_v depends on the same set of Bernoulli indicators as matrix S_ω and, thus, the corresponding restrictions on vectors ω and v are imposed simultaneously. Note that the unrestricted model is specified by setting $S_\omega = I_N$ which automatically fixes $S_v = I_{NT}$.

Expression (18) entails the possibility of a reduced number of innovations driving a higher dimension of state variables, that is some states might be perfectly correlated. The latter is the statistical hypotheses we want to verify given the conditions in Corollary 1. Given the possibility of perfectly collinear states, statistical verification of these conditions can be carried out by adopting disturbance smoothing techniques proposed by De Jong & Shephard (1995), and Durbin & Koopman (2002), and further developed by Uzeda (2018). In our model, imposing perfect collinearity through equality restrictions in the state-equation error terms introduces an easy way of estimating the restricted models.

The restrictions from equation (13) are imposed when any of the indicators κ_i is equal to 1 for $i \in \{1, \ldots, N\} \backslash \{n\}$. In our framework we assume an independent Bernoulli prior distribution for each κ_i and estimate these indicators via the MCMC simulation obtaining a sample of draws of the Bernoulli indicators sampled from the posterior distribution, denoted by $\left\{\kappa_i^{(s)}\right\}_{s=1}^{S}$. From this outcome of the estimation procedure we compute the posterior probability of a hypothesis that the n^{th} shock is identified through heteroskedasticity by computing a fraction of the MCMC draws for which $\kappa_i^{(s)} = 0$ holds for all $i \in \{1, \ldots, N\} \backslash \{n\}$. This probability is confronted with the posterior probability of the alternative hypothesis that the n^{th} shock is not identified through heteroskedasticity computed as a fraction of the MCMC draws for which $\kappa_i^{(s)} = 1$ for any $i \in \{1, \ldots, N\} \backslash \{n\}$. The mechanism of setting the restrictions through Bernoulli indicators formally introduces the

Stochastic Search Specification Selection (S^4) where the specification selection is performed by linking the form of the selection matrices S_ω and S_v to the indicators. Estimation of the indicators is discussed in Section 5.

5. Bayesian inference

5.1. Prior distributions

The prior distribution for the unrestricted elements of the matrix B_0 collected in the vector b_0, conditionally on a hyper-parameter γ_b, is an r-variate normal distribution with mean vector zero and a diagonal covariance matrix $\gamma_b I_r$. The hyper-parameter γ_b is interpreted as the level of shrinkage imposed on the structural parameters b_0 and is also estimated. For that purpose, we define the marginal prior distribution of γ_b to be inverse gamma 2, denoted by $\mathcal{IG}2$, with parameters $\underline{a}$ and $\underline{b}$ set to 1. See Bauwens, Richard & Lubrano (1999, Appendix A) for the detailed specification of the distribution and the random sampling algorithm.

The conditional prior distribution of the vector of constant terms, μ_n, $n \in \{1, \ldots, N\}$, given a constant term specific hyper-parameter γ_μ, is an N-variate normal distribution with mean zero and variance $\gamma_\mu I_N$. The marginal prior distribution for γ_μ is $\mathcal{IG}2$ with parameters $\underline{a}$ and $\underline{b}$ set to 1.

To specify the prior distribution of the structural VAR slope parameters, let $\underline{P} = \begin{bmatrix} D & \mathbf{0}_{N \times N(p-1)} \end{bmatrix}$, where D is an $N \times N$ diagonal matrix and $\mathbf{0}_{a \times b}$ is an $(a \times b)$ matrix of zeros. Typically the diagonal elements of the matrix D are zeros for stationary variables and ones for persistent variables, as in the Minnesota prior, but they could also be other known quantities (see Doan, Litterman & Sims, 1984). Then the conditional prior distribution of the equation-specific autoregressive parameters, $\beta_n = \left(B_{1.n}, \ldots, B_{p.n}\right)$, where $B_{l.n}$ is the n^{th} row of matrix B_l for $l \in \{1, \ldots, p\}$, is a pN-variate normal distribution. It is conditioned on an equation-specific autoregressive hyper-parameter $\gamma_{\beta.n}$ and $B_{0.n}$ and has prior mean equal to $B_{0.n}\underline{P}$ and a covariance matrix equal to $\gamma_{\beta.n}\underline{H}$. The diagonal matrix $\underline{H}$ has the main diagonal set to the vector $\left((1^2)^{-1}\iota_N', (2^2)^{-1}\iota_N', \ldots, (p^2)^{-1}\iota_N'\right)'$, and thus it facilitates imposing

a decaying pattern of prior variances for the subsequent lags as in the Minnesota prior of Doan et al. (1984). The prior distribution for $\gamma_{\beta.n}$ is $\mathcal{IG}2$ with parameters $\underline{a}$ and $\underline{b}$ set to 1.

The parameters of the SV process have the following prior distributions. Similarly as in Frühwirth-Schnatter & Wagner (2010) and Chan (2016), the vector $\tilde{\omega}$ follows an N-variate normal distribution with the mean set to zero and the diagonal covariance matrix with an N-vector $\underline{D}_{\omega}$ on the main diagonal. The vector $\bar{h}_0$ follows an N-variate normal distribution with the mean set to a zero vector and the diagonal covariance matrix with vector $\underline{D}_{h_0}$ on the main diagonal. Finally, each of the $N-1$ indicators follows a Bernoulli distribution with the probability of success set to $\underline{\pi}$. We set $\underline{\pi} = \sqrt[N-1]{0.5}$, which implies that the prior probability of the event that $\kappa_i = 0$ holds for all $i \in \{1,\dots,N\}\backslash\{n\}$ is 0.5. This way, we do not discriminate between the hypotheses of partial identification and the lack thereof giving them equal chances *a priori*.

To summarize, the prior specification takes the following form:

$$p(\theta) = p\left(b_0|\gamma_b\right) p\left(\gamma_b\right) p\left(\mu|\gamma_\mu\right) p\left(\gamma_\mu\right) \left(\prod_{n=1}^{N} p\left(\beta_n|B_{0.n}, \gamma_{\beta.n}\right) p\left(\gamma_{\beta.n}\right) \right) p(\tilde{\omega}) p(\bar{h}_0) \left(\prod_{i\in\{1,\dots,N\}\backslash\{n\}} p(\kappa_i) \right), \tag{20}$$

where the specific prior distributions are:

$$\begin{aligned}
b_0|\gamma_b &\sim \mathcal{N}_r\left(\mathbf{0}_r, \gamma_b I_r\right) \\
\mu|\gamma_\mu &\sim \mathcal{N}_N\left(\mathbf{0}_N, \gamma_\mu I_N\right) \\
\beta_n'|B_{0.n}, \gamma_{\beta.n} &\sim \mathcal{N}_{pN}\left(B_{0.n}\underline{P}, \gamma_{\beta.n}\underline{H}\right) \\
\gamma_b &\sim \mathcal{IG}2\left(\underline{a}, \underline{b}\right) \\
\gamma_\mu &\sim \mathcal{IG}2\left(\underline{a}, \underline{b}\right) \\
\gamma_{\beta.n} &\sim \mathcal{IG}2\left(\underline{a}, \underline{b}\right) \\
\tilde{\omega} &\sim \mathcal{N}_N\left(\mathbf{0}_N, \operatorname{diag}\left(\underline{D}_{\omega}\right)\right) \\
\bar{h}_0 &\sim \mathcal{N}_N\left(\mathbf{0}_N, \operatorname{diag}\left(\underline{D}_{h_0}\right)\right) \\
\kappa_i &\sim \mathcal{B}ernoulli\left(\underline{\pi}\right)
\end{aligned}$$

for $n \in \{1, \ldots, N\}$ and $i \in \{1, \ldots, N\} \backslash \{n\}$.

The above choice of the prior distributions is practical and priority is given to distributions that result in convenient and proper full conditional posterior distributions, and therefore, allow for the derivation of an efficient Gibbs sampler. The hierarchical prior distributions for the constant terms, autoregressive slope parameters, and the structural matrix constitute a flexible framework in which the impact of the choice of the hyper-parameters of the prior distribution on the inference is reduced, in line with Giannone, Lenza & Primiceri (2015).

5.2. Estimation procedure

We conduct Bayesian estimation of the model through an iterative procedure by sampling from the full conditional posterior distributions of the subgroups of parameters. All but one of these distributions have the form of known statistical distributions. The nonstandard distribution is the full conditional posterior distribution of the unrestricted contemporaneous effects b_0. For this parameter vector we apply a Hamiltonian Monte Carlo (HMC) sampler and, therefore, we call this algorithm the Hamiltonian Monte Carlo within Gibbs sampler. We report all of the computational details of the estimation procedure in Appendix B and focus on the novel elements of the algorithm developed for our model below.

Sampling contemporaneous effects. The full conditional posterior distribution of the unrestricted parameters of the contemporaneous effects matrix collected in vector b_0 is proportional to:

$$|\det(B_0)|^T \exp\left\{-\frac{1}{2}\left(b_0 - \bar{b}_0\right)' \overline{V}_{b_0}^{-1}\left(b_0 - \bar{b}_0\right)\right\}, \tag{21}$$

where $\bar{b}_0$ and $\overline{V}_{b_0}$ are the location and scale parameters of the distribution for which the exact formulae given our model are given in Appendix B. The relationship between B_0 and b_0 is presented in equation (9).

Sampling from the distribution in a form given by equation (21) is nontrivial due to a non-standard form of the kernel of this distribution. At the same time, efficient

samplers are in demand because this is a general form of the posterior distribution of the contemporaneous effects of the SVAR models identified by exclusion restrictions (see Waggoner & Zha, 2003), sign restrictions (see e.g. Uhlig, 2005), a combination of thereof (as in Arias, Rubio-Ramírez & Waggoner, 2018), or utilising stochastic restrictions through the prior distribution (see e.g. Baumeister & Hamilton, 2015). The problem of sampling from the posterior distribution from equation (21) with a flexible manner of imposing exclusion restrictions as in equation (9), that allows for non-recursive identification patterns and cross-equation restrictions, was discussed in Canova & Pérez Forero (2015). They proposed to use a random walk Metropolis-Hastings sampling algorithm building on a similarity of the distribution in (21) to a multi-variate normal distribution. Canova & Pérez Forero (2015) choses the latter with appropriately constructed covariance matrix as the candidate sampling density. Their approach does not cover the case of identification through heteroskedasticity. The generalisation of the approach of Canova & Pérez Forero (2015) to this case is proposed by Lütkepohl & Woźniak (2017) for the SVAR models identified by Markov-switching conditional heteroskedasticity where a similar Metropolis-Hastings sampler is used. However, both of the approaches above work best if the determinant of B_0 is equal to 1, an unlikely case for models identified through heteroskedasticity.

In this paper, we adapt the Hamiltonian Monte Carlo (HMC) technique by Neal (2011) to sample from the distribution in equation (21). In our implementation, the HMC is used to produce a single candidate draw of vector b_0 that is subsequently accepted, or rejected in favor of the previous state of the MCMC for b_0, proportionally to the Metropolis-Hastings acceptance probability for an expanded parameter vector (see Hoffman & Gelman, 2014, Section 2). The expansion of the parameter space is a feature of the HMC. This draw is treated as the most up to date draw of b_0 and is subsequently used to sample from the full conditional posterior distributions of the remaining sub-groups of parameters within the Gibbs sampler iteration. Our adaptation, therefore, constrains to some extent simulating the Hamiltonian dynamics (see Neal, 2011) remaining the improved efficiency of drawing from the posterior distribution of b_0. The efficiency gains relative to using the random

walk Metropolis-Hastings sampler are enormous as documented in Section **??**.

We use the particular version of the HMC called the No-U-Turn sampler (NUTS) proposed by Hoffman & Gelman (2014).[2] It constitutes a fully automated algorithm as it eliminates the need to choose two HMC-specific parameters, namely the number-of-steps and the step size parameter. The latter is obtained via dual averaging proposed by Nesterov (2009).

Sampling conditional volatilities. As explained in Section 3 we estimate the SV model in its non-centered parameterization introduced by Frühwirth-Schnatter & Wagner (2010) and adapted for the SV model by Chan (2016). This parameterization allows us to estimate the model for the verification of the partial identification of a structural shock through heteroskedasticity discussed in Section 4 as well as its unrestricted version with every structural shock following its own independent heteroskedastic process as in Section 3. Such a flexibility of the proposed framework is achieved by allowing the innovations of the state equations to be perfectly correlated across structural shocks, in the restricted case, with disturbance smoothing techniques proposed by De Jong & Shephard (1995), and Durbin & Koopman (2002), and further developed by Uzeda (2018). Finally, to derive a general, computationally feasible and efficient estimation algorithm implementing the features mentioned above we combine the auxiliary mixture sampler approach of Kim, Shephard & Chib (1998) in its particular implementation proposed by Omori, Chib, Shephard & Nakajima (2007) with the precision sampling techniques of Chan & Jeliazkov (2009).[3] The computational details of the Gibbs sampler for the SV model are given in Appendix B.

Sampling Bernoulli indicators. We estimate Bernoulli indicators by applying the S^4 technique that was proposed by Korobilis (2016), Koop & Korobilis (2016), and used by Schnücker

[2]We implement the NUTS algorithms through the Stan software package by Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li & Riddell (2017) and in particular we utilize functions from R package rstan (see Stan Development Team, 2018)

[3]We use R package SparseM by Koenker & Ng (2002) to implement operations on sparse matrices efficiently.

(2016) that is based on the ideas by George & Mcculloch (1997) and George, Sun & Ni (2008). To verify the partial identification of the n^{th} structural shock let $\boldsymbol{\kappa}_n$ denote the set of Bernoulli indicators κ_i for $i \in \{1, \ldots, N\} \backslash \{n\}$. Then sample each of its elements, κ_i, independently one after another from the Bernoulli distribution in which the probabilities of success and failure are proportional to the product of the corresponding values of the likelihood function and the prior probability. Some more details are given in Appendix B.

References

Amisano, G., & Giannini, C. (1997). *Topics in Structural VAR Econometrics*. Springer-Verlag, Berlin.

Arias, J. E., Rubio-Ramírez, J. F., & Waggoner, D. F. (2018). Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications. *Econometrica*, *86*, 685–720.

Baumeister, C., & Hamilton, J. D. (2015). Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica*, *83*, 1963–1999.

Bauwens, L., Richard, J., & Lubrano, M. (1999). *Bayesian inference in dynamic econometric models*. Oxford University Press, USA.

Bertsche, D., & Braun, R. (2018). *Identification of Structural Vector Autoregressions by Stochastic Volatility*. Technical Report 2018-01 University of Konstanz.

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

Canova, F., & Pérez Forero, F. J. (2015). Estimating Overidentified, Nonrecursive, Time-varying Coefficients Structural Vector Autoregressions. *Quantitative Economics*, *6*, 359–384.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*, 1–32.

Chan, J. (2016). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, (pp. 1–17).

Chan, J., & Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, .

Chan, J., & Jeliazkov, I. (2009). Efficient Simulation and Integrated Likelihood Estimation in State Space Models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*, 101–120.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, *29*, 327–341.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic Forecasting Performance Under Alternative Specification of Time-Varying Volatility. *Journal of Applied Econometrics*, *30*, 551–575.

Cogley, T. (2005). How fast can the new economy grow? A Bayesian analysis of the evolution of trend growth. *Journal of Macroeconomics*, *27*, 179–207.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*, 262–302.

De Jong, P., & Shephard, N. (1995). The Simulation Smoother for Time Series Models. *Biometrika*, *82*, 339–350.

Doan, T., Litterman, R. B., & Sims, C. A. (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews*, *3*, 37–41.

Durbin, J., & Koopman, S. J. (2002). A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika*, *89*, 603–615.

Eisenstat, E., & Strachan, R. W. (2016). Modelling inflation uncertainty. *Journal of Applied Econometrics*, *31*, 805–820.

Frühwirth-Schnatter, S., & Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, *154*, 85–100.

George, E. I., & Mcculloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, *7*, 339–373.

George, E. I., Sun, D., & Ni, S. (2008). Bayesian Stochastic Search for VAR Model Restrictions. *Journal of Econometrics*, *142*, 553–580.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*, 436–451.

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies*, *61*, 247–264.

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

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, *12*, 371–389.

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

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, *65*, 361–393.

Koenker, R., & Ng, P. (2002). SparseM: A Sparse Matrix Package for R.

Koop, G., & Korobilis, D. (2016). Model uncertainty in Panel Vector Autoregressive models. *European Economic Review*, *81*, 115–131.

Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics and Data Analysis*, *101*, 110–120.

Kroese, D. P., Taimre, T., & Botev, Z. I. (2013). *Handbook of Monte Carlo Methods* volume 706. John Wiley & Sons.

Lanne, M., & Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility. *Journal of*

Money, Credit and Banking, *40*, 1131–1149.

Lanne, M., & Lütkepohl, H. (2014). A statistical comparison of alternative identification schemes for monetary policy shocks. In J. Knif, & B. Pape (Eds.), *Contributions to Mathematics, Statistics, Econometrics, and Finance - Essays in Honour of Professor Seppo Pynnönen* ACTA WASAENSIA 296, STATISTICS 7 (pp. 137–152). Vaasa, Finland: University of Vaasa.

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

Lütkepohl, H. (2013). Identifying structural vector autoregressions via changes in volatility. *Advances in Econometrics*, *32*, 169–203.

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

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

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

Lütkepohl, H., & Velinov, A. (2016). Structural vector autoregressions: Checking identifying long-run restrictions via heteroskedasticity. *Journal of Economic Surveys*, *30*, 377–392.

Lütkepohl, H., & Woźniak, T. (2017). *Bayesian Inference for Structural Vector Autoregressions Identified by Markov-Switching Heteroskedasticity*. Technical Report DIW Berlin.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* chapter 5. (pp. 113–162). Chapman & Hall/CRC Press.

Nesterov, Y. (2009). Primal-Dual Subgradient Methods for Convex Problems. *Mathematical Programming*, *120*, 221–259.

Netšunajev, A. (2013). Reaction to technology shocks in Markov-switching structural VARs: Identification via heteroskedasticity. *Journal of Macroeconomics*, *36*, 51–62.

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

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics*, *140*, 425–449.

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

Rigobon, R., & Sack, B. (2003). Measuring the reaction of monetary policy to the stock market. *Quarterly Journal of Economics*, *118*, 639–669.

Schnücker, A. (2016). *Restrictions Search for Panel VARs*. Technical Report DIW Berlin.

Sentana, E., & Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic

factor models. *Journal of Econometrics*, *102*, 143–164.

Siddhartha Chib, Nardari, F., & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, *108*, 281–316.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.17.3.

Uhlig, H. (1997). Bayesian Vector Autoregressions with Stochastic Volatility. *Econometrica*, *65*, 59–73.

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.

Uzeda, L. (2018). *State Correlation and Forecasting : A Bayesian Approach Using Unobserved Components Models*. Technical Report Bank of Canada Ottawa, Ontario.

Waggoner, D. F., & Zha, T. (2003). A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics & Control*, *28*, 349–366.

Woźniak, T., & Droumaguet, M. (2015). Assessing Monetary Policy Models: Bayesian Inference for Heteroskedastic Structural VARs. *University of Melbourne Working Papers Series*, *2017*.

Appendix A. Proofs

Appendix A.1. Proof of Theorem 1

Let B_* be a matrix that satisfies:

$$\Sigma_t = B_* \Lambda_t B_*', \qquad t = 0, 1, \ldots.$$

It will be shown that, under the conditions of Theorem 1, the n^{th} column of B_* must be the same as that of B, except perhaps for a reversal of signs. Without loss of generality it is assumed in the following that $n = 1$ because this simplifies the notation. In other words, it is shown that the first columns of B and B_* are the same except for a reversal of signs, if $\sigma_1^2 \neq \sigma_i^2$, $i = 2, \ldots, N$.

There exists a nonsingular $(N \times N)$ matrix Q such that $B_* = BQ$. Using condition (4) for $t = 0$, Q has to satisfy the relation

$$BB' = BQQ'B'.$$

Multiplying this relation from the left by B^{-1} and from the right by $B^{-1'}$ implies that $QQ' = I_N$ and, hence, Q is an orthogonal matrix.

The relations

$$B\Lambda_t B' = BQ\Lambda_t Q'B'$$

imply $\Lambda_t = Q\Lambda_t Q'$ and, hence, $Q\Lambda_t = \Lambda_t Q$ for all $t = 0, 1, \ldots$.

Denoting the $(i.j)^{\text{th}}$ element of Q by q_{ij}, the latter equation implies that

$$q_{n1}\sigma_1^2 = q_{n1}\sigma_n^2, \qquad n = 1, \ldots, N.$$

Hence, since σ_n^2 is different from σ_1^2 for $n = 2, \ldots, N$, we must have $q_{n1} = 0$ for $n = 2, \ldots, N$. Since, Q is orthogonal, the first column must then be

$$(1, 0, \ldots, 0)' \qquad \text{or} \qquad (-1, 0, \ldots, 0)'$$

which proves the theorem. Q.E.D.

Appendix A.2. Proof of Corollary 1

First suppose that the main diagonal of B_0 does not consist of ones and consider the setup of Theorem 1 with $B = B_0^{-1}$. Then the arguments in the proof of Theorem 1 show that $B_{0*}^{-1} = B_0^{-1}Q$, where Q is as in the proof of Theorem 1. Hence, $B_{0*} = Q'B_0$, which shows that B_{0*} and B_0 have the same n^{th} row up to sign. Normalizing the n^{th} element in the n^{th} row of B_0 to unity implies then uniqueness of the n^{th} row. Note that normalizing elements of B_0 implies that the variances $\sigma_{n.t}^2$ change. However, the relative variances $\tilde{\boldsymbol{\sigma}}_n^2$ are not affected. Q.E.D.

Appendix A.3. Proof of Corollary 2

To show that uniqueness of the n^{th} row of B_0 implies a unique n^{th} column of B_0^{-1} we focus without loss of generality on the first row. If the first row of B_0 is unique, any other admissible B_0 matrix must be of the form QB_0, where Q is an orthogonal matrix of the form:

$$\begin{bmatrix} 1 & \mathbf{0}_{(1\times(N-1))} \\ \mathbf{0}_{((N-1)\times 1)} & Q_* \end{bmatrix},$$

with Q_* being an orthogonal $((N-1)\times(N-1))$ matrix. This fact is an easy implication of Theorem 1. Thus, any admissible inverse has the form $B_0^{-1}Q'$ and, hence, has the same first column as B_0^{-1}. Clearly, the same argument applies for any other row of B_0, meaning that the impact effects of the n^{th} shock are unique if the n^{th} row of B_0 is unique. This fact allows us to do impulse response analysis for a partially identified model. For each identified shock, unique impulse responses are obtained and can be easily computed in the usual way. Q.E.D.

Appendix B. Estimation of SVAR-SV model with restrictions

Appendix B.1. Notation

Let the $(N \times T)$ matrix $Y = [y_1, \ldots, y_T]$ collect all the observations of the time series considered and set the initial conditions $y_0, \ldots, y_{p-1}$ to the first p observations of the available dataset.. Let $K = 1 + pN$ and define the $(K \times 1)$ vector $x_t = \left(1, y'_{t-1}, y'_{t-1}, \ldots, y'_{t-p}\right)'$. It collects all the vectors of observations on the right-hand side of equation (1). Moreover, let $X = [x_1, \ldots, x_T]$ be a $(K \times T)$ matrix. Similarly, collect the residual terms in the matrix $W = [w_1, \ldots, w_T]$, and denote by W_n its n^{th} row. Define a $(N \times K)$ matrix $\tilde{\beta} = [\mu, B_1, \ldots, B_p]$ collecting all the autoregressive parameters on the right-hand side of equation (1) and its n^{th} row - a $(1 \times K)$ vector $\tilde{\beta}_n$. For convenience we also denote by θ all of the parameters of the model. Then equation (1) can be written in matrix notation as:

$$B_0 Y = \tilde{\beta} X + W, \tag{B.1}$$

and the n^{th} row of (B.1) can be written as:

$$B_{0.n} Y = \tilde{\beta}_n X + W_n, \tag{B.2}$$

for $n = 1, \ldots, N$.

Appendix B.2. A Hamiltonian Monte Carlo within Gibbs sampling algorithm

Appendix B.2.1. Hamiltonian Monte Carlo sampler for contemporaneous effects b_0

To sample the unrestricted elements of B_0 collected in the vector b_0 (see equation 9), rewrite the SVAR model from equation (1) as $\bar{y}_t = \bar{x}_t b_0 + w_t$, where $\bar{y}_t = \left(y'_t \otimes I_n\right) s_b - \tilde{\beta} x_t$, and $\bar{x}_t = -\left(y'_t \otimes I_n\right) S_b$. The same equation using the matrix notation is $\bar{Y} = \bar{X} b_0 + W$, where $\bar{Y} = \left(Y' \otimes I_n\right) s_b - \text{vec}(\tilde{\beta} X)$, and $\bar{X} = -\left(Y' \otimes I_n\right) S_b$. Define also an NT-vector $\sigma^2 = \left(\sigma^2_{1.1}, \ldots, \sigma^2_{N.1}, \ldots, \sigma^2_{1.T}, \ldots, \sigma^{2'}_{N.T}\right)'$ and an N-vector $\gamma_\beta = (\gamma_{\beta.1}, \ldots, \gamma_{\beta.N})'$. Then, the location and scale parameters of the full conditional posterior distribution for vector b_0 given in

equation (21) are:

$$\overline{V}_{b_0} = \left[\gamma_{b_0}^{-1} I_r + \operatorname{diag}\left(\tilde{\gamma}_\beta\right)^{-1} + \bar{X}' \operatorname{diag}\left(\sigma^2\right)^{-1} \bar{X}\right]^{-1}$$
$$\overline{b}_0 = \overline{V}_{b_0}\left[\operatorname{diag}\left(\tilde{\gamma}_\beta\right)^{-1} S_b' \operatorname{vec}(B_1) + \bar{X}' \operatorname{diag}\left(\sigma^2\right)^{-1} \bar{Y}\right]$$

where an r-vector $\tilde{\gamma}_\beta = S_b' \operatorname{diag}\left(\iota_N \otimes \gamma_\beta\right) \iota_{N^2}$.

Appendix B.2.2. Sampling autoregressive parameters $\tilde{\beta}$

We sample the constant term and the autoregressive parameters jointly by sampling vectors $\tilde{\beta}_n$ independently equation by equation from the K-variate normal distribution with the mean $B_{0.n}\overline{P}_n$ and covariance matrix $\overline{H}_n$ with the parameters given by:

$$\overline{H}_n = \left[X \operatorname{diag}\left(\sigma_{n.1}^2, \ldots, \sigma_{n.T}^2\right)^{-1} X' + \underline{\widetilde{H}}_n^{-1}\right]^{-1}, \tag{B.3}$$
$$\overline{P}_n = \left[Y \operatorname{diag}\left(\sigma_{n.1}^2, \ldots, \sigma_{n.T}^2\right)^{-1} X' + \underline{\widetilde{P}}\,\underline{\widetilde{H}}_n^{-1}\right]\overline{H}_n, \tag{B.4}$$

where $\underline{\widetilde{P}} = [\mathbf{0}_{N\times 1} \;\; \underline{P}]$ and: $\underline{\widetilde{H}}_n = \begin{bmatrix} \gamma_\mu & \mathbf{0}_{1\times pN} \\ \mathbf{0}_{pN\times 1} & \gamma_{\beta.n}\underline{H} \end{bmatrix}$.

Appendix B.2.3. Sampling shrinkage parameters γ_b, γ_μ, $\gamma_{\beta.n}$

The defined shrinkage parameters are sampled independently from the following inverse gamma 2 distributions:

$$\gamma_b | Y, b_0 \sim \mathcal{IG}2\left(\underline{a} + r, \underline{b} + b_0' b_0\right),$$
$$\gamma_\mu | Y, \mu \sim \mathcal{IG}2\left(\underline{a} + N, \underline{b} + \mu'\mu\right),$$
$$\gamma_{\beta.n} | Y, B_0, \beta_n \sim \mathcal{IG}2\left(\underline{a} + pN, \underline{b} + \left[\beta_n - B_{0.n}\underline{P}\right]' \underline{H}^{-1}\left[\beta_n - B_{0.n}\underline{P}\right]\right).$$

Appendix B.3. Sampling SV process

To sample log-volatilities consider the following representation of the SVAR in (B.1):

$$\operatorname{vec}\left(\left[B_0 Y - \tilde{\beta} X\right]'\right) = \operatorname{vec}(W'). \tag{B.5}$$

Next, note that equations (10) and (11) imply that:

$$w_{n.t} = \exp\left\{\frac{1}{2}\left(\omega_n h_{n.t} + \bar{h}_{n.0}\right)\right\}\tilde{w}_{n.t}, \tag{B.6}$$

where $\tilde{w}_{n.t} \sim \mathcal{N}(0,1)$. It is then easy to see that by squaring and then taking natural logarithms of each entry of the matrices in equation (B.5), i.e. setting $\tilde{y}_{n.t} = \log\left(B_{0.n}y_t - \tilde{\beta}_n x_t\right)^2$ and:

$$\log\left(w_{n.t}\right)^2 = \log\left(\exp\left\{\frac{1}{2}\left(\omega_n h_{n.t} + \bar{h}_{n.0}\right)\right\}\tilde{w}_{n.t}\right)^2 = \left(\omega_n h_{n.t} + \bar{h}_{n.0}\right) + \log\tilde{w}^2_{n.t} \tag{B.7}$$

leads to a convenient SVAR representation given below.[4] Let $\widetilde{Y} = (\tilde{y}_{1.1}, \ldots, \tilde{y}_{1.T}, \ldots, \tilde{y}_{N.1}, \ldots, \tilde{y}_{N.T})'$ and $\widetilde{W}^* = (\log\tilde{w}^2_{1.1}, \ldots, \log\tilde{w}^2_{1.T}, \ldots, \log\tilde{w}^2_{N.1}, \ldots, \log\tilde{w}^2_{N.T})'$. Then:

$$\widetilde{Y} = \left(\operatorname{diag}\left(S_\omega\tilde{\omega}\right)\otimes I_T\right)h + F\bar{h}_0 + \widetilde{W}^* \tag{B.8}$$

$$h = L^{-1}S_v\tilde{v}, \tag{B.9}$$

where:

$$L = I_N \otimes H,$$

$$F = I_N \otimes \imath_T,$$

$$S_v = S_\omega \otimes I_T.$$

The system in (B.8)-(B.9) is a standard linear state space model except for the fact that, after the transformations of the measurement equation in (B.5) discussed above, the elements of $\tilde{W}^*$ follow a log chi-squared distribution. To transform the state space system into a (conditionally) Gaussian linear state space model studies such as Kim et al. (1998) and Omori et al. (2007) propose approximating the distribution of $\log\tilde{w}^2_{n.t}$ by a mixture of Normal densities. In this paper we follow the approach in Omori et al. (2007) who

[4] In practice it is common to set $\tilde{y}_{n.t} = \log\left(B_{0.n}y_{n.t} - \tilde{\beta}_n x_{n.t} + \bar{c}\right)^2$ where $\bar{c}$ is a small constant, say, 10^{-4} to avoid numerical issues when $w_{n.t}$ is close to zero.

suggest a ten-component Gaussian mixture as follows:

$$\log \tilde{w}_{n.t}^2 \sim \Pr[s_{n.t} = 1] \mathcal{N}(\alpha_{s_{n.t}=1}, \Sigma_{s_{n.t}=1}) + \cdots + \Pr[s_{n.t} = 10] \mathcal{N}(\alpha_{s_{n.t}=10}, \Sigma_{s_{n.t}=10}), \tag{B.10}$$

where moments as well as the probabilities associated with each component-density above are are fixed and given by Omori et al. (2007). $s_{n.t} \in \{1, \ldots, 10\}$ is a discrete-valued state variable which is further estimated in our model. Conditional on draws for $s = (s_{1.1}, \ldots, s_{1.T}, \ldots, s_{N.1}, \ldots, s_{N.T})$ the measurement equation (B.8) can be expressed as:

$$\widetilde{Y}_\alpha = (\text{diag}(S_\omega \tilde{\omega}) \otimes I_T) h + F\bar{h}_0 + \widetilde{W}_s^*, \tag{B.11}$$

where $\tilde{Y}_\alpha = \tilde{Y} - \alpha_s$, and $\alpha_s = (\alpha_{s_{1.1}}, \ldots, \alpha_{s_{1.T}}, \ldots, \alpha_{s_{N.1}}, \ldots, \alpha_{s_{N.T}})'$. $\widetilde{W}_s^* = \widetilde{W}^* | s$ are, conditionally on s, normally distributed with the mean set to a vector of zeros and with a diagonal covariance matrix $\Sigma_s = \text{diag}(\Sigma_{s_{1.1}}, \ldots, \Sigma_{s_{1.T}}, \ldots, \Sigma_{s_{N.1}}, \ldots, \Sigma_{s_{N.T}})$.

An MCMC algorithm to estimate the SV model in our SVAR setting requires sampling from the full conditional posterior densities of s, $\tilde{v}$, $\tilde{\omega}$, and $\bar{h}_0$. The three latter distributions, conditionally on s are multivariate normal. For the computational efficiency of the algorithm it is essential to apply the precision sampler by Chan & Jeliazkov (2009) with operations on sparse or band matrices in the computer code for the computations of the mean vectors and the covariance matrices of these distributions. Next, each $s_{n.t}$ is sampled from a multinomial distribution. Finally, for the model with restrictions κ_i indicators are sampled independently from a Bernoulli distribution.

Appendix B.3.1. Sampling latent process s

The auxiliary sampler of Omori et al. (2007) allows each element of s to be drawn independently *a posteriori*. The full conditional posterior probabilities $\Pr\left[s_{n.t} = i | \tilde{y}_{n.t}, h_{n.t}, \bar{h}_{n.0}, \tilde{\omega}\right]$ are computed using Bayes' rule and are given by:

$$\Pr\left[s_{n.t} = i | \tilde{y}_{n.t}, h_{n.t}, \bar{h}_{n.0}, \tilde{\omega}\right] = \frac{\Pr[s_{n.t} = i] \mathcal{N}\left(\tilde{y}_{n.t} | s_{n.t} = i, h_{n.t}, \bar{h}_{n.0}, \tilde{\omega}\right)}{\sum_{j=1}^{10} \Pr[s_{n.t} = j] \mathcal{N}\left(\tilde{y}_{n.t} | s_{n.t} = j, h_{n.t}, \bar{h}_{n.0}, \tilde{\omega}\right)}, \tag{B.12}$$

for $i = 1, \ldots, 10$, where the normal form of the likelihood function used in the equation above for a given $s_{n.t}$ follows directly from equation (B.11). Given the probabilities above each $s_{n.t}$ is drawn independently from a multinomial distribution through an inverse transform method (see Kroese, Taimre & Botev, 2013).

Appendix B.3.2. Sampling SV state equation disturbances $\tilde{v}$

The disturbances of the SV state equation are sampled jointly from a NT-variate normal distribution with the mean $\overline{d_v}$ and the covariance matrix $\overline{D_v}$ that are given by:

$$\overline{D_v} = \left[S_v' L^{-1\prime} \left(\operatorname{diag}\left(S_\omega \tilde{\omega} \right) \otimes I_T \right)' \Sigma_s^{-1} \left(\operatorname{diag}\left(S_\omega \tilde{\omega} \right) \otimes I_T \right) L^{-1} S_v + I_{NT} \right]^{-1}, \tag{B.13}$$

$$\overline{d_v} = \overline{D_v} S_v' L^{-1\prime} \left(\operatorname{diag}\left(S_\omega \tilde{\omega} \right) \otimes I_T \right)' \Sigma_s^{-1} \left(\tilde{Y}_\alpha - F \bar{h}_0 \right). \tag{B.14}$$

Appendix B.3.3. Sampling SV process scale parameters $\tilde{\omega}$

Rewrite (B.11) as:

$$\widetilde{Y}_\alpha = h^* S_\omega \tilde{\omega} + F \bar{h}_0 + \widetilde{W}_s^*, \tag{B.15}$$

where:

$$\underset{(NT \times N)}{h^*} = \sum_{n=1}^{N} \tilde{S}_n h \tilde{s}_n, \tag{B.16}$$

where $\tilde{S}_n$ is an $(NT \times NT)$ matrix filled with zeros except for its columns from $(n-1)T+1$ to $(n-1)T+T$ that are set to the corresponding columns of I_{NT}, and $\tilde{s}_n$ is a $(1 \times N)$ matrix equal to the n^{th} row of I_N. Then, sample $\tilde{\omega}$ jointly from a N-variate normal distribution with the mean $\overline{d_\omega}$ and the covariance matrix $\overline{D_\omega}$ given by:

$$\overline{D_\omega} = \left[S_\omega' h^{*\prime} \Sigma_s^{-1} h^* S_\omega + \operatorname{diag}\left(\underline{D}_\omega \right)^{-1} \right]^{-1}, \tag{B.17}$$

$$\overline{d_\omega} = \overline{D_\omega} S_\omega' h^{*\prime} \Sigma_s^{-1} \left(\tilde{Y}_\alpha - F \bar{h}_0 \right). \tag{B.18}$$

Appendix B.3.4. Sampling initial conditions $\bar{h}_0$

The vector of initial values of the SV process, $\bar{h}_0$, are sampled jointly from the N-variate normal distribution with the mean $\overline{d}_{h_0}$ and the covariance matrix $\overline{D}_{h_0}$ given by:

$$\overline{D}_{h_0} = \left(F'\Sigma_v^{-1}F + \operatorname{diag}\left(\underline{D}_{h_0}\right)^{-1}\right)^{-1}, \tag{B.19}$$

$$\overline{d}_{h_0} = \overline{D}_{h_0}F'\Sigma_v^{-1}\left[\widetilde{Y}_\alpha - \left(\operatorname{diag}\left(S_\omega\tilde{\omega}\right)\otimes I_T\right)h\right]. \tag{B.20}$$

Appendix B.4. Sampling Bernoulli indicators κ

To verify the partial identification of the n^{th} structural shock let $\boldsymbol{\kappa}_n$ denote the set of Bernoulli indicators κ_i for $i \in \{1,\ldots,N\}\backslash\{n\}$. Then sample each of its element κ_i independently one after another for $i \in \{1,\ldots,N\}\backslash\{n\}$ from the Bernoulli distribution in which the probabilities of success and failure proportional to the product of the corresponding values of the likelihood function and the prior probability. In other words, draw κ_i from the Bernoulli distribution with the probability of success given by $\overline{\pi}_i = \frac{p_{1.i}}{p_{0.i}+p_{1.i}}$ where:

$$p_{1.i} = p\left(\tilde{Y}_\alpha|\kappa_i = 1, \kappa_{(i,n)}, s, \bar{h}_0, \tilde{v}, \tilde{\omega}, B_0, \ldots, B_p\right)\underline{\pi}, \tag{B.21}$$

$$p_{0.i} = p\left(\tilde{Y}_\alpha|\kappa_i = 0, \kappa_{(i,n)}, s, \bar{h}_0, \tilde{v}, \tilde{\omega}, B_0, \ldots, B_p\right)(1-\underline{\pi}). \tag{B.22}$$