MARKOV-SWITCHING STRUCTURAL VECTOR AUTOREGRESSIONS: THEORY AND APPLICATION

JUAN F. RUBIO-RAMIREZ, DANIEL WAGGONER, AND TAO ZHA

ABSTRACT. This paper develops a new and easily implementable necessary and sufficient condition for the exact identification of a Markov-switching SVAR model. The theorem applies to models with both linear and some nonlinear restrictions on the structural parameters. We also derive efficient MCMC algorithms to implement sign and long-run restrictions in Markov-switching SVARs. Using our methods, four well-known identification schemes are used to study whether monetary policy has changed in the euro area since the introduction of the European Monetary Union. We find that models restricted to only time-varying shock variances dominate the other models. We find a persistent post-1993 regime that is associated with low volatility of shocks to output, prices, and interest rates. Finally, the output effects of monetary policy shocks are small and uncertain across regimes and models. These results are robust to the four identification schemes studied in this paper.

I. INTRODUCTION

A recent debate on whether it is bad monetary policy or bad luck that explains the U.S. inflation-unemployment dynamics in the 1970s has motivated a number of empirical works. Boivin (1999), Clarida, Gali, and Gertler (2000), Lubik and Schorfheide (2004)), and Boivin and Giannoni (2005), on the one hand, find that bad monetary policy is the main explanation for the volatile and high inflation of the 1970s. Primiceri (2005), Sargent, Williams, and Zha (2005), Bernanke and Mihov (1998), Cogley and Sargent (2005), and Canova and Gambetti (2004), on the other hand, find little evidence in favor of the view that the monetary policy rule has changed drastically.

In order to shed some light on the debate Sims and Zha (2005) extend the seminal work of Hamilton (1989) and use Markov-switching structural vector autoregressions (SVARs) to disentangle between the two possible explanations. Sims and Zha (2005) develop novel and efficient Markov-Chain Monte Carlo (MCMC) methods for Markov-switching SVARs identified with linear restrictions on each structural equation. Their methods, however, cannot be applied directly to models identified in other ways. In particular, long-run restrictions on impulse responses, as introduced by Blanchard and Quah (1993), impose restrictions on nonlinear functions of the sum of the structural coefficients. Sign restrictions on impulse response functions, as proposed by Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) imply nonlinear restrictions on the model parameters.

This paper extends the MCMC method of Sims and Zha (2005) to Markov-switching SVARs with short-run, long-run, and sign restrictions on impulse responses. We show that if the model with short-run and long-run restrictions is exactly identified, there exists a unique rotation of the parameter matrices under a recursive SVAR system that are mapped onto the structural parameters of the original model. We derive an efficient algorithm for finding such a rotation. For Markov-switching SVARs

---

Date: October 26, 2005.

Key words and phrases. Markov-switching; regime changes; volatility; identification.

We thank Fabio Canova, Jon Faust, Ellis Tallman, Harald Uhlig, and especially Jim Nason for helpful discussions and comments. Eric Wang provided excellent research assistance. We greatly acknowledge the technical support of parallel computing from the Computing College of Georgia Institute of Technology. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

with sign restrictions, we develop a more efficient version of Uhlig's (2005) algorithm. This development is particularly important, since the MCMC computation is in general quite time-consuming.

The SVAR literature often employs the widely used necessary condition of Rothenberg (1971) to ascertain if models are exactly identified.[1] Rothenberg's necessary condition is easy to implement by simply counting enough restrictions in total. We give examples of SVARs that satisfy this necessary condition, but are *not* identified. This motivates us to develop a new and easily implementable necessary and sufficient condition for exactly identified SVARs. Our necessary and sufficient conditions not only count the number of restrictions equation by equation, but also require the restrictions to follow a particular pattern. Our theorem applies to SVARs with both linear restrictions and some nonlinear restrictions on the parameters of each equation.

We apply our procedures to analyzing whether monetary policy and the volatility of euro area macroeconomic variables have changed since the introduction of the EMU. In the last decade, the observed volatility of aggregate euro area variables has decreased significantly. For example, from the late 1970s and early 1980s, inflation has decreased from about 10 percent to under 5 percent, output growth volatility has fallen while the average annual growth rate has remained unchanged, and short-term nominal interest rates and money growth have decreased, which are now at record lows. Is the reduction in observed euro area aggregate volatility a result of regime changes in monetary policy? Or does it simply reflect the decreasing magnitude of the shocks that impinge on the euro area economy? To answer these questions,we study Markov-switching SVARs with four different identification schemes: (1) the recursive identification of Christiano, Eichenbaum, and Evans (1996), (2) the non recursive approach of Gordon and Leeper (1994) and Sims and Zha (2005), (3) the identification with a combination of contemporaneous and long-run restrictions on impulse responses as introduced by Blanchard and Quah (1993) and Galí (1992), and (4) the identification governed by sign restrictions proposed by Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005).

Our approach to Markov-switching differs from that of Sims (1993), Uhlig (1997), Cogley and Sargent (2005), and Primiceri (2005). These papers allow parameters to drift in a continuous state space. We can approximate the continuous drift arbitrarily well by putting greater prior weight on the diagonal of the transition matrix. At the same time, our approach offers additional flexibility by allowing for a large, discrete jump in the SVAR coefficients.[2]

Our empirical results reveal that the source of time variation embedded in euro area aggregate variables can be attributed to changes in shock variances. According to the relative marginal likelihoods (i.e., posterior odds ratios), Markov-switching SVARs based solely on time-varying shock variances are strongly favored compared to models in which slope coefficients also change with the regime. This result is robust across the four identification schemes we study and consistent with the findings of Stock and Watson (1996, 2003), Primiceri (2005), Canova and Gambetti (2004), and Sims and Zha (2005) for the U.S. data. We also find a stable and persistent post-1993 regime. This regime is associated with low volatility of the shocks to output, prices, and the short-term interest rate. Finally, the real effects of monetary policy shocks are small, or at most uncertain, relative to other shocks. These results are robust across identifications and regimes.

The rest of the paper is organized as follows. Section II lays out the general framework. Section III describes our application and the four identifications we use in the paper. Section IV reports and explains our key findings. Section V concludes the paper. Detailed proofs of the theorems are provided in the appendices.

---

[1] Rothenberg (1971) also provides a sufficient condition, but his condition is not implementable for SVARs.

[2] Sims and Zha (2005) discuss these issues in greater detail.

II. The Framework

In this section we present a framework to analyze Markov-switching SVAR models. We take the Bayesian approach, but all the theoretical results apply to the classical framework as well. We begin by presenting the general form of Markov-switching SVARs studied in this paper. Next, we introduce a class of linear identifying restrictions on transformations of the structural parameters. We derive the necessary and sufficient conditions for the Markov-switching SVAR to be exactly identified and show how to generate Markov-chain Monte Carlo (MCMC) draws from the posterior distribution. Finally, we illustrate our methods with an example.

II.1. **The Structural Model.** Following Hamilton (1989) and Sims and Zha (2005), we consider Markov-switching SVARs of the following form

$$y_t' A_0(s_t) = \sum_{\ell=1}^{p} y_{t-\ell}' A_\ell(s_t) + z_t' C(s_t) + \varepsilon_t' \tag{1}$$

where p is the lag length, T is the sample size, y_t is an $n \times 1$ vector of endogenous variables, z_t is equal to one[3], and ε_t is an $n \times 1$ vector of structural shocks. The conditional distribution ε_t is normal with mean 0 and covariance matrix I_n (the $n \times n$ identity matrix). The value of s_t is an element of $\{1, \cdots, h\}$ and s_t evolves according to a Markov process with transition matrix $\Pi = (\pi_{i,j})_{1 \le i,j \le h}$, where $\pi_{i,j}$ is the probability that s_t equals i given that s_{t-1} is j. For $0 \le \ell \le p$ and $1 \le k \le h$, $A_\ell(k)$ is an $n \times n$ matrix of parameters. For $1 \le k \le h$, $C(k)$ is a $1 \times n$ vector of parameters. The initial conditions, $y_0, \cdots, y_{1-p}$, are taken as given.

Let

$$A_+'(k) = [A_1(k)', \ldots, A_p(k)', C(k)']$$

for $1 \le k \le h$ and

$$x_t' = [y_{t-1}', \ldots, y_{t-p}', z_t']$$

for $1 \le t \le T$. The model (1) can be compactly written as

$$y_t' A_0(s_t) = x_t' A_+(s_t) + \varepsilon_t'. \tag{2}$$

The parameters of the structural model are $(A_0(k), A_+(k))$ for $1 \le k \le h$. The reduced-form representation implied by the structural model (2) is

$$y_t' = x_t' B(s_t) + u_t'(s_t)$$

where

$$B(s_t) = A_+(s_t) A_0^{-1}(s_t),\ u_t'(s_t) = \varepsilon_t' A_0^{-1}(s_t), E\left(u_t(s_t) u_t'(s_t)\right) = \Sigma(s_t) = \left(A_0(s_t) A_0'(s_t)\right)^{-1}.$$

The parameters of the reduced-form model are $(B(k), \Sigma(k))$ for $1 \le k \le h$.

[3] It is straightforward to include other exogenous variables in our framework.

II.2. **Identifying Restrictions.** Without restrictions the structural system (2) is not identified. If P is an orthogonal matrix,[4] the reduced-form representation derived from $(A_0(k),A_+(k))$ and $(A_0(k)P,A_+(k)P)$ are identical and hence the structural models are observationally equivalent. Sims and Zha (2005) describe how to identify the model using linear restrictions on the contemporaneous parameter matrix $A_0(k)$ and develop Bayesian methods for simulating the posterior distribution of the structural parameters. This class of restrictions includes recursive schemes as described by Christiano, Eichenbaum, and Evans (1996) and non-recursive schemes as described by Gordon and Leeper (1994) and Sims and Zha (2005).

Two alternative identification schemes have also been widely used. Blanchard and Quah (1993) and Galí (1992) use both contemporaneous and long-run restrictions on impulse responses; Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) use sign restrictions on impulse responses. In this section we extend the results of Sims and Zha to the former class of restrictions; sign restrictions are of a different nature and will be analyzed later in Section III.6.

For SVARs with both short-run and long-run restrictions, the methods used in the existing literature typically involve a system of nonlinear equations to be solved in order to obtain the maximum likelihood estimates or the posterior estimates if a prior is used.[5] When simulating from the posterior distribution, however, solving a system of nonlinear equations (or minimizing a nonlinear function) for each posterior draw is time-consuming and practically infeasible if a large number of MCMC draws are required to get accurate results. In this subsection, we show, however, that contemporaneous and long-run restrictions on impulse responses can be represented as linear restrictions on transformations of the structural parameters. This innovation is the key to the efficient MCMC methods developed later in this paper.

A transformation $X(\cdot)$ of the structural parameters is defined as follows.

Definition 1. Let $X(\cdot)$ be a transformation from the set of structural parameters to the set of $m \times n$ matrices, with $n \leq m$, such that either

(1a) $X(A_0(k)C, A_+(k)C) = X(A_0(k), A_+(k))C$, for every invertible matrix C.

or

(1b) $X(A_0(k)C, A_+(k)C) = X(A_0(k), A_+(k))(C')^{-1}$, for every invertible matrix C.

Condition (1a) applies when the restrictions are on the structural parameters themselves. Condition (1b) applies when the restrictions are on the impulse responses. This case includes both restrictions at finite horizons and long-run restrictions. Almost all identifying restrictions used in the existing SVAR literature can be presented as linear restrictions on the columns of some transformation $X(A_0(k), A_+(k))$. In particular, for $1 \leq j \leq n$ there exist $q_j \times m$ matrices Q_j of rank q_j such that $(A_0(k), A_+(k))$ satisfy the restrictions if and only if:

$$Q_j X(A_0(k), A_+(k))\, e_j = 0 \tag{3}$$

where e_j is the j^{th} column of the $n \times n$ identity matrix.[6]

---

[4]By definition P is an orthogonal matrix if and only if $PP' = I$

[5]The 2SLS estimate, as used by Galí (1992), is an approximation to the maximum likelihood estimate. The accuracy of this approximation depends on how good the instruments are in the first stage of the estimation.

[6]In addition to condition (1a) or (1b), one needs $X(\cdot)$ to be of full rank with respect to the restrictions Q_i. The technical condition is that there exists a set of structural parameters $(A_0(k), A_+(k))$ such that

$$rank(Q_i X(A_0(k), A_+(k))) = rank(Q_i).$$

In the examples considered in this paper, since the image of $X(\cdot)$ is dense in the set of all $m \times n$ matrices, this condition will be satisfied. In general, since this condition needs to hold only for a single set of parameter values, one can simply test the ranks of several arbitrarily chosen parameter values.

The recursive and non recursive restrictions on the contemporaneous parameter matrix $A_0(k)$ as used in the literature can be defined as linear restrictions on the columns of

$$X(A_0(k), A_+(k)) = A_0(k). \tag{4}$$

Conditional on the k^{th} state, the contemporaneous impulse responses to the j^{th} shock correspond to the j^{th} column of $\left(A_0^{-1}(k)\right)'$. When the i^{th} variable of the structural model is in first difference, the long-run impulse response of the i^{th} variable to the j^{th} shock conditional on the k^{th} state is the element in the i^{th} row and j^{th} column of $L'(k)$ where

$$L(k) = \left(A_0(k) - \sum_{\ell=1}^{p} A_\ell(k)\right)^{-1}.$$

Thus, Definition 1 allows us to represent contemporaneous and long-run restrictions on impulse responses as linear restrictions on the columns of

$$X(A_0(k), A_+(k)) = \begin{bmatrix} \left(A_0^{-1}(k)\right)' \\ L'(k) \end{bmatrix}. \tag{5}$$

Clearly, transformations (4) and (5) belong to the class of transformation functions $X(\cdot)$ in Definition 1.

II.3. **Normalization.** Since the identifying restrictions given by (3) do not uniquely determine the sign of any equation, a sign normalization rule is needed. While the theory developed in this paper will work for any choice of sign normalization, a poor choice of sign normalization may distort statistical inference concerning impulse responses (see Waggoner and Zha 2003b for details). In our applications, we follow the likelihood-preserving normalization proposed by Waggoner and Zha (2003b).

For Markov-switching models, there is an additional type of normalization. Any permutation of the states will result in an observationally equivalent set of parameters. Intuitively, permuting the states can be thought of as an arbitrary renaming of the states, i.e., permuting the first and second states can be interpreted as renaming the first state as the second and vice versa. Since the names of the states clearly do not affect the properties of the model, there will be observationally equivalent sets of parameters. We follow the Wald normalization as described in Hamilton, Waggoner, and Zha (2003), which minimizes the distance, in the appropriate metric, between the observationally equivalent parameter sets and some reference set of parameters, usually the maximum likelihood estimate. Since there are only a finite number of permutations, there are only a finite number of comparisons to make. The models with both sign and permutation normalizations are called normalized models. The Markov-switching SVAR models considered in this paper are normalized.

II.4. **Is the Model Exactly Identified?** A large part of the SVAR literature deals with exactly identified models. The precise definition of exact identification is given below.

Definition 2. A Markov-switching SVAR is exactly identified if and only if for almost every reduced-form parameter $(B(k), \Sigma(k))$ there exists a unique set of structural parameters $(A_0(k), A_+(k))$ with $B(k) = A_+(k)A_0^{-1}(k)$ and $\Sigma(k) = (A_0(k)A_0'(k))^{-1}$ that satisfies the identifying restrictions (3).

In an important article Rothenberg (1971) gives a necessary condition for exact identification, which requires $n(n-1)/2$ restrictions.[7] Except for a recursive system, however, the model may not be identified even if there are $n(n-1)/2$ linear

[7]Rothenberg (1971) also provides a sufficient condition, but his condition is not implementable for SVARs.

restrictions.[8] The following theorem gives us a new and easily implementable necessary and sufficient condition for a Markov-switching SVAR system to be exactly identified.[9]

Theorem 3. A Markov-switching SVAR is exactly identified if and only if there exists a permutation σ of $1,\cdots,n$ such that $rank\,(Q_i) = q_i = n - \sigma(i)$ for $1 \leq i \leq n$.

Proof. The proof is provided in Appendix A. □

Notice that we can always permute the equations in the original system, (1), so that $\sigma(i) = i$. Theorem 3 allows us to check if a Markov-switching SVAR is exactly identified. Rothenberg's (1971) necessary condition simply counts the total number of restrictions. Our necessary and sufficient condition not only counts the number of restrictions but also requires that they follow a certain pattern equation by equation. Any linear restrictions on $X(\cdot)$ allow for certain nonlinear restrictions on $A_0(k)$ and $A_+(k)$ themselves. Thus Theorem 3 applies to a wide range of identification schemes, including both linear and nonlinear restrictions on $A_0(k)$ and $A_+(k)$ as implied by (3).

Given restrictions on $X(\cdot)$ that exactly identify the model, how do we find the set of structural parameters such that the restrictions are satisfied? The following theorem tells us how to do it:

Theorem 4. A Markov-switching SVAR is exactly identified if and only if for almost every structural parameter $(A_0(k), A_+(k))$, there exists a unique orthogonal matrix $P(k)$ such that

$$(A_0(k)P(k), A_+(k)P(k))$$

satisfy the restrictions.

Proof. If $(A_0(k), A_+(k))$ and $(\tilde{A}_0(k), \tilde{A}_+(k))$ are two sets of structural parameters such that $(A_0(k)A_0'(k))^{-1} = (\tilde{A}_0(k)\tilde{A}_0'(k))^{-1}$, it follows that $(\tilde{A}_0^{-1}(k)A_0(k))(\tilde{A}_0^{-1}(k)A_0(k))'$ is an identity matrix, so $P(k) = \tilde{A}_0^{-1}(k)A_0(k)$ is an orthogonal matrix. □

Definition 2 gives the relationship between the reduced-form and the structural parameters that must hold in order for the model to be exactly identified. Theorem 4 gives the conditions for exact identification in terms of the structural parameters alone.

Theorem 4 is the key for an efficient MCMC algorithm for statistical inference and model comparison. If the model is exactly identified, one simply makes a posterior draw of the structural parameters in a recursive (triangular) system using the existing MCMC method (see Sims and Zha 1999 and 2005). Theorem 4 then guarantees the existence of an orthogonal matrix $P(k)$ that transforms this draw into a draw of the structural parameters that satisfy the restrictions given by (3).[10]

II.5. **An Algorithm to Find** $P(k)$**.** The bottleneck of the MCMC algorithm is to find the rotation matrix $P(k)$ for any posterior draw from a recursive system.[11] The following algorithm gives a step-by-step description of how to find this rotation efficiently. To simplify the notation, we assume, without loss of generality, that the equations in the original system have been permuted so that $rank\,(Q_i) = q_i = n - i$.[12]

---

[8]Examples will be shown later in this paper. See Sims and Zha (1999) for other examples.

[9]Of course, our necessary and sufficient condition also works for constant parameter SVARs.

[10]This procedure applies to the maximum likelihood estimation as well. We first obtain the maximum likelihood estimates of the parameters in a recursive system and then use $P(k)$ to rotate them to get the estimates of the structural parameters.

[11]Or equivalently any posterior draw from the reduced-form parameters that have been transformed to structural parameters via the Cholesky decomposition.

[12]This assumptions is equivalent to assuming that $\sigma(i) = i$ in Theorem 3

Algorithm 1. Let a Markov-switching SVAR be exactly identified and $(A_0(k), A_+(k))$ be any set of structural parameters drawn from a recursive system.

(Step 1) Set $i = 1$.

(Step 2) Form the matrix

$$\tilde{Q}_i(k) = \begin{bmatrix} Q_i X(A_0(k), A_+(k)) \\ p_1(k)' \\ \vdots \\ p_{i-1}(k)' \end{bmatrix}.$$

If $i = 1$, then $\tilde{Q}_i(k) = Q_i X(A_0(k), A_+(k))$.

(Step 3) Let $p_i(k)$ be any unit length vector such that $\tilde{Q}_i(k) X(A_0(k), A_+(k)) p_i(k) = 0$. Such a vector exists because $rank(Q_i) = n - i$ and hence $rank\left(\tilde{Q}_i\right) < n$. Use the LU decomposition of $\tilde{Q}_i(k)$ to find this vector.

(Step 4) If $i = n$ stop; otherwise, set $i = i + 1$ and go to step 2.

The above algorithm produces the orthogonal matrix

$$P(k) = [p_1(k), \cdots, p_n(k)]$$

that is required by Theorem 4. If the restrictions implied by matrices Q_j can be permuted to a triangular system,[13] the algorithm of finding $P(k)$ becomes even more efficient as one needs to use only a single QR decomposition (see Appendix B for details).

The restrictions given by (3) are more general than those considered by Sims and Zha (2005) and their method may not always be applicable. Algorithm 1 extends the Sims and Zha method. Suppose we wish to simulate from the posterior distribution for an exactly identified system with restrictions given by (3). We begin with a posterior draw of the model parameters in any exactly identified Markov-switching SVAR that the method of Sims and Zha (2005) can handle (for example, a recursive system). Denote this draw by $(A_0(k), A_+(k))$. We then use Algorithm 1 to find the rotation matrix $P(k)$ such that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the restrictions given by (3).

II.6. **An Example.** In this section we provide an example to illustrate how Theorem 4 and Algorithm 1 work in practice. To maximize the clarity of exposition, we consider a three-variable constant-parameter SVAR with one lag.[14] The three variables are output growth ($\Delta logY$), the interest rate (R), and inflation ($\Delta \log P$). For simplicity, we consider only a single lag so that $A_+ = A_1$. There are three identifying restrictions: aggregate demand (AD) shocks have no long-run effect on output, monetary policy (MP) shocks have no long-run effect on output, and MP shocks have no contemporaneous effect on output.

These impulse response restrictions can be expressed as the restrictions on the columns of $X(\cdot)$:

$$X(A_0, A_+) = \begin{bmatrix} \left(A_0^{-1}\right)' \\ (L)' \end{bmatrix} = \begin{matrix} \\ \Delta logY \\ R \\ logP \\ \Delta logY \\ R \\ logP \end{matrix} \begin{matrix} \begin{matrix} MP & AD & AS \end{matrix} \\ \begin{bmatrix} 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \end{matrix}$$

[13]A detailed description of such a permutation is given in Appendix B.

[14]The illustration can be easily extended to a Markov-switching SVAR.

where zeros indicate exclusion restrictions and $\times$ indicates no restrictions. It follows from Theorem 3 that this system is exactly identified. The importance of Theorem 3 lies in its checkable necessary and sufficient condition for determining whether the model is exactly identified. If, for example, we replaced the restriction that MP shocks have no long-run effect on output with the restriction that aggregate supply (AS) shocks have no contemporaneous effect on the interest rate, Theorem 3 would tell us that the model is not identified. Since this alternative identification scheme has three restrictions, a direct use of the necessary condition given by Rothenberg (1971) would lead to the incorrect conclusion that the model is exactly identified.

Returning to the original identification, we can write the matrices Q_j as

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } Q_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Since there is no restriction on the third column of $X(\cdot)$, there is no Q_3.

For the purpose of walking through Algorithm 1, suppose that reduced-form parameters B and Σ are

$$B = \begin{bmatrix} 0.5 & -1.25 & -1 \\ 0.5 & 0.25 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 1 & 0.5 & 1 \\ 0.5 & 4.25 & 2.5 \\ 1 & 2.5 & 3 \end{bmatrix}$$

To implement Algorithm 1 we need to first compute A_0 from A_+ implied from a recursive identification scheme and then rotate them so that restrictions defined by Q_j and $X(\cdot)$ hold. The Cholesky decomposition of Σ, $A_+ = BA_0^{-1}$, and $L = (A_0 - A_+)^{-1}$ implies that $(A_0^{-1})'$ and $(L)'$ are given by:

$$(A_0^{-1})' = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } (L)' = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus, $X(\cdot)$ equals:

$$X(A_0(k), A_+(k)) = \begin{bmatrix} \left(A_0^{-1}(k)\right)' \\ (L(k))' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

It follows from $X(\cdot)$ and Q_1 that

$$\tilde{Q}_1 = Q_1 X(A_0, A_+) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

The first step in Algorithm 1 is to find a unit length vector p_1 such that $\tilde{Q}_1 p_1 = 0$. The most computationally efficient method of doing this is to employ the LU decomposition of $\tilde{Q}_1(1)$. From a programming standpoint, however, a convenient method is to employ the QR decomposition of $\tilde{Q}_1'$.[15] Let $\tilde{Q}_1' = QR$ where Q is orthogonal and R is upper triangular. If we choose p_1 to be

[15] In Matlab, the function qr() applied to an $m \times n$ matrix will return an $m \times m$ orthogonal matrix and an $m \times n$ upper triagular matrix. In some applications where $m < n$, however, the "orthogonal" matrix will be $m \times n$ and the triangular matrix will be $n \times n$. In this case, one needs to pad the matrix $\tilde{Q}_i$ with a row of zeros before proceeding as usual.

the last row of Q, then

$$\tilde{Q}_1 p_1 = R' Q' p_1 = R' \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$$

since R' is lower triangular. Therefore we set p_1:

$$p_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

To obtain p_2, we form

$$\tilde{Q}_2 = \begin{bmatrix} Q_2 X(A_0, A_+) \\ p_1' \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

As before, take p_2 to be the last row of the orthogonal component of the QR decompostion of $\tilde{Q}_2'$ to get

$$p_2 = \begin{bmatrix} 0.7071 \\ -0.7071 \\ 0 \end{bmatrix}.$$

To obtain p_3, we form

$$\tilde{Q}_3 = \begin{bmatrix} p_1' \\ p_2' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0.7071 & -0.7071 & 0 \end{bmatrix}.$$

Again, take p_3 to be the last row of the orthogonal component of the QR decomposition of $\tilde{Q}_3'$ to get

$$p_3 = \begin{bmatrix} -0.7071 \\ -0.7071 \\ 0 \end{bmatrix}.$$

Combining these steps, one obtains the orthogonal matrix

$$P = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} = \begin{bmatrix} 0 & 0.7071 & -0.7071 \\ 0 & -0.7071 & -0.7071 \\ 1 & 0 & 0 \end{bmatrix}.$$

It is straightforward to verify that

$$X(A_0 P, A_+ P) = X(A_0, A_+) P$$

satisfies the restrictions. In Appendix B we show that the restrictions defined in this example can be permuted into a triangular system. We then show how a faster algorithm using a single QR decomposition can be applied to this example.

III. The Application

In this section, we apply our methods to answering the question of whether monetary policy in the euro area has changed since the introduction of the European Monetary Union (EMU) using post-1970 euro-area data.

The process toward forming the EMU was initiated more than 25 years ago. In March 1979 the European Monetary System (EMS) was established with the objectives of reducing inflation and preparing for monetary integration. Ten years later the Delors Report set out a plan to introduce the EMU over three stages. The first stage was launched in 1990 to increase cooperation among central banks in the euro area. In January 1994 the second stage began with the establishment of the European Monetary

Institute (EMI) as the forerunner to the European Central Bank (ECB). The third and final stage began in January 1999 when the euro became the single currency for the member states of the euro area and a single monetary policy was introduced under the authority of the ECB.

In the last decade we have observed that annual inflation has been under 5 percent while it was well above 10 percent in the late 1970s and early 1980s, the volatility of output has decreased while its average annual growth rate has remained the same, and both short-term nominal interest rates and money growth have been at a record low. Figure 1 displays these facts.

The coincidence of both events (an introduction of the EMU and the lower volatility of prices and output) motivates us to ask the following questions: Is the decrease in volatility linked to (1) changes in monetary policy in the euro area or (2) changes in the magnitude of shocks hitting the economy?

A researcher giving an affirmative answer to the first question could argue that monetary policy has been better in the euro area since the early 1990s. A researcher giving a positive answer to the second question could maintain the hypothesis that shocks hitting the euro area have been less volatile in the last decade. Our methodology allows us to distinguish a model with time-varying shocks only and a model with time-varying coefficients. The results will shed some light on the debate.

Some previous work has studied the effects of the EMU on monetary policy and macroeconomic volatility in the euro area. Peersman and Smets (2003) use a SVAR to conclude that the overall macroeconomic effects of monetary policy in the euro area have been stable over time. Ciccarelli and Rebucci (2003), however, find that the monetary transmission mechanism has changed since the late 1990s. Similarly, Angeloni and Ehrmann (2003) find evidence that the monetary transmission mechanism has become more potent and homogeneous across countries in the EMU, and De Bondt (2002) documents a quicker pass-through process since the introduction of the euro.

There are two potential shortcomings in the previous studies. First, most of them (except Ciccarelli and Rebucci, 2003) did not consider models with time-varying parameters. Instead they used pre- and post-EMU data by searching for a structural break. Second, none of these studies allowed for the time-varying volatility of shocks.

Dividing the sample into pre- and post-EMU data exacerbates the small sample problems. Moreover, the structural break analysis used to divide the sample rests on the unrealistic assumption that the probability of a regime change is either one or zero. An event as institutionally complicated as introducing the EMU may involve a number of transitional periods with uncertainty about the new monetary system. Equally important, allowing for heteroscedastic shocks in regime-switching models is crucial to eliminating the bias toward finding changes in the parameters (see Sims and Zha 2005).

The Markov-switching SVARs studied in this paper are suited to avoiding these shortcomings. The regimes are treated stochastically; the sample does not split because a large number of parameters remain constant across regimes; and heteroscedastic shocks are an integral part of the model.

At the same time, our methodology allows us to study a large class of identification schemes to check the robustness of our results. Specifically, we consider the four widely used identification schemes: (1) a recursive system as in Christiano, Eichenbaum, and Evans (1996), (2) a non recursive system as in Gordon and Leeper (1994) and Sims and Zha (2005), (3) a system with both contemporaneous and long-run restrictions on impulse responses as in Blanchard and Quah (1993) and Galí (1992), and (4) a system with sign restrictions on impulse responses as in Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005).

The first two schemes impose linear restrictions only on the columns of $A_0(k)$, and therefore, we can use the MCMC method of Sims and Zha (2005) directly. The third identification scheme belongs to a more general class of Markov-switching SVARs and the Sims and Zha method cannot be applied. Instead, we use the methods developed in Section II.2. A new MCMC method for the SVARs identified using sign restrictions is developed in Section III.6.

III.1. **Variation Across Regimes.** In our applications we consider three cases of time variation for the Markov-switching SVARs.[16] The first case is the constant-parameter SVAR, which is commonly used in the existing literature.

The second case allows $\Sigma(k)$ to vary but keeps the reduced-form coefficients $B(k)$ constant across regimes. For structural parameters in this case, $A_0(k)$ is allowed to vary but $A_+(k)$ must be of the form $A_+(k) = BA_0(k)$, where B is constant across regimes. We call this case the variance-only case. The exact restrictions needed to get the variance-only case are discussed in Appendix C.

The third case is the one where both $B(k)$ and $\Sigma(k)$ are allowed to differ across regimes. In general, the corresponding parameters $A_0(k)$ and $A_+(k)$ also change across regimes. If we let every parameter vary across regimes, we would have the over-parameterization problems associated with few degrees of freedom. Thus, we place restrictions on time variation in the manner that only part of $A_+(k)$ can differ across regimes. The nature of such restrictions is discussed in detail in Sims and Zha (2005) and, for completeness, also in Appendix C. We call this case the all-change case.

The comparison of these three types of time variation is important because it will allow us to determine whether the data reflect some structural changes and, if so, whether the regime change is due to the volatility of shocks or to the shift in parameter values.

III.2. **Data.** We use quarterly data form 1970:1 to 2003:4 from the Area-Wide Model (AWM) database released by the European Central Bank.[17] All the variables used in our applications are listed along with the variable symbols used by the AWM database. Output is real GDP in millions of euros with the base year 1995 (YER). The price level is the GDP deflator with the base year 1995=100 (YED). The measure of the money stock is M3 in millions of euros.[18] The nominal interest rate is the short-term interest rate (STN). The nominal exchange rate is euro/$ (EEN). Figure 1 reports the annualized quarterly percent changes for output, prices, M3, and the exchange rate. The interest rate is plotted as percent in level. Each identified model has five lags and includes a constant term. Five lags are included to control for possible seasonal effects that may not have been captured by seasonally adjusted data. In the rest of the section, we describe each of the four identification schemes in detail.

III.3. **CEE Identification.** Christiano, Eichenbaum, and Evans (1996) use a recursive identification scheme to identify monetary policy. We call this identification CEE. Under this identification, the contemporaneous matrix $A_0(k)$ is assumed to be triangular for $k = 1, ..., h$. In our application, we use five variables: log GDP (log Y), log GDP deflator (log P), the nominal short-term interest rate (R), log M3 (log M), and log nominal exchange rate (log Ex). We follow Christiano, Eichenbaum, and Evans and place the variables in the order of log Y, log P, log R, log M, and log Ex such that an output shock will affect output only, a shock to inflation will affect output and inflation, and so on.[19] Since this identification scheme imposes linear restrictions only on the columns of $A_0(k)$, we can use the method of Sims and Zha (2005) directly.

[16]The methods developed in this paper apply to many other types of time variations. See Sims and Zha (2005) for details.

[17]See Fagan, Henry, and Mestre (2004) for details.

[18]The M3 measure of the money stock is not included in the Area-wide Model (AWM) database. We obtained this variable from the reference series on monetary aggregates reported by the ECB.

[19]Appendix D describes in detail this identification scheme, using the notation of Section II.2.

III.4. **GLSZ Identification.** Gordon and Leeper (1994) and Sims and Zha (2005) propose an alternative identification scheme. We call this identification GLSZ. Their identification focuses on the interpretation of the structural equations themselves. In particular, they separate the monetary policy equation from the money demand equation and other non policy equations. The restrictions used to achieve this identification require the simultaneous (non recursive) relationships between financial variables such as the interest rate and money. The identification scheme is described in Table 1 where the same variables are used as in the CEE identification. An $\times$ in Table 1 means that the corresponding parameter in $A_0(k)$ for $k = 1,...,h$ is unrestricted, while zeros indicate exclusion restrictions. The monetary policy (MP) column in Table 1 represents the Federal Reserve contemporaneous behavior, the information (Inf) column describes the financial sector, the MD column corresponds to the money demand equation, and the block consisting of the last two columns represents the production sector (PS), whose variables are arbitrarily ordered to be upper triangular.[20] As in the CEE case, this identification scheme imposes linear restrictions only on the columns of $A_0(k)$, and therefore, the MCMC method of Sims and Zha (2005) can be applied directly.

III.5. **BGQ Identification.** Instead of the short-run restrictions discussed above, Blanchard and Quah (1993) use restrictions on the long-run impulse responses to achieve exact identification of an SVAR. When the system consists of more than two or three equations, we often do not have enough long-run restrictions that are economically justifiable to achieve exact identification. Galí (1992) suggests a combination of contemporaneous and long-run restrictions on impulse responses to get the SVAR identified. We call this identification schemeBGQ.[21]

Almost all SVARs with long-run restrictions use the variables in first difference. Following Peersman and Smets (2003), we consider a four-variable SVAR with three contemporaneous and three long-run restrictions on impulse responses. The four endogenous variables are quarterly output growth ($\Delta \log Y$), quarterly inflation (ΔP), the nominal short-term interest rate (R), and quarterly change of the nominal exchange rate euro/dollar ($\Delta \log Ex$). The contemporaneous restrictions are:

- Monetary policy shocks have no contemporaneous effect on output.
- Exchange rate shocks have no contemporaneous effect on output.
- Exchange rate shocks have no contemporaneous effect on the interest rate.

The long-run restrictions on impulse responses are:

- Aggregate demand shocks have no long-run effect on output.
- Monetary policy shocks have no long-run effect on output.
- Exchange rate shocks have no long-run effect on output.

Recall that $A_0^{-1}(k)$ and $L(k)$ represent the contemporaneous and long-run impulse responses, respectively. Thus, the above restrictions imply the following exclusion restrictions on $A_0^{-1}(k)$ and $L(k)$:

$$A_0^{-1}(k) = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & 0 & \times \end{bmatrix}, \quad L(k) = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}, \tag{6}$$

[20] Appendix D uses the notation in Section II.2 to describe this identification scheme in detail.

[21] Evans and Marshall (2004) use an SVAR model with long-run restrictions as a benchmark for their general equilibrium model.

where the symbol $\times$ means no restriction imposed and zero means an exclusion restriction.[22] It can be easily seen from Theorem 3 that the Markov-switching SVAR with the restrictions given by (6) is exactly identified. Using Theorem 3 to check whether the model is exactly identified should always be a first step.

To emphasize the importance of Theorem 3, consider that, instead of assuming that exchange rate shocks have no contemporaneous effect on output, we assume that demand shocks have no contemporaneous effect on output. This alternative identification scheme implies the following set of restrictions on $A_0^{-1}(k)$:

$$A_0^{-1}(k) = \begin{bmatrix} 0 & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ \times & \times & 0 & \times \end{bmatrix}.$$

In this case, Theorem 3 implies that the system would *not* be exactly identified. Since this alternative identification scheme imposes the same number of identification restrictions as the one described by (6), a direct use of the necessary condition given by Rothenberg (1971) would lead to the incorrect conclusion that this alternative identification scheme exactly identifies the model.

Because the identification (6) imposes the restrictions on $A_0^{-1}(k)$ and $L(k)$, the method of Sims and Zha (2005) no longer applies. We instead use the techniques developed in Section II.2 by drawing the parameters of a recursive system and then rotating each draw of these parameters to satisfy the contemporaneous and long-run restrictions.

III.6. **CDFU Identification.** The identification schemes described in Section II.2 are based on linear restrictions on transformations of the structural parameters. An objective in applying this class of restrictions is to identify monetary policy shocks. According to the conventional wisdom, a contractionary monetary policy shock should raise the interest rate and lower prices. Successful identification would produce impulse responses that conform to this conventional wisdom, but sometimes this class of identifying restrictions does not generate such responses. Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) propose an alternative approach. Their basic idea is to use sign restrictions directly on impulse responses to identify SVARs. In response to a contractionary monetary shock, for example, the interest rate should rise, while money and prices should fall. We call this identification scheme CDFU.

The methods developed in Section II.2 cannot be applied here, because a Markov-switching SVAR with sign restrictions on impulse responses is *not* exactly identified. According to Theorem 4, a necessary and sufficient condition for a Markov-switching SVAR to be exactly identified is that for any starting value of $(A_0(k), A_+(k))$ the unique $P(k)$ exists such that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the restrictions. In the case of sign restrictions, however, there exist a number of such $P(k)$'s.

To solve this problem, we develop an efficent MCMC algorithm, which can be viewed as a modified version of Uhlig's (2005) method.[23] We begin with any Markov-switching SVAR that is exactly identified and let $(A_0(k), A_+(k))$ be the model parameters. We search for an orthogonal matrix $P(k)$ such that the impulse responses implied by $(A_0(k)P(k), A_+(k)P(k))$ satisfy the sign restrictions. The main difference between Uhlig's approach and ours is one of efficiency and ease of implementation. Uhlig generates a random orthogonal matrix recursively column by column, while we use the following theorem to obtain a random orthogonal matrix using a single QR decomposition.

[22]See Appendix D for a detailed description of identification implied by matrices (6)

[23]Uhlig's (2005) method, together with the algorithms of Faust (1998) and Canova and De Nicoló (2002), is briefly described in Appendix E.

Theorem 5. Let X be an $n \times n$ random matrix with each element having an independent standard normal distribution. Let $X = QR$ be the QR decomposition of X with the diagonal of R normalized to be positive. Then Q has the uniform (or Haar) distribution.

Proof. The proof follows directly from Stewart (1980).[24] □

Theorem 5 gives us a convenient way of implementing a random selection of orthogonal matrices to obtain impulse responses that satisfy the sign restrictions. The following algorithm describes this implementation.

Algorithm 2. (Step 1) Let $(A_0(k), A_+(k))$ be a draw from the posterior distribution of any Markov-switching SVAR that is exactly identified.

(Step 2) Draw an independent standard normal $n \times n$ matrix X and let $X = QR$ be the QR decomposition of X with the diagonal of R normalized to be positive.

(Step 3) Let $P(k) = Q$ and generate impulse responses from $A_0(k)P(k)$ and $B(k)$.

(Step 4) If these impulse responses do not satisfy the sign restrictions, return to Step 3.

As a byproduct of this algorithm, $(A_0(k)P(k), A_+(k)P(k))$ is a posterior draw of the structural parameters for the Markov-switching SVAR with the sign restrictions.[25]

Our algorithm differs from Uhlig's (2005) method in two aspects: (1) all the posterior draws are kept in practice and (2) the orthogonal matrix is simply a draw from uniform (or Haar) distribution, whereas Uhlig (2005) searches for it recursively. These two differences make our algorithm more efficient, especially for an SVAR system of more than three or four variables.

We consider an SVAR model with the CDFU identification, using the same five variables as in the CEE and GLSZ cases. The sign restrictions are:

- In response to an expansionary monetary policy shock, the interest rate falls while money and prices rise.
- In response to a positive shock to money demand, both the interest rate and money increase.
- In response to a positive demand shock, both output and prices rise.
- In response to a positive supply shock, output rises but prices fall.
- In response to a positive external shock, the exchange rate devaluates and output increases.

All the sign restrictions hold for only two quarters. We begin with the CEE identification and use the Sims and Zha method to generate posterior draws of the model parameters. For each draw we use Algorithm 2 to rotate the draw such that the impulse responses satisfy the sign restrictions.

III.7. **A Comment on the Variance-Only Models.** The variance-only case has received considerable attention in the literature (Stock and Watson, 1996, 2003; Canova and Gambetti, 2004; Primiceri, 2005; Sims and Zha, 2005, for example). Sims and Zha (2005) develop the MCMC method for the variance-only SVARs with the CEE and GLSZ identifications. For the BGQ and CDFU schemes, our MCMC method begins with posterior draws of $(A_0(k), A_+(k))$ under a recursive system with $A_+(k) = BA_0(k)$, using the method of Sims and Zha (2005). For each draw, we use the algorithms developed in this paper

---

[24] Stewart (1980) has even more efficient algorithms for generating uniform random orthogonal matrices, but they are less straightforward and more difficult to implement.

[25] In theory the algorithm is not guaranteed to terminate. In practice, we set a maximum number of iterations to be 1000, in which steps (2) through (4) are repeated. If the maximum is reached, the algorithm should move to step (1) to draw another set of parameter values. In our application this maximum was never reached for millions of posterior draws.

to find the rotation matrix $P(k)$ so that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the restrictions. Is the resulting rotated draw of the variance-only type? Since $P(k)$ is orthogonal, we have that

$$B(k) = A_+(k)P(k)(A_0(k)P(k))^{-1} = A_+(k)(A_0(k))^{-1} = B.$$

and, thus, the answer is yes. This result is crucial because it allows us to consider variance-only cases under the BGQ and CDFU schemes.

IV. Empirical Results

In this section we use the methods described in Section II to estimate Markov-swtiching SVARs identified with the four identification schemes discussed in Section III. For each of the four identification schemes we report (1) marginal likelihoods of the models for the three different types of variation across regimes and different number of states, (2) the posterior probabilities of the estimated regimes, (3) changes in variances of reduced-form residuals across regimes, and (4) the impulse responses to a monetary shock for the best-fit model under each identification scheme.

IV.1. **Marginal Likelihood.**

Table 2 reports log marginal likelihoods for the three types of models under the CEE and GLSZ identifications: the constant-parameter model, the variance-only and all-change models with different numbers of states.[26] The symbol $*$ means that at least one of the states is redundant so that there is no posterior probability for the redundant state. In such an over-fitting situation, the standard error for the marginal likelihood is quite large but the marginal likelihoods are all below the marginal likelihood for the constant-parameter model.

For both identifications the variance-only model with 2 states has the highest marginal likelihood and thus is favored by the data. As can be seen in Table 2, the 2-state variance-only model outperforms all other models by the difference of at least 7 in log value for the CEE identification and more than 10 in log value for the GLSZ identification. The difference of 1 to 4 in log value means that the two models are competitive, but the difference of 7 or more implies strong evidence in favor of the model with a higher marginal likelihood. Therefore, the data clearly imply only two regimes in the euro area between 1970 and 2004 and supports the hypothesis that only the variance of the shocks, not the coefficients, vary across regimes.

We obtain similar results for the BGQ and CDFU identification schemes. Table 3 reports log marginal likelihoods for the same three models under the BGQ and CDFU identifications. The results for the CDFU identification are identical to those of the CEE identification because the model parameters with the CDFU identification are simply an orthogonal rotation of the model parameters with the CEE identification as discussed in Section III.6.

For the BGQ identification, the variance-only model with 3 states is favored by the data. All the variance-only models reported in Table 3 outperform the constant parameter model and the all-change model by the difference of at least 8 in log marginal likelihood. We interpret this result as strong evidence in favor of the variance-only specification. Within the set of

[26] All the marginal likelihoods reported in this paper are computed with a sequence of 6 million MCMC draws, which takes about 20 hours on a Pentium-IV PC. With 100 repeated runs of sequences from different starting points, the computed maximum of numerical standard errors for all marginal likelihoods is less than 0.7 in log value. Using the Newey-West (1987) approximation procedure, we obtain much smaller numerical standard errors. The marginal likelihood for the constant VAR model is computed using the algorithms described by Chib (1996) and Waggoner and Zha (2003a). The Matlab code can be downloaded from home.earthlink.net/ tzha02/programCode.html. Since the MCMC algorithm for the Markov-switching SVARs is not a Gibbs sampler, the marginal likelihoods for these models are computed with the modified harmonic means procedure discussed by Geweke (1999). We have also studied the models with other types of time variation. For example, we have let the coefficients in one or more structural equations (including the monetary policy equation) vary across regimes. Although the results are not reported in Table 2, the marginal likelihoods for all these models are substantially lower than those of the variance-only model with 2 states.

variance-only models, on the other hand, evidence in favor of three regimes is not as strong, since the differences among log marginal likelihoods are less than 2. Because of space limitation, we present only the 3-state variance-only model in the rest of this paper, although the other two variance-only models are equally good.

Overall, evidence from the four identification schemes uniformly supports regime changes in the euro area. More important, we find that regime change can be fully characterized by the variance of the shocks changing across regimes. This result is robust to different identifications.

IV.2. **Regimes.** In this section we analyze the estimated posterior probabilities of regimes for the best-fit model under each of the four identifications studied so far. Figures 2 and 3 display the posterior probability of each regime for the 2-state variance-only model with the CEE and GLSZ identifications. Figure 4 displays the posterior probability of each regime for the 3-state variance-only model with the BGQ identification.[27]

For both CEE and GLSZ identifications there is one state whose estimated probability is high and persistent after 1993. We call this state the EMS regime. The other state is called the non-EMS regime. There are two important features. First, the non-EMS regime is concentrated in the 1970-1980 period, while the EMS regime has high probability after 1993. Second, although the EMS regime remains with high probability for some years before 1993, it periodically switches to the non-EMS regime, probably reflecting periods of uncertainty about implementing the new monetary system (see Ungerer et al. 1990 for details).

For both identifications, we observe that the non-EMS regime has high posterior probability in periods of high output and inflation volatility, i.e., during both the 1970s and the turbulent period of 1992-93 (see Figure 1). On the other hand, the EMS regime has high posterior probability in times of low volatility of both output and inflation (i.e., the periods after 1993). Another feature of the data that distinguishes both regimes is the behavior of M3 growth and short-term interest rates. During the 1970s and a large part of the 1980s when the non-EMS regime is more prominent, both interest rates and money growth are well above 10 percent; since 1993, both variables are below 10 percent, and the EMS regime prevails in most of this later period.

Between 1980 and 1993 the two regimes switch several times, reflecting the uncertainty associated with the intuitional changes taking place in the euro area. The probability of the EMS regime, however, begins to increase after 1980, coinciding with the fact that inflation in the euro area declined sharply during the 1980s.

There are three regimes under the BGQ identification. The first regime, called the transitional EMS regime, occurs between the late 1970s and the early 1990s. The second regime, called the EMS regime, becomes dominant after 1993. The third regime, called the non-EMS regime, appears sporadically in the early 1970s. As is the case for the CEE and GLSZ identifications, the EMS regime is associated with low output and inflation volatility. On the other hand, the transitional EMS regime reflects more of the transition from high inflation in the late 1970s to the early 1990s. Finally, the non-EMS regime is related to several isolated events: the rapid increase of inflation during the early 1970s, the recession in 1975, and the impact of the 9/11 event.

In summary, an EMS regime is consistently estimated since the early 1990s for all the identification schemes. Since this regime is associated with low volatility of both output and inflation, the following section analyzes how variances of reduced-form shocks vary across regimes.

IV.3. **Shock Variances across Regimes.** For the 2-state variance-only model with the CEE identification, Table 4 reports the variance of a reduced-form shock to each variable under each of the two regimes, along with the relative variance across

[27] As discussed previously, the CDFU regimes are identical to the CEE regimes, and therefore, we report only the results for the CEE, GLSZ, and BGQ identifications.

regimes. The EMS regime is associated with lower volatility of all the reduced-form shocks. In particular, most of the fall in volatility is due to the fall in the variances of shocks to the interest rate and prices.

Table 5 reports the shock variances for the 2-state variance-only model with the GLSZ identification. As in the CEE case, the EMS regime is associated with a lower volatility of shocks to all the variables, and again, most of the fall in volatility occurs in shocks to the interest rate and prices.

Similar results hold for the shock variances in the 3-state variance-only model with the BGQ identification. As can be seen from Table 6, the variances of shocks to all the variables are smaller for the EMS regime than for the other regimes, and the larger reduction in variances is due to shocks to the interest rate and inflation.[28]

The EMS regime, a state that we robustly find after 1993, is the regime that has the lowest shock variances in all the macroeconomic variables. Our empirical results clearly support the hypothesis that the lower macroeconomic volatility observed in the euro area since the early 1990s is due to smaller shocks hitting the economy and, in particular, smaller shocks to the interest rate and inflation.

IV.4. **Impulse Responses.** In this section we analyze how the impulse responses to a contractionary monetary shock vary across regimes for the best-fit model under each identification. We also compare these impulse responses to those implied by the constant-parameter model. In all the figures discussed below, the posterior median estimates of impulse responses to a one-standard-deviation contractionary monetary shock are displayed, along with the error bands containing two-thirds of the posterior probability.

Figure 5 displays the impulse responses for the CEE identification. The first column corresponds to the impulse responses generated from the constant-parameter model. The second column corresponds to the EMS regime and the third to the non-EMS regime. All the columns have a similar pattern of dynamic responses. In response to a contractionary shock to monetary policy, the interest rate rises and money falls (liquidity effect), output falls (output effect), but the price level rises somewhat (price puzzle).[29] The increase in the price level, although statistically significant, is not economically important as compared to the responses of prices to other shocks. The impulse responses for the non-EMS regime are larger than those for the EMS regime, while the responses for the constant-parameter model are in between. The most important difference between the two regimes is the effect of a contractionary monetary policy shock on the short-term interest rate. Under the non-EMS regime the effect is larger and more lasting.

The above results are robust to the GLSZ identification, as reported in Figure 6. The impulse responses for the non-EMS regime (the first column in Figure 6) are larger than those for the EMS regime (the second column), while the responses for the constant-parameter model (the first column) are in between. In response to a contractionary shock to monetary policy, the interest rate rises and money falls (liquidity effect), output falls (output effect), and the price level falls or stays unchanged (no price puzzle). The main difference from the CEE results is that the uncertainty about the dynamic responses is larger. All the 68 percent posterior probability bands are wider than those in Figure 5. The wider bands imply that the price responses are statistically insignificant for both regimes.

Figure 7 displays the impulse responses for the BGQ identification. The posterior point estimates are very similar to the previous results under different identifications, but the uncertainty around the estimates is much bigger. The point estimates

---

[28] For the CDFU identification, the variances of reduced-form shocks are the same as in the CEE case, since an orthogonal transformation does not change these shock variances.

[29] The price puzzle still exists when we include commodity prices in the models or when we reorder the variables (for example, letting the interest rate respond to commodity prices or the exchange rate or both).

show the usual pattern of responses to a contractionary monetary policy shock, but the error bands seem unusually wide and ill-determined.[30] This exercise reinforces the importance of accurate error bands. Our results show that it would be misleading to rely on the point estimates alone.

The impulse responses for the CDFU identification are displayed in Figure 8. Again, the usual pattern of responses to a contractionary monetary policy holds and there is no price puzzle. The responses under the EMS regime are smaller than those under the non-EMS regime. The output effect is smaller and more uncertain for the 2-state variance-only model than the output effect for the constant-parameter model. The variance decomposition for the output effect is less than 20 percent for all three models. This result is consistent with Uhlig's (2005) finding of little evidence of output effect for the U.S.[31]

In summary, the output effects of monetary policy shocks on the euro area economy are small relative to the effects of other shocks and there is much uncertainty around these effects. This result seems robust across different identifications and across different regimes.

V. Conclusion

Long-run restrictions and sign restrictions on impulse responses have become popular tools in identifying different structural shocks in the data. Most models with these restrictions in the SVAR literature are exactly identified according to the widely used necessary condition of Rothenberg (1971). We show examples in which the necessary condition of Rothenberg (1971) is satisfied but the model is unidentified. This paper develops a new and implementable necessary and sufficient condition for Markov-switching SVARs to be exactly identified. This theorem is straightforward to use in practice.

We also develop new and efficient methods for implementing long-run and sign restrictions in Markov-switching SVARs. These methods are important for MCMC algorithms in which a long sequence of posterior draws is typically needed for the Markov-switching SVAR model.

We apply our methodology to the euro area data using four widely used identification schemes. Markov-switching SVARs based solely on time-varying shock variances are strongly favored compared to the rest of the time-varying specifications. A persistent regime is found after 1993. This regime is associated with low volatility of key macroeconomic variables such as output, prices, and the interest rate. The real effects of monetary policy shocks are small and uncertain across models and across regimes. All these results are robust to all the identifications we study in this paper.

Our methodology and results suggest some directions for future research. One direction is to use our methodology to study monetary transmission processes across countries in the euro area. Another is to build and estimate a DSGE with time-varying parameters and variances where Markov-switching SVARs can be used as benchmark models for model comparison.[32]

---

[30]The error bands reported by Peersman and Smets (2003) are much better behaved. Note that they have a different sample period and their bands are generated by only 100 draws. We find that this particular identification is quite fragile. For example, if the data for 2003 were excluded, the characteristics of the estimated impulse responses would be completely different. This finding is consistent with that of Faust and Leeper (1997).

[31]Uhlig (2005) uses a different set of variables, however. He identifies only monetary policy shocks while we identify five different shocks. But our results do not change much when we restrict our identification to monetary policy shocks only.

[32]See Fernández-Villaverde and Rubio-Ramírez (2005) for some details.

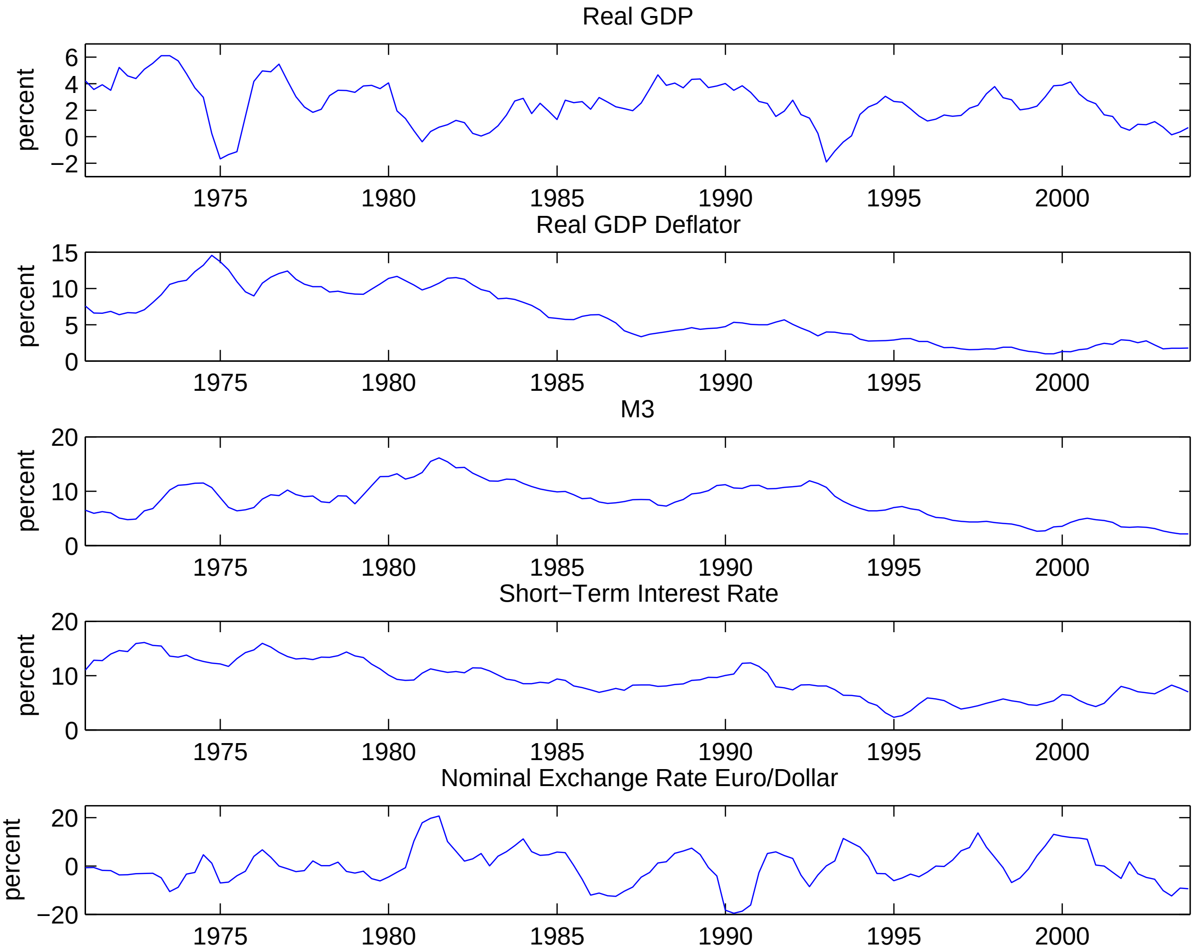

FIGURE 1. Quarterly Percent Change for Real GDP, Real GDP Deflator, M3, Short-Term Interest Rates, and Nominal Exchange Rate euro/Dollar from 1970:1 to 2003:4

	MD	Inf	MP	PS	PS
Y	$\times$	$\times$	0	$\times$	$\times$
P	$\times$	$\times$	0	$\times$	0
R	$\times$	$\times$	$\times$	0	0
M	$\times$	$\times$	$\times$	0	0
Ex	0	$\times$	0	0	0

TABLE 1. Identification Scheme for GLSZ.

	CEE		GLSZ	
Constant	2271.00		2273.60	
States	Variance-only	All-change	Variance-only	All-change
2	2283.83	2257.49	2291.08	2264.18
3	2277.91	*	2280.47	*
4	2274.93	*	2275.96	*

TABLE 2. Marginal log likelihoods for the three types of across regime variation for different number of states under the CEE and the GLSZ identification schemes.

	BGQ		CDFU	
Constant	1697.10		2271.00	
States	Variance-only	All-change	Variance-only	All-change
2	1731.57	1723.71	2283.83	2257.49
3	1733.00	*	2277.91	*
4	1731.77	*	2274.93	*

TABLE 3. Marginal log likelihoods for the three types of across regime variation for different number of states under the BGQ and the CDFU identification schemes.

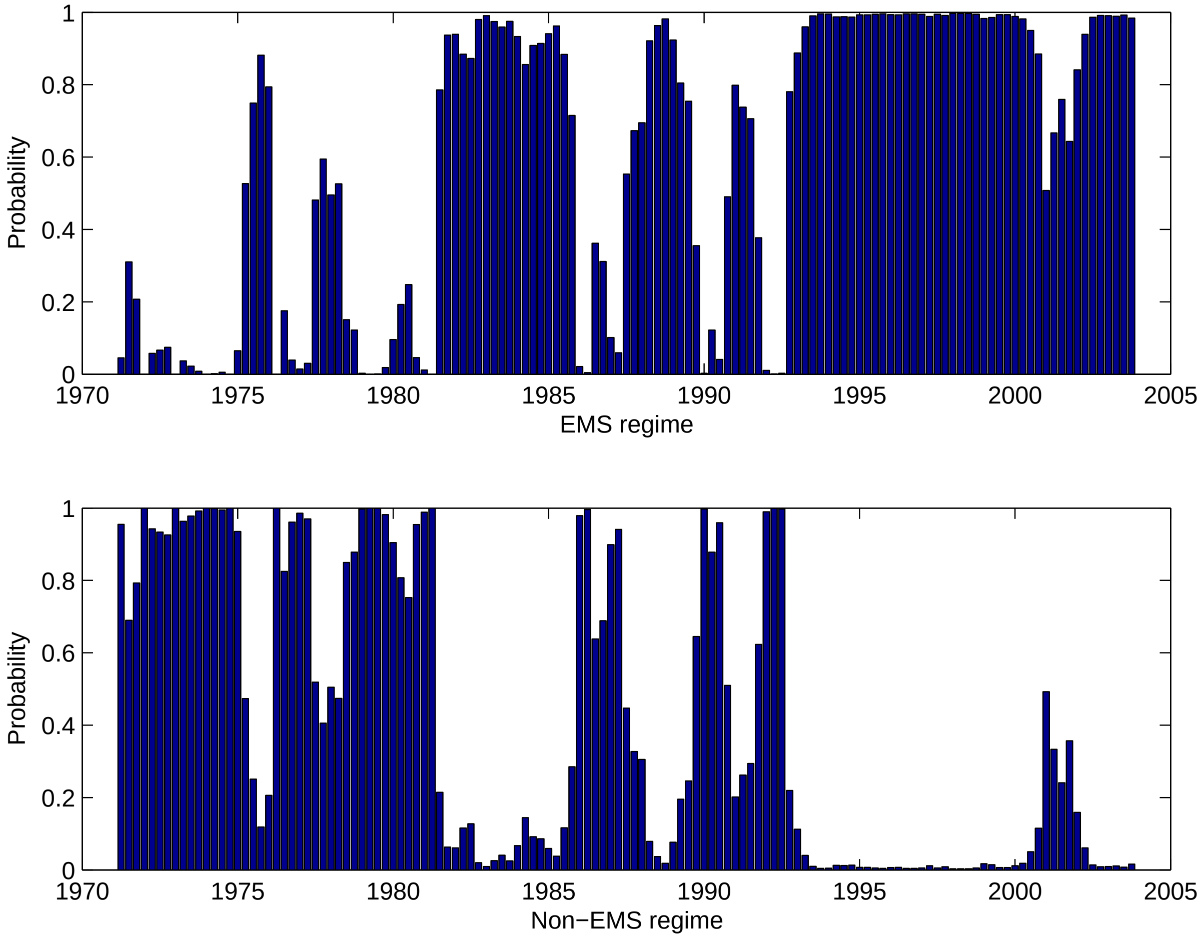

FIGURE 2. Posterior probabilities of states for the 2-state variance-only specification model under the CEE identification.

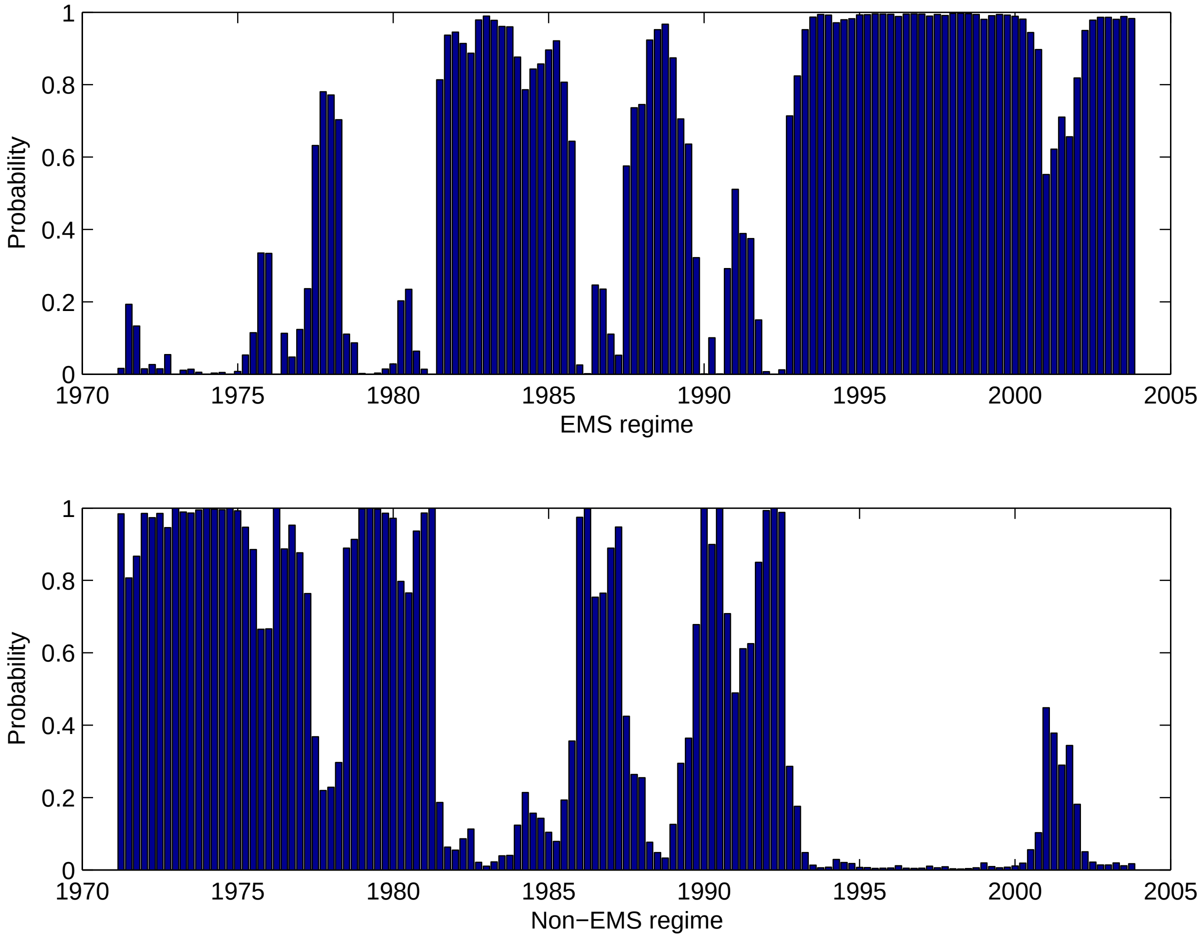

FIGURE 3. Posterior probabilities of states for the 2-state variance-only specification model under the GLSZ identification.

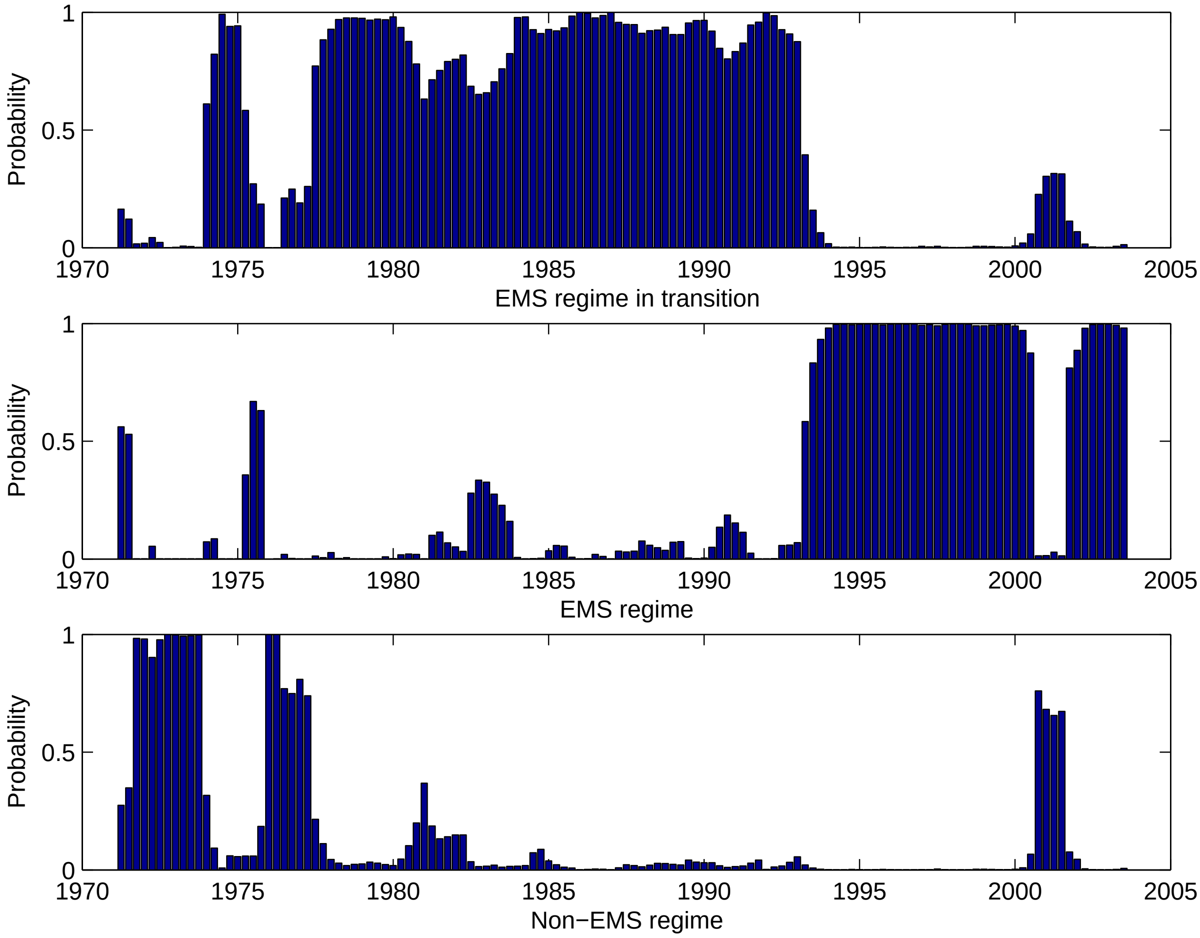

FIGURE 4. Posterior probabilities of states for the 3-state variance-only specification model under the BGQ identification.

Variables	EMS	Non-EMS	Relative volatility
Y	0.014E-03	0.042E-03	3.039
P	0.003E-03	0.016E-03	4.235
R	0.015E-03	0.064E-03	4.063
M	0.015E-03	0.028E-03	1.854
Ex	0.553E-03	1.087E-03	1.963

TABLE 4. Residual variance of the shocks for the 2-state variance-only model under the CEE scheme.

Variables	EMS	Non-EMS	Relative volatility
Y	0.013E-03	0.039E-03	2.797
P	0.004E-03	0.014E-03	3.459
R	0.018E-03	0.052E-03	2.755
M	0.012E-03	0.028E-03	2.340
Ex	0.551E-03	1.200E-03	2.175

TABLE 5. Residual variance of the shocks for the 2-state variance-only model under the GLSZ scheme.

Variables	Early EMS	EMS	Non-EMS
ΔY	0.056E-03	0.013E-03	0.028E-03
ΔP	0.009E-03	0.003E-03	0.064E-03
R	0.053E-03	0.013E-03	0.081E-03
ΔEx	1.216E-03	0.663E-03	1.439E-03

TABLE 6. Residual variance of the shocks for the 3-state variance-only model under the BGQ scheme.

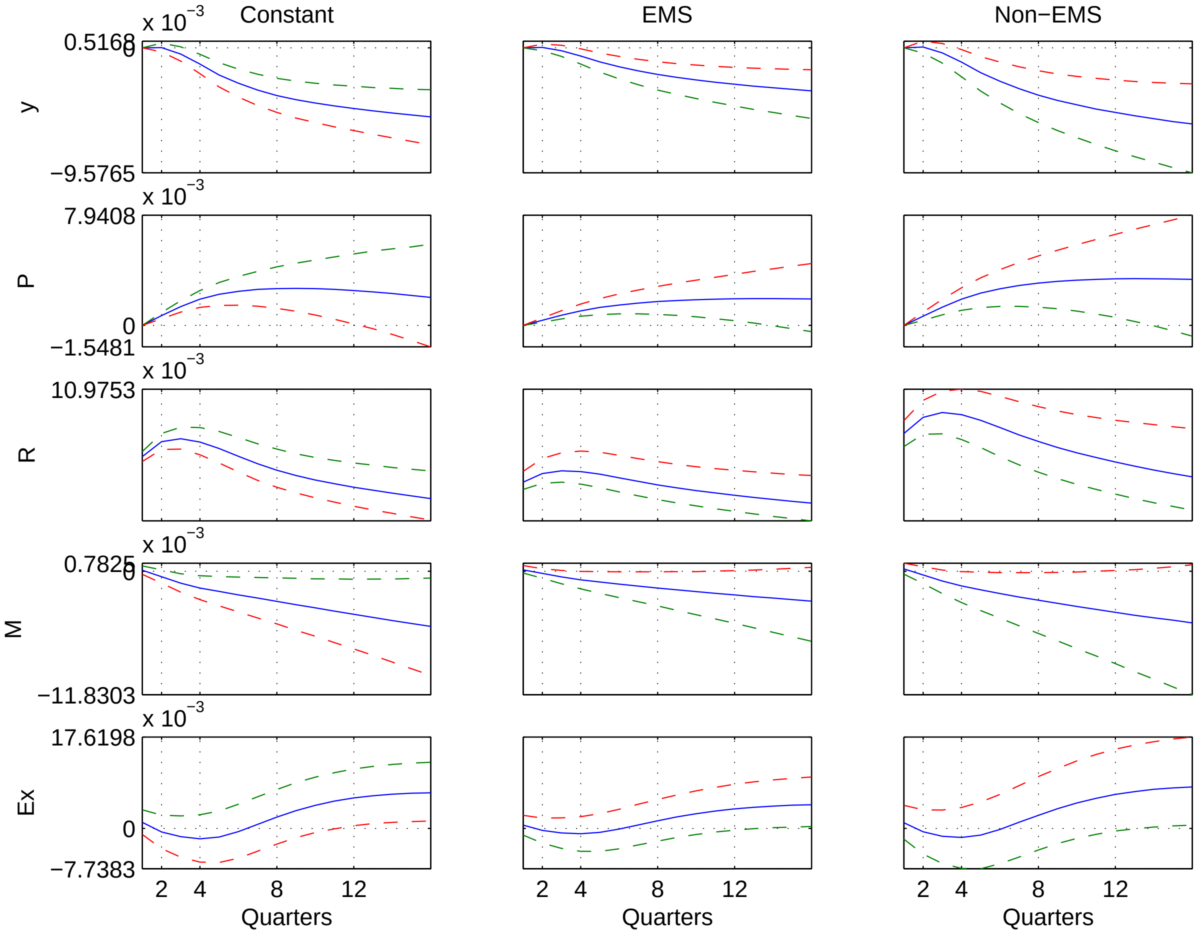

FIGURE 5. Impulse responses to a one-standard-deviation monetary policy shock under the CEE identification scheme. The solid line represents the posterior median estimate and the two dashed lines contain the 68 percent probability based on 500,000 MCMC draws.

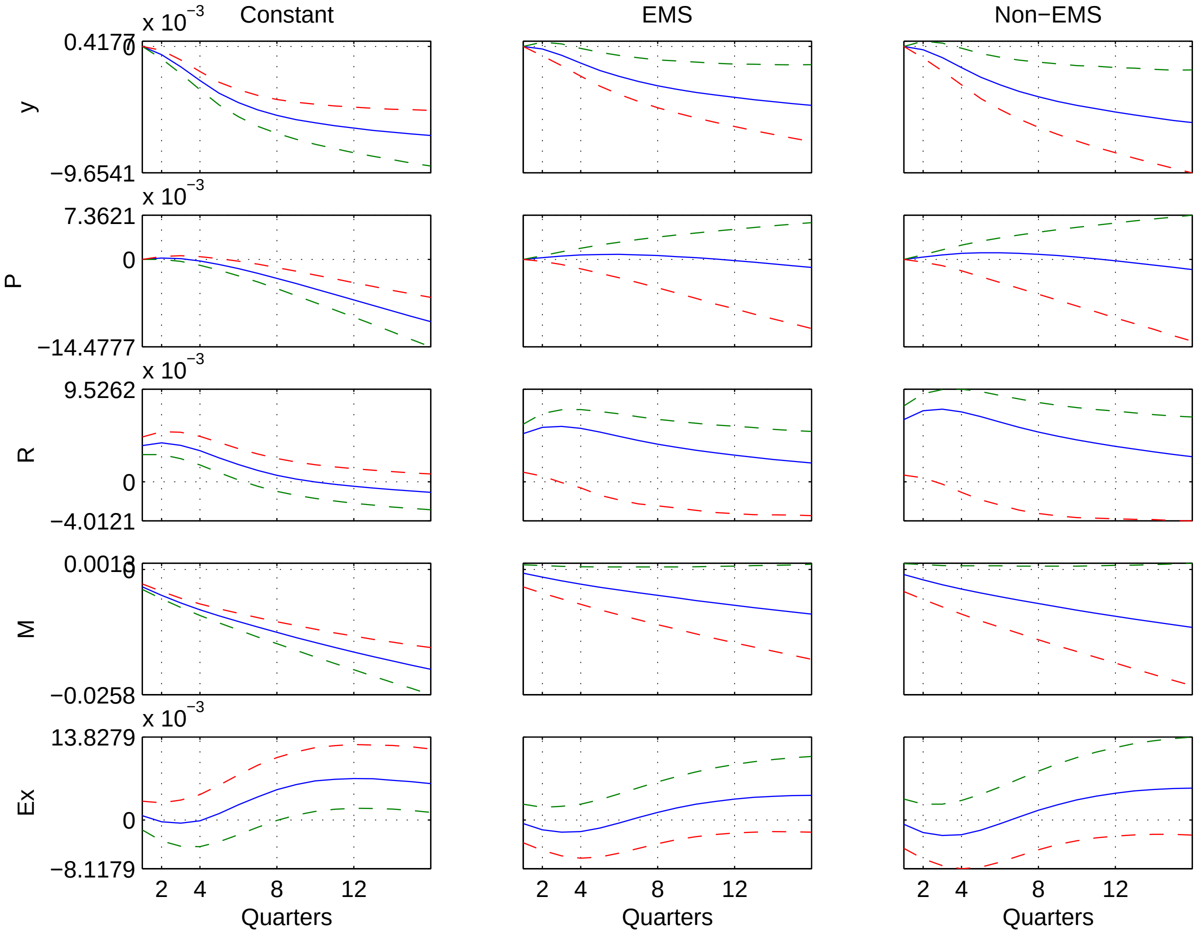

FIGURE 6. Impulse responses to a one-standard-deviation monetary policy shock under the GLSZ identification. The solid line represents the posterior median estimate and the two dashed lines contain the 68 percent probability based on 500,000 MCMC draws.

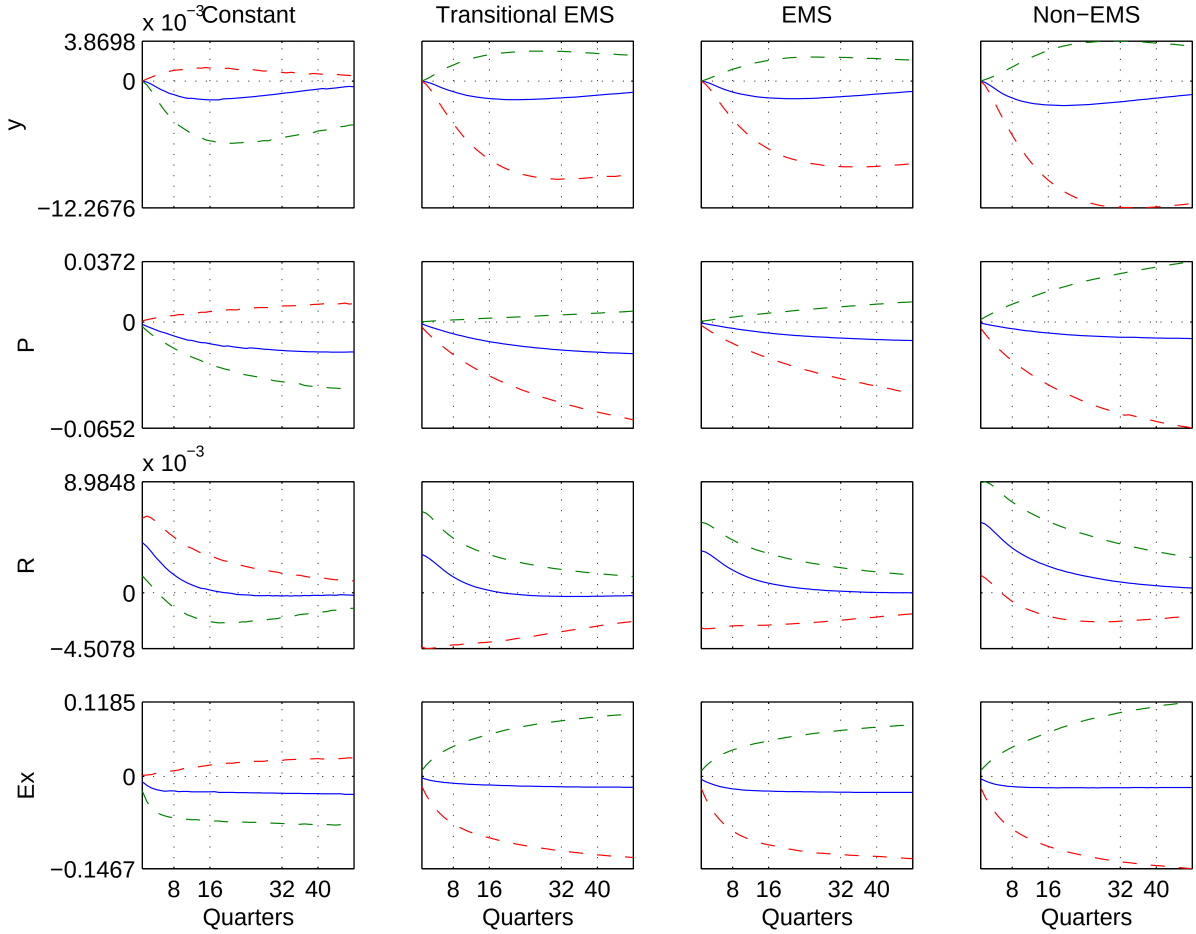

FIGURE 7. Impulse responses to a one-standard-deviation monetary policy shock under the BGQ identification. The solid line represents the posterior median estimate and the two dashed lines contain the 68 percent probability based on 500,000 MCMC draws.

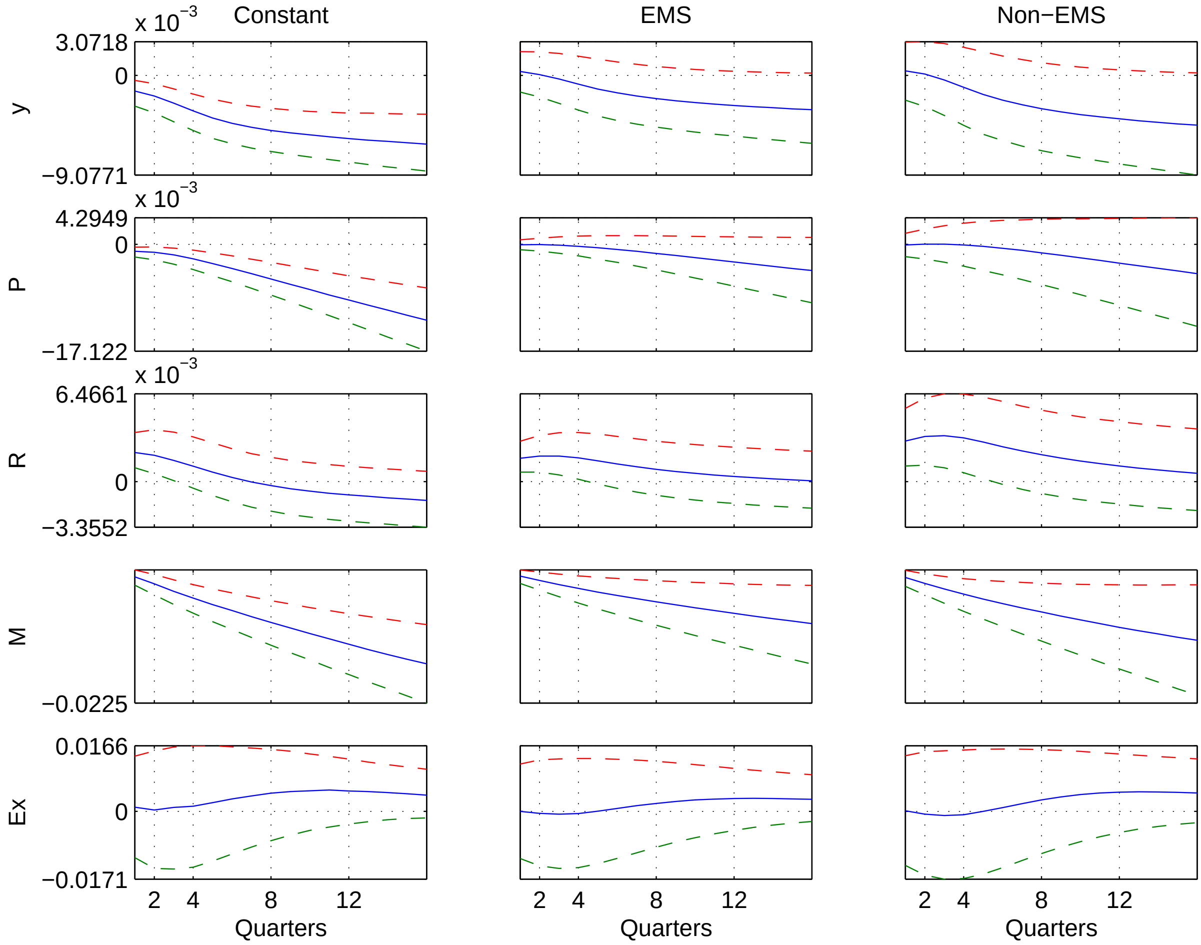

FIGURE 8. Impulse responses to a one-standard-deviation monetary policy shock under the CDFU identification. The solid line represents the posterior median estimate and the two dashed lines contain the 68 percent probability based on 500,000 MCMC draws.

APPENDIX A. PROOF OF THEOREM 3

Algorithm 1 shows how to construct an orthogonal matrix $P(k)$ satisfying the requirements of Theorem 3. That algorithm required that $rank\left(\tilde{Q}_i(k)\right) < n$, and the orthogonal matrix produced will be unique, up to the sign of each column, if and only if $rank\left(\tilde{Q}_i(k)\right) = n-1$. The rank of $\tilde{Q}_i(k)$ will be $n-1$ if and only if the span of vectors $p_1(k), \cdots, p_{i-1}(k)$ constructed in Algorithm 1 intersect with the span of the rows of $\tilde{Q}_i(k)$ only at the origin. One can show, though we do not explicitly do so here, that this condition will be violated on at most a set of measure zero. This proves that if $rank(Q_i) = n-i$, then the normalized Markov-switching SVAR is exactly identified. What remains to be shown is that if the normalized Markov-switching SVAR is exactly identified, then $rank(Q_i) = n-i$. We proceed via a sequence of lemmas and corollaries.

Theorem 4 implies that if a normalized linearly identified Markov-switching SVAR is exactly identified, then for almost all structural parameters $(A_0(k), A_+(k))$ there exists a unique orthogonal matrix $P(k)$ such that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the linear restrictions. The following lemma implies that existence holds for all structural parameters, while uniqueness is still only guaranteed for almost all structural parameters.

Lemma 6. If a normalized linearly identified Markov-switching SVAR is exactly identified, then for every structural parameter $(A_0(k), A_+(k))$ there exists an orthogonal matrix $P(k)$ such that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the linear restrictions.

Proof. Suppose that there exists a structural parameter $(A_0(k), A_+(k))$ such that there is no orthogonal matrix $P(k)$ such that $(A_0(k)P(k), A_+(k)P(k))$ satisfy the linear restrictions. We show that there is an open set U about $(A_0(k), A_+(k))$ such that for every $\left(\tilde{A}_0(k), \tilde{A}_+(k)\right) \in U$ there is no $P(k)$ such that $\left(\tilde{A}_0(k)P(k), \tilde{A}_+(k)P(k)\right)$ satisfy the linear restrictions. Since open sets have positive measure, this contradicts Theorem 4.

If there were no such open set U, then there would exist sequences $\left(A_0^i(k), A_+^i(k)\right)$ and $P^i(k)$ such that $\left(A_0^i(k), A_+^i(k)\right)$ converges to $(A_0(k), A_+(k))$, $P^i(k)$ is orthogonal, and $\left(A_0^i(k)P^i(k), A_+^i(k)P^i(k)\right)$ satisfy the linear restrictions. Since the $P^i(k)$ are elements of a compact space, some subsequence converges to an orthogonal matrix $P(k)$. But since $\left(A_0^i(k)P^i(k), A_+^i(k)P^i(k)\right)$ satisfy the linear restrictions, so will $(A_0(k)P(k), A_+(k)P(k))$, which is a contradiction. □

Lemma 7. For $1 \leq i \leq k \leq n$, let V_i be a subspace of $\mathbb{R}^n$. If for every invertible $n \times n$ matrix A, there exists an orthonormal set $\{v_1, \cdots, v_k\}$ in $\mathbb{R}^n$ such that $v_i \in AV_i$, then there exists a j with $1 \leq j \leq k$ and $\dim(V_i) \geq k$.

Proof. Given a subspace W of $\mathbb{R}^n$ and $\varepsilon \in \mathbb{R}$, let $A_{W,\varepsilon}$ be the linear transformation that fixes W and maps each u in the perpendicular component of W to εu. If $\dim(V_i) < k$ for $1 \leq i \leq k$, then using the following three statements a W and $\varepsilon > 0$ can be constructed such that $A_{W,\varepsilon}$ violates the conditions of the lemma. So it suffices to prove the following.

(1) If $\dim(V_i) < k$, then there exists a subspace U of $\mathbb{R}^n$ of dimension $n-k+1$ such that $U \cap V_i = \{0\}$.

(2) Let W be a $k-1$ dimensional subspace of $\mathbb{R}^n$. There exists a $\delta > 0$ such that there cannot be k orthonormal vectors in the set

$$S_{W,\delta} = \{w + u \in \mathbb{R}^n \,|\, w \in W \text{ and } \|u\| < \delta\}.$$

(3) Let U and V be subspaces of $\mathbb{R}^n$ such that $U \cap V = \{0\}$ and let W be the perpendicular complement of U. For every $\delta > 0$ there exists a $\gamma > 0$ such that for all $\varepsilon < \gamma$ if $v \in A_{W,\varepsilon}V$ and $\|v\| = 1$, then $v \in S_{W,\delta}$.

(1) If $\dim(V_i) < k \leq n$, then each V_i is of measure zero in $\mathbb{R}^n$, as will be the union of the V_i. So there exists a u_1 that is not contained in any V_i. If $k = n$, then the one-dimensional subspace generated by u_1 is the required subspace. If $k < n$, then let $\tilde{V}_i$ be the subspace generated by V_i and u_1. Since $\dim\left(\tilde{V}_i\right) < k+1 \leq n$, by the same measure argument as before, there will exist

a u_2 that is not contained in the union of the $\tilde{V}_i$. If $k = n-2$, then the two-dimensional subspace generated by u_1 and u_2 is the required subspace. This argument can be continued until a basis $u_1, \cdots, u_{n-k+1}$ has been constructed for the required subspace.

(2) Suppose there were $v_1, \cdots, v_k$ in $S_{W,\delta}$ that were orthonormal. Since the v are in $S_{W,\delta}$, write $v_i = w_i + u_i$ where $w_i \in W$ and $\|u_i\| < \delta$. Let X be the $n \times k$ matrix $\begin{bmatrix} w_1 & \cdots & w_k \end{bmatrix}$ and let Y be the $n \times k$ matrix $\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}$. Because the w are in a $k-1$ dimensional space, the matrix $X'X$ is singular and because the v are orthonormal, $Y'Y$ is the $k \times k$ identity matrix. Because δ can be chosen to be arbitrarily small, $X'X$ can be made to be arbitrarily close to the identity matrix, which is a contradiction.

(3) If this were not true, then there would exist a $\delta > 0$ and a sequence of v_i and ε_i such that the ε_i tend to zero and $v_i \in A_{W,\varepsilon_i}V$, $\|v_i\| = 1$, and $v_i \notin S_{W,\delta}$. Because U and W are perpendicular components, v_i can be uniquely written as $v_i = u_i + w_i$ where $u_i \in U$ and $w_i \in W$. Since $\|v_i\| = 1$ and u_i and w_i are orthogonal, $\|w_i\| \leq 1$. Since $v_i \notin S_{W,\delta}$, $\|u_i\| > \delta$. Since $v_i \in A_{W,\varepsilon_i}V$, $\frac{1}{\varepsilon_i}u_i + w_i \in V$. Dividing by the norm, we see that

$$\frac{u_i + \varepsilon_i w_i}{\sqrt{\|u_i\|^2 + \varepsilon_i^2 \|w_i\|^2}} \in V$$

Since this is a bounded sequence, some subsequence must converge. Since $\|u_i\|$ is bounded away from zero, $\|w_i\|$ is bounded above, and V is closed, the convergent subsequence must converge to a non zero element of $U \cap V$, which is a contradiction. □

Lemma 8. For $1 \leq i \leq k \leq n$, let V_i be a subspace of $\mathbb{R}^n$ with $\dim(V_1) \leq \cdots \leq \dim(V_k)$. The following statements are equivalent.
(1) For every invertible $n \times n$ matrix A there exists an orthonormal set $\{v_1, \cdots, v_k\}$ such that $v_i \in AV_i$.
(2) For $1 \leq i \leq k$, $\dim(V_i) \geq i$.

Proof. (1)⇒(2). Proceed by finite induction on k. When $k = 1$, the result is trivially true. Now suppose that (1)⇒(2) for some $k < n$. Let $(V_1, \cdots V_{k+1})$ be a sequence of subspaces of non decreasing dimension such that (1) holds. By Lemma 7, we know that $\dim(V_{k+1}) \geq k+1$. Since (1) holds for $(V_1, \cdots V_{k+1})$, (1) will also hold for $(V_1, \cdots V_k)$. This implies that for $1 \leq i \leq k$, $\dim(V_i) \geq i$. This, combined with the fact that $\dim(V_{k+1}) \geq k+1$, shows that (2) holds.

(2)⇒(1). Assume that (2) holds and let A be any invertible $n \times n$ matrix. Since $\dim(AV_1) \geq 1$, there exists a vector $v_1 \in AV_1$ of unit length. Now suppose that an orthonormal set $\{v_1, \cdots v_j\}$ has been chosen so that $v_i \in AV_i$ for $1 \leq i \leq j$. Let U be the $n-j$ dimensional subspace of $\mathbb{R}^n$ consisting of vectors orthogonal to $\{v_1, \cdots v_j\}$. Since $\dim(AV_{j+1}) \geq j+1$, the intersection of U and AV_{j+1} contains a non zero vector. Let v_{j+1} be any element of $U \cap AV_{j+1}$ of unit length. Then $\{v_1, \cdots v_{j+1}\}$ is a set of orthonormal vectors with $v_i \in AV_i$ for $1 \leq i \leq j+1$. So (1) holds.

□

Corollary 9. For $1 \leq i \leq n \leq m$, let Q_i be a matrix with m columns with $rank(Q_1) \geq \cdots \geq rank(Q_n)$. Let X be a full rank $m \times n$ matrix such that $rank(Q_iX) = rank(Q_i)$. The following are equivalent.
(1′) For every invertible $n \times n$ matrix A there exists a $n \times n$ orthogonal matrix P such that $Q_iXA^{-1}Pe_i = 0$.
(2′) For $1 \leq i \leq n$, $rank(Q_i) \leq n-i$.

Proof. The corollary is a simple restatement of Lemma 8 when $k = n$. Define

$$V_i = \{v \in \mathbb{R}^n \,|\, Q_iXv = 0\}.$$

Since $Q_iXA^{-1}Pe_i = 0$ if and only if $Pe_i \in AV_i$, (1′) is equivalent to (1) of Lemma 8. Since $\dim(V_i) = n - rank(Q_iX) = n - rank(Q_i)$, (2′) is equivalent to (2) of Lemma 8.

□

We can now complete the proof of Theorem 3. Assume that the normalized linearly identified Markov-switching SVAR is exactly identified. Let there be a set of structural parameters such that $rank\left(Q_i X\left(A_0(k), A_+(k)\right)\right) = rank\left(Q_i\right)$. By permuting the equations of the original system, we can assume without loss of generality that the linear restrictions on the columns of $X(\cdot)$ satisfy the condition $rank(Q_1) \geq \cdots \geq rank(Q_n)$. Since the model is assumed to be exactly identified, Lemma 6 implies that for every invertible matrix C there exists an orthogonal matrix P such that $(A_0(k)CP, A_+(k)CP)$ satisfy the linear conditions. So

$$\begin{aligned} 0 &= Q_i X\left(A_0(k)CP, A_+(k)CP\right) e_i \\ &= Q_i X\left(A_0(k), A_+(k)\right) DPe_i \end{aligned}$$

where $D = C$ if condition (1a) holds and $D = (C')^{-1}$ if condition (1b) holds. In either case, condition (1′) of Corollary (9) holds, so $rank(Q_i) \leq n - i$. The rank conditions of Rothenberg (1971) imply that in fact $rank(Q_i) = n - i$. This completes the proof of Theorem 3.

APPENDIX B. ANALYSIS OF TRIANGULAR SYSTEMS

Algorithm 1 gives us a way to find the matrix $P(k)$ for a general class of linear restrictions. Most restrictions used in the literature are exclusion restrictions. If these restrictions meet certain conditions, we have an even more efficient algorithm for determining the matrix $P(k)$. Such conditions are described by the following definition.

Definition 10. Identifying restrictions of the form of (3) are *triangular* if the following condition holds: $Q_j X\left(A_0(k), A_+(k)\right) e_j = 0$ if and only if there is a permutation matrix P_1 of the rows of $X\left(A_0(k), A_+(k)\right)$ and a permutation matrix P_2 of the columns of $X\left(A_0(k), A_+(k)\right)$, such that the permuted matrix $P_1 X\left(A_0(k), A_+(k)\right) P_2$ is lower triangular.

If exclusion restrictions are triangular, Algorithm 1 can be further improved, so that the orthogonal matrix given by Theorem 4 can be found using a single QR decomposition as described in the following theorem.

Theorem 11. Suppose the identifying restrictions are triangular and let P_1 and P_2 be the permutation matrices that make the restrictions triangular. For $1 \leq k \leq h$, let $\left(A_0(k), A_+(k)\right)$ be a set of structural parameters coming from the recursive identification. Using the QR decomposition for $\left(P_1 X\left(A_0(k), A_+(k)\right)\right)'$, write $P_1 X\left(A_0(k), A_+(k)\right) = T_L(k) P_3(k)$ where $P_3(k)$ is an orthogonal matrix and $T_L(k)$ is lower triangular. The structural parameters $\left(A_0(k)P(k), A_+(k)P(k)\right)$ for $P(k) = P_3(k)' P_2'$ satisfy the restrictions.

Proof. Because of condition (1) on the transformation $X(\cdot)$,

$$X\left(A_0(k)P(k), A_+(k)P(k)\right) = X\left(A_0(k), A_+(k)\right) P(k)$$

So

$$\begin{aligned} P_1 X\left(A_0(k)P(k), A_+(k)P(k)\right) P_2 &= P_1 X\left(A_0(k), A_+(k)\right) P(k) P_2 \\ &= T_L(k) P_3(k) P_3(k)' P_2' P_2 \\ &= T_L(k) \end{aligned}$$

which implies that the rotated parameters $X\left(A_0(k)P(k), A_+(k)P(k)\right)$ satisfy the restrictions. □

The illustration of section II.6 is continued. In that example, the restrictions were of the form

$$\begin{array}{c} \\ \Delta logY \\ R \\ logP \\ \Delta logY \\ R \\ logP \end{array} \begin{array}{c} \begin{array}{ccc} MP & AD & AS \end{array} \\ \left[\begin{array}{ccc} 0 & x & x \\ x & x & x \\ x & x & x \\ 0 & 0 & x \\ x & x & x \\ x & x & x \end{array}\right] \end{array}$$

To rotate this into a triangular form, the first and third columns need to be interchanged and the fourth row needs to be made the first row. This implies that

$$P_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } P_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

are the required permutation matrices. As in that example, we walk through the algorithm for a single state and so suppress the index k. The reduced-form parameters B and Σ were given and then A_0 and A_+ were computed via a Cholesky decomposition, i.e., the A_0 and A_+ are a set of structural parameters coming from a recursive identification. The transformation $X\left(A_0, A_+\right)$ is

$$X\left(A_0, A_+\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The QR decomposition of $\left(P_1 X\left(A_0, A_+\right)\right)'$ gives

$$P_3 = \begin{bmatrix} -0.7071 & -0.70710 & 0 \\ -0.7071 & 0.7071 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T_L = \begin{bmatrix} -1.4142 & 0 & 0 \\ -0.7071 & -0.7071 & 0 \\ -1.7678 & 1.0607 & 0 \\ -1.4142 & 0 & 1 \\ 0 & 1.4142 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The required rotation is $P = P_3 P_2$, which is equal to

$$P = P_3 P_2 = \begin{bmatrix} 0 & -0.70710 & -0.7071 \\ 0 & 0.7071 & -0.7071 \\ 1 & 0 & 0 \end{bmatrix}$$

It is easy to verify that $X\left(A_0 P, A_+ P\right)$ satisfies the restrictions.

APPENDIX C. THE REGIMES

In this section we describe the three types of parameter variation across regimes considered in this paper. First, we show the likelihood function of model (1) when identified using linear restrictions on columns of $A_0(k)$. Second, we detail the three types of parameter variation across regimes. Third, we describe the three prior distributions that implement the variation across regimes. Finally, we describe the three implied posterior distributions and how to draw from them.

It is important to note that, in principle, the three types of variation of parameters across regimes can only be considered when the model is identified using linear restrictions on columns of $A_0(k)$, i.e., when it is possible to use the methods described by Sims and Zha (2005). As shown in section IV.2, when a draw from a Markov-switching SVAR exactly identified using linear restrictions on columns of $A_0(k)$ is rotated to consider a more general set of restrictions, the three types of variation across regimes still hold. Therefore, these three types of parameter variation across regimes are general enough to be applied to all the identification schemes analyzed in this paper.

C.1. **The Likelihood Function.** In this section we describe how to evaluate the likelihood for the Markov-switching SVAR defined by (1) when identified using linear restrictions on columns of $A_0(k)$.

For $j = 1,\ldots,n$ and $k = 1,\ldots,h$, let the $q_j \times n$ matrix Q_j, where $q_j \leq n$, define the q_j restrictions over the j^{th} column of A_0, then linear restrictions on columns of $A_0(k)$ can be written as:

$$Q_j A_0\,(k)\,e_j = 0.$$

Let $a_{j,0}(k)$ be a column of $A_0(k)$, then the former restriction can be written in the following way:

$$Q_j\, a_{j,0}(k) = 0,$$

Let U_j be the $n \times q_j$ matrix whose columns form the orthonormal basis for the null space of Q_j. Then, $Q_j\, a_{j,0}(k) = 0$ if and only if $\exists$ a $q_j \times 1$ vector $b_j(k)$ such that

$$a_{j,0}(k) = U_j\, b_j(k).$$

Finally, for $j = 1,\ldots,n$, let

$$b_j = [b_j\,(1)',\ldots,b_j\,(h)']',$$

$$b = [b_1,\ldots,b_n,]$$

and

$$U = [U_1',\ldots,U_n']'.$$

Note the following three points. First, any set of $a_{j,0}(k)$ and Q_j for $j = 1,\ldots,n$ and $k = 1,\ldots,h$ implies a set of U_j and $b_j(k)$ for $j = 1,\ldots,n$ and $k = 1,\ldots,h$ and vice versa. Therefore, it is equivalent to defining the linear restrictions using either $a_{j,0}(k)$ and Q_j or U_j and $b_j(k)$. This implies that we can evaluate the likelihood function either using Q_j and $a_{j,0}(k)$ or U_j and $b_j(k)$. We follow the second approach.

Let us define:

$$d_j\,(k) = a_{j,+}\,(k) - \overline{S} a_{j,0}\,(k)\,,$$

for $k = 1,\ldots,h$ and $j = 1,\ldots,n$, where

$$\overline{S} = [I_{n\times n}', 0_{(m-n)\times n}']'$$

and $m = np + 1$.

Now let

$$d_j = [d_j(1)', \ldots, d_j(h)']'$$

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

Finally, let

$$d = [d_1, \ldots, d_n].$$

Note that any A_0 and d imply a matrix A_+. Therefore, for any given U, the matrices b and d imply the matrices A_0 and A_+. Thus, we can write the likelihood function using either A_0 and A_+ or b and d. We choose the first option.

Now, if we define

$$Y^t = [y_1 \ldots y_t]', \text{ and}$$

for all t, we can write the conditional likelihood function as follows.

Given the restriction matrix U, the conditional likelihood function, $\pi\left(y_t|Y^{t-1}, s_t, b, d\right)$, is:

$$\begin{aligned}\pi\left(y_t|Y^{t-1}, s_t, b, d\right) \quad \propto \quad & \det|\,[U_1 b_1(s_t) \ldots U_n b_n(s_t)]|\exp\left[-\frac{1}{2}\sum_{j=1}^{n} b_j'(s_t)\,U_j' S_t U_j b_j(s_t)\right] \\ & \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\left(d_j(s_t) + \left(\overline{S} - P_t\right)U_j b_j(s_t)\right)' H_t \left(d_j(s_t) + \left(\overline{S} - P_t\right)U_j b_j(s_t)\right)\right],\end{aligned}$$

where

$$H_t = x_t' x_t,$$

$$P_t = H_t^{-1} x_t' y_t,$$

and

$$S_t = y_t' y_t - P_t' H_t P_t.$$

Next, following Kim and Nelson (1998), we can write the likelihood function $\pi(Y_T|b,d)$. Hence, given the restriction matrix U, the likelihood function, $\pi\left(Y^T|b,d,\Pi\right)$, is:

$$\pi(Y^T|b,d,\Pi) \propto \prod_{t=1}^{T}\left\{\sum_{s_t=1}^{h}[\pi(y_t|Y^{t-1}, s_t, b, d)\Pr(s_t|Y^{t-1}, b, d, \Pi)]\right\}$$

where

$$\Pr(s_t|Y^{t-1}, b, d, \Pi) = \sum_{s_t=1}^{h}\pi(s_t|s_{t-1})\Pr(s_{t-1}|Y^{t-1}, b, d, \Pi)$$

and $\Pr(s_{t-1}|Y^{t-1}, b, d, \Pi)$ is updated using the Bayes rule.[33]

[33] We initialize the system setting $\Pr(s_0|Y^0, b, d, \Pi) = \Pr(s_0|b, d, \Pi) = 1/h$.

C.2. **Modelling Regimes.** If we let all the parameters vary across regimes, b and d can be estimated independently across regimes. Therefore, we could use the methods used by Chib (1996) to perform the model estimation. The problem is that a Markov-switching SVAR with four to seven endogenous variables and one-year lag length would suffer the over-parameterization problems associated with few degrees of freedom. Hence, we define three sets of priors that restrict the variation of parameters across regimes. First, we consider priors that impose a constant-parameters model, i.e., no cross-regime variation. Second, we contemplate priors that only allow the reduced-form variances of the shocks to change across regimes. Finally, we also use priors that imply that both structural parameters and structural variances can change across regimes. The actual priors for each of the cases are defined in subsection C.3. In this section we just highlight the main differences among the three sets of priors and their implications for across regime variation. In order to do that, we first rewrite the parameters defining model (1) in the following way:

$$a_{i,j,0}(k) = \overline{a}_{i,j,0}\xi_j(k)\,\phi_{i,j}(k),$$

$$d_{i,j,\ell}(k) = \overline{d}_{i,j,\ell}\xi_j(k)\,\lambda_{i,j}(k),$$

and

$$c_j(k) = \overline{c}_j\xi_j(k)\,\mu_j(k)$$

for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. Notice that writing the parameters this way already imposes a restriction on cross-regime variation. We restrict the cross-regime variation of d, since we do not allow for variation between lags (i.e., $(d_{i,j,\ell}(k) = d_{i,j,\ell'}(k)$ for $\ell',\ell = 1,\ldots,p)$). This restriction is common to the three cases considered here.

- *Constant-Parameters Case*. These priors impose $\xi_j(k)=1$, $\phi_{i,j}(k)=1$, $\lambda_{i,j}(k)=1$, and $\mu_j(k)=1$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. Therefore $a_{i,j,0}(k)=\overline{a}_{i,j,0}(k)$, $d_{i,j,\ell}(k)=\overline{d}_{i,j,\ell}(k)$, and $c_j(k)=\overline{c}_j$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. This case corresponds to the constant-parameters VARs widely used in the literature.
- *Variance-Only Case*. These priors impose $\phi_{i,j}(k)=1$, $\lambda_{i,j}(k)=1$, and $\mu_j(k)=1$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. Therefore, we can write $a_{i,j,0}(k)=\overline{a}_{i,j,0}\xi_j(k)$, $d_{i,j,\ell}(k)=\overline{d}_{i,j,\ell}\xi_j(k)$, and $c_j(k)=\overline{c}_j\xi_j(k)$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. These restrictions imply that $B(k)$ does not change across regimes.
- *All-Change Case*. These priors impose $\xi_j(k)=1$, $\overline{a}_{j,0}=1$, and $\overline{c}_j=1$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. Therefore $a_{i,j,0}(k)=\phi_{i,j}(k)$, $d_{i,j,\ell}(k)=\overline{d}_{i,j,\ell}\lambda_{i,j}(k)$, and $c_j(k)=\overline{c}_j(k)\,\mu_j(k)$ for $i,j=1,\ldots,n$ and $k=1,\ldots,h$. These restrictions imply that the reduced-form parameters and variances change across regimes.

C.3. **The Priors.** In this appendix we specify the details of the priors used in the paper. First, we describe the priors on Π, common to the three cases. Then, we describe the priors on the parameters that differ across the three cases.

The priors on the transition matrix, Π, take a Dirichlet form, as suggested by Chib (1996). For the kth column of Π, π_k, the prior density is $\pi(\pi_k)=\pi(\pi_{1k},\ldots\pi_{nk}) \propto \pi_{1k}^{\alpha_{1k}-1}\ldots\pi_{nk}^{\alpha_{nk}-1}$. We choose α_{ij} for $i,j=1,\ldots,n$ as described in Sims and Zha (2004). Basically, we set α_{ij} for $i,j=1,\ldots,n$ such that the average duration of each state is around seven quarters, independently of the number of regimes.

Now let us describe the priors on the parameters that differ across the three cases. Before proceeding, we introduce a few new notations. Let ς_n be a column vector of n ones. Let

$$\overline{A}_0 = [\overline{a}_{1,0},\ldots,\overline{a}_{n,0}],$$

where $\overline{a}_{j,0}$ is a $n \times 1$ vector of the form:

$$\overline{a}_{j,0} = [\overline{a}_{1,j,0}, \ldots, \overline{a}_{n,j,0}]' \text{ for all } j.$$

Now let

$$\xi = [\xi_1, \ldots, \xi_n],$$

where ξ_j is a $h \times 1$ vector of the form:

$$\xi_j = [\xi_j(1), \ldots, \xi_j(h)]' \text{ for all } j.$$

Let

$$\phi = [\phi_1, \ldots, \phi_n],$$

where ϕ_j is a $nh \times 1$ vector of the form:

$$\phi_j = \left[\phi_j'(1), \ldots, \phi_j'(h)\right]', \text{ for all } j,$$

where $\phi_j(k)$ is a $n \times 1$ vector of the form:

$$\phi_j(k) = \left[\phi_{1,j}(k), \ldots, \phi_{n,j}(k)\right]', \text{ for all } k \text{ and all } j.$$

Define also

$$\overline{d} = [\overline{d}_1', \ldots, \overline{d}_n']',$$

where $\overline{d}_j$ is a $m \times 1$ vector of the form:

$$\overline{d}_j = [\overline{d}_{j,1}', \ldots, \overline{d}_{j,p}', \overline{c}_j]' \text{ for all } j,$$

where $\overline{d}_{j,\ell}$ is a $n \times 1$ vector of the form:

$$\overline{d}_{j,\ell} = \left[\overline{d}_{1,j,\ell}, \ldots, \overline{d}_{n,j,\ell}\right]' \text{ for all } \ell \text{ and all } j.$$

Let

$$\lambda = [\lambda_1, \ldots, \lambda_n],$$

where λ_j is a $nh \times 1$ vector of the form:

$$\lambda_j = \left[\lambda_j'(1), \ldots, \lambda_j'(h)\right]' \text{ for all } j,$$

where $\lambda_j(k)$ is a $n \times 1$ vector of the form:

$$\lambda_j(k) = \left[\lambda_{1,j}(k), \ldots, \lambda_{n,j}(k)\right]' \text{ for all } j \text{ and all } k.$$

Let

$$\mu = [\mu_1, \ldots, \mu_n],$$

where μ_j is a $h \times 1$ vector of the form:

$$\mu_j = [\mu_j(1), \ldots, \mu_j(h)]' \text{ for all j.}$$

Then we can write

$$a_{j,0} = \Phi_j\left(\xi_j \otimes \overline{a}_{j,0}\right),$$

where

$$\Phi_j = diag\left(\left\{\Phi_j(k)\right\}_{k=1}^h\right)$$

and

$$\Phi_j(k) = diag\left(\left\{\phi_{i,j}(k)\right\}_{i=1}^n\right).$$

Finally, we can also write

$$d_j = \Lambda_j \left(\xi_j \otimes \overline{d}_j\right),$$

where

$$\Lambda_j = diag\left(\left\{\Lambda_j(k)\right\}_{k=1}^{h}\right),$$

$$\Lambda_j(k) = \begin{bmatrix} I_p \otimes \Delta_j(k) & 0_{np\times 1} \\ 0_{1\times np} & \mu_j(k) \end{bmatrix},$$

and

$$\Delta_j(k) = diag\left(\left\{\lambda_{i,j}(k)\right\}_{i=1}^{n}\right).$$

We are now ready to specify the priors corresponding to all the cases. We begin with the all-change case and work backward to the constant-parameters case.

C.3.1. *All-Change Case.* Let $\xi_j = \varsigma_h$, $\overline{a}_{j,0} = \varsigma_n$, and $\overline{c}_j = 1$ for all j. Then

$$a_{j,0} = \Phi_j \varsigma_{hn} = \phi_j \text{ for all } j,$$

$$d_j = \Lambda_j \left(\varsigma_h \otimes \overline{d}_j\right) \text{ for all } j,$$

and

$$c_j = \mu_j \text{ for all } j.$$

Let the priors on the contemporaneous parameters of the model, $a_{j,0}$, now be:

$$\pi\left(a_{j,0}\right) = \pi\left(\phi_j\right) = \aleph\left(0, I_h \otimes H_{j,0}\right) \text{ for all } j,$$

where $H_{j,0}$ is set following the procedure described in Sims and Zha (2004).

Since

$$\phi_j = \left(I_h \otimes U_j\right) b_j, \text{ for all } j,$$

that implies priors on b_j of the form:

$$\pi\left(b_j\right) = \aleph\left(0, \overline{H}_{j,0}\right),$$

where

$$\overline{H}_{j,0} = \left(U_j'\left(I_h \otimes H_{j,0}^{-1}\right) U_j\right)^{-1}.$$

Let the priors on the lagged and constant parameters of the model, $\overline{d}_j$, now be:

$$\pi\left(\overline{d}_j\right) = \aleph\left(0, H_{j,+}\right) \text{ for all } j,$$

$$\pi\left(\lambda_j\right) = \aleph\left(0, \left(I_h \otimes I_n\right)\sigma_\lambda^2\right) \text{ for all } j,$$

and

$$\pi\left(\mu_j\right) = \aleph\left(0, I_h \otimes \sigma_{j,\mu}^2\right) \text{ for all } j,$$

where $H_{j,+}$ is set following the procedure described in Sims and Zha (2004), $\sigma_\lambda = 50$, and $\sigma_{j,\mu}$ is set in the same way as $H_{j,+}$.

C.3.2. *Variance-Only Case.* Let $\phi_j = \varsigma_n$, $\lambda_j = \varsigma_{nh}$, and $\mu_j = 1$ for all j, then

$$a_{j,0} = \xi_j \otimes \overline{a}_{j,0} \text{ for all } j,$$

$$d_j = \xi_j \otimes \overline{d}_j \text{ for all } j,$$

and

$$c_j = \overline{c}_j \xi_j \text{ for all } j.$$

Let the priors on the contemporaneous parameters of the model, $\overline{a}_{j,0}$, now be:

$$\pi\left(\overline{a}_{j,0}\right) = \aleph\left(0, H_{j,0}\right) \text{ for all } j,$$

where $H_{j,0}$ is set following the procedure described in Sims and Zha (2004).

Since

$$\xi_j \otimes \overline{a}_{j,0} = (I_h \otimes U_j)\, b_j, \text{ for all } j,$$

that implies priors on b_j of the form:

$$\pi(b_j|\xi_j) = \aleph\left(0, \widetilde{H}_{j,0}\right),$$

where

$$\widetilde{H}_{j,0} = \widetilde{\Upsilon}_{j,h} \otimes \left(U_j' H_{j,0}^{-1} U_j\right)^{-1},$$

and

$$\widetilde{\Upsilon}_{j,h} = \begin{bmatrix} \xi_j(1)^2 & \xi_j(1)\xi_j(2) & \dots & \xi_j(1)\xi_j(h) \\ \xi_j(2)\xi_j(1) & \xi_j(2)^2 & \dots & \xi_j(2)\xi_j(h) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_j(h)\xi_j(1) & \xi_j(h)\xi_j(2) & \dots & \xi_j(h)^2 \end{bmatrix}.$$

Let us define the priors on the lagged and constant parameters of the model. We have:

$$\pi\left(\overline{d}_j\right) = \aleph\left(0, H_{j,+}\right) \text{ for all } j,$$

and:

$$\pi\left(\overline{c}_j\right) = \aleph\left(0, \sigma_{j,\overline{c}}\right) \text{ for all } j,$$

where $H_{j,+}$ and $\sigma_{j,\overline{c}}$ are set following the procedure described in Sims and Zha (2004).

Finally, let priors on $\xi_j(k)$ be defined over $\zeta_j(k) = \xi_j^2(k)$ as:

$$\pi\left(\zeta_j(k)\right) = \Gamma\left(\alpha_\zeta, \beta_\zeta\right) \text{ for all } k \text{ and } j.$$

where $\alpha_\zeta = 1$ and $\beta_\zeta = 1$.

C.3.3. *Constant-Parameters Case.* Let $\xi_j = \varsigma_h$, $\phi_j = \varsigma_n$, $\lambda_j = \varsigma_{nh}$, and $\mu_j = 1$ for all j, then

$$a_{j,0} = \varsigma_h \otimes \overline{a}_{j,0} \text{ for all } j,$$

$$d_j = \varsigma_h \otimes \overline{d}_j \text{ for all } j,$$

and

$$c_j = \overline{c}_j \text{ for all } j.$$

Let the priors on the contemporaneous parameters of the model, $\overline{a}_{j,0}$, now be:

$$\pi\left(\overline{a}_{j,0}\right) = \aleph\left(0, H_{j,0}\right) \text{ for all } j,$$

where $H_{j,0}$ is set following the procedure described in Sims and Zha (2004).

Since

$$\varsigma_h \otimes \overline{a}_{j,0} = \left(I_h \otimes U_j\right) b_j, \text{ for all } j,$$

that implies priors on b_j of the form:

$$\pi\left(b_j\right) = \aleph\left(0, \widehat{H}_{j,0}\right),$$

where

$$\widehat{H}_{j,0} = \widehat{\Upsilon}_{j,h} \otimes \left(U_j' H_{j,0}^{-1} U_j\right)^{-1},$$

and

$$\widehat{\Upsilon}_{j,h} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}.$$

Let the priors on the lagged and constant parameters of the model, $\overline{d}_j$, now be:

$$\pi\left(\overline{d}_j\right) = \aleph\left(0, H_{j,+}\right) \text{ for all } j,$$

where $H_{j,+}$ is set following the procedure described in Sims and Zha (2004). Finally, let us consider the priors on $\overline{c}_j$:

$$\pi\left(\overline{c}_j\right) = \aleph\left(0, \sigma_{j,\overline{c}}\right) \text{ for all } j,$$

where $\sigma_{j,\overline{c}}$ is set in the same way as $H_{j,0}$.

C.4. **Posterior.** In this section we briefly describe our procedure to draw from the posterior of the structural parameters of model (2). Because of space considerations we refer the reader to Sims and Zha (2004) for a detailed discussion of the posterior distributions implied by the priors described in section C.3 and how to draw from them. Suffice it to say that we are interested in the following posterior distributions:

$$\pi(S^T|Y^T,b,\overline{d},\phi,\lambda,\mu,\Pi)$$
$$\pi(\Pi|Y^T,b,\overline{d},\phi,\lambda,\mu,S^T)$$
$$\pi(\phi,\lambda,\mu|Y^T,b,\overline{d},S^T,\Pi)$$
$$\pi(b|Y^T,\overline{d},\phi,\lambda,\mu,S^T,\Pi),$$

where $S^T = (s_1,\ldots,s_T)$ and $\overline{d}$, ϕ, λ, and μ are defined in section C.3. We also use standard MCMC to draw from these posterior distributions and the modified harmonic mean (MHM), described in Gelfand and Dey (1994) and Geweke (1999), to compute the marginal likelihood. We use Geweke's (2005) procedures to check the convergence of the posterior draws.

Appendix D. The Identification Schemes

In this section we use the notation of section II.2 to describe three of the identification schemes used in the paper.

D.1. **CEE.** In order to write the CEE in the notation of section II.2 we let $X(\cdot)$ be equal to (4) and consider the following set of Q_j:

$$Q_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$Q_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

D.2. **GLSZ.** In order to write the GLSZ in the notation of section II.2 we let $X(\cdot)$ be equal to (4) and consider the following set of Q_j:

$$Q_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_4 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$Q_5 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

D.3. **BGQ.** In order to write the BGQ in the notation of section II.2 we let $X(\cdot)$ be equal to (5) and consider the following set of Q_j:

$$Q_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$Q_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Appendix E. Existing Method to Estimate Signed Restricted SVARs

Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) have proposed an alternative approach to impose sign restrictions directly on impulse responses themselves to identify SVARs. For example, in response to a contractionary monetary shock the interest rate should rise, while money and prices should fall. Although Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) start from the same idea, they implement it in different ways. In this section, we first briefly describe the approaches of Faust, Canova and De Nicoló, and Uhlig, highlighting the problems of applying these approaches to our switching model.

E.1. **Faust Method.** Faust (1998) presents a way to check the robustness of any claim from an SVAR. All possible identifications are checked searching for the one that is worst for the claim, subject to the restriction that the identified SVAR produces the impulse response functions with the "correct" sign.

Faust (1998) shows that this problem is equivalent to solving an eigenvalue problem $\sum_{i=0}^{M} \frac{R!}{i!(R-i)!}$ times, where R is the number of sign restrictions and $M = \max(n-1, R)$. As Faust (1998) recognizes, this method may not be feasible for large problems, like the one analyzed here.

E.2. **Canova and De Nicoló Method.** Canova and De Nicoló (2002) also identify SVARs using impulse response sign restrictions. Their method is based on the following theorem:

Theorem 12. Let P $(n \times n)$ be an orthogonal matrix. Then a unique series $\{\{\theta_{i,j}\}_{j=i+1}^{n}\}_{i=1}^{n-1}$ exists, where $0 \leq \theta_{i,j} < 2\pi$ if $j = i+1$ and $-\pi/2 \leq \theta_{i,j} \leq \pi/2$ if $j > i+1$, such that:[34]

$$P = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} Q_{i,j}(\theta_{i,j})$$

or

$$P = S \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} Q_{i,j}(\theta_{i,j})$$

where

$$S = \begin{bmatrix} 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & -1 \end{bmatrix}$$

and

$$Q_{i,j}(\theta_{i,j}) = \begin{bmatrix} & & & \text{col } i & & \text{col } j & & \\ & & & \downarrow & & \downarrow & & \\ & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ \text{row } i \rightarrow & 0 & \cdots & \cos(\theta_{i,j}) & \cdots & -\sin(\theta_{i,j}) & \cdots & 0 \\ & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ \text{row } j \rightarrow & 0 & \cdots & \sin(\theta_{i,j}) & \cdots & \cos(\theta_{i,j}) & \cdots & 0 \\ & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}.$$

Proof. The proof follows from Algorithm 5.2.2 of Golub and Van Loan (1996). □

Using theorem 12, Canova and De Nicoló (2002) identify SVARs with the following algorithm:

Algorithm 3.

(1) Begin with a triangular SVAR system.

(2) Draw the system parameters $A_0(k)$ and $B(k)$ from the posterior distribution.

(3) Determine a grid on the set of all orthogonal matrices.

(4) Perform a grid search to find an orthogonal matrix $P(k)$, such that the impulse responses generated from $A_0(k)P(k)$ and $B(k)$ satisfy all the sign restrictions.

[34] In Canova and De Nicoló (2002), the notation $Q_{i,j}(\theta)$ is used where θ is implicitly assumed to vary with different i and j.

Theorem 12 allows for different ways to design a grid, but because the space of all orthogonal $n \times n$ matrices is a $n(n-1)/2$ dimensional space, any grid that divides the interval $[-\pi/2, \pi/2]$ in M points[35] implies a search over $2hM^{n(n-1)/2}$ points in the space of all orthogonal $n \times n$ matrices.[36] Thus, it is not feasible to perform this grid search for large values of n and/or h.

E.3. **Uhlig's Methods.** Uhlig (2005) proposes another method to identify SVARs based on impulse response sign restrictions. His method also draws from the set of posterior orthonormal matrices, such that the impulse response sign restrictions hold, using the following algorithm:

Algorithm 4.

(1) Begin with a triangular SVAR system.

(2) Draw the system parameters $A_0(k)$ and $B(k)$ from the posterior distribution.

(3) Draw n independent standard normal vectors of length n and recursively orthonormalize them. Call $P(k)$ the resulting orthonormal matrix.

(4) Generate the impulse responses from $A_0(k)P(k)$ and $B(k)$.

(5) If these impulse responses do not satisfy the sign restrictions, keep the draw. Otherwise discard it.

This method is feasible for large models like the one we are dealing with in this paper. In fact, the method we propose is just a more efficient version of Uhlig's approach.

[35] Or the interval $[-\pi, \pi]$ in $2M$ points.

[36] The h term comes from the fact that we have to find the $P(k)$ for h regimes.

REFERENCES

ANGELONI, I., AND M. EHRMANN (2003): "Monetary Policy Transmission in the Euro Area: Any Changes after EMU?," *Economic Policy*, 37, 470–492.

BERNANKE, B. S., AND I. MIHOV (1998): "Measuring Monetary Policy," *Quarterly Journal of Economics*, 113(3), 869–902.

BLANCHARD, O. J., AND D. QUAH (1993): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 83(3), 655–673.

BOIVIN, J. (1999): "Revisiting the Evidence on the Stability of Monetary VARs," Manuscript, Columbia University.

BOIVIN, J., AND M. P. GIANNONI (2005): "Has Monetary Policy Become More Effective?," Manuscript, Columbia University.

CANOVA, F., AND G. DE NICOLO (2002): "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics*, 49(6), 1131–1159.

CANOVA, F., AND L. GAMBETTI (2004): "Structural Changes in the US Economy: Bad Luck or Bad Policy?," Manuscript, Universitat Pompeu Fabra.

CHIB, S. (1996): "Calculating Posterior Distributions and Model Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1996): "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics*, 78(1), 16–34.

CICCARELLI, M., AND A. REBUCCI (2003): "Measuring Contagion with a Bayesian, Time-Varying Coefficient Model," ECB Working Paper No. 263.

CLARIDA, R. H., J. GALÍ, AND M. GERTLER (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

COGLEY, T., AND T. SARGENT (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics*, 8, 262–302.

DE BONDT, G. (2002): "Retail Bank Interest Rate Pass-Through: New Evidence at the Euro Area Level," ECB Working Paper No. 136.

EVANS, C. L., AND D. A. MARSHALL (2004): "Economic Determinants of the Nominal Treasury Yield Curve," Manuscript, Federal Reserve Bank of Chicago.

FAGAN, G., J. HENRY, AND R. MESTRE (2004): "An Area-Wide Model (AWM) for the Euro Area," *Economic Modelling*.

FAUST, J. (1998): "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Conference Series on Public Policy*, 48, 207–244.

FAUST, J., AND E. M. LEEPER (1997): "When Do Long-Run Identifying Restrictions Give Reliable Results?," *Journal of Business and Economic Statistics*, 15, 345–353.

FERNÁNDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMÍREZ (2005): "Estimating Macroeconomic Models: A Likelihood Approach," Federal Reserve Bank of Atlanta Working Paper.

GALÍ, J. (1992): "How Well Does the IS-LM Model Fit Postwar U.S. Data?," *Quarterly Journal of Economics*, 107(2), 709–738.

GELFAND, A. E., AND D. K. DEY (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society (Series B)*, 56, 501–514.

GEWEKE, J. F. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, 18, 1–126.

——— (2005): "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association, forthcoming*.

GOLUB, G. H., AND C. F. VAN LOAN (1996): *Matrix Computations (Third Edition)*. The Johns Hopkins University Press, Baltimore and London.

GORDON, D. B., AND E. M. LEEPER (1994): "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy*, 102(6), 1228–1247.

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

HAMILTON, J. D., D. F. WAGGONER, AND T. ZHA (2003): "Normalization in Econometrics," Manuscript, University of California (San Diego) and Federal Reserve Bank of Atlanta.

KIM, C.-J., AND C. R. NELSON (1998): *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Cambridge.

LUBIK, T. A., AND F. SCHORFHEIDE (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190–217.

NEWEY, W. K., AND K. K. WEST (1987): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PEERSMAN, G., AND F. SMETS (2003): *The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis*. Cambridge University Press, Cambridge.

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–852.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

SARGENT, T. J., N. WILLIAMS, AND T. ZHA (2005): "Shocks and Government Beliefs: the Rise and Fall of American Inflation," *American Economic Review, forthcoming*.

SIMS, C. A. (1993): "A 9 Variable Probabilistic Macroeconomic Forecasting Model," *Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycle*, 28, 179–214.

SIMS, C. A., AND T. ZHA (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1155.

——— (2004): "MCMC Method for Markov Mixture Simultaneous-Equation Models: A Note," Federal Reserve Bank of Atlanta Working Paper 2004-15.

——— (2005): "Were There Regime Switches in U.S. Monetary Policy?," *American Economic Review, forthcoming*.

STEWART, G. W. (1980): "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators," *SIAM Journal on Numerical Analysis*, 17(3), 403–409.

STOCK, J. H., AND M. W. WATSON (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

STOCK, J. H., AND M. W. WATSON (2003): "Has the Business Cycle Changed? Evidence and Explanations," Monetary Policy and Uncertainty: Adapting to a Changing Economy, Federal Reserve Bank of Kansas City Symposium, Jackson Hole, Wyoming, August 28-30.

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

——— (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*.

UNGERER, H., J. J. HAUVONEN, A. LOPEZ-CLAROS, AND T. MAYER (1990): "The European Monetary System: Developments and Perspectives," *International Monetary Fund Occasional Paper 73*.

WAGGONER, D. F., AND T. ZHA (2003a): "A Gibbs Sampler for Structural Vector Autoregressions," *Journal of Economic Dynamics and Control*, 28(2), 349–366.

——— (2003b): "Likelihood Preserving Normalization in Multiple Equation Models," *Journal of Econometrics*, 114(2), 329–347.

FEDERAL RESERVE BANK OF ATLANTA