Assessing Monetary Policy Models: Bayesian Inference for Heteroskedastic Structural VARs

Tomasz Woźniak [☆a], Matthieu Droumaguet[b]

[a]*Department of Economics, University of Melbourne*
[b]*Goldman Sachs, Hong Kong*

November 20, 2015

Abstract

We present a flexible structural vector autoregressive modeling framework with identification via heteroskedasticity. It encompasses a range of volatility models and allows for imposing over-identifying restrictions. The latter feature is of particular importance because conventional identifying restrictions can be imposed on the contemporaneous effects matrix identified via heteroskedasticity. As a result, statistical methods can be used for comparing models with alternative sets of restrictions that just-identify a conventional SVAR model. Efficient Bayesian algorithms are derived for estimating larger models that are difficult to handle in a frequentist analysis. We propose an appropriate marginal data density estimator for model selection and comparison. These tools can be used for models that are not globally identified and for comparing non-nested hypotheses. The new method is applied in a comparison of three classical monetary policy models for the U.S. We find that models for so-called non-conventional monetary policy outperform the model that considers a shock to federal funds rate as a monetary policy instrument.

Keywords: Identification Through Heteroskedasticity, Monetary Policy, Markov-switching Models, Mixture Models, Regime Change, Bayesian Hypotheses Assessment

JEL classification: C11, C12, C32, E52

1. Introduction

Our objective is to establish empirical support for three popular monetary policy models for the U.S., all expressed as Structural Vector Autoregressions (SVAR), and to let data decide on a ranking amongst them. Specifically, we consider models that identify monetary policy shocks in a system of seven variables by alternative definitions of the monetary policy instrument. These models induce policy shocks via different targeting variables used by the monetary authority. These are the federal funds rate (FF), as motivated by works of Bernanke & Blinder (1992), Sims (1986, 1992) and others, non-borrowed reserves (NBR), as proposed by Christiano & Eichenbaum (1992), and total reserves (NBR/TR), as suggested by Strongin (1995) (see also Christiano, Eichenbaum & Evans, 1999, for a more detailed discussion of these monetary policy models). In these conventional analyses, assessing the impact of monetary policy shocks on the economy is done through just-identifying restrictions.

In this paper, we identify the SVARs through heteroskedasticity (see e.g. Rigobon, 2003), and therefore, we can test the identifying restrictions behind the monetary policy models. Moreover, posterior probabilities

[☆]*Corresponding author:* Department of Economics, The University of Melbourne, Level 4, FBE Building, 111 Barry Street, Carlton 3053, Victoria, Australia, *Phone:* +61 3 8344 5310, *Fax:* +61 3 8344 6899, *Email address:* tomasz.wozniak@unimelb.edu.au, *URL:* http://bit.ly/tomaszw.

of models given data are used in order to discriminate between the models. Consequently, not only does our analysis formally test the restrictions, but it delivers positive arguments supporting the use of one of the models for the analysis of the effects of monetary policy actions on the real economy.

This paper proposes a flexible SVAR model with heteroskedastic structural shocks. Changes in the volatility of the structural shocks follow a Markov-switching process, and therefore, we call the model SVAR with Markov-switching Heteroskedasticity (hereafter SVAR-MSH). This model enables us to estimate all of the elements of the contemporaneous effects matrix. Any exclusion restrictions imposed on this matrix, such as those identifying the monetary policy shocks in the aforementioned papers, over-identify the model, and thus, can be tested. Moreover, our novel algorithm allows for the estimation of a variety of heteroskedastic processes that were used to identify SVARs through heteroskedasticity. Particularly, it encompasses a Markov-switching heteroskedastic model used by Lanne, Lütkepohl & Maciejowska (2010) and its particular version with endogenously determined monotonic regime changes (as proposed by Chib, 1998), a mixture of normal distributions model used by Lanne & Lütkepohl (2010), exogenously determined changes in volatility as in Lanne & Lütkepohl (2010), as well as other specifications (see Sims, Waggoner & Zha, 2008). This great variety of volatility processes is used to prove our results robust.

We use quarterly time series beginning in 1960 and finishing in 2007 in order to assess and discriminate between the three monetary policy models. Our empirical analysis delivers a strong argument in favor of models for so-called non-conventional monetary policy that outperform the model that considers a shock to federal funds rate as a monetary policy instrument. In particular, the model of Strongin (1995) that defines this instrument as an interaction between the non-borrowed and total reserves gains most of the posterior probability mass. All of the three models, however, sustain their validity when compared with the unrestricted model in which all of the elements of the structural matrix are estimated. These findings contradict the results of Lanne & Lütkepohl (2008, 2013) and Normandin & Phaneuf (2004) that were obtained using a maximum likelihood based inference for exogenously determined change points for the volatility regimes, mixed normal distributions, and a GARCH process respectively. Lanne & Lütkepohl (2008) reject all models, while Lanne & Lütkepohl (2013) find that the FF scheme is the only one not rejected by the data. Instead, Normandin & Phaneuf (2004) clearly reject a model with interest rate targeting. While the differences between our conclusions and those presented by other authors could be attributed to different sample periods, data frequencies and model specifications[1], our results are qualitatively more appealing: they are based on marginal data densities that are appropriate for small samples and, most importantly, are robust with respect to the parameter uncertainty and the heteroskedastic process specification. Amongst multiple alternatives for modeling heteroskedasticity, a Markov-switching model with three states occurs to have the best in-sample fit.

One of the challenges in performing our comparisons is relative size of the sample given the number of parameters in our model. We address this issue by applying Bayesian inference that includes the derivation of a flexible estimation algorithm for the models with shrinkage prior distributions and applying appropriate inference tools. More precisely, we derive a Gibbs sampler for the estimation of models with potentially large numbers of variables and with any number of volatility states that is supported by the data. The algorithm allows for estimating models with various heteroskedasticity patterns as well as for imposing flexible restrictions on the short-run effects of the shocks.

Studies on identification through heteroskedasticity published so far use frequentist methods for the estimation and the statistical inference. The models used are relatively simple in the sense that they include a very limited number of variables only or they use very simple models for the volatility changes. An important reason for these limitations is that frequentist methods are problematic for more complex models. For example, optimizing the likelihood function of an SVAR model with residual volatility modeled with an Markov-switching mechanism is a formidable computational challenge. It becomes infeasible when there are more than three or four variables and the number of volatility regimes is large. In order to explain the problems with the maximization of the likelihood function for such models we use the permutation

[1] All of the three papers use monthly data. Lanne & Lütkepohl (2008, 2013) use sample period from 1965 to 1995, whereas Normandin & Phaneuf (2004) choose the sample from 1982 to late 1990's. All of the papers are based on partial identification in which the restrictions are imposed only in equations corresponding monetary variables.

augmented sampling proposed by Geweke (2007). As a result, we discover the shape of the posterior distribution (and implicitly of the likelihood function) for the parameters of the structural-form globally illustrating its multi-modality. The number of modes increases rapidly with the number of variables in the system and the number of heteroskedastic states.

The problems described above are caused by the lack of global identification of the structural system due to possible changes of the ordering and signs of the structural shocks, and the relabelling of the heteroskedastic states. This identification issue constitutes a challenge for Bayesian inference. As shown by Celeux, Hurn & Robert (2000) and Frühwirth-Schnatter (2004), the lack of identification of the mixture models leads to a biased estimation of the marginal data densities (MDD). We show that the same solution that allows the unbiased estimation of the MDDs for mixture models exposed to the label switching problem solves the problem of making inference on globally unidentified structural models. Therefore, for the purpose of the unbiased estimation of the MDDs for the SVAR-MSH models, firstly, we apply the permutation sampling introduced by Frühwirth-Schnatter (2001), and then, adapt the estimator of Chib (1995) with the fix of Marin & Robert (2008) to our structural system. Moreover, we generalize the estimator of Marin & Robert (2008) to the case in which the parameters of the model are divided in convenient groups allowing for a feasible Gibbs sampler.

In the following parts of the paper we first introduce the three monetary policy models (Section 2), and then, the family of Structural VARs with Markov-switching heteroskedasticity and consider the problem of statistical inference in a model that is not globally identified (Section 3). In Section 4, we show that a range of different patterns of heteroskedasticity are nested within our general specification. The Bayesian analysis is presented in Section 5. All the details of the estimation of the marginal data densities are given in Section 6. Section 7 empirically investigates which monetary policy model is supported by the data. Section 8 concludes.

2. Three Models for U.S. Monetary Policy Analysis

Christiano et al. (1999) (henceforth CEE) review a number of identification schemes for structural vector autoregressive (SVAR) models which have been used in the related literature to specify monetary policy shocks and analyze their impact on the economy. The structural restrictions in each of the schemes just identify the monetary policy shocks. Thus, in the conventional SVAR setup, the different schemes cannot be compared with statistical tests. Therefore, we use changes in the volatility of the residuals for comparing the different identification schemes with statistical methods.

Seven variables are used in CEE's benchmark monetary models for the U.S.: gdp_t - log of real GDP, p_t - log of GDP deflator, $pcom_t$ - log of index of commodity prices, FF_t - federal funds rate, nbr_t - log of non-borrowed reserves plus extended credit, tr_t - log of total reserves, m_t - log of money stock M1. For this system of macroeconomic aggregates CEE consider an N-dimensional SVAR(p) model:

$$A_0 y_t = \mu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \tag{1}$$

$$u_t \sim \mathcal{N}\left(\mathbf{0}, \operatorname{diag}(\lambda)\right) \tag{2}$$

where μ is a constant term, A_i $(i = 1, \ldots, p)$ are $N \times N$ coefficient matrices and u_t is a Gaussian white noise error term with a diagonal covariance matrix. The diagonal elements are collected in the $N \times 1$ vector λ.

CEE consider the following three alternative identification schemes that are based on different assumptions about the targeting variable for monetary policy and have been used in the related literature:

FF policy shock The federal funds rate FF_t is used as policy instrument. This identification scheme is motivated by work of Bernanke & Blinder (1992), Sims (1986, 1992) and others.

NBR policy shock nbr_t is assumed to be the policy instrument, which is suggested by assumptions made in a study by Christiano & Eichenbaum (1992).

NBR/TR policy shock Strongin (1995) assume that the Fed accommodates demand shocks to total reserves and threats the interactions between nbr_t and tr_t as the monetary policy instrument.

Table 1: Restrictions imposed on A_0 matrix for alternative monetary policy models

FF policy shock model	NBR policy shock model	NBR-TR policy shock model	y_t
$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & 0 & 0 \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & 0 \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} \end{bmatrix}$	$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & 0 & 0 \\ a_{51} & a_{52} & a_{53} & 0 & a_{55} & 0 & 0 \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & 0 \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} \end{bmatrix}$	$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & 0 \\ a_{51} & a_{52} & a_{53} & 0 & a_{55} & a_{56} & 0 \\ a_{61} & a_{62} & a_{63} & 0 & 0 & a_{66} & 0 \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} \end{bmatrix}$	$\begin{bmatrix} gdp_t \\ p_t \\ pcom_t \\ FF_t \\ nbr_t \\ tr_t \\ m_t \end{bmatrix}$

In the original papers, the different identifications were achieved by assuming a lower-triangular matrix A_0 and appropriately reordering variables FF_t, nbr_t and tr_t in vector y_t. For our purposes, however, it is useful to maintain the same ordering of the variables for all identification schemes. Using vector

$$y_t = (gdp_t, p_t, pcom_t, FF_t, nbr_t, tr_t, m_t)'$$

as in the FF scheme, we represent the three monetary policy models by different sets of restrictions on matrix A_0. These restrictions are presented in Table 1.

In the SVAR models of monetary policy the restrictions on A_0 just-identify the monetary policy shocks. There are no over-identifying restrictions that could be tested against the data. The SVAR-MSH model presented in Section 3 overcomes this problem and enables us to compare the three identification schemes with statistical tools.

Related comparisons based on models with heteroskedasticity in a frequentist setting were performed by Lanne & Lütkepohl (2008, 2013) and Normandin & Phaneuf (2004). The first paper assumes exogenously given change points for the volatility regimes and the second study also considers mixed normal distributions while Normandin & Phaneuf (2004) model volatility changes by a GARCH process. Both Normandin & Phaneuf (2004) and Lanne & Lütkepohl (2008) use six variable models only. They use monthly data and partly consider slightly different parameter restrictions. In their frequentist setting, they can test the various sets of restrictions but cannot test the models directly against each other. Their results are quite heterogeneous. While Lanne & Lütkepohl (2013) find that the FF scheme is the only one not rejected by the data, Normandin & Phaneuf (2004) clearly reject a model with interest rate targeting. Lanne & Lütkepohl (2008) reject all models.

In our Bayesian framework we can use various volatility models including a full SVAR-MSH model. Moreover, by analyzing posterior odds we can not only test the restricted models against the unrestricted one, but also discriminate between different monetary policy schemes and investigate which of them is the most likely one given the data.

3. Heteroskedastic Structural Vector Autoregressions

3.1. The model

In this section, we introduce a Structural VAR model for the N-dimensional vector of observable variables y_t in which the structural shocks are heteroskedastic, namely the SVAR-MSH model. The structural form of the model is given by equation (1). However, let the diagonal elements of the diagonal covariance matrices of the structural shocks change over time according to a latent process s_t, for $t = 1, \dots, T$. Conditionally on this process the structural shocks are normally distributed with the mean set to a vector of zeros and diagonal covariance matrix:

$$u_t | s_t \sim \mathcal{N}\left(\mathbf{0}, \operatorname{diag}\left(\lambda_{s_t}\right)\right), \tag{3}$$

for $t = 1, \dots, T$. Further in this section, we use the heteroskedasticity feature of the data in order to identify the matrix of contemporaneous effects A_0.

The process s_t for each t can take a discrete number from values $m = 1, \dots, M$. It is referred to as a hidden Markov process because it is unobservable and it possesses a Markov property: both, the conditional probabilities of states at time t, denoted by $\Pr[s_t|\mathbf{y}_{t-1}, s_{t-1}, \theta]$, and the conditional density of vector of observations, denoted by $p(y_t|\mathbf{y}_{t-1}, s_{t-1}, \theta)$, depend only on the previous state of the Markov process, s_{t-1}, and not on any of its former states (see e.g. Frühwirth-Schnatter, 2006, Chapter 10). The properties of the Markov process are fully governed by a $M \times M$ transition probabilities matrix $\mathbf{P}$, with its element (i, j) being the probability of switching to state j at time t, given that at time $t-1$ the process was in state i:

$$p_{ij} = \Pr[s_t = j|s_{t-1} = i], \tag{4}$$

for $i, j = 1, \dots, M$ and $\sum_{j=1}^{M} p_{ij} = 1$. Since the hidden Markov process has M states, also M vectors of variances of the structural shocks, $\lambda_1, \dots, \lambda_M$, are estimated. Such a flexible Markov-switching heteroskedastic process gives a lot of possibilities of modeling particular patterns of changes in volatility in economic data. This will be further discussed in Section 4.

The Markov-switching process for modeling heteroskedasticity has been successfully implemented to macroeconomic time series analysis. Markov-switching VAR models have been used in multiple studies (see e.g. Sims & Zha, 2006; Sims et al., 2008), also for the purpose of identification of the contemporaneous effects matrix A_0 in equation (1) (see Lanne et al., 2010; Herwartz & Lütkepohl, 2014; Kulikov & Netšunajev, 2013; Markun, 2011). In numerous empirical studies MS-VAR models proved not only to extremely well fit the data, but also they allowed for proposing a historical description of sub-periods in various economies corresponding to persistent states of the model with higher and lower volatilities.

The SVAR-MSH model presented so far is identical to the model introduced by Lanne et al. (2010). In this model the heteroskedastic error term allows for the statistical identification of all of the N^2 elements of the matrix of contemporaneous effects A_0 according to the identification proposed by Rigobon (2003). Therefore, the whole matrix A_0 can be estimated in an identified structural form model given by equations (1) and (3). Any further restrictions imposed on the matrix A_0 are over-identifying the system, and thus, can be tested. In particular, the restrictions implied by the three different identification schemes for the monetary policy shocks can be compared to one another as well as against the model with unrestricted matrix A_0. In effect, the assumptions of the theoretical structural models motivating the zero restrictions for SVARs can be verified using the feature of heteroskedasticity of the error term.

In order to gain more insights into the identification through heteroskedasticity we analyze the implied reduced form model and its relations to the structural form. The reduced form of the model is obtained by multiplying the structural form model given by equation (1) by A_0^{-1} from the left hand side. Therefore, the reduced form residuals, ϵ_t, are related to the structural form residuals by $A_0\epsilon_t = u_t$, if A_0 is uniquely determined. Suppose that the M covariance matrices of the reduced form residuals are Σ_m for $m = 1, \dots, M$. Then, there exists a decomposition $A_0\Sigma_m A_0' = \text{diag}(\lambda_m)$ for $m = 1, \dots, M$. Note that M symmetric matrices Σ_m have $M \cdot N(N+1)/2$ unique elements, whereas matrix A_0 and M vectors λ_m have $N^2 + MN$ elements.

Furthermore, we impose restrictions by standardizing the first vector of variances of the structural shocks to a vector of ones: $\lambda_1 = \imath_N$. In such a case matrix A_0 remains unrestricted. The conditions for the uniqueness of this decomposition for any number of states M are derived in Lanne et al. (2010). Another possibility is to standardize the diagonal elements of matrix A_0 to ones, $\text{diag}(A_0) = \imath_N$, leaving vectors λ_m unrestricted, for $m = 1, \dots, M$ (see Section 5 as well as Markun, 2011, Chapter 2 for more details). These two decompositions are observationally equivalent. However, we motivate our choice by the possibility of deriving a fast and efficient estimation algorithm. Taking this standardization into account the heteroskedastic structural system is either exactly identified, for $M = 2$, or over-identified, for $M > 2$.

The interpretation of the parameters of the structural form model comes directly from the choice of the standardization. They can be presented in the following form:

(i) $A_0\Sigma_1 A_0' = I_N$ and $A_0\Sigma_m A_0' = \text{diag}(\lambda_m)$ for $m = 2, \dots, M$,

(ii) $B_0\Sigma_m B_0' = \text{diag}(\tau_m)$ for $m = 1, \dots, M$ and $\text{diag}(B_0) = \imath_N$.

The first pattern is the one used in the current study, whereas the second one restricts the diagonal elements of the matrix of the contemporaneous effects B_0 to ones, leaving the variances of the structural shocks, τ_m unrestricted, for $m = 1, \dots, M$. Pattern (ii) explicitly models the variances of the structural shocks. Its elements can be transformed to pattern (i) by the following transformations: $\lambda_m = \tau_m / \tau_1$ and $A_0 = \operatorname{diag}(\tau_1)^{-1/2} B_0$, where "/" denotes element-wise division for two vectors. As we standardize the λ_m such that $\lambda_1 = \imath_N$, all the remaining variances of the structural shocks, λ_m for $m = 2, \dots, M$, should be interpreted as ratios of variances in the m^{th} state, relative to the variances in the first state. For instance, some value of $\lambda_{m.i}$ ($m = 2, \dots, M$ and $i = 1, \dots, N$) that is greater (less) than one says, how many times the variance of the i^{th} structural shock is larger (smaller) than its variance in the first state ($m = 1$). Moreover, the variances of the structural shocks in the first state are captured by the matrix A_0.

We emphasize that the identification of A_0 using heteroskedasticity is only a statistical identification that allows us to estimate all the elements of this matrix without imposing any further restrictions on the model. However, the uncorrelated shocks do not have economic interpretations as such. In order to call any of the structural shocks, say a monetary policy shock, economic theory needs to be employed. Still, it is extremely useful to use such an identification of the structural shocks as it gives the possibility of testing any of the further restrictions imposed on the model.

Imposing restrictions on A_0. In order to identify economic structural shocks in many studies exclusion restrictions are imposed on the matrix of contemporaneous effects. We adapt the framework of Waggoner & Zha (2003) in order to impose the zero restrictions on some of the elements of matrix A_0. This task can be handled much easier if one analyzes the system in equation (1) for each equation separately:

$$A_{0.n} y_t = \mu_n + A_{1.n} y_{t-1} + \cdots + A_{p.n} y_{t-p} + u_{t.n}, \tag{5}$$

for $t = 1, \dots, T$ and $n = 1, \dots, N$, where $A_{i.n}$ denotes the n^{th} row of matrix A_i.

The vectors $A_{0.n}$ can be represented in terms of their unrestricted elements:

$$A_{0.n} = a_n V_n, \tag{6}$$

for $n = 1, \dots, N$, where a $1 \times r_n$ vector a_n contain the unrestricted elements of $A_{0.n}$ and V_n is a fixed suitable $r_n \times N$ matrix containing zeros and ones. For instance, consider a 2×2 matrix A_0 that is restricted to an upper-triangular matrix such that $A_{0.21} = 0$. Then, for the first row: $a_1 = \begin{bmatrix} A_{0.11} & A_{0.12} \end{bmatrix}$ and $V_1 = I_N$, whereas for the second row: $a_2 = A_{0.22}$ and $V_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

Using this notation, equations (1) and (5) can be written in terms of the unrestricted parameters as:

$$a_n V_n y_t = \mu_n + A_{1.n} y_{t-1} + \cdots + A_{p.n} y_{t-p} + u_{t.n}. \tag{7}$$

Note, that by imposing restrictions on matrix A_0 as in equation (6), we obtain a multitude of possible identification schemes that include: an unrestricted A_0 matrix, recursive and non-recursive identification patterns, just-identifying and over-identifying restrictions.

3.2. *The problem of global identification of the structural system*

The structural system in the case of normalization (i) consists of the (potentially unrestricted) matrix A_0, the restricted vector $\lambda_1 = \imath_N$, and the unrestricted vectors $\lambda_2, \dots, \lambda_M$. Usual identification conditions ensure local identification only. We shall now consider the problem of global identification of the system and its consequences for inference. The structural system is identified up to:

- relabeling of heteroskedastic states,
- sign change of structural shocks,
- reordering of structural shocks.

The first problem has its origins in the invariance of the likelihood function to the relabelling of the regimes of the hidden Markov process as described by Celeux et al. (2000). The problem can be easily illustrated for the variances of structural shocks. Consider a simple case, in which there are two volatility states, $M = 2$, say high and low volatility state. Then, irrespectively of whether we call τ_1 a high volatility state and τ_2 a low volatility state or conversely the value of the likelihood function remains the same. In order to show that the problem is also relevant in standardization (i) we express λ_2 and A_0 ($M = 2$ and $\lambda_1 = \imath_N$) in relation to the parameters of standardisation (ii). We obtain two possible labelings of the states. In the first possible labelling the system is standardized by τ_1, and thus, can be defined by: $A_0 = \text{diag}(\tau_1)^{-1/2}B_0$ and $\lambda_2 = \tau_2/\tau_1$. Whereas in the second labelling it is standardized by τ_2, therefore, it becomes: $A_0 = \text{diag}(\tau_2)^{-1/2}B_0$ and $\lambda_2 = \tau_1/\tau_2$. Thus, if for some n, $\tau_{1.n} > \tau_{2.n}$, then $\lambda_{2.n}$ in the first labelling is less than 1, and in the second one it is greater than 1.

The second problem is a standard feature of SVAR models for which the matrix of contemporaneous effects is identified up to signs of the rows of that matrix. If one changes the sign of any number of rows of matrix A_0, then all such matrices will lead to the same value of the likelihood function. It is a characteristic of a system in which one identifies from the data variance-covariance matrices Σ_m and identifies the structural system by $\Sigma_m = A_0^{-1}\text{diag}(\lambda_m)A_0^{-1\prime}$, for $m = 1, \ldots, M$.

Another feature of the structural system identified via heteroskedasticity, in which all N^2 elements of the A_0 matrix are estimated, is that it is only identified up to permutations of rows of A_0 and elements of λ_m (see Rigobon, 2003). This problem results from the fact that the ordering of the shocks is a priori arbitrary. This feature can be described as the invariance of the likelihood function with respect to the reordering of the rows of A_0 and the associated reordering the elements of the vectors λ_m.

Inference and global identification. All the three aforementioned challenges for identification are a manifestation of the same problem, namely, the invariance of the likelihood function with respect to different rearrangements of the structural system. In effect, the likelihood function as well as the posterior distribution have $M! \cdot N! \cdot 2^N$ modes ($M!$ is attributed to heteroskedastic states' label switching, 2^N to the sign change and $N!$ to the reordering of of the structural shocks).

Figure 1 illustrates the shape of the joint posterior distribution of two parameters from the same column of matrix A_0 for the case of $N = 2$ and $M = 2$. Graphs from (a) to (c) present the distribution if only the local identification with respect to the change of signs of rows of A_0, equation permutation or labels switching of heteroskedastic states are taken into account respectively. Consequently, the distribution in graph (a) is symmetric around both of the zero lines, whereas the one in graph (b) is symmetric around the 45 degrees line. State labels permutation results in a distribution with two modes when one is a rescaled version of the other by the structural shock variance from the second state. Finally, graph (d) of Figure 1 illustrates the global shape of the marginal posterior distribution for the two parameters. It is obtained after applying the full permutation augmentation to the draws. Consequently, it has 16 modes.

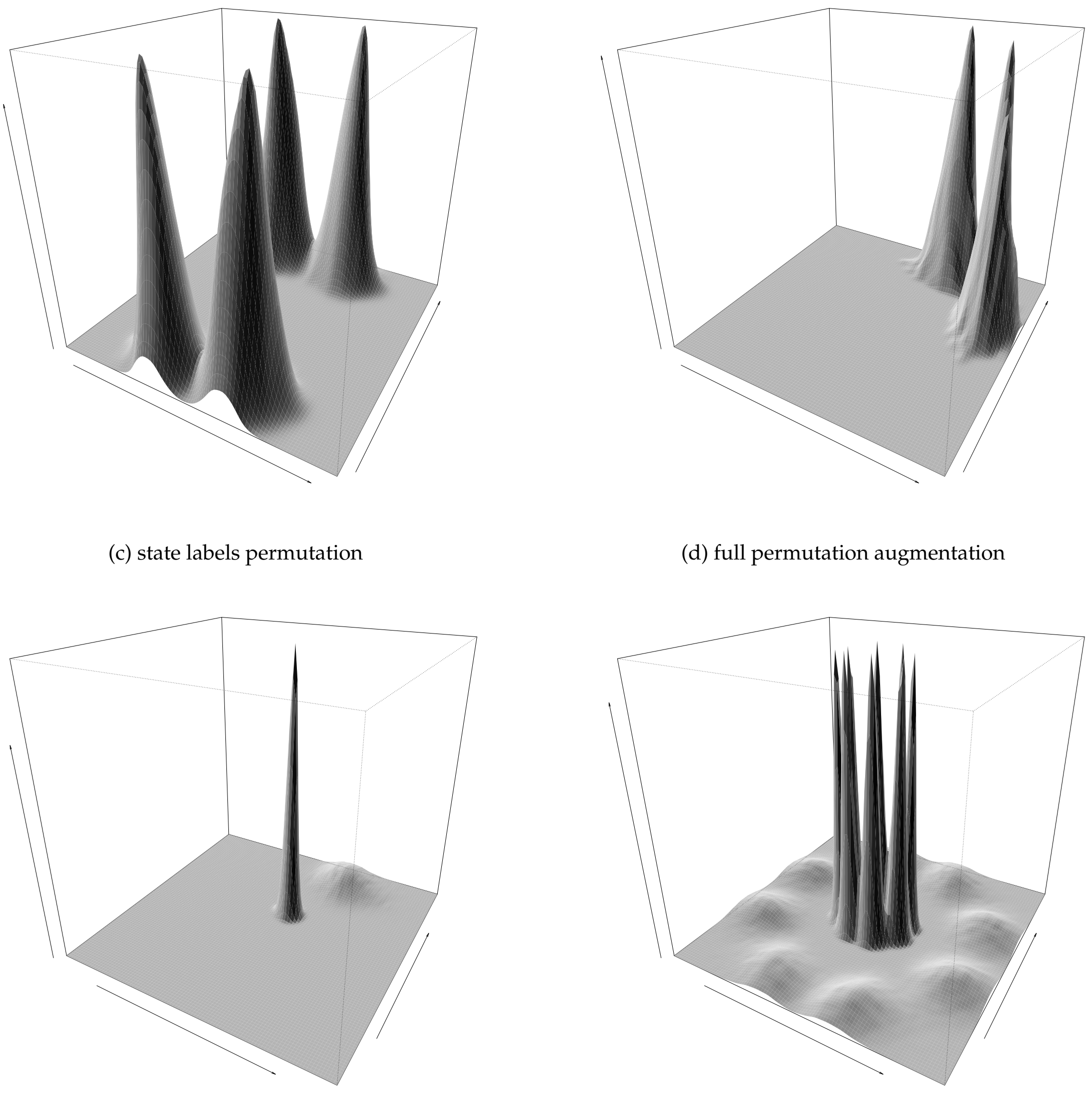
In such a case, the lack of the global identification of the structural model has particularly serious consequences for the maximum likelihood estimation of the SVAR-MSH. The number of global maxima increases with the number of volatility states as well as with with the number of variables in the system. The main obstacle is the maximization of the log-likelihood function. With a multiplicity of global maxima that are not necessarily well separated the estimation appears hardly possible with increasing N or M. Similar difficulties are encountered during the assessment of the curvature of the likelihood function around the maximum likelihood estimator. In the case of multiple maxima being close to one another, the assessment of the variability of the estimator may appear heavily distorted.

In order to perform the statistical inference on the restrictions imposed on the structural models we use marginal data densities. In Bayesian inference, by specifying prior distributions we discover the shape to the likelihood function globally in the form of the posterior distribution of the parameters given the data. This is in contrast to the likelihood inference in which we discover the shape of the likelihood function only locally, around the maximum likelihood estimator (compare the argument to that of Sims & Uhlig, 1991; Sims, 2010). By using the information from the globally known likelihood function, we conduct inference that is equivalent to inference made on a globally identified model (compare with the argument of Frühwirth-Schnatter, 2004).

Figure 1: Local identification of the structural system

(a) sign change (b) equation permutation

(c) state labels permutation (d) full permutation augmentation



Note: The graphs illustrate the shape of the joint posterior distribution of two parameters from the same column of matrix A_0 exposed to different sources of the lack of global identification for the case of $N = 2$ and $M = 2$. Graph (a) represents the distribution when only signs of rows of matrix A_0 are changed. Graphs (b) and (c) illustrate the distributions with either equations of heteroskedastic states' labels are permuted respectively. Graph (d) presents the joint posterior distribution of these two parameters after the full permutation augmentation. The graphs were prepared using the estimation output of the unrestricted model with two Markov-switching heteroskedastic states.

Numerical tools for inference. It is, therefore, essential for Bayesian inference on SVAR-MSH models to use numerical methods that that discover the posterior distribution globally. In particular, a sampling algorithm has to be capable of simulating draws from all the $M!N!2^N$ modes of the posterior distribution.

Two existing extensions of the Gibbs sampling algorithm (derived in Section 5) can provide an appropriate output for the permutation problem. A permutation augmented sampling proposed by Geweke (2007) perfectly reproduces the symmetry of the posterior modes, and therefore constitutes a conceptually optimal solution. In this algorithm, each draw from the posterior distribution is used to construct $J = M!N!2^N$ draws from each of the possible permutations. Thus, the number of draws from such a procedure is SJ, where S is a number of draws from a Gibbs sampler. This solution might appear highly impractical due to the potentially extremely large number of draws. Alternatively, a random permutation algorithm proposed by Frühwirth-Schnatter (2001), that rearranges each draw from the posterior distribution to a draw from a randomly chosen permutation, can be used.

Both of the solutions require a feasible transformation of a draw from the posterior distribution, that is generated from one of the possible permutations, to a draw from any other permutation. Performing the state labels permutation requires computing the parameters of standardisation (ii) from the parameters of standardisation (i) that uses vector $\tau_1 = \left(A_{0.11}^{-2}, \dots, A_{0.NN}^{-2}\right)'$ and the formulae given earlier in this section. Then the label permutation is performed for vectors τ_m, for $i = 1, \dots, M$, and the parameters of standardisation (i) are retrieved from the permuted τ_ms and matrix B_0.

4. Other Patterns of Heteroskedasticity

Although we have focussed the discussion in Section 3 on the SVAR-MSH model, the setup of state-dependent volatility also covers exogenously specified changes in volatility as in Lanne & Lütkepohl (2008) and residuals with mixtures of normal distributions as used by Lanne & Lütkepohl (2010). We extend the framework of restricting the transition probabilities matrix by Sims et al. (2008) to our structural-from setup to cover also these volatility models. We show that these different models for heteroskedasticity can be presented in terms of restricting the transition probabilities matrix, $\mathbf{P}$.

Consider a single row of the transition probabilities matrix, denoted by $\mathbf{P}_m$ for $m = 1, \dots, M$, which can be represented in terms of its unrestricted elements:

$$\mathbf{P}_m = \mathbf{p}_m W_m, \tag{8}$$

for $m = 1, \dots, M$, where a $1 \times z_m$ vector $\mathbf{p}_m$ contain the unrestricted transition probabilities from $\mathbf{P}_m$ such that $\mathbf{p}_m \imath_M = 1$ and W_m is a suitable $z_m \times M$ matrix containing zeros and ones. All the three mentioned patterns for modeling heteroskedasticity will now be presented in terms of the above framework of restricting the transition probabilities matrix. We also consider other such extensions.

Markov-switching heteroskedasticity. The basic characteristic of the SVAR-MSH model is that the error term is heteroskedastic with $M - 1$ diagonal covariance matrices of the structural shocks being estimated (λ_1 is set to $\imath_N$). If the hidden Markov process is assumed to be irreducible and aperiodic, then all the M^2 elements of matrix $\mathbf{P}$ are estimated (see Hamilton, 1989). This choice is represented by setting z_m to M and matrices W_m to identity matrices of order M, for $m = 1, \dots, M$. Moreover, irrespectively of the state in which the process is in some period, it is allowed to switch to any of the remaining states or to remain in the same state in a period that follows. Although, matrix $\mathbf{P}$ decides on this property, the posterior inference about the hidden Markov process focuses on the marginal posterior probabilities of states, $\Pr[s_t|\mathbf{y}]$ for $t - 1, \dots, T$. These probabilities can be computed directly from the output of the Bayesian estimation.

Exogenous regime changes in volatility. Heteroskedasticity modeled with regime changes in volatility assumes that at a certain point in time the covariance matrix of the residual term changes. Moreover, once a certain state in the level of volatility is finished it does not appear again. Such an approach is sometimes proposed in empirical studies that document one (see e.g. Clarida, Galí & Gertler, 2000; Lubik & Schorfheide, 2004) or multiple (see e.g. Sims & Zha, 2006) changes in the volatility of the U.S. macroeconomic data. Lanne & Lütkepohl (2008) used this pattern of heteroskedasticity in order to identify matrix A_0. In fact, any pattern for exogenous changes in volatility could be represented by delivering any sequence of state realizations $\{s_t\}_{t=1}^{T}$.

Our Bayesian simulation algorithm is directly applicable for this type of specification. In fact, it is significantly simplified, because there is no need to estimate the hidden Markov process or the transition probabilities matrix, $\mathbf{P}$, that govern the properties of the Markov process. Given s_t, for $t = 1, \dots, T$, all other parts of the algorithm presented in Section 5 stay unchanged.

Endogenous regime changes in volatility. The previous two patterns for modeling heteroskedasticity impose very strong assumption that the state realisations or state probabilities are determined exogenously, and thus, known. This assumption, of course, lowers the fit of the model to the data whenever the dates of the changes are not chosen optimally. In order to introduce endogenous regime changes in the volatility we adapt the framework of Chib (1998) (see Sims et al., 2008, for the implementation for MS-VAR models) to the SVAR model and use them to identify all the elements of matrix A_0.

Consider a model with just two states of the volatility ($M = 2$) and one regime change. The first state of the volatility corresponds to the beginning of the sample. This state is persistent, but at some point it switches to the second state, which identifies the regime change. Then, the second state of the volatility prevails until the end of the sample data, and the state does not switch back to the first state at any point in time. Such a pattern in volatility can be represented by the transition probability matrix as:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ 0 & 1 \end{bmatrix},$$

and consequently within the restricted Markov-switching framework by setting $W_1 = I_2$ and $W_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$. The transition probabilities matrix corresponding to the case of three volatility states and two regime changes together with matrices W_m are:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & 0 \\ 0 & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}, \quad W_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad W_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}.$$

Despite the fact that the last row of matrix $\mathbf{P}$ contains the degenerate case in which the last state is sampled with probability 1, the whole framework remains consistent with the definition of the Dirichlet distribution, and thus, can be easily implemented in our algorithms. Moreover, the endogenous regime changes introduce nonstationarity of the Markov process. Therefore, the ergodic probabilities of states represent the degenerate case with the unconditional probability of the last state is equal to one. The filtering and smoothing algorithm is initiated by setting $s_0 = 1$ rather than at the ergodic probabilities. This choice breaks the dependence of $\Pr[s_0]$ on $\mathbf{P}$ and in consequence the Gibbs draws for $\mathbf{p}_m$, for $m = 1, \dots, M$ do not require a Metropolis-Hastings step.

The proposed way of modeling the structural change in the volatility differs from the one used by Lanne & Lütkepohl (2008) and in fact has not yet been used to identify full matrix A_0. The difference is that the structural changes are endogenously determined within the model and, in fact, are estimated. Again the analysis of the posterior probabilities of states, $\Pr[s_t|\mathbf{y}]$, is informative about the dates of the structural changes. The current setting is therefore way much more flexible and less arbitrary than the model of Lanne & Lütkepohl (2008).

Mixture models. Mixture of normal distributions models are nested within the Markov-switching by setting all the rows of the transition probability matrix $\mathbf{P}$ to the same vector. This restriction can be represented by $\mathbf{P} = \imath_M \mathbf{p}$, where $\imath_M$ is a $M \times 1$ vector of ones and $\mathbf{p}$ is a $1 \times M$ random vector following a M-dimensional Dirichlet distribution ($\mathbf{p}\imath_M = 1$). This restriction can be simply presented in the form of restriction of equation (8), by setting $\mathbf{p}_m = \mathbf{p}$ and $W_m = I_M$ for $m = 1, \dots, M$.

Lanne & Lütkepohl (2010) used non-normal residuals modeled with a mixture of normal distributions to identify matrix A_0. Each component of of the mixture is a normal distribution centered at a vector of zeros.

The components differ, however, by the values of the variances. This case, however interesting, left alone might seem not that appealing for the economic applications. In this model the forecasted probabilities of the states are time-invariant and equal to $\mathbf{p}$.

Notice that, in the time series analysis, despite the fact that the probabilities $\mathbf{p}$ define a latent process s_t in which the probabilities of belonging to a particular state are independent of t, the model often results in highly serially correlated posterior probabilities of states $\Pr[s_t|\mathbf{y}]$. Therefore, a mixture model may be considered a way of modeling heteroskedasticity with a simple (and misspecified in case of autocorrelated state probabilities) parametrization.

Other patterns for volatility. The idea of modeling different patterns for volatility by specifying other models within the Markov-switching model by restricting or expanding the transition probabilities matrix has its origins in the framework proposed by Sims et al. (2008). Sims et al. do not consider a mixture model in their specification, however, they relate a Markov-switching model to structural breaks (equivalent to our regime change), incremental and discontinuous shifts and time-dependent transition probabilities. The two additional models could, of course, also be used in order to identify matrix A_0 through heteroskedasticity. We, however, leave this possibility for further research.

5. Bayesian Inference on SVAR-MSH models

5.1. Motivation

Estimating structural form of the model. There are two ways of estimating the SVAR models. One method consists of the estimation of the model in the reduced form (hereafter RF) and then retrieving from these estimates the values of the structural form (hereafter SF) parameters. The other is to estimate directly the parameters of the model in the SF. The former is often used as it usually results in simple algorithms for the estimation (see e.g. Rubio-Ramírez, Waggoner & Zha, 2010), although it is highly problematic for the inference on over identified models (Sims & Zha, 1999). The latter requires nonstandard estimation methods (see Waggoner & Zha, 2003; Villani, 2009). We argue that in order to make inference on the SVAR-MSH model it is essential to estimate the model of the SF. This point is very important given that our main objective is to discriminate between different schemes of identifying the monetary policy shock within the SVAR framework.

From the econometric perspective the problem of comparing the monetary policy models presented in Section 7 is to select the restrictions on matrix A_0 the most supported by the data and to test them against the unrestricted model with a full matrix A_0 being estimated. Only models estimated in the SF are not observationally equivalent. Such models lead to different implied RFs, and therefore, they can be tested against one another.

Moreover, the impulse response analysis based on the over identified SVAR models retrieved from the RF VAR lead to an improper inference about the shape and uncertainty of the impulse responses (see Sims & Zha, 1999). The main reason for this phenomenon is that the conversion of the RF to an over-identified SF is not a one to one transformation, and consequently, the parameter space is changed in the two specifications. Therefore, Sims & Zha (1999) argue for the estimation of the SF specification.

In order to gain more insights in the choice of the SF, it is illustrative to compare this approach to the one by Kulikov & Netšunajev (2013) who proposed a Bayesian framework for the estimation of the unrestricted matrix A_0 from the RF MS-VAR. The main limitation of Kulikov & Netšunajev's approach is that it can make use only of the just-identified system, and thus, only two states for the covariance matrix are allowed for. In order to compute matrices A_0 and λ_2 (notation from the current paper is applied here) from the RF covariance matrices Σ_1 and Σ_2 they implement a one-to-one transformation by Horn & Johnson (2013). There, however, does not exist in the literature an algebraic transformation that would allow to compute matrix A_0 and variances of the structural shocks from three or more RF covariance matrices (There, however, exist conditions for unique identification of A_0 in that case derived by Lanne et al., 2010). The SVAR-MSH model presented in Section 3 may include any number of states M, potentially fitting the data better.

Also the arguments from the beginning of this section constitute limitations for Kulikov & Netšunajev's approach. Kulikov & Netšunajev are right when they consequently compute impulse responses from the just-identified SF model (with unrestricted matrix A_0) due to the critique of Sims & Zha (1999). However, their approach does not allow for performing model comparison or hypotheses testing. In conclusion, for the purpose of hypotheses testing or the analysis of the over identified models (including any restrictions on matrix A_0) the choice of the modeling framework of the current paper is superior.

Applying Bayesian Inference. Bayesian methods have gained popularity in the empirical macroeconomic research because they also work for very large models and for highly nonlinear models that are difficult to handle in a frequentist framework. The application of a class of shrinkage prior distributions makes the Bayesian methods extremely useful for the analysis of large systems of economic variables. This feature is widely used in the VAR models literature (see e.g. Doan, Litterman & Sims, 1984; Bańbura, Giannone & Reichlin, 2010; Giannone, Lenza & Primiceri, 2012), but also applies to a general class of Markov-switching VAR models (see e.g. Frühwirth-Schnatter, 2006; Sims et al., 2008).

These arguments remain valid also for the SVAR-MSH model analyzed in this study. One of the most severe difficulty in the estimation of such models is the maximization of the likelihood function that becomes often impossible with increasing number of the states or dimensionality of time series. In consequence, so far in the empirical studies systems of up to four variables are only analyzed (see e.g. Lanne & Lütkepohl, 2010; Lanne et al., 2010; Lütkepohl & Netšunajev, 2014; Netšunajev, 2013). Our proposition overcomes these difficulties.

Introduced in the current paper Bayesian estimation of the SVAR-MSH models is feasible, also for large models with multiple states, because we implement an efficient Gibbs sampler for the estimation of the matrix of contemporaneous effects A_0. In another study on the Bayesian estimation of the SVAR-MSH models (see Markun, 2011, Chapter 2) a Metropolis-Hastings algorithm is adapted. That solution is feasible for small systems of variables (Markun analyses system of three variables, $N = 3$), however, in our application including seven variables it is way too expensive in terms of the time and computational resources required for the estimation. The same costs apply to another sampling technique that we tried, namely the adaptive griddy Gibbs sampler of Ritter & Tanner (1992), which is highly impractical for the estimation of 49 elements of (7×7) matrix A_0. We, therefore, adapt the Gibbs sampler of Villani & Warne (2003) and Villani (2009) to the estimation of the full as well as the restricted matrix A_0. In the two other works this algorithm is applied to the homoskedastic SVAR models with just-identified, meaning restricted, matrix A_0, but the same algorithm applies with minor changes to the unrestricted SVAR-MSH model.

Finally, Bayesian inference is most suitable for discriminating between different identification patterns applying to the same system of variables which is the setting presented in Section 2. The three monetary policy models are defined by zero restrictions on matrix A_0. Note that none of these specifications is nested in any other. Therefore, the strategy implemented in empirical studies using frequentist testing methods is to test each of these models against the unrestricted specification (see Lanne & Lütkepohl, 2008, 2013). Such an analysis, however informative, does not allow to discriminate between the monetary models, because they are not directly compared. Dufour (1989) shows how testing of non-nested hypotheses can be performed in the frequentist framework, however, his solution depends on simulation techniques which in the current setting would be extremely costly, especially taking into account the problems with the classical estimation of the SVAR-MSH models. The analysis of Bayes factors or posterior probabilities of models/hypotheses constitutes an easy to implement and valid alternative to the frequentist approach, and gives positive arguments *in favor of* models or hypotheses.

5.2. Notation and likelihood function

Let $\{y_t\}_{t=1}^{T}$ be a realization of a N-dimensional stochastic process, and let $N \times T$ matrix $\mathbf{y} = (y_1, \ldots, y_T)$ collect all the observations on the considered time series. Let $K = 1 + pN$ and define the $K \times 1$ vector $x_t = \left(1, y_{t-1}', y_{t-1}', \ldots, y_{t-p}'\right)'$ that collects all the vectors of observations on the RHS of equation (7) and define a $K \times T$ matrix $X = (x_1, \ldots, x_T)$, where the initial conditions $y_0, \ldots, y_{p-1}$ are treated as given and set to the first p observations of the available dataset. Similarly, collect the residual terms in matrix $U = (u_1, \ldots, u_T)$, and denote by U_n its n^{th} row. Let $\mathbf{y}_m$, X_m, U_m and $U_{n.m}$ denote matrices corresponding to matrices $\mathbf{y}$, X, U and U_n collecting only state specific columns for which $s_t = m$ for $m = 1, \ldots, M$. The column dimension of these matrices is then T_m and $\sum_{m=1}^{M} T_m = T$. Matrix $\mathbf{S} = (s_1, \dots, s_T)$ is the realization of the hidden Markov process for periods from 1 to T. Define a $N \times K$ matrix $\alpha = \left(\mu, A_1, \dots, A_p\right)$ collecting all the autoregressive parameters on the RHS of equation (**??**) and its n^{th} row - a $1 \times K$ vector α_n collecting the parameters on the RHS of equation (7). For convenience we also form matrix $\Lambda = (\lambda_1, \dots, \lambda_M)$, where $\lambda_1 = \imath_N$, and denote by θ all the parameters of the model. Then equation (**??**) can be written in the matrix notation as:

$$A_0\mathbf{y} = \alpha X + U, \tag{9}$$

and equation (7) as:

$$a_n V_n \mathbf{y} = \alpha_n X + U_n, \tag{10}$$

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

Given the assumption in equation (3) regarding the residual term of the SVAR-MSH model the likelihood function is given by (see also Markun, 2011, Section 2.3):

$$p\left(\mathbf{y}|\mathbf{S}, \theta\right) = (2\pi)^{-\frac{TN}{2}} \left|\det\left(\left[a_1 V_1, \dots, a_N V_N\right]\right)\right|^T \prod_{m=1}^{M}\prod_{n=1}^{N} \lambda_{m.n}^{\frac{-T_m}{2}} \times \\ \times \exp\left\{-\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[a_n V_n \mathbf{y}_m - \alpha_n X_m\right]\lambda_{m.n}^{-1}\left[a_n V_n \mathbf{y}_m - \alpha_n X_m\right]'\right\}. \tag{11}$$

The likelihood function written in this form emphasizes the feature of the SVAR models that equations of the model can be analyzed one by one leading to convenient full conditional posterior distributions.

The conditional distribution of $\mathbf{S}$ given $\mathbf{P}$ is:

$$p\left(\mathbf{S}|\mathbf{P}\right) = \Pr\left(s_0|\mathbf{P}\right)\prod_{i=1}^{M}\prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})}, \tag{12}$$

where $p\left(s_0|\mathbf{P}\right)$ is a distribution of the initial condition s_0 and $N_{ij}(\mathbf{S}) = \#\{s_{t-1} = j, s_t = i\}$ is a number of transitions from state i to state j, for $i, j = 1, \dots, M$. The full data likelihood function given by:

$$p\left(\mathbf{y}, \mathbf{S}|\theta\right) = p\left(\mathbf{y}|\mathbf{S}, \theta\right) p\left(\mathbf{S}|\mathbf{P}\right), \tag{13}$$

together with the prior distribution for the parameters θ allows us to perform the posterior analysis.

5.3. *Prior analysis*

The prior distributions for the unrestricted elements of the rows of the transition probabilities matrix, $\mathbf{p}_m$, are set independently for each row and are given by r_m-dimensional Dirichlet distributions as in Sims et al. (2008). We denote the parameters of each of such distributions by e_k for $k = 1, \dots, z_m$.

The variances of the structural shocks, $\lambda_{m.n}$ for $m = 2, \dots, M$, are independently *a priori* distributed with an inverse gamma 2 distribution with parameters a and b (see Bauwens, Richard & Lubrano, 1999, Appendix A, for the detailed specification of the distribution and the random sampling algorithm).

The rows of matrix A_0 are normally distributed with the mean set to a vector of zeros and a diagonal covariance matrix $\underline{\underline{S}}_n$. Then, the implied prior distribution for the unrestricted elements of each such row is also normally distributed with the mean set to a vector of zeros and a diagonal covariance matrix $\underline{S}_n = V_n \underline{\underline{S}}_n V_n'$.

Each vector α_n' of autoregressive parameters is assumed to independently *a priori* follow a conditional K-variate normal distribution given a_n. The mean of this distribution is set to $\underline{P}' V_n' a_n'$, where $N \times K$ matrix $\underline{P}$ is equal to $\begin{bmatrix}\mathbf{0}_{N\times 1} & I_N & \mathbf{0}_{N\times N(p-1)}\end{bmatrix}$ in order to express the prior belief that the variables are unit-root nonstationary following Doan et al. (1984). The prior covariance matrix, $\underline{H}_n$, is diagonal with with the diagonal elements set according to those of Doan et al. (1984) and Litterman (1986). Many other prior specifications, including hierarchical structures, are feasible here.

To summarize, the prior specification takes the detailed form of:

$$p(\theta) = \prod_{n=1}^{N} p\left(\alpha_n | a_n\right) p\left(a_n\right) \prod_{m=1}^{M} p\left(\lambda_{m.n}\right) p\left(\mathbf{p}_m\right), \tag{14}$$

where each of the prior distributions is as assumed:

$$\begin{aligned} \alpha_n' | a_n &\sim \mathcal{N}_K\left(\underline{P}' V_n' a_n', \underline{H}_n\right) \\ a_n' &\sim \mathcal{N}_{r_n}\left(\mathbf{0}_{r_n}, \underline{S}_n\right) \\ \lambda_{m.n} &\sim \mathcal{IG}2\left(\underline{a}, \underline{b}\right) \\ \mathbf{p}_m &\sim \mathcal{D}_{r_m}\left(\underline{e}_1, \ldots, \underline{e}_{z_m}\right), \end{aligned}$$

for $n = 1, \ldots, N$ and $m = 1, \ldots, M$.

The above choice of the prior distributions is practical and the priority is given to distributions that result in convenient and proper full conditional distributions, and therefore, allow for the derivation of an efficient Gibbs sampler. Note that the prior distributions of a_n and $\lambda_{m.n}$ are invariant to either label switching, equation permutation or sign reversion of rows of matrix A_0. These assumptions enable us to perform the Bayesian hypotheses assessment using marginal data densities for the class of SVAR-MSH models, as described in Section 6, without further complications.

5.4. Gibbs Sampler

Simulating hidden Markov process. In order to estimate the states of the hidden Markov process we apply the algorithms presented in Frühwirth-Schnatter (2006, Section 11.2).

Sampling transition probabilities matrix. For the models for which $\Pr[s_0|\mathbf{P}] = \Pr[s_0]$ in equation (12) – see more details in Section 4 – the unrestricted elements of the rows of matrix $\mathbf{P}$, $\mathbf{p}_m$, are sampled independently from a z_m-dimensional Dirichlet distribution given $\mathbf{S}$:

$$\mathbf{p}_m | \mathbf{S} \sim \mathcal{D}_{r_m}\left(\underline{e}_1 + N_{m1}(\mathbf{S}), \ldots, \underline{e}_{z_m} + N_{mz_m}(\mathbf{S})\right),$$

for $m = 1, \ldots, M$. The parameters of the prior Dirichlet distributions are updated by the count of the transitions from one state to another given $\mathbf{S}$.

For the estimation of the stationary Markov chain, i.e. for the Markov-switching and mixture models, a Metropolis-Hastings step is required due to the dependency introduced by the prior distribution $\Pr(s_0|\mathbf{P})$ being set to a vector of ergodic probabilities. For more details the reader is referred to (see Frühwirth-Schnatter, 2006, Section 11.5.5) or, for the case of the restricted matrix $\mathbf{P}$, Droumaguet & Woźniak (2012, Section 5.2).

Sampling structural shocks variances. Each $\lambda_{m.n}$ given $\mathbf{y}, \mathbf{S}, \alpha_n$ and a_n is independently drawn from the inverse gamma 2 distribution:

$$\lambda_{m.n} | \mathbf{y}, \mathbf{S}, \alpha_n, a_n \sim \mathcal{IG}2\left(\underline{a} + T_m, \underline{b} + \left(a_n V_n \mathbf{y}_m - \alpha_n X_m\right)\left(a_n V_n \mathbf{y}_m - \alpha_n X_m\right)'\right) \tag{15}$$

for $m = 2, \ldots, M$ and $N = 1, \ldots, N$. The definition and random number generating algorithm for the inverse gamma 2 distribution can be found in Bauwens et al. (1999, Appendix A).

Sampling matrix of contemporaneous effects. The full conditional posterior distribution of the unrestricted parameters of the matrix of contemporaneous effects is:

$$a_1,\dots,a_N|\mathbf{y},\mathbf{S},\Lambda,\alpha \propto \left|\det\left(\left[a_1V_1,\dots,a_NV_N\right]\right)\right|^T \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\left(a_n-\mu_{a_n}\right)\Omega_{a_n}^{-1}\left(a_n-\mu_{a_n}\right)'\right\}, \tag{16}$$

where:

$$\Omega_{a_n} = \left[V_n\left(\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m'/\lambda_{m.n} + \underline{P}\underline{H}_n^{-1}\underline{P}'\right)V_n' + \underline{S}_n^{-1}\right]^{-1},$$

$$\mu_{a_n} = \alpha_n\left(\sum_{m=1}^{M}X_m\mathbf{y}_m'/\lambda_{m.n} + \underline{H}_n^{-1}\underline{P}'\right)V_n'\Omega_{a_n}.$$

The full conditional distribution of a_n is of a similar form as analyzed in Villani & Warne (2003) and Villani (2009, Appendix C), but extended by the heteroskedasticity of the error term. Another difference is that Villani & Warne (2003) and Villani (2009) analyze the model with at least $N(N-1)/2$ restrictions imposed in order to identify matrix A_0, whereas we are interested also in estimating a model without any such restrictions. The same sampling algorithm, however, can be adopted to our model without further adjustments beyond those presented above. In order to independently sample the unrestricted elements, a_n, for $n = 1,\dots,N$, we first sample from the absolute normal distribution and then transform these draws to draws from the full conditional distribution from equation (16). For all the details of this sampling algorithm the reader is referred to Villani & Warne (2003, Section 4) and Villani (2009, Appendix C).

Sampling autoregressive parameters. The convenient form of the prior distribution and the likelihood function allow for sampling the autoregressive parameters independently equation by equation from a multivariate normal distribution:

$$\alpha_n'|\mathbf{y},\mathbf{S},\Lambda,a_n \sim \mathcal{N}_K\left(\overline{P}_n'V_n'a_n',\overline{H}_n\right), \tag{17}$$

for $n = 1,\dots,N$, where:

$$\overline{H}_n = \left[\left(\sum_{m=1}^{M}X_mX_m'/\lambda_{m.n}\right) + \underline{H}_n^{-1}\right]^{-1},$$

$$\overline{P}_n = \left[\left(\sum_{m=1}^{M}\mathbf{y}_mX_m'/\lambda_{m.n}\right) + \underline{P}\underline{H}_n^{-1}\right]\overline{H}_n.$$

6. Marginal Data Densities for Unidentified Models

Bayesian hypotheses assessment relies on the marginal data densities (MDD), denoted by $p(\mathbf{y}|\mathcal{M}_i)$ for some model $\mathcal{M}_i$, that are used in order to compute the posterior probabilities of models, $\Pr(\mathcal{M}_i|\mathbf{y})$, and hypotheses represented by them. The computation of the MDDs for the Markov-switching models is non-trivial due to the invariance of the likelihood function to relabelling of the states, which was described in Celeux et al. (2000). For our structural model this problem is generalized to the invariance of the likelihood function to the relabelling of the heteroskedastic states, the reordering of the rows of A_0 and elements of λ_ms, as well as the change of the sign of the rows of A_0 as described in Section 3.2.

For the purpose of the unbiased estimation of the MDDs for the SVAR-MSH models we adapt the estimator of Chib (1995) with the fix of Marin & Robert (2008) to our structural system. Moreover, we generalize the estimator of Marin & Robert (2008) to the case in which the parameters of the model are divided in convenient groups allowing for a feasible Gibbs sampler.

Our estimator is based on the so called *basic marginal likelihood identity* (see Chib, 1995) and we present it in a logarithm form adjusted to the SVAR-MSH models (and neglecting the conditioning on $\mathcal{M}_i$) as:

$$\ln \hat{p}(\mathbf{y}) = \ln \hat{p}(\mathbf{y}|\theta^*) + \ln \hat{p}(\theta^*) \\ - \ln \hat{p}(\alpha^*|\mathbf{y}) - \ln \hat{p}(\Lambda^*|\mathbf{y}, \alpha^*) - \ln \hat{p}\left(A_0^*|\mathbf{y}, \alpha^*, \Lambda^*\right) - \ln \hat{p}\left(\mathbf{P}^*|\mathbf{y}, \alpha^*, \Lambda^*, A_0^*\right). \quad (18)$$

The first two elements on the RHS of equation (18) are the ordinates of the log-likelihood function and the logarithm of the prior distribution respectively, both evaluated at θ^*. The remaining elements are the logarithms of the ordinates of appropriate conditional posterior distributions evaluated at θ^* that jointly represent the logarithm of the ordinate of the posterior distribution, $\ln \hat{p}(\theta^*|\mathbf{y})$. θ^* is any value of the parameters, however, as recommended by Chib (1995), it should represent a region of high posterior probability mass.

The estimator defined in equation (18) is a simple application of the estimator by Chib (1995) to the SVAR-MSH model. We shall now explain how we adapt the fix by Marin & Robert (2008). The likelihood function and the prior distribution are invariant to the rearrangements of the structural system. Therefore, their ordinates evaluated at θ^* are correctly computed. However, the posterior distribution is not invariant to these rearrangements, which makes is biased if computed simply as in Chib (1995). The correction is implemented by averaging the ordinates of the conditional posterior distributions over the possible rearrangements of the structural system. The correction of Marin & Robert (2008) was applied only to state label permutation. We, however, apply this solution to all the sources of the invariance listed in Section 3.2.

Denote by $\theta^{*(j)}$, for $j = 1, \ldots, J$, a value of a vector of all the parameters of the model, θ^*, subject to one of the rearrangement of the structural system and coherently, by $\alpha^{*(j)}, \Lambda^{*(j)}, A_0^{*(j)}$ and $\mathbf{P}^{*(j)}$ the corresponding rearrangements of its elements. J is the number of rearrangements. For instance, the number of all of the possible permutations of the structural system for the unrestricted SVAR-MSH model is $J = M!N!2^N$. Denote by $\theta^{(j.s)}$, for $j = 1, \ldots, J$ and $s = 1, \ldots, S$, the sth draw from the posterior distribution with it's elements rearranged according to jth permutation (use similar superscripts for the elements of θ).

Then, the ordinate of the marginal posterior distribution distribution for α is computed by applying the Rao-Blackwell tool (see Gelfand & Smith, 1990):

$$\hat{p}(\alpha^*|\mathbf{y}) = \frac{1}{JS} \sum_{s=1}^{S} \sum_{j=1}^{J} \prod_{n=1}^{N} \hat{p}\left(\alpha_n^{*(j)}|\mathbf{y}, \mathbf{S}^{(j.s)}, \Lambda_n^{(j.s)}, a_n^{(j.s)}\right),$$

where on the RHS there is the ordinate of the full conditional posterior distribution for α_n defined in equation (17). The formula above uses a sample of S draws from the full Gibbs sampling algorithm, i.e. the output from the estimation of the model.

The ordinate of the conditional posterior distribution for Λ is computed by:

$$\hat{p}(\Lambda^*|\mathbf{y}, \alpha^*) = \frac{1}{JS_1} \sum_{s=1}^{S_1} \sum_{j=1}^{J} \prod_{n=1}^{N} \prod_{m=1}^{M} \hat{p}\left(\lambda_{m.n}^{*(j)}|\mathbf{y}, \alpha_n^{*(j)}, \mathbf{S}^{(j.s)}, a_n^{(j.s)}\right),$$

using the full conditional posterior distribution for $\lambda_{m.n}$ defined in equation (15). Note that in order to compute $\hat{p}(\Lambda^*|\mathbf{y})$ the sample from the full Gibbs sampler cannot be used as was the case for the computations of $\hat{p}(\alpha^*|\mathbf{y})$. Instead, a sample of S_1 draws from a reduced Gibbs sampler is used. It can be obtained by iterative sampling from the following distributions:

$$p(\mathbf{S}|\mathbf{y}, \alpha^*, \mathbf{P}, A_0, \Lambda), \qquad p(\mathbf{P}|\mathbf{S}), \qquad p(\Lambda|\mathbf{y}, \mathbf{S}, \alpha^*, A_0), \qquad p(A_0|\mathbf{y}, \mathbf{S}, \alpha^*, \Lambda).$$

Similarly, we compute the ordinate of the conditional posterior distribution of A_0:

$$\hat{p}\left(A_0^*|\mathbf{y},\alpha^*,\Lambda^*\right) = \frac{1}{JS_2}\sum_{s=1}^{S_2}\sum_{j=1}^{J}\hat{p}\left(A_0^{*(j)}|\mathbf{y},\alpha^{*(j)},\Lambda^{*(j)},\mathbf{S}^{(j.s)}\right),$$

where this step requires an iterative sampling from: $p\left(\mathbf{S}|\mathbf{y},\alpha^*,\Lambda^*,\mathbf{P},A_0\right)$, $p\left(\mathbf{P}|\mathbf{S}\right)$ and $p\left(A_0|\mathbf{y},\mathbf{S},\alpha^*,\Lambda^*\right)$ in order to obtain a sample of S_2 draws from the reduced Gibbs sampler. The difficulty in evaluating $\hat{p}(A_0^*|\mathbf{y},\alpha^*,\Lambda^*)$ is that the distribution of the RHS is known only up to a normalizing constant and is given in equation (16). Therefore, for each unique $\mathbf{S}^{(\cdot,s)}$ we compute this constant by applying the generalized harmonic mean estimator of Gelfand & Dey (1994) as follows:

$$c_{A_0|\mathbf{S}^{(\cdot,s)}} = M!N!S_3\left[\sum_{s_3=1}^{S_3}\frac{\hat{q}\left(A_0^{(s_3)}\right)}{\hat{p}\left(A_0^{*(j.s_3)}|\mathbf{y},\alpha^{*(j)},\Lambda^{*(j)},\mathbf{S}^{(j.s)}\right)}\right]^{-1}. \tag{19}$$

The S_3 draws used in formula (19) come from a reduced Gibbs algorithm iteratively sampling from $p(A_0^{*(j)}|\mathbf{y},\alpha^{*(j)},\Lambda^{*(j)},\mathbf{S}^{(j.s)})$ as described in Section 5.4. $\hat{q}$ is a multivariate normal distribution with the mean set to the sample mean of the S_3 draws and the covariance matrix set to a 0.9 times the sample covariance matrix of these draws. The RHS of equation (19) includes the multiplication by $M!N!$, following the recommendation by Frühwirth-Schnatter (2004), because we do not take into account permutations of the possible arrangements of A_0 in our computations of $c_{A_0|\mathbf{S}}$. The conditioning in the target distribution in practice allows us to draw from a well defined area of the parameter space of A_0.

Finally, we approximate the last element on the RHS of equation (18) by computing:

$$\hat{p}\left(\mathbf{P}^*|\mathbf{y},\alpha^*,\Lambda^*,A_0^*\right) = \frac{1}{JS_4}\sum_{s=1}^{S_4}\sum_{j=1}^{J}\prod_{m=1}^{M}\hat{p}\left(\mathbf{p}_m^{*(j)}|\mathbf{S}^{(j.s)}\right),$$

where the probability density function on the RHS is a pdf of an appropriate Dirichlet distribution, and a sample of S_4 draws comes from an iterative reduced Gibbs algorithm on $p(\mathbf{S}|\mathbf{y},\mathbf{P},\alpha^*,\Lambda^*,A_0^*)$ and $p(\mathbf{P}|\mathbf{S})$.

We conclude this section with a couple of remarks regarding the computations of the MDDs for different models estimated in this study:

- For the unrestricted SVAR-MSH model the number of all of the possible rearrangements of the structural system is $J = M!N!$. For systems with high M or N it might not be a practical choice to use all of the permutations for the computations. Instead, following Marin & Robert (2008), we recommend using a random sample of permutations, which decreases J and the time required for the estimations.
- Models with restricted matrix A_0, when unique restrictions are used for each equation (as in the three monetary policy models that we use), have uniquely defined rows of A_0. Consequently, for such models the estimation of the MDDs does not require rearranging of the structural system with respect to the ordering of the rows of A_0 and elements of λ_ms. Only the state labels' permutation needs to be taken into account.
- In models with endogenously determined regime changes the state labels uniquely determine the states. For these models the estimation of the MDDs does not require permuting the state labels.
- The estimation of the MDDs is significantly simplified for models with exogenously determined regime changes due to the fact that the realisations of the states, $\mathbf{S}$, are known. In these models, the matrix of transition probabilities, $\mathbf{P}$, is neither defined or estimated. Therefore, the estimator defined in equation (18) does not have the last element on the RHS. Instead all of the elements include conditioning on $\mathbf{S}$. Consequently, all of the reduced Gibbs sampling algorithms should be adjusted for these features and simplified.

Table 2: Data information

Variable	Description	Seasonally Adjusted	Source	Index
gdp_t	log of real gross domestic product	yes	BEA	GDPC1
p_t	log of gross domestic product: implicit price deflator	yes	BEA	GDPDEF
$pcom_t$	log of nominal non-energy commodities price index: monthly data, value at the end of quarter	no	WB	INONFUEL
FF_t	effective federal funds rate	no	FED	FEDFUNDS
nbr_t	log of non-borrowed reserves of depository institutions	yes	FED	BOGNONBR
tr_t	log of board of governors total reserves, adjusted for changes in reserve requirements	yes	FED	TRARR
m_t	log of M1 money stock	yes	FED	M1SL

Note: The data is quarterly, starting in quarter 1, 1960, and ending in quarter 4, 2007. Legend:
BEA - U.S. Department of Commerce: Bureau of Economic Analysis
FED - Board of Governors of the Federal Reserve System
WB - World Bank

7. An Empirical Example

7.1. Data and Setting of the Empirical Analysis

In this section we perform an empirical assessment of the three structural identification schemes introduced in Section 2.

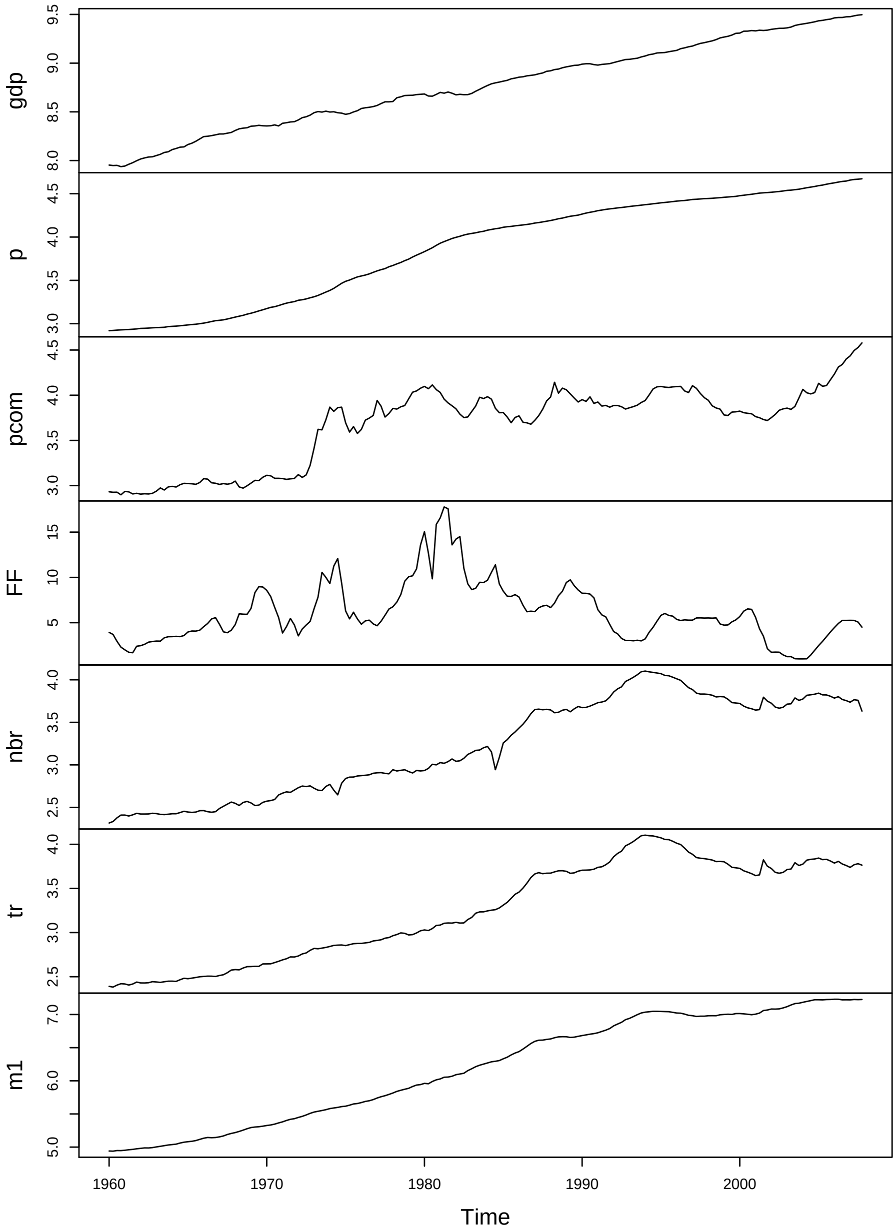
Data. We use the same seven, $N = 7$, macroeconomic variables as defined in Section 2. The data are quarterly beginning in 1960 and ending in the last quarter of 2007. Their transformation and sources are given in Table 2. The beginning of the sample period is determined by the availability of the monetary aggregates, whereas its end is determined by the financial crisis beginning in 2008. In that year some of the variables, especially non-borrowed reserves and total reserves, changed their levels dramatically. This may have led to structural changes in monetary policy that cannot be captured by changing volatility alone. Notice that in our setup only the residual variances are allowed to change and not the other VAR coefficients. Hence, these models may not capture the changes occurring in 2008. To avoid such complications in our empirical modeling we terminate the sample at the end of 2007. Figure 2 presents the sample data.

Estimated models. The objective of the investigation is to determine which of the restrictions implied by the three structural models are supported by the data and to determine which model should be used for the analysis of the impact of the monetary policy on the real economy. We, therefore, estimate models with three different schemes for the identification of the structural shocks as well as models with unrestricted A_0. The restrictions imposed in the four types of structural models in our competition are presented in Table 1.

Moreover, we compare the model fit as well as the robustness of the assessment of the structural identification schemes by estimating models with various patterns for heteroskedasticity. Structural VARs with Markov-switching heteroskedasticity, mixture of normal distributions for the error term, as well endogenous regime changes in volatility are estimated with two and three states. We estimate models with exogenously determined volatility regimes with one, two and three regime breaks. Models with two exogenously determined volatility regimes have the structural break set to quarter 4 of 1979. Models with three such states have an additional break in quarter 1 of 1984, and consequently, models with four regimes have an additional heteroskedastic regime change in quarter 4, 1988.

That gives a total of 36 models estimated in our empirical investigation. All of them have four autoregressive lags, $p = 4$, and consistent prior distributions.

Figure 2: Data plot



Prior distributions and posterior computations details. All the models assume normal prior distributions for the unrestricted elements of each row of the matrix A_0 with prior mean set to a zero vector and the covariance matrix, $\underline{S}_n$, set to a diagonal matrix of appropriate dimension with tens on the diagonal, that is, the priors for the rows of A_0 are:

$$a_n \sim \mathcal{N}(\mathbf{0}, 10 I_{r_n}), \qquad n = 1, \ldots, N.$$

The parameters of the inverse gamma prior distributions for the variances of the structural shocks are: $\underline{a} = 1$ and $\underline{b} = 1$ which makes the distributions quite diffuse.

For the parameters of the vector autoregression, μ and $A_1, \ldots, A_p$, we assume the Litterman prior introduced by Doan et al. (1984) and Litterman (1986). The prior distribution is set for the row vectors α_n, and therefore, *a priori* these parameters follow a conditional $(1 + pN)$-variate Normal distribution given a_n, with mean equal to a vector $a_n V_n \underline{P}$ and a diagonal covariance matrix, $\underline{H}_n$. The elements on the diagonal of the covariance matrix are determined by a set of of hyper-parameters, (ϕ_1, ϕ_2, ϕ_3) and are as follows:

$$\left(\varsigma_i \phi_3\right)^2 \text{ for } \mu_i, \qquad \left(\frac{\phi_1}{k}\right)^2 \text{ for } A_{k.ii}, \qquad \left(\frac{\phi_1 \phi_2}{k} \frac{\varsigma_i}{\varsigma_j}\right)^2 \text{ for } A_{k.ij}, \tag{20}$$

for $i, j = 1, \ldots, N$ and $i \neq j$, and $k = 1, \ldots, p$. We scale the variances of the prior distribution using the OLS estimators of variances of the residuals of autoregressions of order 17 for each of the variables, ς_i, for $i = 1, \ldots, N$. We set the value of the hyper-parameter responsible for shrinking of the constant terms, ϕ_3, to 10. The overall shrinking hyper-parameter for the autoregressive parameters, ϕ_1, is set to 0.3 as in Adolfson, Lindé & Villani (2007). The values of the variances of the prior distributions decrease with the indicator for lag, k, according to the harmonic pattern for quarterly data (see Doan et al., 1984). Finally, we set the value of ϕ_2 to 0.1.

For the rows of the transition probabilities matrix of the Markov-switching models we assume *a priori* a M-variate Dirichlet distributions with parameters set to 1 for all the transition probabilities except the diagonal elements $\mathbf{P}_{ii}$, for $i = 1, \ldots, M$, for which it is set to 10. Thereby, we express our expectation of a high probability of remaining in a given state, that is, we assume some persistence of the heteroskedastic states. The expected duration of the states implied by such prior assumptions depends on the number of states. For instance, for the models with two states, $M = 2$, the prior distribution implies an expected state duration of around eleven periods, whereas for the model with three states, $M = 3$, the expected duration of the states is around six periods. We assume similar parameters for the Dirichlet prior distribution of the models with endogenous regime changes: 10 for the probabilities of remaining in a particular regime and 1 for the probabilities of switching to the next one. For the mixture models all the elements of vector e_k are set to ones.

In order to perform the posterior simulations we first sampled a total of 7.2 million draws from the posterior distribution for each model and the last million was left for the subsequent analysis. In the reduced Gibbs sampling for the computations of the elements of the MDDs, we sampled 5,000 draws from the full conditional distributions. We found the latter number sufficient for the numerical integration purposes because of the conditioning on fixed values of the parameters, i.e. the elements of θ^*. In order to compute the constant that normalizes the full conditional posterior distribution of the unrestricted elements of matrix A_0, which is defined in equation (19), we used samples of 500 draws from this distribution.

For the models with restricted matrix of contemporaneous effects, A_0, we used a full set of possible permutations because we could perform these computations in a reasonable time. However, for the models with unconstrained matrix A_0 we used a random sample of the possible rearrangements of the structural system to keep JS not too large and the computational time required for the computations of the MDDs manageable. In order to check the numerical stability of the estimators computed using a random sample of permutations we used three sample sizes. They consisted of one, two or four percent of a total number of J possible rearrangements. Despite small variations in the value of the MDD estimators, all the conclusions in our empirical investigation remain the same, irrespectively of the fraction of rearrangements used. Therefore, the reliability of our model assessment is assured. The details of these checks are available on request.

7.2. Assessment of Structural Models

In Table 3 we report the natural logarithms of the MDDs for all of the estimated models. Consider first the model selection with respect to the preferred heteroskedasticity pattern and the number of volatility regimes. Markov-switching heteroskedastic models have the values of the logarithms of the MDDs higher

Table 3: Empirical assessment of structural models via logarithms of marginal data densities

M	Unrestricted	FF policy shock	NBR policy shock	NBR-TR policy shock
		Markov-switching heteroskedasticity		
2	2551.3	2627.2	2632.1	**2636.4**
3	2621.3	2670.1	2672.9	**2703.1**
		Mixture of normal distributions		
2	2545.3	2617.8	2625.1	**2630.1**
3	2591.0	2653.3	2652.8	**2678.2**
		Endogenous regime changes in volatility		
2	1620.3	**2435.3**	2407.1	2411.1
3	**2389.0**	2345.0	2107.4	2306.6
		Exogenous regime changes in volatility		
2	2095.9	2177.2	2176.4	**2177.3**
3	2166.2	2249.0	2252.7	**2253.4**
4	2166.6	2233.7	2235.3	**2260.6**

Note: The reported natural logarithms of marginal data densities are computed with the estimator described in Section 6. M denotes the number of states of the Markov process. Values marked bold denote the maximum value in each of the rows.

than these values for other corresponding models (those with the same number of states and imposed restrictions on A_0 matrix). Therefore, they open the ranking of the volatility processes and are followed respectively by mixture models, endogenous and exogenous regime changes. The MS models taken jointly gain the posterior probability mass that is by nearly 11 orders of magnitude[2] greater than the joint posterior probability of mixture models by Lanne & Lütkepohl (2010), and by 116 orders of magnitude greater than that of endogenous regime change models. With that respect, the models with exogenous regime changes of Lanne & Lütkepohl (2008) gain negligible posterior probabilities. The SVAR-MSH model with unrestricted transition probabilities matrix **P** is, therefore, preferred over models with restrictions.

The larger the number of heteroskedastic states, the higher the value of the MDDs for the considered models. This regularity holds for all of the models, also within models with the same volatility patterns[3]. Models with 3 states jointly have the posterior probability by 28 orders of magnitude larger than models with 2 states. If only models with exogenous regime changes were considered, then the models with 4 states are supported by data although this result does not necessarily hold for each particular model[4]. These findings clearly reject the approach proposed by Kulikov & Netšunajev (2013) which requires exactly two heteroskedastic states. To conclude, we find a strong evidence that Markov-switching models with three states outperform other models.

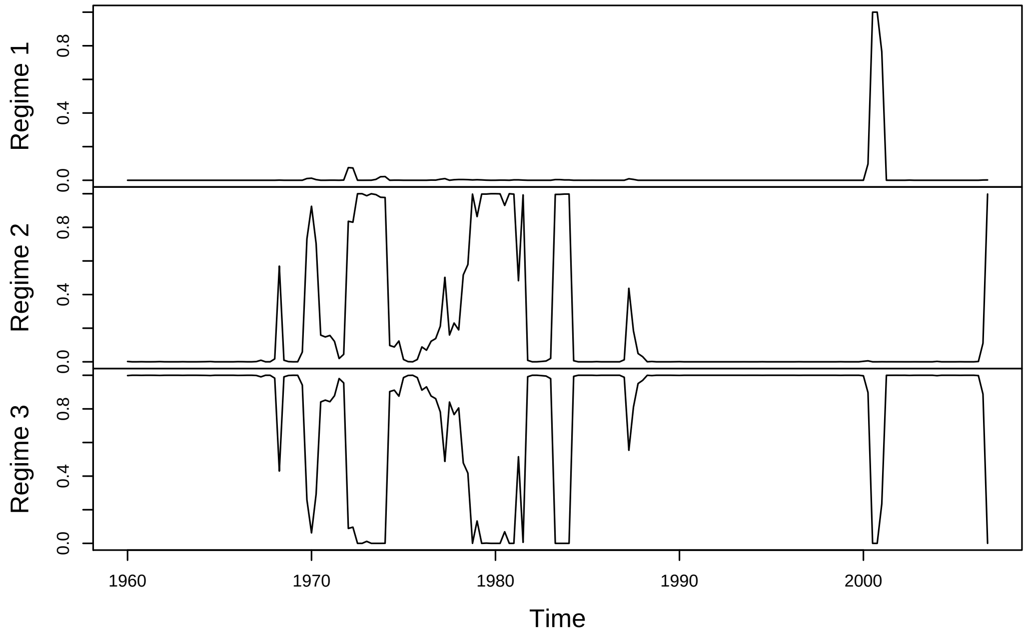
The assessment of the considered monetary policy models gives a clear indication in favor of the

[2]By stating that some class of models gains the posterior probability mass that is of some orders of magnitude greater than that of another class of models, we refer to the value of the base 10 logarithm of the posterior odds ratio comparing these probabilities. Posterior probabilities of models are computed using the Bayes formula assuming equal prior probabilities for all models equal to 36^{-1}. The joint posterior probability mass for a class of models is computed by summing the posterior probabilities of models belonging to this particular class.

[3]This regularity does not hold for the endogenous regime changes models with three states for which we did not manage to obtain reliable estimates of the MDDs.

[4]We failed to estimate models with more states due to a very low occurrence of the additional state. The redundancy of a state contradicts the assumption of Markov process stationarity for MS and mixture models. It is also highly cumbersome for the numerical stability of our estimation algorithms. Models with endogenous regime change appeared the most problematic with that respect. We find the structure of the model that excludes the possibility that the Markov process returns to its previous state not suitable for our data.

Figure 3: Marginal posterior probabilities of states, $\Pr[s_t|\mathbf{y}]$, for the NBR policy shock model with Markov-switching heteroskedasticty with 3 states



non-conventional monetary policy models. The NBR-TR policy shock model by Strongin (1995) finds the strongest support in the data and attracts over 99 percent of the posterior probability mass. This result is robust and takes into account the uncertainty with respect to different specifications of the heteroskedasticity including its formulation and the number of states. The second best model is the NBR policy shock model by Christiano & Eichenbaum (1992) with the base 10 logarithm of the Bayes factor equal to 13 in favor of the NBR-TR model. Other models, including the FF policy shock model attributed to Bernanke & Blinder (1992) and Sims (1986, 1992) and the unrestricted model, gain negligible posterior probabilities. This ordering holds for all of the models irrespectively of the heteroskedasticity pattern or the number of states.

Note that each of the three monetary policy models is preferred by the data over the unrestricted model, which is a new finding contradicting previous studies (compare to Lanne & Lütkepohl, 2008, 2013; Normandin & Phaneuf, 2004). The difference could potentially be attributed to different model specifications and a different sample frequency of the data used in these papers.

We found that models with endogenous regime changes are particularly not suitable for our data. In these models regime change is constrained to be monotonic, such that old regimes are constrained never to recur. This feature is clearly rejected in the estimation (see the posterior probabilities of states for the NBR-TR model plotted in Figure 3), a conclusion identical to the one by Sims & Zha (2006). Moreover, in our estimations of these models one of the states was always occurring for just a few (if not just one) periods, which makes the estimation of the state-specific parameters for that state cumbersome. In our view, this explains the instability in the MDD estimates for these models.

7.3. Comparison of the Monetary Policy Models

To summarize the assessment of of the structural models, we find a robust support for the NBR-TR policy shock model. We shall analyze it in more details below, which is feasible mainly because the restricted

Table 4: Posterior estimates of matrices A_0, λ_2 and λ_3 for models with Markov-switching volatility with 3 states

y_t	A_0							λ_2	λ_3
				FF policy shock model					
gdp_t	20.464							0.141	0.026
	(0.425)							(0.047)	(0.003)
p_t	3.097	25.703						0.120	0.017
	(0.746)	(1.485)						(0.042)	(0.002)
$pcom_t$	-4.177	-4.075	9.428					0.610	0.199
	(1.448)	(2.584)	(1.548)					(0.318)	(0.069)
FF_t	-1.046	0.437	-1.104	1.106				6.776	0.228
	(1.818)	(2.849)	(0.862)	(0.205)				(3.602)	(0.098)
nbr_t	0.180	-2.740	0.147	0.161	10.957			0.342	0.055
	(1.407)	(2.247)	(0.393)	(0.036)	(1.195)			(0.217)	(0.013)
tr_t	-0.984	0.129	-0.213	-0.091	-11.392	15.601		0.318	0.029
	(1.014)	(1.852)	(0.313)	(0.027)	(1.086)	(1.215)		(0.168)	(0.005)
m_t	-3.094	-4.283	-0.324	0.037	0.006	-8.348	21.725	0.084	0.024
	(1.015)	(1.715)	(0.255)	(0.020)	(0.855)	(1.194)	(1.489)	(0.004)	(0.031)
				NBR policy shock model					
gdp_t	21.224							0.145	0.027
	(0.669)							(0.048)	(0.004)
p_t	3.729	26.341						0.117	0.018
	(0.688)	(1.493)						(0.040)	(0.003)
$pcom_t$	-1.478	-2.665	9.577					0.577	0.212
	(1.647)	(2.714)	(1.577)					(0.311)	(0.075)
FF_t	-3.102	-1.522	-0.826	1.013	6.053			4.892	0.163
	(1.527)	(2.775)	(0.750)	(0.177)	(1.393			(2.480)	(0.066)
nbr_t	1.611	-1.813	0.453		9.966			0.402	0.053
	(1.179)	(2.117)	(0.397)		(1.116)			(0.216)	(0.013)
tr_t	-3.081	-1.352	-0.254	-0.080	-11.213	15.269		0.350	0.028
	(0.778)	(1.800)	(0.302)	(0.025)	(1.008)	(1.192)		(0.167)	(0.005)
m_t	-5.139	-6.369	-0.302	0.046	-0.001	-7.882	20.871	0.094	0.022
	(1.002)	(1.596)	(0.253)	(0.019)	(0.715)	(1.052)	(1.555)	(0.034)	(0.004)
				NBR-TR policy shock model					
gdp_t	21.530							0.124	0.026
	(0.495)							(0.037)	(0.003)
p_t	3.215	25.652						0.094	0.017
	(0.685)	(1.482)						(0.028)	(0.002)
$pcom_t$	-3.192	-3.542	9.391					0.532	0.191
	(1.600)	(2.609)	(1.556)					(0.256)	(0.069)
FF_t	-8.252	-3.390	-0.699	0.994	7.749	-4.048		3.547	0.122
	(1.940)	(2.872)	(0.720)	(0.176)	(1.839)	(2.028)		(1.715)	(0.050)
nbr_t	5.741	2.497	0.428		15.007	-13.784		1.107	0.030
	(0.939)	(2.050)	(0.358)		(1.326)	(1.398)		(0.375)	(0.006)
tr_t	-3.174	-3.000	-0.017			13.999		0.059	0.087
	(1.214)	(2.176)	(0.369)			(1.523)		(0.018)	(0.023)
m_t	-3.343	-4.540	-0.319	0.035	0.165	-8.887	21.953	0.057	0.024
	(0.885)	(1.642)	(0.246)	(0.019)	(0.614)	(0.991)	(1.394)	(0.018)	(0.004)

Note: The table summarizes the posterior means and the posterior standard deviations (in brackets). In this model $\lambda_1 = \imath_N$.

models have distinct restrictions for each row of matrix A_0. By analysing the posterior draws for these models we come to the conclusion that we obtained samples for identified models. Equation permutation seems to not have occurred in our Gibbs output. Moreover, although we did not impose ordering restrictions on the states of the latent Markov process (similar to those recommended by Frühwirth-Schnatter, 2001), the posterior draws are such that we could easily find such ordering restrictions holding. The modes of the posterior distribution of state-specific parameters are, therefore, well separated. That enables us to report posterior characteristics of the parameters although we do not know the conditions for global identification

Table 5: Posterior estimates of transition probability matrices **P** for models with Markov-switching volatility with 3 states

	FF model			NBR model			NBR-TR model		
Regime 1	0.776	0.075	0.149	0.787	0.079	0.134	0.791	0.072	0.137
	(0.101)	(0.069)	(0.089)	(0.102)	(0.072)	(0.087)	(0.103)	(0.067)	(0.087)
Regime 2	0.042	0.827	0.131	0.048	0.806	0.146	0.024	0.808	0.168
	(0.037)	(0.077)	(0.069)	(0.039)	(0.079)	(0.072)	(0.024)	(0.070)	(0.067)
Regime 3	0.016	0.031	0.953	0.011	0.037	0.952	0.013	0.050	0.937
	(0.012)	(0.017)	(0.018)	(0.010)	(0.017)	(0.018)	(0.009)	(0.020)	(0.022)

Note: The table summarizes the posterior means and the posterior standard deviations (in brackets).

of the model.

Table 4 reports the estimation results for matrices A_0, λ_2 and λ_3 for the FF, NBR and NBR-TR policy shock models for Markov-switching heteroskedasticity process with three states, whereas Table 5 reports the posterior estimates of transition probability matrices **P** of the corresponding models. Notice that the posterior means of the diagonal elements of matrix A_0 have similar values in all of the models. These parameters are related to the structural shock variances in the first state. There is a variation in the values of the posterior means for other elements of this matrix. The estimates of the structural shock variances in states 2 and 3 are very similar for all of the models.

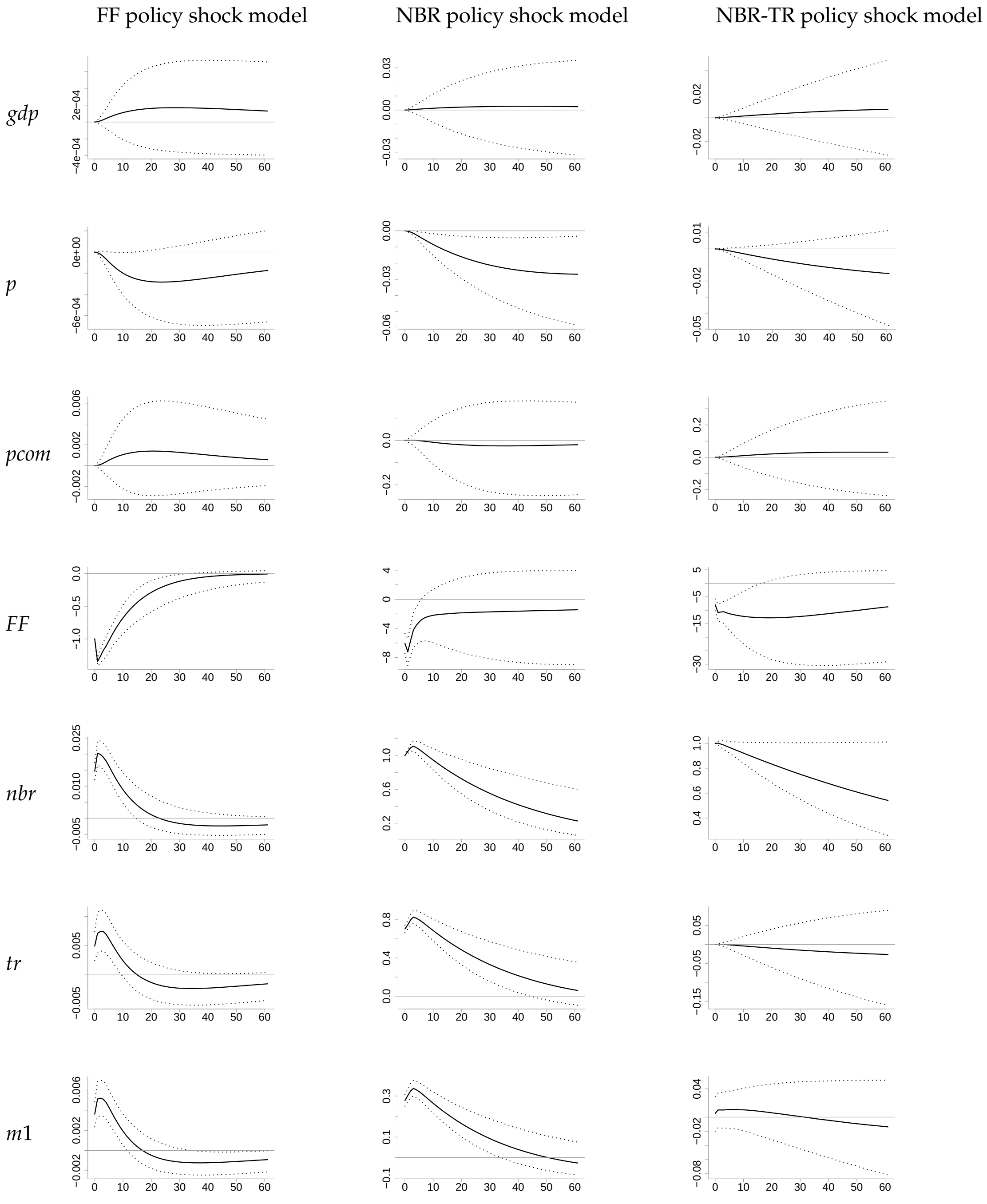
Volatility of structural shocks. In order to characterize the volatility states we analyze the estimates of vectors λ, the transition probabilities matrices **P** and the posterior probabilities of states, $\Pr[s_t|\mathbf{y}]$, that are plotted in Figure 3 for the NBR-TR model (very similar plots were obtained for the NBR and FF models). Regime 1 occurs only for three quarters starting from Q3 2000 to Q1 2001 and it is the state of the highest variances of most of the structural shocks except for the federal funds rate and the non-borrowed reserves shocks. This regime is also the least persistent one, with the probability of staying in the state for the next period not exceeding 0.8.

The shocks to federal funds rate and the non-borrowed reserves have the highest variances in regime 2 that has the high probability of occurrence starting from late 1960's up to mid-1980's (with temporal breaks in favor of regime 3) as well as in Q4 2007 denoting the beginning of the financial crisis. Regime 2 gains high probabilities of occurrence in periods corresponding to the first years of the terms of Arthur F. Burns and Paul Volcker as chairmen of the Federal Reserve. This state's expected duration slightly exceeds 5 quarters, and its unconditional probability is just over 21 percent. Other shocks have their variances higher in this state than in regime 3 but lower than in regime 1.

Finally, regime 3 with the lowest values of the structural shocks variances occurs throughout most of the 1960's, in mid-1970's, and then from mid-1980's up to the beginning of the financial crisis. Therefore, the state covers e.g. the period of the Alan Greenspan Federal Reserve chaimanship. With respect to the posterior state probabilities, regime 3 resembles the *Greenspan regime* of Sims & Zha (2006). This regime is the most persistent one with the persistence probability exceeding 0.93, the expected duration of around 4 years and the unconditional probability of nearly 72 percent. Regimes 2 and 3 are much more related to each other than to regime 1 in all of the models as documented by the values of the transition probabilities of the states.

Impulse response functions. Figure 4 shows the impulse responses to the expansionary monetary policy shocks, as defined by the three monetary policy models, together with 16 and 84 percent quantiles of their posterior distributions. We reported the responses for models with the highest posterior probabilities, i.e. models with Markov-switching heteroskedsticity with three states. In order to plot the impulse responses, we standardized the diagonal elements of A_0 matrices to ones. Therefore, the impulse responses to one standard deviation monetary policy shocks in particular states would be just a rescaled versions of the reported ones.

Figure 4: Impulse response functions: the effect of expansionary monetary policy shocks



Note: The graph reports the impulse responses to the expansionary monetary policy shocks, as defined by the three monetary policy models, together with 16 and 84 percent quantiles of their posterior distributions. The plots are obtained for models with four autoregressive lags and Markov-switching heteroskedsticity with three states for which the diagonal elements of A_0 matrices are standardized to ones.

The reaction of the real GDP to an expansionary monetary policy shock is positive, however, not significant. The zero effect line lays within the bands of high posterior probability. Still, with probabilities from 0.56 to 0.6 at all horizons the response is positive. The conclusion is to some extent in line with the effects estimated by Uhlig (2005) who used sign restrictions in order to identify the monetary policy shock.

The NBR-TR policy shock model, which gained the highest posterior probability, is the only model with insignificant price puzzle. This stands in contrast to the FF and NBR policy shock models in which the reaction of the GDP deflator to an expansionary monetary policy shock is negative and significant. Therefore, ordering the federal funds rate after the non-borrowed and total reserves in a traditional Cholesky-type identification in addition to modelling heteroskedasticity of the structural shocks might diminish the unfavorable feature on many empirical models. The response of the commodity prices is positive in the NBR-TR and FF policy shock models and insignificant in all of the considered specifications.

Finally, the responses of the federal funds rate and the monetary aggregates are in line with the theory and in particular they are of similar shapes as the responses reported by Strongin (1995). Similarly, unproblematic in terms of shapes are the corresponding responses in the FF and NBR policy shock models. All of the models include correctly specified liquidity effect.

8. Conclusions

This paper proposed a Bayesian methodology to estimate and discriminate between heteroskedastic structural VARs. The method is implementable and reliable even for models that are not globally identified. This feature is particularly relevant for the analysis of structural models with Markov-switching heteroskedasticity in which the structural system is only identified up to: the sign change and the reordering of the structural shocks, and the permutations of heteroskedastic states' labels. As argued, the identification problem leads to multiple modes of the likelihood function and the posterior distribution of the parameters. Moreover, while maximum likelihood estimation becomes infeasible with the increasing number of variables in the system and/or volatility states, Bayesian estimation and inference remains operational. Moreover, the same solution that allows the unbiased estimation of the marginal data densities for mixture models exposed to the label switching problem is shown to solve the problem of making inference on globally unidentified structural models.

Our empirical findings regarding the volatility of structural shocks confirm those already stated in related papers: multiple volatility states that are not monotonic are required in order to capture the heteroskedasticity of multivariate macroeconomic systems. When it comes to the assessment of the monetary policy models, we found strong and robust support for a model that defines an interplay between the non-borrowed and total reserves as a monetary policy instrument. This result is robust to the number of heteroskedastic states and the considered processes for volatility as well as with respect to the parameters uncertainty. An interesting concurrence is that the monetary policy model best supported by the data is the only one that has only a negligible price puzzle.

Acknowledgments

The authors acknowledge an excellent research assistance by Yang Song. The authors thank in particular Professor Helmut Lütkepohl for his support and multiple discussions as well as Efrem Castelnuovo, Siddhartha Chib, Matthew Greenwood-Nimmo, Marek Jarociński, Bartosz Maćkowiak, Vance Martin, Błażej Mazur, Aleksei Netšunajev, Jacek Osiewalski, Mateusz Pipeń and Bruce Preston. They are grateful also to the participants of the seminars at the European Central Bank, Università di Bologna, Università degli Studí di Milano-Bicocca, European University Institute, Cracow University of Economics and Freie Universität Berlin, as well as of the presentations during the *Identification in Macoroconomics* Workshop 2014 organized by the National Bank of Poland, the Rimini Bayesian Econometrics Workshop 2014, the Melbourne Bayesian Econometrics Workshop 2014 and the FBE Time Series Econometrics Workshop. Tomasz Woźniak acknowledges the support from the Faculty of Business and Economics at the University of Melbourne through the faculty research grant. Financial support was provided by the Deutsche Forschungsgemeinschaft through the SFB 649 *Economic Risk*.

References

Adolfson, M., Lindé, J., & Villani, M. (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometric Reviews*, *26*, 289–328.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics*, *92*, 71– 92.

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

Bernanke, B. S., & Blinder, A. S. (1992). The Federal Funds Rate and the Channels of Monetary Transmission. *American Economic Review*, *82*, 901–921.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and Inferential Difficulties with Mixture Posterior Distributions. *Journal of the American Statistical Association*, *95*, 957970.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, *90*, 1313–1321.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, *86*, 221–241.

Christiano, L. J., & Eichenbaum, M. (1992). Identification and the Liquidity Effect of a Monetary Policy Shock. In A. Cukierman, Z. Hercowitz, & L. Leiderman (Eds.), *Political Economy, Growth, and Business Cycles* (pp. 335–370). Cambridge: MIT PRESS.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary Policy Shocks: What Have We Learned and to What End? In J. B. Taylor, & M. Woodford (Eds.), *Handbook of Macroeconomics* chapter Chapter 1. (pp. 65–148). Elsevier.

Clarida, R., Galí, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, *115*, 147–180.

Doan, T., Litterman, R. B., & Sims, C. A. (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews*, *3*, 1–100.

Droumaguet, M., & Woźniak, T. (2012). *Bayesian Testing of Granger Causality in Markov-Switching VARs*. Working Paper Series European University Institute Florence, Italy.

Dufour, J.-M. (1989). Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica*, *54*, 335–355.

Frühwirth-Schnatter, S. (2001). Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association*, *96*, 194–209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal*, *7*, 143–167.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Gelfand, A. E., & Dey, D. K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*, 501–514.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*, 398–409.

Geweke, J. (2007). Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis*, *51*, 3529–3550.

Giannone, D., Lenza, M., & Primiceri, G. E. (2012). Prior Selection for Vector Autoregressions.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*, 357–384.

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

Horn, R. A., & Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press.

Kulikov, D., & Netšunajev, A. (2013). Identifying Monetary Policy Shocks via Heteroskedasticity : a Bayesian Approach.

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. (2010). Structural Vector Autoregressions With Nonnormal Residuals. *Journal of Business & Economic Statistics*, *28*, 159–168.

Lanne, M., & Lütkepohl, H. (2013). A Statistical Comparison of Alternative Identification Schemes for Monetary Policy Shocks.

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

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics*, *4*, 25–38.

Lubik, T. A., & Schorfheide, F. (2004). Testing for indeterminacy: An application to us monetary policy. *American Economic Review*, (pp. 190–217).

Lütkepohl, H., & Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs. *Journal of Applied Econometrics*, *29*, 479–496.

Marin, J.-M., & Robert, C. P. (2008). Approximating the marginal likelihood in mixture models. *Bulletin of the Indian Chapter of ISBA*, *V*, 2–7.

Markun, M. (2011). *Bayesian Vector Autoregressive Analysis*. Ph.D. thesis European University Institute.

Netšunajev, A. (2013). Reaction to technology shocks in markov switching structural vars: identification via heteroscedasticity. *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.

Rigobon, R. (2003). Identification Through Heteroskedasticity. *The Review of Economics and Statistics*, *85*, 777–792.

Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, *87*, 861– 868.

Rubio-Ramírez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies*, *77*, 665–696.

Sims, C. A. (1986). Are Forecasting Models Usable for Policy Analysis? *Quarterly Review. Federal Reserve Bank of Minneapolis.*, *10*, 2–16.

Sims, C. A. (1992). Interpreting the macroeonomic time series facts. the effects of monetary policy. *European Economic Review*, *36*, 975–1011.

Sims, C. A. (2010). Understanding non-bayesians.

Sims, C. A., & Uhlig, H. (1991). Understanding Unit Rooters: A Helicopter Tour. *Econometrica*, *59*, 1591–1599.

Sims, C. A., Waggoner, D. F., & Zha, T. (2008). Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics*, *146*, 255–274.

Sims, C. A., & Zha, T. (1999). Error bands for impulse responses. *Econometrica*, *67*, 1113–1155.

Sims, C. A., & Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, *96*, 54–81.

Strongin, S. (1995). The identification of monetary policy disturbances explaining the liquidity puzzle. *Journal of Monetary Economics*, *35*, 463–497.

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.

Villani, M. (2009). Steady-State Priors for Vector Autoregressions. *Journal of Applied Econometrics*, *24*, 630– 650.

Villani, M., & Warne, A. (2003). Monetary Policy Analysis in a Small Open Economy Using Bayesian Cointegrated Structural VARs.

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