Identifying Shocks in Structural VAR Models via Heteroskedasticity: a Bayesian Approach

Dmitry Kulikov and Aleksei Netšunajev*

Abstract

This paper contributes to the literature on statistical identification of macroeconomic shocks by proposing a Bayesian VAR with time–varying volatility of the residuals that depends on a hidden Markov process, referred to as an MS-SVAR. With sufficient statistical information in the data and certain identifying conditions on the variance–covariance structure of the innovations, distinct volatility regimes of the reduced form residuals allow all structural SVAR matrices and impulse response functions to be estimated without the need for conventional *a priori* identifying restrictions. We give mathematical identification conditions and propose a novel combination of the Gibbs sampler and a Bayesian clustering algorithm for the posterior inference on MS-SVAR parameters. The new methodology is applied to US macroeconomic data on output, inflation, real money and policy rates, where the effects of two real and two nominal shocks are clearly identified.

JEL Code: C11, C32, C54

Keywords: Markov switching models, Volatility regimes, Statistical identification, Bayesian inference, Clustering methods, SVAR analysis

Author's e-mail address: dmitry.kulikov@eestipank.ee

*We would like to thank participants in the research seminars at Eesti Pank and Lietuvos bankas, and attendees of the IAAE 2015 Applied Econometrics Conference and the Econometrics Society 2015 World Congress for their useful comments on this paper, and Helmut Lütkepohl, Tomasz Woźniak, Karsten Staehr, Lenno Uusküla and Svetlana Makarova for their direct and valuable feedback. Naturally, all errors and omissions in this paper remain our own. The second author gratefully acknowledges support from Deutsche Forschungsgemeinschaft (DFG) through SFB649 "Economic Risk". Parts of the paper were completed while the second author was a visiting researcher at Eesti Pank, whose support is gratefully acknowledged.

Non-technical summary

Structural vector autoregressions are popular tools in modern empirical macroeconomics, and they are widely used in monetary policy analysis and other applications for examining dynamic interactions and the effects of various shocks on real and nominal macroeconomic aggregates. A central issue of this research is the identification of structural shocks, such as monetary policy, aggregate demand and aggregate supply shocks, for which the conventional approach used in most of the empirical research in the area requires restrictions to be placed directly on the structural vector autoregression parameters. However, recent literature has raised some criticism of the conventional approach, because even if the identifying assumptions that are imposed are based on a widely accepted economic idea, there may still be a gap between the data and the theoretical model, leading to a potentially biased inference of the dynamic reactions of the model variables, and confounding the development of new theories. In this study we depart from the traditional approach to shock identification and use the additional statistical information that is available in many macroeconomic data series in the form of the time-varying volatility of error terms to help in identifying structural parameters and the interpretation of shocks. We model the volatility states of the errors with a hidden Markov process, referring to the new framework as the Markov switching structural vector autoregression. We show how existing mathematical results allow a statistical identification of the structural parameters for at least two volatility regimes, without the need to impose any a priori identifying assumptions. We apply Bayesian statistical inference for parameter estimation and shock identification in the new framework. The new methodology is validated using the medium-scale monetary policy systems for the US and the euro area, and a small-scale model with an interest rate premium for the Estonian economy. Previous empirical research has shown that the US macroeconomic data since mid-1960s are notable for the time-varying volatility of macroeconomic shocks, while the remaining model parameters can be considered stable. A similar set of results applies for the euro area macroeconomic data starting from the early 1970s. We find sufficient volatility information in our data samples to be able to identify and disentangle a full set of shocks for every estimated model in our empirical applications. Furthermore, we undertake a careful economic interpretation of the shocks that are identified, by looking at their short-run impacts and impulse responses, comparing them with the existing literature, and finding consistent economic narratives for every shock in our empirical models. The shock identification in our models is achieved without the a priori identifying restrictions that are common in other empirical studies. Although we are mostly interested in monetary policy and risk premium shocks, our statistical identification methodology enables us simultaneously to disentangle and attach economic interpretations to other structural macroeconomic innovations, such as aggregate demand, aggregate supply, and money demand shocks. We also point out that the results of our statistical shock identification procedure are

not always compatible with the traditional short-run and sign identification schemes used in much of the recent empirical literature, which warrants further careful validation and checking of the existing results using the new identification methodology in this paper together with other alternative approaches.

Contents

1 Introduction

Structural vector autoregressions (SVARs) firmly belong in the toolbox of modern empirical macroeconomists and are widely applied in monetary policy research and related analyses to study interactions between the real-world data and the set of hypothetical structural shocks that govern the observed macroeconomic cycles. At the core of this vast econometric literature are the mathematical and statistical properties of different identification schemes for inferring the effect of structural shocks on the endogenous variables from the set of estimated reduced-form parameters. The conventional identification approaches widely used in the empirical macroeconomic research typically require some restrictions to be imposed directly on the SVAR system matrices; see Sims (1980), Bernanke (1986), Blanchard and Watson (1986), Blanchard and Quah (1989), Canova and De Nicoló (2002), Uhlig (2005), among many others. However, although the conventional identification assumptions are often based on well-understood ideas from macroeconomics, a gap may still exist between the real-world data and the theoretical models, leading to possible biases in the estimated dynamic responses of endogenous system variables and potentially confounding the development of new theories. This type of criticism of the conventional just-identifying and sign restrictions is now new and was recently re-asserted in contributions by Lütkepohl (2012) and Lütkepohl and Netšunajev (2014).

In this study we eschew the conventional identification assumptions in favour of additional statistical information, typically found in long stretches of macroeconomic data in the form of the time-varying volatility of reduced-form errors, to identify structural shocks in a modified SVAR framework along the lines of Rigobon (2003) and Lanne and Lütkepohl (2008). Specifically, we postulate a number of discrete volatility states for the reduced-form residuals driven by a hidden Markov process, akin to Lanne, Lütkepohl and Maciejowska (2010), and refer to the overall framework as a "Markov-switching structural vector autoregression" (MS-SVAR). Our structural form identification approach proceeds from the reduced-form parameters of a Bayesian VAR model with a Markov switching volatility structure that satisfies a minimal set of *a priori* restrictions to the full set of structural SVAR matrices that characterise the nature and effects of the hypothetical structural shocks supported by the available sample of real-world macroeconomic data. Building on the existing results in Lanne et al. (2010), the first half of this paper contains two new propositions on the structural-form identification in the class of MS-SVAR models with an arbitrary number of volatility states. In particular, we derive sufficient conditions on the reduced-form variance-covariance matrices that guarantee model identification up to an arbitrary permutation order of the structural shocks. We then propose a new method of pinning down the specific order of the structural shocks from the posterior simulations of the reduced-form parameters by employing a computationally-intensive Bayesian clustering algorithm from the statistical literature; see Xu and Wunsch (2005).

Recently, a handful of empirical studies dealing with identification of structural shocks in SVAR models using the time-varying volatility of the residuals have appeared in the literature, all of them applying the conventional maximum likelihood estimator to this particular setting; see Bacchiocchi and Fanelli (2012), Netšunajev (2013) and Lütkepohl and Netšunajev (2014), among many others. In the present study we depart from the conventional maximum likelihood estimator in favour of the Bayesian statistical approach for the parameter inference and shock identification in the MS-SVAR framework. In our view, the latter offers a number of advantages over the conventional frequentist methods. Firstly, going Bayesian yields a fuller picture of the posterior effects of different shocks, even before their economic interpretations are ascertained, proving invaluable in the empirical validation of the proposed MS-SVAR identification methodology and in searching for suitable economic narratives for the observed structural innovations. Secondly, the modelling philosophy pursued in this paper, proceeding from the given macroeconomic data sample to the fully-fledged statistical model with only a minimal set of required *a priori* assumptions for attaining the structural shock identification, is quintessentially Bayesian. And lastly, from the practical perspective, Bayesian methods are often less sensitive to the likelihood irregularities and numerical maximization difficulties that can be particularly daunting for the hidden Markov models; see Herwartz and Lütkepohl (2014). On the downside, the full-fledged Bayesian inference requires a choice of suitable priors, makes use of non-standard numerical algorithms and tends to be computationally demanding.

The second half of this paper presents an empirical application of the new methodology to the quarterly data series of the US output, inflation, real money and monetary policy rates over the last 45 years. As pointed out by Primiceri (2005) and Sims and Zha (2006), the US macroeconomic cycle since the mid-1960s has been characterized by the time-varying volatility of macroeconomic shocks, while the remaining reduced-form VAR parameters appear to be stable over the sample period. We find sufficient volatility information in these data to disentangle four fundamental shocks, two of which additionally require the application of the proposed statistical clustering technique to achieve full identification. A successful economic interpretation of these shocks rests on a careful analysis of their posterior short-run impacts and impulse response functions, which are detailed in the empirical section of the paper, allowing us to label two real (supply and demand) and two nominal (monetary policy and money multiplier) structural shocks that govern the US macroeconomic dynamics over the past 45 years. To reiterate, the statistical identification of the four structural shocks in our empirical MS-SVAR example is achieved without any conventional *a priori* restrictions on the SVAR system matrices, though these are common in the recent applied macroeconomic literature.

The rest of the paper is organized as follows. Section 2 describes the MS-SVAR model and discusses shock identification issues using the auxiliary statistical information contained in the time-varying volatility of the reduced-form residuals. Section 3 provides a detailed overview of the Gibbs sampler and Bayesian clustering algorithms required for the full statistical inference on the effects of hypothetical structural shocks in a given data sample. Section 4 presents an empirical application of the MS-SVAR model to the US data sample and illustrates the practical identification steps for structural shocks. Finally, we conclude and provide some possible future directions for the theoretical and applied research in this area.

2 Structural VAR identification via discrete volatility states

Let the dynamics of an $n \times 1$ vector $\boldsymbol{y}_t$ of endogenous variables be given by the following vector autoregressive model:

$$\mathbf{A}_0 \boldsymbol{y}_t = \boldsymbol{k}_0 + \boldsymbol{k}_1 t + \mathbf{A}_1 \boldsymbol{y}_{t-1} + \ldots + \mathbf{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t(s_t)\,, \tag{1}$$

where $\boldsymbol{k}_0$ and $\boldsymbol{k}_1$ are (optional) deterministic intercept and linear trend parameters respectively, $\mathbf{A}_0$ is an unrestricted $n \times n$ contemporaneous parameter matrix, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are autoregressive matrices, and $\boldsymbol{\epsilon}_t(s)$ is a vector of serially uncorrelated structural innovations that depends on the hidden discrete state parameter $s \in \{1, \ldots, m\}$. We assume the following conditional distribution of the structural innovations:

$$\boldsymbol{\epsilon}_t(s) \,|\, s \sim \text{Normal}(\,\mathbf{0}, \mathbf{D}(s)\,)\,,$$

where $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ is a family of suitably distinct $n \times n$ diagonal matrices, and where $\mathbf{D}(1) \equiv \mathbf{I}$ is assumed for the identification purposes. When $m = 1$, this model reduces to the conventional textbook SVAR case; see Hamilton (1994), Amisano and Giannini (1997) and Lütkepohl (2005). When the hidden discrete state process s_t is Markov and the number of states is greater than one, we refer to model (1) as the "Markov-switching structural vector autoregression".

Assuming the non-singularity of the contemporaneous parameter matrix $\mathbf{A}_0$, the model can be written in the familiar reduced-form VAR representation with time-varying volatility of errors:

$$\boldsymbol{y}_t = \boldsymbol{c}_0 + \boldsymbol{c}_1 t + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t(s_t)\,, \tag{2}$$

where $\boldsymbol{c}_i = \mathbf{A}_0^{-1} \boldsymbol{k}_i$ for each $0 \leq i \leq 1$, $\boldsymbol{\Phi}_j = \mathbf{A}_0^{-1} \mathbf{A}_j$ for each $1 \leq j \leq p$, and:

$$\boldsymbol{u}_t(s) \,|\, s \sim \text{Normal}(\,\mathbf{0}, \boldsymbol{\Sigma}(s)\,)\,,$$

where $\boldsymbol{\Sigma}(s) = \mathbf{A}_0^{-1} \mathbf{D}(s) \mathbf{A}_0'^{-1}$ for each volatility state $s \in \{1, \ldots, m\}$, giving rise to the family of reduced-form variance-covariance matrices $\{\boldsymbol{\Sigma}(s) : 1 \leq s \leq m\}$.

Model (2) provides the basis for writing down the likelihood function of the sample data and obtaining statistical inference on the parameters; see Section 3 and the discussion therein.

The well-known SVAR identification issue in this framework arises due to the essential non-uniqueness of the structural form (1) from the perspective of the likelihood function and the associated reduced-form model (2). To make this point precise, consider the structural model (1) premultiplied by an arbitrary conformable unitary matrix $\mathbf{U}$:[1]

$$\mathbf{U}\mathbf{A}_0\boldsymbol{y}_t = \mathbf{U}\boldsymbol{k}_0 + \mathbf{U}\boldsymbol{k}_1 t + \mathbf{U}\mathbf{A}_1\boldsymbol{y}_{t-1} + \ldots + \mathbf{U}\mathbf{A}_p\boldsymbol{y}_{t-p} + \mathbf{U}\boldsymbol{\epsilon}_t(s_t)\,. \tag{3}$$

The structural form identification problem in the case of $m = 1$ manifests itself in the observational equivalence of the continuum of SVAR models in (3) from the likelihood function standpoint, since all of them give rise to exactly the same reduced-form VAR representation (2). Essentially, the matrix $\mathbf{U}$ in (3) acts as an additional free model parameter, which needs to be pinned down to achieve a unique SVAR identification from the reduced-form coefficients. The conventional identification assumptions in the SVAR literature make use of certain *a priori* restrictions on the system matrices $\mathbf{A}_0, \ldots, \mathbf{A}_p$, broadly falling into one of the following three categories:

- The orthogonalisation of reduced-form errors in Sims (1980) and the recursive identification schemes of Christiano, Eichenbaum and Evans (1999), and general non-recursive approach of Sims (1986), Bernanke (1986) and Blanchard and Watson (1986) impose *a priori* exclusion (zero) restrictions on the elements of $\mathbf{A}_0$ or $\mathbf{A}_0^{-1}$, which are usually motivated by the macroeconomic theory without prior testing of their statistical validity;
- The long–run identification schemes of Blanchard and Quah (1989) restrict certain functions of $\mathbf{A}_0$ and the autoregressive matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ based on *ex ante* theoretical considerations, but this approach alone does not suffice for the complete structural form identification and is difficult to test for in the data;
- The sign restrictions of Canova and De Nicoló (2002) and Uhlig (2005) use theory-based prescriptions about the likely directions of short-run responses to certain shocks, leaving the underlying structural model essentially unidentified; see Rubio-Ramírez, Waggoner and Zha (2005) and Fry and Pagan (2011).

In contrast to these conventional approaches, the burden of structural form identification in the class of MS-SVAR models is shifted from *a priori* restrictions on the system matrices $\mathbf{A}_0, \ldots, \mathbf{A}_p$ to certain conditions on the volatility

[1] A square matrix $\mathbf{U}$ is said to be unitary, if $\mathbf{U}^*\mathbf{U} = \mathbf{I}$, where an asterisk denotes a Hermitian adjoint. If $\mathbf{U}$ consists of real elements, it is called real orthogonal; see Horn and Johnson (2013).

structure of structural innovations in (1), which are made precise below. Consider again the model (3) with $m > 1$; for each $s \in \{1, \ldots, m\}$ the variance-covariance matrix of the reduced-form errors $\mathbf{\Sigma}(s)$ is given by:

$$\mathbb{E}\,\mathbf{A}_0^{-1}\mathbf{U}'\mathbf{U}\boldsymbol{\epsilon}_t(s)\boldsymbol{\epsilon}_t(s)'\mathbf{U}'\mathbf{U}\mathbf{A}_0'^{-1} = [\mathbf{A}_0^{-1}\mathbf{U}']\cdot[\mathbf{U}\mathbf{D}(s)\mathbf{U}']\cdot[\mathbf{U}\mathbf{A}_0'^{-1}] =: \tilde{\mathbf{A}}_0^{-1}\cdot\tilde{\mathbf{D}}(s)\cdot\tilde{\mathbf{A}}_0'^{-1} \tag{4}$$

In order words, instead of directly observing the structural matrices $\mathbf{A}_0^{-1}$ and $\mathbf{D}(s)$ for $s > 1$ as they appear in (1), the statistical information contained in the reduced-form errors will only provide evidence about matrices $\tilde{\mathbf{A}}_0^{-1}$ and $\tilde{\mathbf{D}}(s)$ defined in the expression above. For an arbitrary unitary $\mathbf{U}$ and given $s > 1$, the matrix $\tilde{\mathbf{D}}(s)$ in (4) is generally not diagonal and different from the structural $\mathbf{D}(s)$ matrix; the same applies to $\tilde{\mathbf{A}}_0^{-1}$ in this expression and its structural counterpart $\mathbf{A}_0^{-1}$ in (1). By imposing mild *a priori* assumptions on the family of diagonal matrices $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ in (1), we seek to pin down the matrices $\tilde{\mathbf{A}}_0^{-1}$ and $\tilde{\mathbf{D}}(s)$, such that all essential dynamic features and the economic interpretation of the original structural innovations in (1) are preserved intact.[2] In other words, the statistical identification in the class of MS-SVAR models proposed in this paper proceeds by recovering the family $\{\tilde{\mathbf{D}}(s) : 1 \leq s \leq m\}$ and the matrix $\tilde{\mathbf{A}}_0^{-1}$ from the reduced–form variance–covariance matrices $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$ according to (4) and checking if necessary model identification conditions are satisfied. The final recovery of the structural matrix $\mathbf{A}_0$ then proceeds by addressing some possibly remaining indeterminacies as appropriate, which guarantees the full resolution of the dynamic effects of all the structural shocks in the model. The remaining part of this section deals with the necessary identification conditions and remaining indeterminacy cases in the class of MS-SVAR models.

At the very minimum, we will require the family $\{\tilde{\mathbf{D}}(s) : 1 \leq s \leq m\}$ in (4) to be diagonal and $\tilde{\mathbf{D}}(1) = \mathbf{I}$, by analogy to the original assumptions in the structural MS-SVAR model in (1). To attain this, we build on the previous findings of Lanne et al. (2010), who provide a set of mild conditions on $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ for the mapping $\{\mathbf{\Sigma}(1), \ldots, \mathbf{\Sigma}(m)\} \mapsto \mathbf{A}_0^{-1}$ to be unique up to the column signs of $\mathbf{A}_0^{-1}$ in the case $m > 1$ in their Proposition 1 on page 130.[3] In particular, they require each pair of structural innovations in (1) to have distinct variances in at least one of the m volatility states of

[2] We seek to avoid imposing substantial restrictions on the family $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ in the class of MS-SVAR models because *a priori* we do not have enough information to substantiate too many claims on the volatility structure of the innovations in (1). Most of the applied work is driven by the data, and it is difficult to assert claims like "the volatility of a technology shock is larger than the volatility of a preference shock" prior to estimating model parameters and examining impulse responses and variance decompositions for a particular sample.

[3] Similarly to the case of Lanne et al. (2010), a unitary matrix $\mathbf{U}$ in expression (4) consisting of ± 1 on the main diagonal preserves the variance-covariance structure of the fundamental shocks in (3). But this corresponds to the case of flipping the shock signs and presents just a trivial normalisation issue, since the dynamic impact of structural innovations and their economic interpretation remains unchanged.

the model. For example, it rules out a configuration where for each $s > 1$ the diagonal matrix $\mathbf{D}(s)$ is an arbitrary scalar multiple of the identity matrix $\mathbf{D}(1) \equiv \mathbf{I}$. In the former case, the unitary matrix $\mathbf{U}$ in expression (4) remains unrestricted, leaving the structural form of the MS-SVAR model essentially unidentified.

Below, we extend the previous result of Lanne et al. (2010) and explicitly ascertain one of the central indeterminacy cases for the structural form identification in the class of MS-SVAR models:

PROPOSITION 1. *Let the family of diagonal variance-covariance matrices* $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ *of structural innovations in the MS-SVAR model (1) satisfy the Lanne et al. (2010) condition:* $\forall\, 1 \leq k, l \leq n,\ k \neq l,\ \exists\, s \in \{1, \ldots, m\}$ *s.t.* $d_k(s) \neq d_l(s)$, *where* $d_k(s)$ *denotes the k-th diagonal element of* $\mathbf{D}(s)$ *for each* $s \in \{1, \ldots, m\}$. *Then the family* $\{\tilde{\mathbf{D}}(s) : 1 \leq s \leq m\}$ *defined in (4) retains its diagonal structure with the elements* $\tilde{d}_k(s) = d_{\sigma(k)}(s)$ *for each* $s \in \{1, \ldots, m\}$, $1 \leq k \leq n$ *and a permutation mapping* $\sigma : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$ *independent of s. The matrix* $\tilde{\mathbf{A}}_0^{-1}$ *defined in (4) is a column and sign permuted version of* $\mathbf{A}_0^{-1}$, *where the order of column permutation is determined by* σ *and each column can be arbitrary multiplied by* ± 1.

Proposition 1 makes explicit the sense in which the structural parameters of the MS-SVAR model can be recovered from the reduced form VAR coefficients: without any additional *a priori* information about the innovations volatility in (1) or some prior sorting requirements on the diagonal elements of $\{\mathbf{D}(s) : 1 \leq s \leq m\}$, the order of equations and the structural shocks in (1) is indeterminate from the perspective of the reduced-form VAR representation in (2). In the classical inference case, where the point estimates of the parameters are sought, the solution to this issue would be to impose a certain predetermined order of the diagonal elements of $\tilde{\mathbf{D}}(s)$ for a suitable subset of $s \in \{1, \ldots, m\}$ during the estimation. However, in the case of Bayesian inference, which we use in this paper, this approach is not suitable, since the diagonal elements of $\{\mathbf{D}(s) : 1 \leq s \leq m\}$ are diffuse and given by probability densities. In order words, if a predetermined order is imposed on $\tilde{\mathbf{D}}(s)$ during the posterior inference, it is likely to result in an arbitrarily intermixed order of the structural equations and shocks across different posterior draws of $\tilde{\mathbf{A}}_0^{-1}, \tilde{\mathbf{D}}(2), \ldots, \tilde{\mathbf{D}}(m)$.

In order to overcome this difficulty and fully disentangle the order of structural equations and shocks in (1) from the posterior draws of the reduced-form parameters, we propose the use of computationally-intensive clustering methods applied to the posterior impulse response functions obtained from the draws with mixed structural shocks order. This idea relies on the fact that, although the column order of $\tilde{\mathbf{A}}_0^{-1}$ across different posterior draws is different, all structural shocks retain their unique economic interpretation reflected in their specific dynamic impact on the endogenous system variables. By clustering together the posterior impulse response functions, we should finally be able to group the shocks and equations in (1) in a correct and coherent way.[4] Specific details of the statistical clustering technique that we propose to use as the final identification step for the MS-SVAR model are provided in Subsection 3.2.

From the perspective of statistical inference, the main challenge is to recover the matrices $\tilde{\mathbf{A}}_0^{-1}, \tilde{\mathbf{D}}(2), \ldots, \tilde{\mathbf{D}}(m)$ from the reduced-form VAR model (2), specifically from the family of variance-covariance matrices $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$. For the case of two volatility states, a known result in the matrix analysis states that for any pair of Hermitian matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, where at least one is positive definite,[5] there exists a non-singular matrix $\mathbf{A}$ such that both $\mathbf{A}'\mathbf{\Sigma}_1\mathbf{A}$ and $\mathbf{A}'\mathbf{\Sigma}_2\mathbf{A}$ are diagonal; see Theorem 7.6.4 in Horn and Johnson (2013). To the best of our knowledge, no mathematical results of a similar generality exist for the case of more than two matrices; in other words, without further assumptions, no diagonalizaton of a family of (positive definite) Hermitian matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_m$ by joint $*$congruence can be achieved when $m > 2$.[6] Moreover, the order condition of Rothenberg (1971) suggests that without further restrictions on $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$, the structural matrices $\tilde{\mathbf{A}}_0^{-1}, \tilde{\mathbf{D}}(2), \ldots, \tilde{\mathbf{D}}(m)$ in the case $m > 2$ are likely to be over-identified.[7]

However, for the purpose of inferring the structural parameters directly from the reduced-form family of $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$ in the general case $m \geq 1$, it is useful to characterise the Hermitian subspace from which the posterior draws of $\tilde{\mathbf{A}}_0^{-1}, \tilde{\mathbf{D}}(2), \ldots, \tilde{\mathbf{D}}(m)$ can be obtained. This results is given by the following proposition:

PROPOSITION 2. *The mapping between the set of reduced-form variance-covariance matrices* $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$ *defined in (2) and the matrices* $\tilde{\mathbf{A}}_0^{-1}$ *and* $\{\tilde{\mathbf{D}}(s) : 1 \leq s \leq m\}$ *defined in (4) is one-to-one if and only if the family* $\{\mathbf{\Sigma}(k)\,\mathbf{\Sigma}(s)^{-1} : 1 \leq s \leq m\}$ *for a given* $1 \leq k \leq m$ *is commuting.*

It follows from Proposition 2 that the cases $m = 1$ and $m = 2$ require no restrictions on the subspace of Hermitian matrices from which the reduced-form variance-covariance matrices are simulated in the Bayesian context. However, the general case $m > 2$ requires the family of $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$ to satisfy the commutativity property stated in Proposition 2. A suitable computer al-

[4]Although the parameters are statistically identified, the resulting shocks lack economic interpretations, since we do not impose any *a priori* meanings on the structural equations in (1). A suitable narrative for the identified shocks ought to be found by a careful analysis of the estimated short-run impacts and impulse responses, consulting the relevant theoretical and empirical literature as necessary. This procedure is similar to the one used by Lütkepohl and Netšunajev (2014) for checking sign restrictions, and is illustrated in our empirical application in Section 4.

[5]In full generality, it is sufficient that a real linear combination of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ is positive definite.

[6]Two square matrices $\mathbf{B}$ and $\mathbf{C}$ are said to be $*$congruent, if there exists a non-singular matrix $\mathbf{A}$ such that $\mathbf{A}^*\mathbf{B}\mathbf{A} = \mathbf{C}$; see Horn and Johnson (2013).

[7]Lanne et al. (2010) exploit this observation and propose a test for over-identification in $\mathbf{A}_0$ by imposing the necessary restrictions on the reduced-form variance-covariance matrices directly during the estimation.

gorithm to implement a random simulation of $\{\mathbf{\Sigma}(s) : 1 \leq s \leq m\}$ from this restricted subspace, however, remains a practical challenge. Therefore, we limit our empirical application in Section 4 to the case of two volatility states, where the recovery of the contemporaneous parameter matrix $\mathbf{A}_0$ is guaranteed by the existing mathematical results without any additional implementation difficulties.

To recap, the MS-SVAR model (1) can be identified by imposing certain *a priori* requirements on the number of volatility states, m, and the variance-covariance matrices of the structural innovations $\{\mathbf{D}(s) : 1 \leq s \leq m\}$, but otherwise without any additional conventional prior restrictions on the system matrices $\mathbf{A}_0, \ldots, \mathbf{A}_p$. This offers a crucial advantage over the conventional identification schemes widely used in the applied macroeconomic research, where the shape and nature of the orthogonalized impulse response functions may be altered in the identification process. The time-varying volatility regimes of the structural innovations, on the other hand, enable us to obtain a full and unrestricted inference on the contemporaneous parameter matrix $\mathbf{A}_0$ together with its sibling, the short-run impact matrix $\mathbf{A}_0^{-1}$, and to disentangle different shocks by looking at the estimated responses of system variables, in concert with suitable economic narratives, without imposing *a priori* assumptions on how they ought to affect the system dynamics.

3 Bayesian statistical inference for the MS-SVAR model

We use the Bayesian approach for obtaining posterior statistical inference on the structural parameters in the MS-SVAR model (1). The proposed statistical inference procedure in this paper is divided into two phases. Firstly, we obtain the desired number of posterior parameter simulations from the reduced-form VAR model (2) using the Gibbs sampler approach of Chib (1996). Secondly, as outlined in Section 2, the full structural form identification in the MS-SVAR model will often require an additional Bayesian inference step to pin down the order of structural equations and shocks from the initial simulations of the reduced-form parameters. In this section we describe both phases of the Bayesian inference procedure for the MS-SVAR models.

3.1 Gibbs sampler algorithm for the reduced-form model

In this subsection we consider the reduced-form Bayesian VAR model with Markov switching volatility regimes for the error terms. It is well known that the posterior inference based on the full likelihood function in hidden Markov models is complex and computationally expensive; see Marin, Mengersen and Robert (2005). Among several proposed solutions to this issue, the Gibbs sampler combined with data augmentation is the most popular in the applied literature. In this paper we use the Gibbs sampler in the context of a hidden

Markov model for the volatility of innovations; see Carter and Kohn (1994), Chib (1996), Krolzig (1997) and Sims, Waggoner and Zha (2008).

Assume, in the context of the reduced form VAR model (2), that the hidden volatility state evolution over time is governed by:

$$s_t \,|\, s_{t-1} \sim \text{Markov}(\, \mathbf{P}, \boldsymbol{\eta}_0 \,)\,,$$

where the $m \times m$ matrix $\mathbf{P}$ governs the conditional distribution of state transitions, and s_0 is distributed according to the m-dimensional vector $\boldsymbol{\eta}_0$. The trajectory of the hidden states $S_T := \{s_1, \ldots, s_T\}$ is obtained by simulation, where T denotes the sample size and, conditional on a given trajectory, the reduced-form variance-covariance matrices $\{\boldsymbol{\Sigma}(s) : 1 \le s \le m\}$ can be estimated using the VAR model residuals split across different volatility states. Bayesian inference on the remaining parameters $\boldsymbol{\beta} := \text{vec}(\boldsymbol{c}_0, \boldsymbol{c}_1, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$ is similar to the usual GLS estimator of the linear regression model with heteroscedastic innovations; see Geweke (1993) and Krolzig (1997).

More specifically, our Gibbs sampler for the Markov swathing Bayesian VAR model includes the following four steps, repeated over a desired number of iterations:[8]

1. S_T is generated by drawing in reverse time order from the posterior distribution:
$$p(s_t \,|\, Y_T, s_{t+1}) \propto p(s_t \,|\, Y_t) \cdot p(s_{t+1} \,|\, s_t)\,, \tag{5}$$
where the first term in the expression is generated recursively using the Chib (1996) Bayesian simulation algorithm for hidden Markov models. It involves the prediction:
$$p(s_t \,|\, Y_{t-1}) = \sum_{s=1}^{m} p(s_t \,|\, s_{t-1} = s) \cdot p(s_{t-1} = s \,|\, Y_{t-1})\,,$$
and update steps:
$$p(s_t \,|\, Y_t) \propto p(s_t \,|\, Y_{t-1}) \cdot \ell(\boldsymbol{y}_t \,|\, Y_{t-1};\, \boldsymbol{\beta}, \boldsymbol{\Sigma}(s_t))\,,$$
where Y_t denotes sample data up to $1 \le t \le T$, and $\ell(\boldsymbol{y}_t \,|\, Y_{t-1}; \boldsymbol{\beta}, \boldsymbol{\Sigma}(s_t))$ is the Gaussian likelihood function of $\boldsymbol{y}_t$ for a given volatility state $s_t \in \{1, \ldots, m\}$;

2. Given a simulated trajectory S_T from the previous step, the Markov transition kernel $\mathbf{P}$ is updated element-by-element, where for each $s \in$

[8] As usual in the recursive Bayesian simulation algorithms, each step of the Gibbs sampler partially depends on the previous iterative draw; see Robert and Casella (2004). We economize on the notation by only showing dependence on Gibbs draws within the same sampler iteration. In addition, all Gibbs sampler expressions in this section are conditioned on the "pre-sample" observations $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{1-p}$.

$\{1,\ldots,m\}$ the posterior probability of leaving the volatility state s is given by the following discrete distribution:[9]

$$p_s \,|\, Y_T, S_T \sim \text{Dirichlet}(\, \alpha_{s1} + n_{s1}(S_T), \ldots, \alpha_{sm} + n_{sm}(S_T)\,)\,, \qquad (6)$$

where $\{\alpha_{sk} : 1 \leq s, k \leq m\}$ are hyper-parameters of the Dirichlet prior for $\mathbf{P}$, and $n_{sk}(S_T)$ is the number of transitions from state s to state k in the given trajectory S_T;

3. The posterior distributions of the reduced-form variance-covariance matrices for each state $s \in \{1,\ldots,m\}$ are given by:

$$\boldsymbol{\Sigma}^{-1}(s) \,|\, Y_T, S_T \sim \text{Wishart}(\, [\mathbf{C}(s) + \bar{\mathbf{C}}(s)]^{-1}; \, \tau(s) + T(s)\,)\,, \qquad (7)$$

where the family of $n \times n$ non-singular matrices $\{\mathbf{C}(s) : 1 \leq s \leq m\}$ and scalars $\{\tau(s) : 1 \leq s \leq m\}$ are hyper-parameters of the Wishart priors for $\{\boldsymbol{\Sigma}(s) : 1 \leq s \leq m\}$, and $\bar{\mathbf{C}}(s)$ are estimated variance-covariance matrices of model residuals belonging to a particular volatility state s, and $0 \leq T(s) \leq T$ is the number of occurrences of s in S_T:

$$\bar{\mathbf{C}}(s) := \sum_{t=1}^{T} \bar{\boldsymbol{u}}_t(\boldsymbol{\beta})\, \bar{\boldsymbol{u}}_t'(\boldsymbol{\beta}) \cdot \mathbf{1}\{s_t = s\}\,, \quad T(s) := \sum_{t=1}^{T} \mathbf{1}\{s_t = s\}\,,$$

$$\bar{\boldsymbol{u}}_t(\boldsymbol{\beta}) := \boldsymbol{y}_t - \boldsymbol{c}_0 - \boldsymbol{c}_1 t - \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} - \ldots - \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p}\,;$$

4. The posterior distribution of the reduced–form VAR coefficients is Gaussian:

$$\boldsymbol{\beta} \,|\, Y_T, S_T, \{\boldsymbol{\Sigma}(s) : 1 \leq s \leq m\} \sim \text{Normal}(\, \bar{\boldsymbol{b}}, \bar{\mathbf{B}}\,)\,, \qquad (8)$$

where the parameters of this distribution are given by the expressions:

$$\bar{\boldsymbol{b}} = \bar{\mathbf{B}}\,(\mathbf{X}' \otimes \mathbf{I}_n)\, \boldsymbol{\Omega}(S_T)\, \boldsymbol{y}\,, \qquad \bar{\mathbf{B}} = \left[(\mathbf{X}' \otimes \mathbf{I}_n)\, \boldsymbol{\Omega}(S_T)\, (\mathbf{X} \otimes \mathbf{I}_n)\right]^{-1},$$

where the $nT \times nT$ block–diagonal matrix $\boldsymbol{\Omega}(S_T)$ is defined as follows:

$$\boldsymbol{\Omega}(S_T) := \begin{pmatrix} \boldsymbol{\Sigma}^{-1}(s_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \boldsymbol{\Sigma}^{-1}(s_T) \end{pmatrix},$$

where $\boldsymbol{y} := (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$ is a $nT \times 1$ data vector, and each row of a $T \times (2 + np)$ data matrix $\mathbf{X}$ contains the following elements:

$$(1, t, \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}')\,.$$

The most direct and computationally economical way of incorporating normal informative priors about $\boldsymbol{\beta}$ into (8) is to use the dummy variable approach of Theil and Goldberger (1961). We use this method of incorporating prior information into the model in our empirical application in Section 4.

[9]For the mathematical details on Dirichlet and Wishart distributions see Poirier (1995).

A recursive iteration of the Gibbs sampler algorithm described by equations (5) to (8) produces, after a pre-specified number of burn-in steps, a sequence of posterior draws of S_T, $\boldsymbol{\beta}$, and the variance-covariance matrices $\{\boldsymbol{\Sigma}(s) : 1 \leq s \leq m\}$. This concludes the description of the Bayesian statistical inference for the reduced-form VAR model with Markov switching volatility of the error terms given in (2). The remaining part of this section is devoted to the Bayesian statistical inference on the structural parameters of the MS-SVAR model.

3.2 Structural form identification using Bayesian clustering algorithm

This subsection deals with the Bayesian statistical inference on the structural parameters of the MS-SVAR model (1). The posterior sampling of $\tilde{\mathbf{A}}_0$ defined in (4) relies on the matrix decomposition result in Horn and Johnson (2013) for the case of two volatility states:[10]

$$\boldsymbol{\Sigma}^{-1}(1) = \tilde{\mathbf{A}}_0' \tilde{\mathbf{A}}_0 \,, \quad \boldsymbol{\Sigma}^{-1}(2) = \tilde{\mathbf{A}}_0' \, \tilde{\mathbf{D}}(2)^{-1} \tilde{\mathbf{A}}_0 \,, \tag{9}$$

which is guaranteed to exist for any two positive definite Hermitian matrices $\boldsymbol{\Sigma}^{-1}(1)$ and $\boldsymbol{\Sigma}^{-1}(2)$ produced by the Gibbs sampler, and is unique up to the column order and signs of $\mathbf{A}_0$ as detailed in Section 2. Given the posterior simulations of $\tilde{\mathbf{A}}_0$ and the reduced-form parameters $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p$, the impulse responses can be computed in the usual manner; see Hamilton (1994) and Lütkepohl (2005).[11]

As was previously emphasised in Section 2, the posterior simulations of $\tilde{\mathbf{A}}_0, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p$ from the Gibbs sampler do not generally lead to the complete structural form identification in the Bayesian context. In particular, the unique structural $\mathbf{A}_0$ matrix remains in most cases unidentified, since the posterior draws of $\tilde{\mathbf{A}}_0$ from the Gibbs sampler can have arbitrary permuted column orders and signs relative to $\mathbf{A}_0$. However, the underlying structural shocks and their dynamic impacts on the endogenous model variables in the MS-SVAR framework remain intact regardless of the permutations of $\mathbf{A}_0$ in the simulated $\tilde{\mathbf{A}}_0$ matrices. In this section we show how a Bayesian clustering algorithm applied to the set of impulse response functions computed from $\tilde{\mathbf{A}}_0$ pins down the unique order of the structural equations and shocks in the estimated MS-SVAR model.

Specifically, we propose to use Bayesian learning algorithms to group the

[10] The proof of Theorem 7.6.4 (a) on page 487 of Horn and Johnson (2013) can serve as a template for the computer implementation of this matrix decomposition result.

[11] Since our model generally contains multiple volatility regimes, we calculate the impulse response functions with respect to a standardised unitary stimulus. These impulse responses are then used as an input to the Bayesian clustering algorithm. After the final identification of $\mathbf{A}_0$, the estimated posterior impulse responses can be re-adjusted to any desired initial stimulus.

structural shocks automatically by exploiting the initial posterior draws of impulse response functions from the Gibbs sampler. The problem of detecting similarities in the data is not new and has been of interest in a variety of fields such as biology or informatics and is referred to as cluster analysis. Therefore we rely on the existing clustering tools and proceed to learn similar shocks from the data in two steps. First, we group (cluster) the pairs of impulse responses that are most similar in terms of their shape. At this step clustering is meant to separate pairs of shocks with different labels into the most suitable clusters. In this way, pairs of "supply-demand" ordered shocks are separated from the "demand-supply" ordered pairs. Second, clustered pairs of impulse responses should be merged on the basis of the similarity of shocks. In other words, regardless of the ordering of shocks in a cluster, we have to find all the demand shocks and all the supply shocks and merge them.

Proceeding to step one, we separate the impulse responses of the two central shocks into four ($k = 4$) clusters by the means of k-medoids algorithm proposed by Kaufman and Rousseeuw (1987). We use four clusters as potentially there are positive/negative supply and demand shocks to be detected. A medoid here is a pair of impulse responses from Gibbs sampler draws whose average dissimilarity to all other pairs of impulse responses is minimal. Clustering is performed on the output of the Gibbs sampler without normalizing the signs of impact effect of shocks. The divisive algorithm proceeds in the following steps:

1. Randomly select k pairs of impulse responses as the initial medoids.

2. Associate each pair of impulse responses with the closest medoid.

3. For each medoid m and each pair of impulse responses p associated with m, swap m and p and compute the dissimilarity of p to all the data points associated with m. Select the medoid p with the lowest dissimilarity. We use Euclidean distance as a measure of dissimilarity.

4. Repeat steps 2 and 3 until there is no change in the assignment of the medoids.

After obtaining four cluster of impulse responses we further have to agglomerate the impulse responses associated with each shock. Put differently, in this step we merge the shocks that have most similar effects on the variables. Prior to agglomeration we standardize the signs of the impact effects of each shock such that (i) the impact effect on the first two variables with the minimal standard error is positive; the other shock is standardized to have a positive impact on the remaining variable automatically; (ii) the mean reaction of output to both shocks is positive. This facilitates agglomeration as the shocks are set to have a similar effect on output among the four clusters. Note that the ordering of shocks is still unknown. To learn it we agglomerate the impulse responses of the shocks using hierarchical clustering known as the unweighted pair-group method using the centroid approach (UPGMC). This method uses the data from the four clusters and involves merging the shocks from the four clusters with the most similar median vectors (see Everitt, Landau, Leese and Stahl (2011)). The agglomerative algorithm includes the following steps:

1. Select the cluster with the lowest standard deviations of the impulse responses as the initial (merged) cluster;

2. Calculate the distance between the median of the impulse responses of the merged cluster and the means of all the other clusters for each shock separately;

3. Merge the shocks mostly similar to that in the merged cluster. The second shock is merged automatically with the remaining shock in the merged cluster;

4. Repeat steps 2 and 3 until all clusters are merged.

The two-step procedure is iterated until changes in the resulting mean matrix of impact effects are negligible.

4 Empirical application to the US macroeconomic dynamics

In this section we provide an empirical illustration of the new approach for identifying structural shocks in the context of a medium-scale Bayesian MS-SVAR model applied to the US macroeconomic data over the last 45 years. In particular, we demonstrate the effectiveness of the Bayesian clustering method for pinning down the order of structural equations from the initial posterior simulations of the reduced-form parameters. Since the new identification approach does not provide any *a priori* labelling of the observed shocks, this section also strives to illuminate the steps needed to find persuasive economic narratives for the dynamic interactions that are revealed between the shocks and the data.

Stock and Watson (2002), Primiceri (2005), Sims and Zha (2006), Justiniano and Primiceri (2008) and some other researchers hold the view that the US macroeconomic dynamics since the mid-1960s have been characterized by a time-varying volatility of shocks, while the remaining reduced-form VAR parameters can be considered essentially constant.[12] This view is consistent with what is known as the "good luck" explanation of the Great Moderation since the mid-1980s.[13] In this paper we make use of this particular view of the US macroeconomic history for exploring the structural dynamics of the data and

[12] A similar claim is made in Rubio-Ramírez et al. (2005) for the euro area macroeconomic data starting from the early 1970s.

[13] A comprehensive account of the Great Moderation and an up-to-date literature survey on the topic can be found in Davis and Kahn (2008).

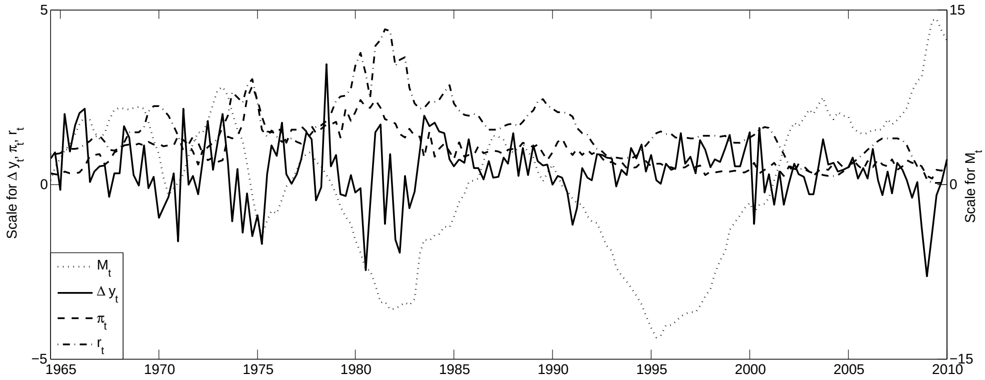

Figure 1: The US macroeconomic data series from 1964Q2 to 2009Q4

statistically identified shocks via the prism of the MS-SVAR model presented in Section 2.

The US macroeconomic data are quarterly and seasonally adjusted, covering the time period from 1964Q2 to 2009Q4. The data are supplied by the Federal Reserve Bank of St. Louis FRED database.[14] Per capita aggregates are computed using the US civilian non-institutional population aged from 16 years up. Figure 1 displays the sample data:[15]

- Output growth rate (Δy_t) is defined as scaled quarter-on-quarter log real GDP per capita changes;
- Inflation rate (π_t) is defined as scaled quarter-on-quarter log changes of the personal consumption expenditures core price index;
- Real money balances (M_t) is defined as the sum of de-trended log inverse money velocity[16] and log real output per capita;
- Monetary policy interest rate (r_t) is defined as the average quarterly federal funds rate.

We now proceed to a detailed presentation of the new approach for identifying structural shocks within the context of a Bayesian MS-SVAR model estimated using the US macroeconomic data in Figure 1. The first phase of our empirical application consists of the initial data exploration and reduced-form model selection and estimation in full isolation from any structural interpretations ascribed to the data by the MS-SVAR model. The most important

[14] All data series are downloaded from `research.stlouisfed.org/fred2`

[15] As a robustness check, in all our applications in this section we have used de-trended output in place of the output growth rates as an alternative business cycle measure, and a GDP deflator-derived inflation measure as a substitute for headline consumer inflation. Our main empirical results and conclusions remain unchanged.

[16] Inverse money velocity is calculated as a ratio of the quarterly sweep-adjusted M2 money stock and quarterly nominal GDP; see Cynamon, Dutkowsky and Jones (2006).

goal of the first phase is to ensure that the data contain sufficient statistical information in the form of time-varying volatility of the reduced–form errors. The second phase of our empirical application will then make use of this additional information for the statistical identification of structural shocks as viewed through lens of the MS-SVAR model and supported by the data.

Using the previously described data, we start by estimating a reduced-form VAR using the Gibbs sampler detailed in Subsection 3.1. The following reduced-form model priors are used for the US macroeconomic data sample: the hyper-parameters of the Dirichlet prior on the elements of the Markov transition kernel $\mathbf{P}$ are $\alpha_{sk} = 10$ for $s = k$ and $\alpha_{sk} = 1$ for $s \neq k$, where $s, k \in \{1, 2\}$. The Wishart prior hyper-parameters on the reduced-form variance-covariance matrices $\mathbf{\Sigma}(1)$ and $\mathbf{\Sigma}(2)$ are given by $\mathbf{C}(1) = \mathbf{C}(2) = \mathbf{I} \cdot (4.80, 0.60, 4.80, 0.12)'$, $\tau(1) = \tau(2) = 12$. For $\boldsymbol{\beta}$, the Minnesota-type dummy observation priors described in Subsection 3.1 are used, with the tightness hyper-parameters set to 1.5.

Table 1: Marginal data density for alternative reduced-form model specifications

Models	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
$m = 1$	-485.648	-445.153	-450.497	-454.256	-456.697	-466.481
$m = 2$	-372.274	-335.396	-323.728	-312.723	-317.406	-322.309
$m = 3$	-360.126	-331.081	-313.793	-303.658	-302.254	-315.769

Notes: Here m denotes the number of volatility states and p denotes the lag augmentation length in the reduced-form VAR model. The marginal data densities are computed using the algorithm proposed in Chib (1995).

The augmentation choice and the number of volatility regimes are driven by the Bayesian model selection procedure based on the estimated marginal data density as follows. The estimated marginal data densities for the reduced-form models with different numbers of volatility regimes and lag augmentation lengths are shown in Table 1. One can observe that models with time-varying volatility of the reduced-form errors are uniformly preferred to standard VAR models for this particular data sample. Comparing two and three-state volatility models, the data clearly support a three-state model with five lags of augmentation. However, we note several shortcomings that are associated with the data-favoured model. Most importantly, our structural identification theory detailed in Section 2 still lacks the full generality needed to work with a number of volatility regimes greater than two. In this application we want to rely on the solid mathematical decomposition results shown in Subsection 3.2, making the two state model preferable.[17] Furthermore, the interpretation of the two state model fits nicely into standard narrative of the post-war US macroeconomic

[17] We also note that the third state in our model tends to capture outliers with extremely high volatility. Across the simulations, the average duration of the third volatility state tends to be around 20 periods, which clearly limits the additional statistical information that can be used for identification purposes.

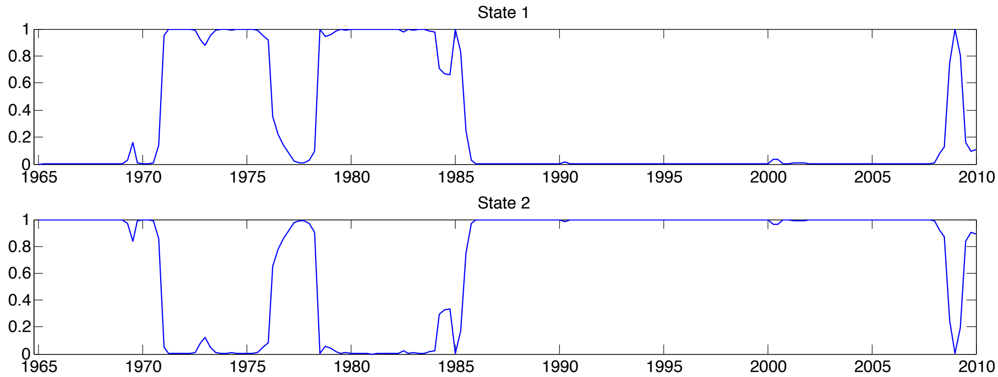

Figure 2: Average of the simulated S_T trajectories for the reduced-form model

history, as we show below. For these reasons we proceed with the reduced-form model with two volatility states, opting for the lag augmentation length $p = 4$ suggested by the marginal data density in Table 1.

The average of the simulated S_T trajectories for the selected reduced-form model is shown in Figure 2. On this figure, the upper panel displays the elevated volatility state, which can clearly be associated with the periods of economic downturn and uncertainty following the first and second oil crises in the early 1970s and 1980s, and the 2007–2009 Global Financial Crisis. The lower panel depicts the normal volatility state, capturing several short tranquil periods in the 1960s, the mid-1970s, and the long Great Moderation period from the mid-1980s until the outbreak of the 2007–2009 Global Financial Crisis. The two volatility states are prominent in the US sample and are well in line with the profession's consensus on US macroeconomic history over the last five decades; see Stock and Watson (2002) and Sims and Zha (2006). Economic history aside, from the perspective of this paper, the well pronounced volatility regimes in the US data should allow us to pin down a full set of statistically identified and economically meaningful shocks in the context of a structural VAR model.

Table 2: Posterior credible sets for the diagonal elements of $\tilde{\mathbf{D}}(2)$

$\tilde{\mathbf{D}}(2)$	*[2.30%*	*16.0%*	*50.0%*	*84.0%*	*97.7%]*
$\tilde{d}_1(2)$	0.033	0.039	0.051	0.065	0.074
$\tilde{d}_2(2)$	0.136	0.159	0.208	0.286	0.343
$\tilde{d}_3(2)$	0.146	0.177	0.252	0.332	0.392
$\tilde{d}_4(2)$	0.519	0.607	0.762	0.961	1.118

Notes: The estimated posterior 96% and 68% confidence sets and the median before the Bayesian clustering procedure are displayed in the columns. The total number of Gibbs simulations is 10000, of which the last 5000 are used for the posterior inference.

We now move from the initial data exploration and the reduced-form model

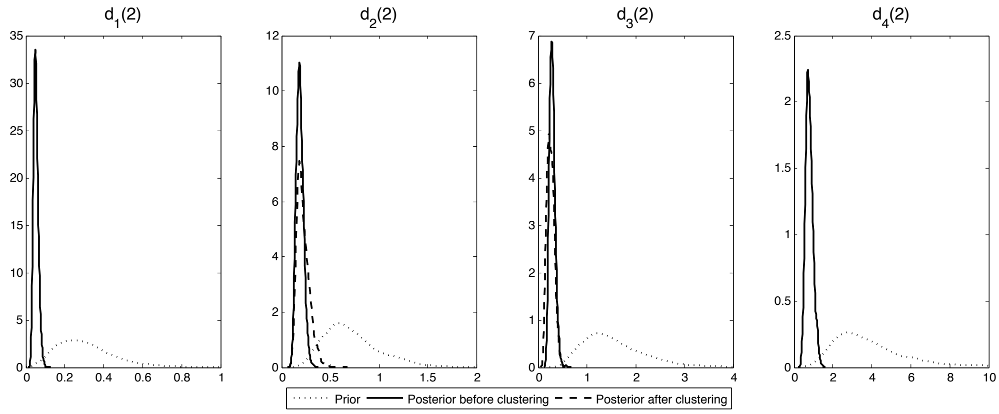

Figure 3: Prior and posterior distributions of $\tilde{\mathbf{D}}(2)$ for the structural model

selection phase of our empirical example to the structural data analysis. We first decompose the reduced-form variance-covariance matrices $\mathbf{\Sigma}(1)$ and $\mathbf{\Sigma}(2)$ generated by the Gibbs sampler using the exact decomposition result in (9). The necessary identification conditions in a two-state MS-SVAR model require all diagonal elements of $\tilde{\mathbf{D}}(2)$, denoted here by $d_1(2), \ldots, d_4(2)$, to be distinct; see Proposition 1 in Section 2 and Lanne et al. (2010). The prior and posterior densities of these elements for the US macroeconomic data are shown in Figure 3. While the data are clearly informative about $d_1(2), \ldots, d_4(2)$, the middle two elements of $\tilde{\mathbf{D}}(2)$ appear to have very similar posterior distributions. The estimated 68% and 90% posterior credible sets and the corresponding medians are shown in Table 2. It follows that $d_1(2)$ and $d_4(2)$ are sufficiently well separated from the middle two elements of $\tilde{\mathbf{D}}(2)$, corroborating the graphical evidence in Figure 3. On the other hand, the posterior credible sets of $d_2(2)$ and $d_3(2)$ overlap in both the 68% and 90% cases. This has a very strong implication for the posterior inference in the structural model: while the first and the fourth shocks are likely to be sufficiently well identified, it may be quite difficult to pin down the second and third shocks in the estimated MS-SVAR model. In order to recover the structural shocks corresponding to the two middle elements of $\mathbf{D}(2)$ we need to resort to the clustering procedure proposed and described in detail in Section 2 and Subsection 3.2 of this paper. The long dash posterior densities displayed in Figures 3 and 4 highlight the effects of this new identification technique and will be discussed further in this subsection.

Additionally, we wish to clarify the Bayesian interpretation of the results in Table 2. Proposition 1 in Section 2 requires the elements of $\tilde{\mathbf{D}}(2)$ to be distinct for any attempt on the subsequent structural data analysis to be meaningful. The Bayesian statistics admits an interpretation of a well-defined statistical model having a unique "true" value of the parameters, but quantifies the inevitable statistical uncertainly about these parameters in any given finite sample using the posterior distributions. Looking through this Bayesian prism, there is ample evidence in Table 2 that the diagonal elements of $\tilde{\mathbf{D}}(2)$ are dis-

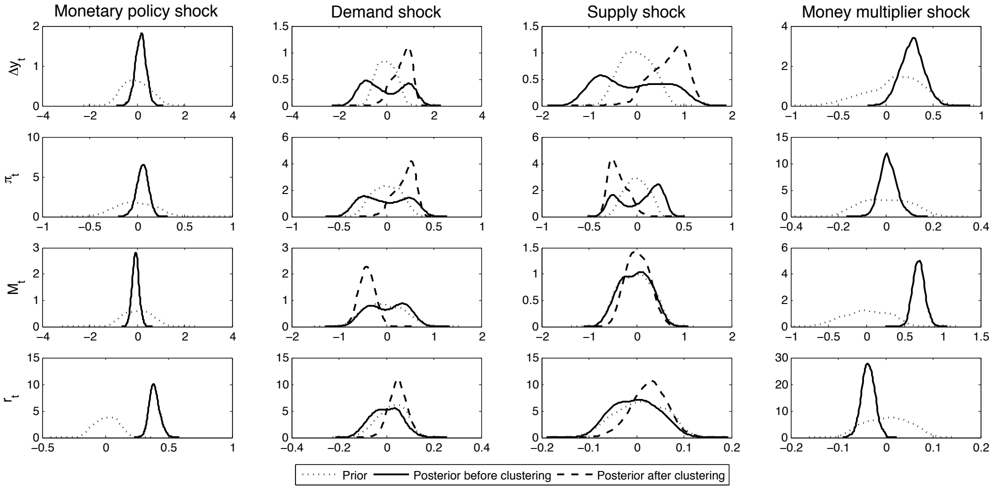

Figure 4: Prior and posterior distributions of $\tilde{\mathbf{A}}_0^{-1}$ for the structural model

tinct,[18] but the posterior uncertainty about the middle two elements of $\tilde{\mathbf{D}}(2)$ implies that the associated shocks are likely to be intermixed across the posterior simulations generated by the Gibbs sampler.

Recall from Section 2 that the short-run impact matrix $\tilde{\mathbf{A}}_0^{-1}$ is generally identified only up to the order and signs of its columns. In the empirical application to the US data, and before the use of the Bayesian clustering procedure, we sign-normalize the short-run impact matrix such that the initial effects of the first shock on r_t (column one of $\tilde{\mathbf{A}}_0^{-1}$), of the second shock on π_t (column two of $\tilde{\mathbf{A}}_0^{-1}$), of the third shock on Δy_t (column three of $\tilde{\mathbf{A}}_0^{-1}$), and of the fourth shock on M_t (column four of $\tilde{\mathbf{A}}_0^{-1}$) are all positive. The prior and posterior distributions of the sign-normalized columns of $\tilde{\mathbf{A}}_0^{-1}$ are shown in Figure 4. We find the US data to be reasonably informative about $\tilde{\mathbf{A}}_0^{-1}$; most clearly this is the case in the first and the last columns of the matrix. On the other hand, before clustering is applied, one clearly observes wide bi-modal posterior densities in the two middle columns of $\tilde{\mathbf{A}}_0^{-1}$ in Figure 4, which is a direct consequence of the structural identification issues associated with the overlapping posterior intervals of $d_2(2)$ and $d_3(2)$ noted before.

By examining the last row of $\tilde{\mathbf{A}}_0^{-1}$ in Figure 4, we can single out a prima facie candidate for the "monetary policy shock" in the estimated model: r_t shows the strongest immediate reaction to the first shock in our system, while in all other cases the biggest initial reactions are associated with different variables.[19] The

[18] For example, the posterior medians of $d_1(2), \ldots, d_4(2)$ in Table 2 are all clearly mutually distinct.

[19] Since the identification in MS-SVAR models is based on statistical arguments with no prior restrictions on the system matrices and associated structural innovations, it is necessary to carefully study all available statistical evidence before attaching economic interpretations to the observed shocks. Any structural data interpretations should be based on the short-run impacts, estimated posterior impulse responses and prior theoretical and empirical knowledge

first shock raises the policy rate as a result of sign normalization, with an estimated immediate positive impact on prices and a slightly negative one on the real money balances, while the instantaneous reaction of real output remains ambiguous. In terms of the short-run impact, our provisional "monetary policy shock" is close to its sibling in Uhlig (2005), apart from the opposite reaction of π_t. More specifically, we also find that the initial real output response is ambiguous and the money supply tends to tighten on impact. The opposite reaction of prices, on the other hand, is likely to be explained by the imposed identification assumption in Uhlig (2005), while in our case the positive initial reaction of π_t appears to be a feature of the US data delivered by our identification methodology. We note that the immediate response of π_t to our provisional "monetary policy shock" in Figure 4 is supported predominantly on the positive half-line, which has strong implications for the sign identification methodology of Uhlig (2005): while the negative initial effects on both π_t and M_t from the monetary policy shock are not ruled out by our results, the posterior probability of this response mix for the US data in our empirical model is rather low.[20]

Similar discrepancies in terms of the immediate monetary policy impact on π_t are observed between the conventional recursive and non-recursive identifications of Christiano et al. (1999) and Gordon and Leeper (1994) on the one hand and our empirical results in Figure 4 on the other. It is common in the conventional approaches to allow the money supply to tighten on impact, while assuming a reaction lag of both real output and prices to a contractionary monetary policy shock. Our statistical identification approach, however, yields a strong positive reaction of prices on impact, taking us back to the classical "price puzzle" result of Sims (1992). In contrast to the more recent literature, Sims (1992) allows for an immediate response of real and nominal quantities to a monetary policy shock and documents a well–defined initial effect on prices that is common across several alternative datasets and model specifications. In light of our empirical analyses for the recent US macroeconomic data, further investigation into the strong positive response of π_t to a contractionary monetary policy shock and its consequences for the conventional identification in the empirical SVAR literature may be warranted.[21]

about the particular sample, all carefully weighted and examined simultaneously. On its own, the strongest initial reaction of r_t to the first observed shock may not be sufficient to justify the labelling; our interpretation of this shock also considers the corresponding posterior impulse responses, which are discussed later in this section.

[20] In a tri-variate monetary SVAR, estimated without the real money balances and identified using sign restrictions, Castelnuovo and Surico (2010) relax the assumption of non-positive response of prices to a monetary policy shock and find strong evidence in favour of the "price puzzle". However, they claim that this finding is limited to the pre-1980 US sample, before the start of the Volcker reserve targeting regime; see also Boivin and Giannoni (2006).

[21] A possible explanation of this empirical result, proposed by Sims (1992), suggests a reaction of the monetary authority to an anticipated inflationary pressure, creating a statistical illusion of a causal link between the monetary policy shocks and subsequent price increases. Although the MS-SVAR framework, presented and empirically illustrated in this paper, al-

The last column of $\tilde{\mathbf{A}}_0^{-1}$ in Figure 4 reveals the immediate impact of the fourth shock on the US macroeconomic variables in our empirical analysis. As shown in Table 2, the last diagonal element of $\tilde{\mathbf{D}}(2)$ that corresponds to this shock is well separated from all the other elements of the matrix, resulting in clear and regularly-shaped posterior densities in the last column of $\tilde{\mathbf{A}}_0^{-1}$. The observed shock raises the real money balances because of the sign normalization and shows a positive impact on the real output, a negative impact on the policy rate, and an ambiguous price effect. We attach a provisional "money multiplier shock" label to this response profile, because its nature bears a strong resemblance to the available empirical evidence in Favara and Giordani (2009) and Peersman and Wagner (2014).

Up to this point we have provisionally identified two out of four structural shocks in our empirical analyses. Before the application of the clustering procedure, the structural model identification issues are visible in the case of two remaining shocks in Figures 4 and 5. As explained earlier in this section when discussing the results in Table 2, this happens due to the randomly permuted order of the two middle columns of $\tilde{\mathbf{A}}_0^{-1}$ when the Gibbs sampler posterior draws are generated by our algorithm. As manifested by wide bi-modal posterior densities in the middle part of Figure 4 and completely uninformative before the clustering posterior impulse responses in the second and fourth columns of Figure 5, the underlying structural nature of the remaining two shocks in our empirical model remains shrouded prior to the full resolution of the identification issues. As suggested by our theoretical results in Section 2, all necessary statistical information needed to group the structural shocks according to their specific economic footprint in the model can be found in the posterior simulations of $\tilde{\mathbf{A}}_0^{-1}$ and the reduced-form parameters $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$. We now carry out the Bayesian clustering procedure, described in detail in Subsection 3.2 of the paper, on the two hitherto unidentified shocks in our empirical model in order to fully resolve the structural dynamics in the US macroeconomic data.

After the clustering, the structural nature of the two previously unidentified shocks clearly emerges in Figures 4 and 5. In particular, after the clustering the posterior impulse responses in the middle part of Figure 5 are sufficiently sharp to confidently discern the statistical effects of the two shocks on real output and prices. Judging by the median responses of y_t and π_t, the third column is consistent with a positive "aggregate demand shock", pushing the real output and prices in the same direction and inducing a pronounced, if somewhat delayed, counteraction from the monetary policy authority. Similarly, the fifth column of Figure 5 conforms to a positive "aggregate supply shock", since the median responses of the real output and prices go in opposite directions. Our empirical results indicate that the Federal Reserve may choose not to react to the aggregate supply shock, as suggested by the wide 68% posterior credible sets around the response profile of the policy rate r_t.

lows us to provide strong statistical evidence in favour of the "price puzzle", it is not intended to test the Sims' hypothesis.

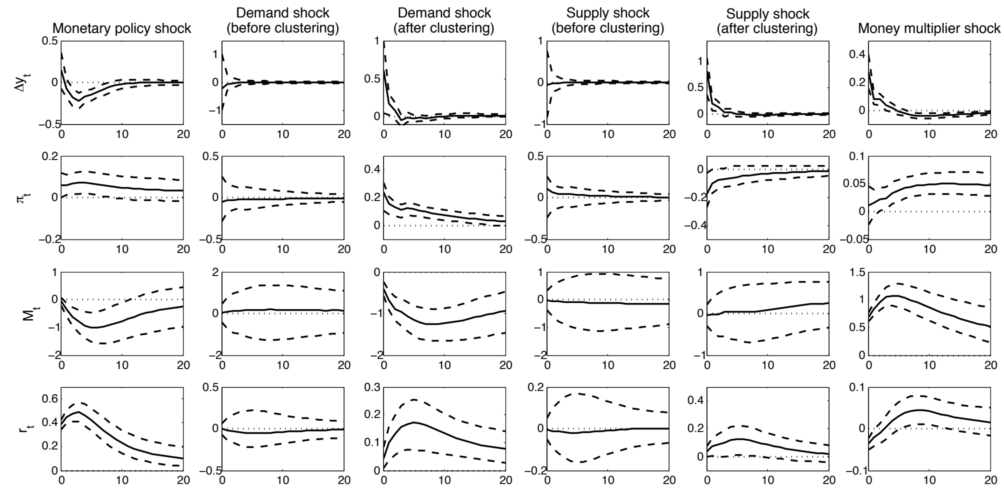

Figure 5: Posterior impulse response functions and 68% credible sets for the structural model

Returning to the other two shocks, provisionally identified using the posterior distribution of $\tilde{\mathbf{A}}_0^{-1}$ in Figure 4, we attempt to cross-validate a "monetary policy shock" and a "money multiplier shock" against the results provided by the estimated posterior impulse responses in our structural model. In the first column of Figure 5, the response profile consistent with a contractionary monetary policy shock emerges: the policy rate increase leads to a reduction in the real money balances, a drop in real output, and the "price puzzle" reaction of the inflation noted earlier.[22] This reaction would be typical for a monetary policy shock considered in earlier studies by Sims (1992), Christiano et al. (1999), Sims and Zha (2006) and many others. The last column of Figure 5 displays the estimated posterior response profiles to the prima facie "money multiplier shock": an increase in the real money balances is associated with an initial drop of the short-term policy rate, which then strongly rebounds as the Federal Reserve reserve reacts to a boom in the real output and a strong and persistent inflationary response of prices. This response profile corroborates the recent finding in Favara and Giordani (2009) and lends further support for the hypothesis that, contrary to the standard New Keynesian framework, the money multiplier shock, which may be linked to the widening of the term spreads, increasing stock returns or an exchange rate deprecation, leads to a well-pronounced response of the real macroeconomic aggregates.[23]

[22]The "price puzzle" in our posterior impulse responses is robust to alternative model specifications. In particular, it remains virtually unchanged if we replace the real output growth with its de-trended version, as suggested in Giordani (2004). Adding a global commodity price index to the list of model variables makes the initial response of inflation to our "monetary policy shock" smaller, so its 68% posterior credible set now contains zero, but it still remains sufficiently large for the subsequent two to four quarters to be taken as credible evidence in favour of the "price puzzle".

[23]Andrés, López-Salido and Nelson (2009) re-visit the role of money in the New Keynesian

In summary, we estimate a Bayesian four-variate two-state MS-SVAR model, the econometrics of which is covered in Sections 2 and 3 of the paper, using a quarterly sample of the US macroeconomic data over the last 45 years. In the initial reduced-form model selection and estimation phase of this empirical application, we uncover a rather familiar constellation of high and low-volatility states spanning the last half century of US macroeconomic history. But most importantly, the switching volatility regimes in the data provide us with sufficient statistical information to fully identify the short-run impacts and impulse responses to four structural shocks, which are found to be consistent with a conventional monetary policy shock, a money multiplier shock and a pair of aggregate supply-demand shocks. Once again, we want to emphasise that the economic interpretations of these shocks emerge from the data by looking at it through the lens of the MS-SVAR model and without the need for any conventional *a priori* identifying restrictions. The first two identified shocks corroborate many previous findings on the monetary policy and aggregate money shocks. In particular, we report a robust and well-pronounced "price puzzle" in response to our "monetary policy shock", while our "money multiplier shock" induces a strong response of the real output and prices in the US data. In order to fully identify the remaining two shocks, we carry out the novel Bayesian clustering procedure applied to the set of estimated posterior impulse responses from the Gibbs sampler. The new clustering method delivers enough statistical evidence for us to confidently discern an "aggregate supply shock" and an "aggregate demand shock" from the data. Finally, we point out that the results of our statistical shock identification procedure are not always compatible with the conventional *a priori* short–run and sign identification approaches used in much of the recent empirical SVAR research.

5 Conclusion

Structural vector autoregressions are popular tools in modern macroeconomics, used as empirical benchmarks for deconstructing and understanding the nature of business cycles and for facilitating the development of new theories. Since their widespread adoption in the early 1980s, the on-going theoretical and empirical research has focused on the most appropriate identification strategies for uncovering the structural macroeconomic dynamics from the real-world data.

In this paper we contribute to the literature on structural shock identification by developing a Bayesian estimator for the vector autoregression with a time-varying volatility of error terms that depend on a hidden Markov process. We refer to this model as a "Markov-switching structural vector autoregression"

framework, finding that certain variations of these structural macroeconomic models imply a forward-looking character of the real money balances, helping to predict future variations in the natural interest rate and other real aggregates.

(MS-SVAR). Given sufficient statistical information in the data, the multiple volatility states of the structural innovations in the MS-SVAR model allow the full identification of all structural matrices and impulse responses without the need for conventional *a priori* parameter restrictions. Building on the existing results in Lanne et al. (2010), the first part of the paper contains a new proposition on the structural-form identification in the class of MS-SVAR models with an arbitrary number of volatility states. In particular, we derive necessary conditions on the reduced-form variance-covariance matrices that guarantee model identification up to an arbitrary permutation order of the structural shocks. We then propose a novel approach for pinning down the specific shock order from the estimated reduced-form parameters by applying a computationally-intensive Bayesian clustering method from the statistical literature.

The new methodology is validated using the US macroeconomic data series, where the nature of different shocks is empirically examined. In particular, we identify the short-run impacts and impulse responses of four structural shocks, which we label a "monetary policy shock", a "money multiplier shock", an "aggregate demand shock" and an "aggregate supply shock". We find a robust and well-pronounced "price puzzle" in response to a "monetary policy shock", while the "money multiplier shock" induces a strong reaction in US real output and prices. The full disentanglement of the "aggregate demand shock" and the "aggregate supply shock" in the US data requires an application of the new clustering method and illustrates its empirical success in this particular data sample.

We also point out that the results of our statistical shock identification procedure are not always compatible with conventional *a priori* short-run and sign identification restrictions used in many of the recent empirical SVAR models, which warrants further careful validation and checking of the existing results using the new MS-SVAR methodology and other alternative approaches.

A Computational appendix

The performance of the Gibbs sampler in our application is assessed in Table A1. Three types of diagnostics are shown: the proposed number of Gibbs sampler draws and the dependence factor (DF) according to Raftery and Lewis (1992), autocorrelation of the draws at the first lag and the Geweke (1992) z-statistic. The statistics are available for each parameter separately, but to conserve space only summarized statistics are shown. Unsurprisingly, the VAR(2) model is more efficient than the model with Markov switching. For the Markov switching model the number of iterations proposed by the Raftery and Lewis (1992) diagnostics is reasonable and the dependence factor indicates low correlation of the draws. Low autocorrelation of the Gibbs draws is further indicated by the estimated autocorrelations at lag one. Autocorrelation decays quickly with increasing time lags. The Geweke (1992) test evaluates the stationarity of the Gibbs draws by comparing the mean of the first 20% of the draws with the last 20% of the draws. The values outside ± 2 indicate the drifting mean of the series. For our Gibbs sampler there may be some drift in individual parameters but the median value indicates no problem. Hence the convergence of the sampler for the US data is satisfactory and further inference can be made.

Table A1: Gibbs sampler diagnostics for the estimated reduced–form model

	Raftery-Lewis		*Autocorrelation at lag 1*			*Geweke z-statistic*		
	N_{draws}	DF	Median	Min	Max	Median	Min	Max
$m=1$, $p=2$	3701	0.99	-0.00	-0.04	0.04	-0.05	-2.85	2.16
$m=2$, $p=4$	3839	1.03	0.06	-0.01	0.24	0.62	-3.02	2.62

References

Amisano, G. and Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Andrés, J., López-Salido, J. D. and Nelson, E. (2009). Money and the natural rate of interest: Structural estimates for the united states and the euro area, *Journal of Economic Dynamics and Control* **33**(3): 758 – 776.

Bacchiocchi, E. and Fanelli, L. (2012). Identification in structural vector autoregressive models with structural change, *Departmental working papers 2012-16*, Universita degli studi di Milano.

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation, *Carnegie-rochester conference series on public policy*, Vol. 25, Elsevier, pp. 49–99.

Blanchard, O. J. and Watson, M. W. (1986). Are business cycles all alike?, *The American business cycle: Continuity and change*, University of Chicago Press, pp. 123–180.

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

Boivin, J. and Giannoni, M. P. (2006). Has monetary policy become more effective?, *The Review of Economics and Statistics* **88**(3): 445 – 462.

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

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrica* **81**(3): 541–553.

Castelnuovo, E. and Surico, P. (2010). Monetary policy, inflation expectations and the price puzzle, *Economic Journal* **120(549)**: 1262–1283.

Chib, S. (1995). Marginal likelihood from the gibbs output, *Journal of the American Statistical Association* **90**(432): 1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics* **75**: 79–97.

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

Cynamon, B. Z., Dutkowsky, D. H. and Jones, B. E. (2006). Redefining the monetary aggregates: a clean sweep, *Eastern Economic Journal* **32**(4): 661 – 673.

Davis, S. J. and Kahn, J. A. (2008). Interpreting the Great Moderation: changes in the volatility of economic activity at the macro and micro levels, *Journal of Economic Perspectives* **22**(4): 155–180.

Everitt, B. S., Landau, S., Leese, M. and Stahl, D. (2011). *Cluster Analysis*, John Wiley and Sons, Ltd.

Favara, G. and Giordani, P. (2009). Reconsidering the role of money for output, prices and interest rates, *Journal of Monetary Economics* **56**(3): 419 – 430.

Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: a critical review, *Journal of Economic Literature* **49**(4): 938–960.

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In J. M. Bernardo, J. O. Berger, A. P. Dawiv, and A. F. M. Smith, eds., Bayesian Statistics, volume 4*, 2nd edn, Clarendon Press, Oxford, UK.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model, *Journal of Applied Econometrics* **8**(Supplement): S19–S40.

Giordani, P. (2004). An alternative explanation of the price puzzle, *Journal of Monetary Economics* **51**(6): 1271 – 1296.

Gordon, D. and Leeper, E. (1994). The dynamic impacts of monetary policy: an exercise in tentative identification, *Journal of Political Economy* **102**(6): 1228–47.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, USA.

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

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*, 2nd edn, Cambridge University Press.

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

Kaufman, L. and Rousseeuw, P. (1987). *Clustering by means of Medoids. In Statistical Data Analysis Based on the L1âĂŞNorm and Related Methods, edited by Y. Dodge*, 2nd edn, North-Holland.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**: 1131–1149.

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

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. (2012). Identifying structural vector autoregressions via changes in volatility, *Discussion Papers 1259*, DIW Berlin.

Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**(3): 479–496.

Marin, J.-M., Mengersen, K. and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions, *in* D. Dey and C. Rao (eds), *Bayesian Thinking: Modeling and Computation*, Vol. 25 of *Handbook of Statistics*, Elsevier, pp. 459 – 507.

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

Peersman, G. and Wagner, W. (2014). Shocks to bank lending, risk-taking, securitization, and their role for U.S. business cycle fluctuations, *CESifo Working Paper No. 4701.*

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*, The MIT Press.

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

Raftery, A. E. and Lewis, S. M. (1992). One long run with diagnostics: Implementation strategies for markov chain monte carlo, *Statistical Science* **7**: 493–497.

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

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn, Springer.

Rothenberg, T. J. (1971). Identification in parametric models, *Econometrica* pp. 577–591.

Rubio-Ramírez, J. F., Waggoner, D. F. and Zha, T. (2005). Markov-switching structural vector autoregressions: Theory and applications, *Discussion Paper*, Federal Reserve Bank of Atlanta.

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

Sims, C. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy, *European Economic Review* **36**(5): 975–1000.

Sims, C. A. (1986). Are forecasting models usable for policy analysis?, *Federal Reserve Bank of Minneapolis Quarterly Review* **10**(1): 2–16.

Sims, C. A., Waggoner, D. F. and Zha, T. (2008). Methods for inference in large multiple-equation markov-switching models, *Journal of Econometrics* **146**(2): 255–274.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**(1): 54–81.

Stock, J. H. and Watson, M. W. (2002). Has the business cycle changed and why?, *in* M. Gertler and K. Rogoff (eds), *NBER Macroeconomics Annual 2002*, Vol. 17, National Bureau of Economic Research, pp. 159–218.

Theil, H. and Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics, *International Economic Review* **2**: 65–78.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**(2): 381–419.

Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms, *Neural Networks, IEEE Transactions on* **16**(3): 645–678.