

Estimating overidentified, nonrecursive, time-varying coefficients structural vector autoregressions

FABIO CANOVA
EUI and CEPR

FERNANDO J. PÉREZ FORERO
BCRP

This paper provides a general procedure to estimate structural vector autoregressions. The algorithm can be used in constant or time-varying coefficient models, and in the latter case, the law of motion of the coefficients can be linear or non-linear. It can deal in a unified way with just-identified (recursive or nonrecursive) or overidentified systems where identification restrictions are of linear or of non-linear form. We study the transmission of monetary policy shocks in models with time-varying and time-invariant parameters.

KEYWORDS. Time-varying coefficient structural VAR models, Metropolis algorithm, identification restrictions, monetary transmission mechanism.

JEL CLASSIFICATION. C11, E51, E52.

1. INTRODUCTION

Vector autoregressive (VAR) models are routinely employed to summarize the properties of the data, and new approaches to the identification of structural shocks have been suggested in the last 10 years (see [Canova and De Nicoló \(2002\)](#), [Uhlig \(2005\)](#), and [Lanne and Lütkepohl \(2008\)](#)). Constant coefficient structural VAR models may provide misleading information when the structure is changing over time. [Cogley and Sargent \(2005\)](#) and [Primiceri \(2005\)](#) were the first to estimate time-varying coefficient (TVC) VAR models, and Primiceri also provides a structural interpretation of the dynamics using recursive restrictions on the matrix of impact responses. Following [Gambetti, Pappa, and Canova \(2008\)](#), the literature nowadays mainly employs sign restrictions to identify structural shocks in TVC-VARs, and the constraints used are, generally, theory based and robust to variations in the parameters of the data generating process (DGP); see [Canova and Paustian \(2011\)](#).

Fabio Canova: canova@eui.eu

Fernando J. Pérez Forero: fernando.perez@bcrp.gob.pe

We would like to thank F. Schorfheide, G. Primiceri, R. Casarin, H. Van Dijk, and three anonymous referees for comments and suggestions. The financial support of the Spanish Ministry of Economy and Competitiveness through Grants ECO2009-08556 and ECO2012-33247, of the Barcelona Graduate School of Economics, and of the European University Institute is gratefully acknowledged. A version of the paper has circulated with the title “A general algorithm for estimating structural VARs.”

Copyright © 2015 Fabio Canova and Fernando J. Pérez Forero. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](#). Available at <http://www.qeconomics.org>.
DOI: 10.3982/QE305

While sign restrictions offer a simple and intuitive way to impose theoretical constraints on the data, they are weak and identify a region of the parameter space. Furthermore, several implementation details are left to the researcher, making comparison exercises difficult to perform. Because of these features, some investigators still prefer to use “hard” nonrecursive restrictions, using the terminology of [Waggoner and Zha \(1999\)](#), even though these constraints are not theoretically abundant. There exist algorithms to estimate nonrecursive structural models [Waggoner and Zha \(2003\)](#) and to estimate recursive overidentified models ([Kociecki, Rubaszek, and Ca’ Zorzi \(2013\)](#)) identified with hard restrictions. However, their extension to TVC models is problematic.

TVC-VAR models are typically estimated using a Gibbs sampling routine, where a state space system is specified, the parameter vector is partitioned into blocks, and draws for the posterior are obtained by cycling through these blocks. When stochastic volatility is allowed for, an extended state space representation is used and one or more parameter blocks are added to the routine. If a recursive contemporaneous structure is assumed, one can sample the block of contemporaneous coefficients equation by equation, taking as given draws for the parameters belonging to previous equations. When the system is nonrecursive, such an approach disregards the cross-equation restrictions. Thus, the sampling must be done differently. To perform standard calculations, one also needs to assume that the covariance matrix of the contemporaneous parameters is block diagonal. When the structural model is overidentified, such an assumption may be implausible. However, relaxing the diagonality assumption complicates the computations, since the conditional distributions used in the Gibbs sampling do not necessarily have a known format.

This paper proposes a general framework to estimate a structural VAR (SVAR) that can handle time-varying coefficient or time-invariant models, identified with hard recursive or nonrecursive restrictions. The procedure can be used in systems that are just-identified or overidentified, and allows for both linear and nonlinear restrictions on the parameter space. Nonrecursive structures have been extensively used to accommodate models that are more complex than those permitted by recursive schemes. As shown, e.g., by [Gordon and Leeper \(1994\)](#), inference may crucially depend on whether a recursive or a nonrecursive scheme is used. In addition, although just-identified systems are easier to construct and estimate, overidentified models have a long history in the literature (see, e.g., [Leeper, Sims, and Zha \(1996\)](#), or [Sims and Zha \(1998\)](#)), and provide a natural framework to test interesting hypotheses.

The algorithm we design exploits the particular format of the structural model and follows [Primiceri’s \(2005\)](#) suggestion to use a Metropolis step within a Gibbs sampling routine to draw the vector of contemporaneous parameters. Because a number of important identification restrictions and a general law of motions of the coefficients imply a nonlinear state space representation for the structural model, we then nest our basic procedure into [Geweke and Tanizaki’s \(2001\)](#) approach to estimate general nonlinear state space models. Thus, we can deal with many structural systems in a compact and unified way without having to pay the computational costs of a full nonlinear simulation methodology.

We use the methodology to identify a monetary policy shock in an overidentified TVC system whose structure is similar to that employed by [Robertson and Tallman \(2001\)](#), [Waggoner and Zha \(2003\)](#), and [Sims and Zha \(2006\)](#). We show that there are time variations in the variance of the monetary policy shock and in the estimated contemporaneous coefficients. These variations translate into important changes in the transmission of monetary policy shocks. We show that time variations in the transmission of policy shocks are reduced when an alternative law of motion for the standard deviation of the shocks is used. We also show that when long- and short-run identification restrictions are employed, the transmission of monetary policy shocks in the 2000s is affected.

The paper is organized as follows. Section 2 builds up intuition describing the algorithm for a static SVAR with time-invariant coefficients and the identification restrictions that are allowed for. Section 3 considers a time-varying coefficients static SVAR. Section 4 presents the general algorithm applicable to nonrecursive, overidentified TVC-VAR models featuring stochastic volatility. Section 5 studies the transmission of monetary policy shocks. Section 6 concludes. Appendices and additional material are available in supplementary files on the journal website, <http://qeconomics.org/supp/305/supplement.pdf> and http://qeconomics.org/supp/305/code_and_data.zip.

2. A CONSTANT COEFFICIENTS STATIC SVAR

To build up the intuition, we start from a static SVAR with constant coefficients,

$$A(\alpha)y_t = \varepsilon_t; \quad \varepsilon_t \sim N(0, I), \quad (1)$$

where $t = 1, \dots, T$, y_t and ε_t are $M \times 1$ vectors, $A(\alpha)$ is a nonsingular $M \times M$ matrix that is assumed to be invertible for almost all α , and α is a vector of structural parameters. The likelihood function of (1) is

$$L(y^T | \alpha) = (2\pi)^{-MT/2} \det(A(\alpha))^T \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (A(\alpha)y_t)' (A(\alpha)y_t) \right\}. \quad (2)$$

Because of $\det(A(\alpha))^T$, the Jacobian of the transformation, (2) is nonlinear in α . Thus, the posterior of α will be nonstandard. Whenever the SVAR is just identified and the restrictions come in a triangular form, posterior draws for α can be obtained using draws of the reduced-form covariance matrix $\Omega(\alpha)^{-1} = A(\alpha)A(\alpha)'$. However, when the system is overidentified, $\Omega(\alpha)^{-1}$ is restricted and proper posterior inference needs to take these restrictions into account (see, e.g., [Sims and Zha \(1998\)](#)).

To describe our approach to sample α from the posterior when restrictions are not necessarily just identifying and recursive, we proceed in two steps. First, we reparameterize the model and present a Metropolis algorithm. Second, we show the type of identification restrictions that are compatible with the setup.

2.1 The reparameterization and the algorithm

Vectorizing (1) produces $\text{vec}(A(\alpha)y_t) = \text{vec}(\varepsilon_t) = \varepsilon_t$. As in Amisano and Giannini (1997), assume that $\text{vec}(A(\alpha)) = S_A\alpha + s_A$, where S_A and s_A are matrices of 1s and 0s. Using $\text{vec}(A(\alpha)y_t) = (y_t' \otimes I)(S_A\alpha + s_A)$, the model can be expressed as

$$\tilde{y}_t = Z_t\alpha + \varepsilon_t, \quad (3)$$

where $\tilde{y}_t \equiv (y_t' \otimes I)s_A$ and $Z_t \equiv -(y_t' \otimes I)S_A$. The likelihood function is

$$L(\tilde{y}^T | \alpha) = (2\pi)^{-MT/2} (\det D(\alpha))^T \exp \left\{ -\frac{1}{2} \sum_{t=1}^T [\tilde{y}_t - Z_t\alpha]' [\tilde{y}_t - Z_t\alpha] \right\}, \quad (4)$$

where $D(\alpha) = \frac{\partial[\text{vec}(A(\alpha)y_t)]}{\partial y_t'} = D_y + D_z(\alpha)$, $\text{vec}(D_y) = s_A$, and $\text{vec}(D_z(\alpha)) = S_A\alpha$.

The reparameterization in (3) makes it easy to design a proposal distribution to be used in a Metropolis routine. Thus, let

$$\alpha^* = \left[\sum_{t=1}^T Z_t' Z_t \right]^{-1} \left[\sum_{t=1}^T Z_t' \tilde{y}_t \right] \quad (5)$$

and

$$P^*(\alpha^*) = \left[\sum_{t=1}^T Z_t' (SSE)^{-1} Z_t \right]^{-1}, \quad (6)$$

where $SSE = \sum_{t=1}^T (\tilde{y}_t - Z_t\alpha^*)(\tilde{y}_t - Z_t\alpha^*)'$. Set $\alpha^0 = \alpha^*$ and, for $i = 1, 2, \dots, G$, perform the following steps:

1. Draw a candidate $\alpha^\dagger \sim p_*(\alpha^i | \alpha^{i-1}) = t(\alpha^{i-1}, rP^*(\alpha^{i-1}), \nu)$, where $r > 0$, $\nu \geq 4$, and t is a t -distribution.

2. Compute $\theta = \frac{p(\alpha^\dagger | \tilde{y}^T) \cdot p_*(\alpha^i | \alpha^{i-1})}{p(\alpha^{i-1} | \tilde{y}^T) \cdot p_*(\alpha^{i-1} | \alpha^i)}$, where $p(\cdot | \tilde{y}^T) = L(\tilde{y}^T | \cdot) [p(\cdot) \mathcal{I}(\alpha)]$ is the posterior kernel of α^\dagger and α^{i-1} , and $\mathcal{I}(\alpha)$ is an indicator function.

3. Draw a $v \sim U(0, 1)$; set $\alpha^i = \alpha^\dagger$ if $v < \theta$ and $\alpha^i = \alpha^{i-1}$ otherwise.

Note a few facts about the algorithm. First, a t -distribution with a small number of degrees of freedom is chosen to explore the tails of the posterior; when ν is large the proposal resembles a normal distribution. Second, and more importantly, the α vector is jointly sampled and the covariance matrix of $P^*(\alpha)$ is nondiagonal. As we discuss later, these features distinguish our algorithm from those present in the literature and provide the flexibility needed to accommodate a variety of structural models. Third, we need a Metropolis step since (5) and (6) ignore the Jacobian term $\det(D(\alpha))$ appearing in (4). In general, our proposal will work well whenever $D(\alpha)$ does not strongly affect

the shape of the posterior, as seems to be the case in the applications discussed in Section 5.

Kociecki, Rubaszek, and Ca' Zorzi (2013) have derived a closed-form solution for the posterior of α under the assumption that $\det(D(\alpha)) = 1$. It turns out that their posterior collapses to our proposal when the prior for α is diffuse. Baumeister and Hamilton (2013) obtain an analytic expression for the posterior for α under a slightly different model setup when sign restrictions are used for identification. They show that, asymptotically, the posterior for α is the prior restricted to the set of structural models that diagonalize the covariance matrix $\Omega(\alpha)$. The algorithm they employ to draw from the posterior of α is similar to the one described in this subsection.

2.2 Identification restrictions

The reparameterized model (3) can deal with linear restrictions on α (both of exclusion and nonexclusion types) and with particular types of nonlinear restrictions on α . We present a few examples for illustration. We focus on overidentified systems because just-identified systems only require adjustments of S_A and of s_A .

2.2.1 *Short-run linear restrictions* Suppose $A(\alpha)$ features both exclusion and nonexclusion linear restrictions:

$$A(\alpha) = \begin{bmatrix} 1 & 0 & -\alpha_2 \\ \alpha_1 & 1 & 0 \\ 0 & \alpha_2 & 1 \end{bmatrix}.$$

Then

$$\text{vec}(A(\alpha)) \equiv \begin{bmatrix} 1 \\ \alpha_1 \\ 0 \\ 0 \\ 1 \\ \alpha_2 \\ -\alpha_2 \\ 0 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{S_A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\alpha} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{s_A}.$$

2.2.2 *Short-run nonlinear restrictions* Suppose $A(\alpha)$ features exclusion restrictions and nonlinear constraints:

$$A(\alpha) = \begin{bmatrix} 1 & 0 & \alpha_3 \\ \alpha_1 & 1 & 0 \\ 0 & (\alpha_2 + 1)^2 & 1 \end{bmatrix}. \tag{7}$$

Then

$$\text{vec}(A(\alpha)) \equiv \begin{bmatrix} 1 \\ \alpha_1 \\ 0 \\ 0 \\ 1 \\ (\alpha_2 + 1)^2 \\ \alpha_3 \\ 0 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S_A} \underbrace{\begin{bmatrix} \alpha_1 \\ (\alpha_2 + 1)^2 \\ \alpha_3 \end{bmatrix}}_{f(\alpha)} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{s_A}.$$

If we define $\tilde{\alpha}_2 \equiv (\alpha_2 + 1)^2$, $f(\alpha) = (\alpha_1, \tilde{\alpha}_2, \alpha_3)'$ is still a linear vector-valued function. Given posterior draws for $\tilde{\alpha}_2$, we can recover $\alpha_2 = \sqrt{\tilde{\alpha}_2} - 1$, provided $\tilde{\alpha}_2 \geq 0$. Hence, certain nonlinear restrictions on α can be handled with an additional accept/reject step. A similar approach can be used in the slightly more general case in which, for example, $f(\alpha) = [\alpha_1, (\alpha_2 + 2\alpha_3)^2, \alpha_3]'$. Here, we set $\tilde{\alpha}_2 \equiv (\alpha_2 + 2\alpha_3)^2 \geq 0$, and use draws of $\tilde{\alpha}_2 \geq 0$ and α_3 to obtain $\alpha_2 = \sqrt{\tilde{\alpha}_2} - 2\alpha_3$.

An accept/reject step will not work in the situation

$$A(\alpha) = \begin{bmatrix} 1 & 0 & \alpha_1\alpha_2 - 1 \\ \alpha_1 & 1 & 0 \\ 0 & \alpha_2 & 1 \end{bmatrix}. \tag{8}$$

Here

$$\text{vec}(A(\alpha)) \equiv \begin{bmatrix} 1 \\ \alpha_1 \\ 0 \\ 0 \\ 1 \\ \alpha_2 \\ \alpha_1\alpha_2 - 1 \\ 0 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S_A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_1\alpha_2 - 1 \end{bmatrix}}_{f(\alpha)} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{s_A}.$$

Adding an inequality constraint does not help, since the third component of $f(\alpha)$ does not have independent variations. Still, if we set $\tilde{f}(\alpha) = (\alpha_1, \alpha_2)$, draws for $\alpha_1\alpha_2 - 1$ can be obtained from the draws of (α_1, α_2) . Thus, the posterior of $f(\alpha)$ can be simulated using the subset of the coefficients with independent variations.

2.2.3 Long-run restrictions Long-run restrictions generally imply nonlinear constraints on the parameters of a VAR. As the editor has pointed out, these restrictions could be dealt with in our framework if $\det(A(\alpha, B)) = \det(A(\alpha))$, that is, if the Jacobian of

the transformation is independent of the matrix of reduced-form autoregressive coefficients B . In this case, draws for B can be made from standard conditional distributions. In general, however, this is not the case. To see this, consider

$$A(\alpha)y_t = A_+y_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0, I). \quad (9)$$

The corresponding VAR is

$$y_t = By_{t-1} + [A(\alpha)]^{-1} \varepsilon_t, \quad (10)$$

where $B \equiv [A(\alpha)]^{-1}A_+$, and the (long-run) cumulative matrix is

$$\mathcal{D} \equiv (I_M - B)^{-1}[A(\alpha)]^{-1}. \quad (11)$$

Let

$$A(\alpha) = \begin{bmatrix} 1 & \alpha_3 & \alpha_5 \\ \alpha_1 & 1 & \alpha_6 \\ \alpha_2 & \alpha_4 & 1 \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} \mathcal{D}_{11} & \mathcal{D}_{12} & \mathcal{D}_{13} \\ \mathcal{D}_{21} & \mathcal{D}_{22} & \mathcal{D}_{23} \\ \mathcal{D}_{31} & \mathcal{D}_{32} & \mathcal{D}_{33} \end{bmatrix} \quad (12)$$

and let b_{ij} be the typical elements of $(I_M - B)^{-1}$. Then

$$\mathcal{D} = \frac{1}{\det[A(\alpha)]} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \times \begin{bmatrix} 1 - \alpha_4\alpha_6 & \alpha_4\alpha_5 - \alpha_3 & \alpha_3\alpha_6 - \alpha_5 \\ \alpha_2\alpha_6 - \alpha_1 & 1 - \alpha_2\alpha_5 & \alpha_1\alpha_5 - \alpha_6 \\ \alpha_1\alpha_4 - \alpha_2 & \alpha_2\alpha_3 - \alpha_4 & 1 - \alpha_1\alpha_3 \end{bmatrix}$$

with $\det[A(\alpha)] \neq 0$. Assume, for example, $\alpha_4 = \mathcal{D}_{21} = \mathcal{D}_{31} = 0$ so that there are both short- and long-run (zero) restrictions. The equality $\mathcal{D}_{21} = 0$ implies $-b_{21}(\alpha_4\alpha_6 - 1) - b_{23}(\alpha_2 - \alpha_1\alpha_4) - b_{22}(\alpha_1 - \alpha_2\alpha_6) = 0$. Since $\alpha_4 = 0$, we have $b_{21} - b_{23}\alpha_2 - b_{22}(\alpha_1 - \alpha_2\alpha_6) = 0$. Similarly, $\mathcal{D}_{31} = 0$ implies $-b_{31}(\alpha_4\alpha_6 - 1) - b_{33}(\alpha_2 - \alpha_1\alpha_4) - b_{32}(\alpha_1 - \alpha_2\alpha_6) = 0$ or $b_{31} - b_{33}\alpha_2 - b_{32}(\alpha_1 - \alpha_2\alpha_6) = 0$. Thus, there are nonlinear constraints that draws of b_{ij} and of α_i must satisfy, and $\det(A(\alpha, B))$ is generally not independent of B . We discuss in Appendix F (in the Supplement) how to deal with these types of systems.

2.2.4 Sign restrictions Although sign restrictions are not the focus of this paper, it is straightforward to show that they can be handled with the algorithm of Section 2.1. Let $A(\alpha)$ be a general matrix with no exclusion restrictions and inequality constraints on, say, the first column. Then one draws α s as above and checks whether the first column satisfies the required constraints. Thus, sign restrictions can be dealt with in the same accept/reject step used in Section 2.2.2. A mixture of zero and sign restrictions, as those suggested by [Arias, Waggoner, and Rubio Ramírez \(2014\)](#), can be handled in the same way.

3. TIME-VARYING COEFFICIENTS STATIC SVAR

Before we move to the standard TVC-SVAR models used in the literature, it is useful to study the intermediate step of a static TVC-SVAR. The model is

$$A(\alpha_t)y_t = \varepsilon_t, \quad (13)$$

$$\alpha_t = \alpha_{t-1} + u_t, \quad (14)$$

where α_0 is given and $A(\alpha_t)$ is assumed to be invertible for almost all α_t , for all t . To handle the nonlinear restrictions described in Section 2.2.2, we consider a slightly more general reparameterization than the one used in Section 2.1. Let $\alpha_t \in X \subset R^k$, where X is “sufficiently large”—we need X to have this feature because simulations may proceed by rejecting draws—and let $f: X \rightarrow R^{k'}$ be continuous, where k' may be different from k . Then the model is reparameterized as

$$\tilde{y}_t = Z_t f_t + \varepsilon_t; \quad \varepsilon_t \sim N(0, I), \quad (15)$$

$$f_t = f_{t-1} + \zeta_t; \quad \zeta_t \sim N(0, V), \quad (16)$$

where V is a full rank, positive definite matrix, $f_t \equiv f(\alpha_t)$, $\tilde{y}_t \equiv (y'_t \otimes I) s_A$, and $Z_t \equiv -(y'_t \otimes I) S_A$. To obtain the joint distribution of $f^T \equiv \{f_t\}_{t=1}^T$ and of V , one can use a Gibbs sampler as long as $p(f^T | \tilde{y}^T, V)$ and $p(V | \tilde{y}^T, f^T)$ are available. Given standard prior assumptions, $p(V | \tilde{y}^T, f^T)$ has an inverted Wishart format and it is easy to draw from.

The conditional posterior $p(f^T | \tilde{y}^T, V)$ cannot be computed in a standard fashion since Z_t is neither exogenous nor predetermined. Still, it is relatively easy to compute the kernel of this conditional posterior, following the same steps outlined in Section 2.1. That is, we compute the likelihood of the reparameterized model, we generate a proposal $p(f_{t+1} | f_t, V)$, and decide whether to accept or reject the draw using a simple Metropolis–Hastings (MH) rule.

The likelihood for the reparameterized model is

$$L(\tilde{y}^T | f^T, V) = (2\pi)^{-MT/2} (\det D(\alpha))^{-T} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\tilde{y}_t - Z_t f_t)' (\tilde{y}_t - Z_t f_t) \right\}, \quad (17)$$

where $\text{vec}(D(\alpha_t)) = s_A + S_A f_t$. Given a prior, the posterior kernel can be easily computed.

The proposal density can be calibrated using Kalman smoothed estimates of f_t and P_t obtained using a version of the model that ignores the Jacobian of the transformation. Thus, given $f_{0|0}$ and $P_{0|0}$, one constructs Kalman filter updated estimates of f_t and of its covariance matrix for each $t = 1, \dots, T$ as

$$f_{t|t} = f_{t|t-1} + K_t [\tilde{y}_t - Z_t f_{t|t-1}], \quad (18)$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} Z'_t \Omega_t^{-1} Z_t P'_{t|t-1}, \quad (19)$$

where $f_{t|t-1} = f_{t-1|t-1}$, $P_{t|t-1} = P_{t-1|t-1} + V$, $K_t = P_{t|t-1} Z'_t \Omega_t^{-1}$, and $\Omega_t = Z'_t P_{t|t-1} Z_t + I$. Smoothed estimates are $f_{T|T}^* = f_{T|T}$, $P_{T|T}^* = P_{T|T}$ and

$$f_{t|t+1}^* = f_{t|t} + P_{t|t} Z'_t P_{t+1|t}^{-1} (f_{t+1|t+2}^* - Z'_t f_{t|t}), \quad (20)$$

$$P_{t|t+1}^* = P_{t|t} - P_{t|t} Z'_t P_{t+1|t}^{-1} Z_t P'_{t|t-1}, \quad t = T-1, \dots, 1. \quad (21)$$

3.1 The basic algorithm

Initially set $f_{0|0}$ and $P_{0|0}$, and, for $i = 1, 2, \dots, G$, perform the following steps.

Step 1. Given (\tilde{y}^T, V^{i-1}) , compute $\{f_{t|t+1}^{*(i-1)}\}_{t=1}^T$ and $\{P_{t|t+1}^{*(i-1)}\}_{t=1}^T$ as in (20) and (21).

1. For $t = 1, \dots, T$, draw a candidate $f_t^\dagger \sim p_*(f_t^i | f_t^{i-1}) = t(f_{t|t+1}^{*(i-1)}, rP_{t|t+1}^{*(i-1)}, \nu)$, $r > 0$, $\nu \geq 4$. Set $(f^\dagger)^T \equiv \{f_t^\dagger\}_{t=1}^T$; $p_*((f^i)^T | (f^{i-1})^T) = \prod_{t=1}^T p_*(f_t^i | f_t^{i-1})$.

2. Compute

$$\theta = \frac{p((f^\dagger)^T | \tilde{y}^T) \cdot p_*((f^{i-1})^T | (f^i)^T)}{p((f^{i-1})^T | \tilde{y}^T) \cdot p_*((f^i)^T | (f^{i-1})^T)},$$

where $p(\cdot | \tilde{y}^T) = L(\tilde{y}^T | (\cdot, V^{i-1})) \cdot [p(\cdot)^T \mathcal{I}(f)]$ is the posterior kernel of $(f^\dagger)^T$ and $(f^{i-1})^T$, and $\mathcal{I}(f)$ is an indicator restricting the prior distribution.

3. Draw $v \sim U(0, 1)$; set $(f^i)^T = (f^\dagger)^T$ if $v < \theta$ and $(f^i)^T = (f^{i-1})^T$ otherwise.

Step 2. Given $(\tilde{y}^T, (f^i)^T)$, draw $(V^i)^{-1} \sim p((V^i)^{-1} | (f^i)^T, \tilde{y}^T) = W(\bar{v}_V, \bar{V}^{-1})$, where

$$\begin{aligned} \bar{v}_V &= T + \underline{v}_V, \\ \bar{V}^{-1} &= \left[\underline{V} + \sum_{t=1}^T (f_t^i - f_{t-1}^i)(f_t^i - f_{t-1}^i)' \right]^{-1} \end{aligned}$$

and \underline{v}_V and \underline{V} are prior parameters.

Given the structure of the problem, if coefficients are constant, $f_t^* = \alpha^*$, $P_t^* = P^*$, all $t = 1, \dots, T$, and the algorithm collapses to the one described in Section 2.1.

4. A STANDARD TIME-VARYING COEFFICIENTS SVAR

Assume that an $M \times 1$ vector of nonstationary variables y_t , $t = 1, \dots, T$, can be represented with a finite order autoregression of the form

$$y_t = B_{0,t}C_t + B_{1,t}y_{t-1} + \dots + B_{p,t}y_{t-p} + u_t, \quad (22)$$

where $B_{0,t}$ is a matrix of coefficients on an $\bar{M} \times 1$ vector of deterministic variables C_t ; $B_{j,t}$, $j = 1, \dots, p$, are square matrices containing the coefficients on the lags of the endogenous variables, and $u_t \sim N(0, \Omega_t)$, where Ω_t is symmetric, positive definite, and full rank for every t . For the sake of presentation, we do not include exogenous variables, but the setup can be easily extended to account for them. Let the structural shocks be $\varepsilon_t \sim N(0, I)$ and let $u_t = A_t^{-1} \Sigma_t \varepsilon_t$, where $A_t \equiv A(\alpha_t)$ is the contemporaneous coefficients matrix, α_t is a vector of free parameters, and $\Sigma_t = \text{diag}\{\sigma_{m,t}\}$ contains the standard deviations of the structural shocks at t in the main diagonal. The SVAR is

$$y_t = X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \quad (23)$$

where $X_t' = I \otimes [C_t', y_{t-1}', \dots, y_{t-p}']$ and $B_t = [\text{vec}(B_{0,t})', \text{vec}(B_{1,t})', \dots, \text{vec}(B_{p,t})']'$ are an $M \times K$ matrix and a $K \times 1$ vector, $K = \bar{M} \times M + pM^2$. It is typical to assume

$$B_t = B_{t-1} + v_t, \quad (24)$$

$$\alpha_t = \alpha_{t-1} + u_t, \quad (25)$$

$$\log(\sigma_{m,t}) = \log(\sigma_{m,t-1}) + \eta_{m,t}. \quad (26)$$

Letting $\eta_t = [\eta_{1t}, \dots, \eta_{Mt}]$, set

$$\mathcal{V} = \text{Var} \left(\begin{bmatrix} \varepsilon_t \\ v_t \\ u_t \\ \eta_t \end{bmatrix} \right) = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \quad (27)$$

where Q , V , and W are full rank matrices.

Thus, the setup captures time variations in (i) the lag structure (see (24)), (ii) the contemporaneous reaction parameters (see (25)), and (iii) the structural variances (see (26)). Common patterns of time variations within blocks are possible if the rank of Q , V , or W is reduced. Models with breaks at a specific date can be accommodated by adding restrictions on (24)–(26); see Canova, Ciccarelli, and Ortega (2012).

4.1 Relaxing standard assumptions

Consider the concentrated model obtained with estimates of the reduced-form VAR coefficients \widehat{B}_t :

$$A(\alpha_t)(y_t - X_t' \widehat{B}_t) \equiv A(\alpha_t) \widehat{y}_t = \Sigma_t \varepsilon_t. \quad (28)$$

Let $\text{vec}(A(\alpha_t)) = S_A f_t + s_A$, where S_A and s_A are matrices with 1s and 0s of dimensions $M^2 \times \dim(f(\alpha))$ and $M^2 \times 1$, respectively, and, as before, $f_t \equiv f(\alpha_t)$. The concentrated model can be reparameterized as

$$(\widehat{y}_t' \otimes I)(S_A f_t + s_A) = \Sigma_t \varepsilon_t$$

and the state space is composed of

$$\tilde{y}_t = Z_t f_t + \Sigma_t \varepsilon_t,$$

$$f_t = f_{t-1} + \zeta_t,$$

and (26), where $\tilde{y}_t \equiv (\widehat{y}_t' \otimes I_M) s_A$ and $Z_t \equiv -(\widehat{y}_t' \otimes I_M) S_A$. Given $(\widehat{B}^T, \Sigma^T, \mathcal{V})$, we need to draw $f^T \equiv \{f_t\}_{t=1}^T$ from $p(f^T | \widehat{y}^T, \Sigma^T, \mathcal{V}, \widehat{B}^T)$.

Standard algorithms (see Primiceri (2005)) partition f_t into blocks associated with each equation, say $f_t = [f_t^{1'}, f_t^{2'}, \dots, f_t^{M'}]'$, and assume that these blocks are indepen-

dent, so that $\mathcal{V} = \text{diag}(\mathcal{V}_1, \dots, \mathcal{V}_M)$. Then

$$p(f^T | \tilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B}^T) = \prod_{m=2}^M p((f^m)^T | (f^{m-1})^T, \tilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B}^T) \times p((f^1)^T | \tilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B}^T). \quad (29)$$

Thus, for each equation m , the coefficients in equation $m - j$, $j \geq 1$, are treated as predetermined and changes in coefficients across equations are uncorrelated. The setup is convenient because equation by equation estimation is possible. Since the factorization does not necessarily have an economic interpretation, it may make sense to assume that the innovations in the f_t blocks are uncorrelated. However, if we insist that each element of α_t has some economic meaning, the diagonality of \mathcal{V} is no longer plausible. For example, if α_t contains policy and nonpolicy parameters, it will be hard to assume that nonpolicy parameters are invariant to changes in the policy parameters (see, e.g., Lakdawala (2012)).

The algorithm described in the previous section relaxes both assumptions, that is, the vector f_t is jointly drawn and \mathcal{V} is not necessarily block diagonal. This modification allows us to deal with recursive, nonrecursive, just-identified, or overidentified structural models in a unified framework. There are, however, computational costs, since systemwide estimation methods are now needed.

4.2 The general algorithm

Set initial values $((B^0)^T, (f^0)^T, (s^0)^T, (\Sigma^0)^T, \mathcal{V}^0)$, where s is a J -dimensional vector of discrete indicator variables described below. Then, for $i = 1, \dots, G$, we have the following steps.

1. Draw $(B^i)^T$ from $p((B^i)^T | \tilde{y}^T, (f^{i-1})^T, (s^{i-1})^T, (\Sigma^{i-1})^T, \mathcal{V}^{i-1}) \cdot I_B(B_i^T)$, where $I_B(\cdot)$ truncates the posterior to insure stationarity of impulse responses. The variable $p(\cdot)$ is normal and can be computed using Kalman filter recursions and a multi-move Carter and Kohn (1994) or a single-move Koop and Potter (2011) strategy.

2. Draw $(f^i)^T$ from $p((f^i)^T | \tilde{y}^T, (s^{i-1})^T, (\Sigma^{i-1})^T, \mathcal{V}^{i-1}, (B^i)^T)$ using the approach described in Section 3.1.

3. Given $(\tilde{y}^T, (B^i)^T, (f^i)^T)$, the model is linear and composed of

$$\widehat{A}(\alpha_t) \widehat{y}_t \equiv y_t^{**} = \Sigma_t \varepsilon_t \quad (30)$$

and (26), where $\text{vec}(\widehat{A}(\alpha_t)) = S_A f_t^i + s_A$, and $\widehat{A}(\alpha_t)$ is the matrix of the contemporaneous coefficients matrix evaluated at the current draw $(f^i)^T$. For the m th equation of the model, we have

$$y_{m,t}^* = \log[(y_{m,t}^{**})^2 + \bar{c}] \approx 2 \log(\sigma_{m,t}) + \log \varepsilon_{m,t}^2, \quad (31)$$

where \bar{c} is a small constant. Since $\varepsilon_{m,t}$ is Gaussian, $\log \varepsilon_{m,t}^2$ is $\log(\chi^2)$ distributed and can be approximated by a mixture of normals. Conditional on s_t , the indicator for the mix-

ture of normals, the model is linear and Gaussian. Thus, as in [Del Negro and Primiceri \(2013\)](#), we proceed as follows:

(a) Given $((y^{**})^T, (B^i)^T, (f^i)^T, (\Sigma^{i-1})^T)$, draw $(s^i)^T$ and compute

$$P(s_{m,t} = j | y_{m,t}^{**}, \log(\sigma_{m,t})) \propto q_j \times \phi\left(\frac{y_{m,t}^{**} - 2\log(\sigma_{m,t}) - \eta_j + 1.2704}{\gamma_j}\right),$$

where $j = 1, \dots, J$, $\phi(x)$ is the normal density, q_j is a set of weights, x is the standardized error term $\log \varepsilon_{m,t}^2$, and η_j and γ_j are the mean and the standard deviation of the j th mixture. Draw $u \sim U(0, 1)$. Set $s_{m,t} = j$ if $P(s_{m,t} \leq j - 1 | y_{m,t}^{**}, \log(\sigma_{m,t})) < u \leq P(s_{m,t} \leq j | y_{m,t}^{**}, \log(\sigma_{m,t}))$.

(b) Given $(\tilde{y}^T, (B^i)^T, (f^i)^T, (s^i)^T)$, use standard Kalman smoother recursions to draw $\{\Sigma_t\}_{t=1}^T$ from (22)–(27), where s^T is obtained in step (a). To ensure independence of the structural variances, each $\sigma_{m,t}$ is sampled assuming a diagonal W .

4. Draw ν^i from $p(\nu^i | \tilde{y}^T, (f^i)^T, (s^i)^T, (\Sigma^i)^T, (B^i)^T)$. The variable ν^i is sampled assuming that each block follows an independent inverted Wishart distribution.

Then use $(B^i)^T, (f^i)^T, (s^i)^T, (\Sigma^i)^T, \nu^i$ as initial values and repeat the sampling.

4.3 Extensions

In the setup used so far, we are constrained about the identification restrictions we can employ; for example, as noted in [Section 2.2.3](#), long-run restrictions produce a nonlinear model for α_t and B_t . Recent identification procedures that restrict certain medium-term multipliers (for example, the maximum effect of a monetary shock on output occurs x months after the disturbances) or the variance decomposition (as it is done in the news shock literature (see, e.g., [Barsky and Sims \(2011\)](#))) also generate a nonlinear model for (α_t, B_t) . In addition, while it is standard to use a log-linear setup for the law of motion of the volatilities, one may want to consider generalized autoregressive conditional heteroscedasticity (GARCH) or Markov switching specifications, which also generate nonlinear or nonnormal laws of motion of the coefficients.

In all these cases, the sequential Monte Carlo methods discussed in [Creel \(2012\)](#) and [Herbst and Schorfheide \(2014\)](#) are the natural candidates to estimate the structural nonlinear model. These methods, however, are computationally intensive and there are still a number of theoretical and practical issues that need to be resolved. Thus, an intermediate approach, which still allows us to deal with all these cases, but is much less computationally demanding, could be of use.

We describe in Appendices E and F how the setup of [Section 4.2](#) needs to be modified to deal with cases where the law of motion of the coefficients or the identification restrictions come in a nonlinear form. Basically, we nest our procedure into [Geweke and Tanizaki's \(2001\)](#) algorithm for estimating nonlinear, non-Gaussian state spaces modified to account for the fact that we are dealing with models with time-varying parameters.

5. THE TRANSMISSION OF MONETARY POLICY SHOCKS

We employ our procedure to study the transmission of monetary policy shocks in an overidentified structural TVC-VAR when short-run zero restrictions are used. We are interested in knowing whether the propagation of policy shocks has changed over time and in what way. To robustify inference, we study how our conclusions change when the law of motion of the standard deviation of the shocks is altered and when mixed long- and short-run restrictions are used to identify policy shocks. We also evaluate the merits of different methods to draw the autoregressive parameters of the model.

5.1 *The SVAR model*

The vector of endogenous variables is $y_t = (\text{GDP}_t, P_t, U_t, R_t, M_t, P\text{com}_t)'$, where GDP_t is a measure of aggregate output, P_t is a measure of aggregate prices, U_t is the unemployment rate, R_t is the nominal interest rate, M_t is a monetary aggregate, and $P\text{com}_t$ represents a commodity price index. The structure of $A(\alpha_t)$ is as in Table 1, where X indicates a nonzero coefficient.

The structural model is identified via exclusion restrictions as follows:

1. *Information equation.* Commodity prices ($P\text{com}_t$) convey information about recent developments in the economy. Therefore, they react contemporaneously to all structural shocks.

2. *Money demand equation.* Within the period money balances are a function of the structural shocks to core macroeconomic variables (R_t, GDP_t, P_t).

3. *Monetary policy equation.* The interest rate (R_t) is used as an instrument for controlling the money supply (M_t). No other variable contemporaneously affects this equation.

4. *Nonpolicy block.* Following [Bernanke and Blinder \(1992\)](#), the nonpolicy variables (GDP_t, P_t, U_t) react to policy, money, or informational changes only with a delay. This setup can be formalized by assuming that the private sector uses only lagged values of these variables as states or that private decisions have to be taken before the current values of these variables are known. The relationship between the variables in the block is left unmodeled and, for simplicity, a recursive structure is assumed.

TABLE 1. Identification restrictions.

Reduced form\Structural	GDP_t	P_t	U_t	R_t	M_t	$P\text{com}_t$
Nonpolicy 1	1	0	0	0	0	0
Nonpolicy 2	X	1	0	0	0	0
Nonpolicy 3	X	X	1	0	0	0
Monetary policy	0	0	0	1	X	0
Money demand	X	X	0	X	1	0
Information	X	X	X	X	X	1

In this setup, it is easy to understand why independence in coefficients of different equations is unappealing: changes in policy and nonpolicy coefficients are likely to be correlated. Let $\varepsilon_t = [\varepsilon_t^1, \varepsilon_t^2, \varepsilon_t^3, \varepsilon_t^{\text{mp}}, \varepsilon_t^{\text{md}}, \varepsilon_t^i]'$ be the vector of structural innovations. The structural model is

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{1,t} & 1 & 0 & 0 & 0 & 0 \\ \alpha_{2,t} & \alpha_{5,t} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11,t} & 0 \\ \alpha_{3,t} & \alpha_{6,t} & 0 & \alpha_{9,t} & 1 & 0 \\ \alpha_{4,t} & \alpha_{7,t} & \alpha_8 & \alpha_{10,t} & \alpha_{12,t} & 1 \end{bmatrix}}_{A(\alpha_t)} \times \begin{bmatrix} \text{GDP}_t \\ P_t \\ U_t \\ R_t \\ M_t \\ \text{Pcom}_t \end{bmatrix} = A_t^+(L) \begin{bmatrix} \text{GDP}_{t-1} \\ P_{t-1} \\ U_{t-1} \\ R_{t-1} \\ M_{t-1} \\ \text{Pcom}_{t-1} \end{bmatrix} + \Sigma_t \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \varepsilon_t^3 \\ \varepsilon_t^{\text{mp}} \\ \varepsilon_t^{\text{md}} \\ \varepsilon_t^i \end{bmatrix}, \quad (32)$$

where $A_t^+(L)$ is a function of $A(\alpha_t)$ and B_t , and we normalize the main diagonal of $A(\alpha_t)$ so that the left-hand side of each equation corresponds to the dependent variable. Finally,

$$\Sigma_t = \begin{bmatrix} \sigma_t^1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_t^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_t^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_t^{\text{mp}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_t^{\text{md}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_t^i \end{bmatrix}$$

is the matrix of standard deviations of the structural shocks.

The structural model (32) is nonrecursive and overidentified by three restrictions. Overidentification obtains because the policy equation is different from the Taylor rule generally employed in the literature. It is easy to check (see Appendix A) that the (constant coefficient version of the) system is globally identified and, therefore, is suitable for interesting policy experiments.

5.2 The prior and computation details

The VAR is estimated with two lags; this is what the Bayes information criterion (BIC) criteria selects for the constant coefficient version of the model. The priors are proper, conjugate for computational convenience, and given by $B_0^{\text{prior}} \sim N(\bar{\mathbf{B}}, 4 \cdot \overline{\mathbf{VB}})$, $Q^{\text{prior}} \sim \text{IW}(k_Q^2 \cdot \overline{\mathbf{VB}}, (1 + K))$, $\alpha_0^{\text{prior}} \sim N(\bar{\alpha}, \text{diag}(\text{abs}(\bar{\alpha})))$, $S^{\text{prior}} \sim \text{IW}(k_S^2 \cdot \text{diag}(\text{abs}(\bar{\alpha})), (1 + \dim \alpha))$, $\log(\sigma_0)^{\text{prior}} \sim N(\bar{\sigma}, 10 \cdot I_M)$, and $W_i^{\text{prior}} \sim \text{IW}(k_W^2, 1 + 1)$, $i = 1, \dots, M$.

To calibrate the hyperparameters, we estimate a constant coefficient version of the model using the first 40 observations as a training sample: $\bar{\mathbf{B}}$ and $\sqrt{\mathbf{V}}\bar{\mathbf{B}}$ are estimated with ordinary least squares (OLS), and $\bar{\alpha}$ and $\bar{\sigma}$ are estimated with maximum likelihood using 100 different starting points and the constant coefficient version of the model. We set $k_Q^2 = 0.5 \times 10^{-4}$, $k_S^2 = 1 \times 10^{-3}$, $k_W^2 = 1 \times 10^{-4}$, and $J = 7$. We generate 150,000 draws, discard the first 100,000, and use 1 out of every 100 of the remaining for inference. The results we present are independent of whether thinning is performed. Convergence was checked using standard statistics. Draws for B_t are discarded if the stability condition fails. The function $I(f)$ used to eliminate outlier draws is uniform over the interval $(-20, 20)$. In our application, all draws were inside the bounds. The acceptance rate for the Metropolis step is 35.6%. Since the structural model has $M = 6$ and $\dim(\alpha_t) = 12$, S_A and s_A are written

$$S_A = \begin{bmatrix} \mathbf{0}_{1 \times \dim(\alpha)} \\ [1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-1)}] \\ [\mathbf{0}_{1 \times (2-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-2)}] \\ \mathbf{0}_{1 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (3-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-3)}] \\ [\mathbf{0}_{1 \times (4-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-4)}] \\ \mathbf{0}_{2 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (5-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-5)}] \\ \mathbf{0}_{1 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (6-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-6)}] \\ [\mathbf{0}_{1 \times (7-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-7)}] \\ \mathbf{0}_{5 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (8-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-8)}] \\ \mathbf{0}_{4 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (9-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-9)}] \\ [\mathbf{0}_{1 \times (10-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-10)}] \\ \mathbf{0}_{3 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (11-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-11)}] \\ \mathbf{0}_{1 \times \dim(\alpha)} \\ [\mathbf{0}_{1 \times (12-1)} \quad 1 \quad \mathbf{0}_{1 \times (\dim(\alpha)-12)}] \\ \mathbf{0}_{6 \times \dim(\alpha)} \end{bmatrix},$$

$$s_A = [e'_1, e'_2, e'_3, e'_4, e'_5, e'_6]'$$

where e_i are vectors in \mathbb{R}^M with

$$e_i = [e_{i,j}]_{j=1}^M \quad \text{such that } e_{i,j} = \begin{cases} 1, & j = i, \\ 0, & j \neq i. \end{cases}$$

5.3 The data

The data come from the *International Financial Statistics* (IFS) data base at the International Monetary Fund and from the Federal Reserve Board (www.imfstatistics.org/imf/about.asp and www.federalreserve.gov/econresdata/releases/statisticsdata.htm, res-

pectively). The sample is 1959:I–2005:IV. We stop at this date to avoid the last financial crisis and to compare our results to those of Sims and Zha (2006), who use a (restricted) Markov switching model over the same sample. The gross domestic product (GDP) deflator, the unemployment rate, the aggregate gross domestic product index (volume, base 2005 = 100), the commodity prices index, for money we use the M2 series from IFS; the Federal Funds rate is from the Fed. All the variables are expressed in year-to-year rate changes (i.e., $y_t^* = \log(y_t) - \log(y_{t-4})$), except for the Federal Funds and the unemployment rate, and are standardized (i.e., we use $(y_t^* - E(y_t^*)) / \text{std}(y_t^*)$) to have all the variables on the same scale.

5.4 Comparing routines for drawing B^T

To draw B^T , one can employ Carter and Kohn's (1994) multi-move strategy where the components of B^T are jointly sampled from normal distributions having moments centered at Kalman smoother estimates. Koop and Potter (2011) argued that multi-move algorithms are inefficient when one requires stationarity of the impulse responses at each t , especially if the VAR is of medium/large dimension. To avoid nonexplosive impulse responses, it is common since Cogley and Sargent (2005) to assume that all the eigenvalues of the companion form matrix associated with B_t lie within the unit circle for $t = 1, \dots, T$. When the multi-move logic is used, if one element of the sequences violates the stationarity restrictions, the entire sequence is discarded, making the algorithm inefficient. As an alternative, Koop and Potter suggest to evaluate the elements of the B^T sequence separately using a single-move algorithm and an accept/reject step. We describe how the algorithm works in our structural system in Appendix B.

To deal with the stationarity issue, one could also consider the shrinkage approach of Canova and Ciccarelli (2009). The approach was originally designed to deal with the curse of dimensionality in large scale panel VAR models, but can also be used in our context. When B_t is of large dimension and each of the components is an independent random walk, the probability that explosive draws for at least one coefficient are obtained is very large. Canova and Ciccarelli make B_t a function of a much lower dimensional vector of factors θ_t , which independently move as a random walk, and this can reduce the inefficiency of the algorithm. We describe how this algorithm works in Appendices C and D.

We compare these three approaches to sample B^T in our medium scale SVAR model to better understand the pros and cons of each routine. The standard multi-move routine is very inefficient if variables are nonstandardized. In fact, the acceptance rate is only 0.46% when year-on-year growth rates are used and is 0.60% when quarterly growth rates are employed. As a referee suggested, the slightly better results obtained with quarterly growth rates is due to the fact that, with this transformation, the data display lower persistence. When the data are standardized, acceptance rates improve with both data transformations (now they are 10.2% and 11.5%, respectively). Standardization reduces inefficiencies because not all variables necessarily have the same units.

In the single-move algorithm, the average acceptance rates for B^T when the data are standardized are 97% and 91% for year-on-year and quarterly growth rates, respectively. However, the higher acceptance rates are more than compensated by the higher

correlations of the draws (serial correlation goes from 0.10 to 0.92–0.94 with the current procedure). Furthermore, there are important computational costs: we need about 12 hours to estimate the model with the multi-move routine, but about 96 hours with the single-move routine. Note that the precision of the two algorithms is roughly the same.

Apart from the constants, the vector B_t has 72 components. To maintain as much as possible the covariance structure of the data unchanged, we estimate the shrinkage model with 15 factors: one common factor, one factor for each equation (6), one factor for each lag (2), and one factor for each variable (6),

$$B_t = \Xi \theta_t + v_t, \quad v_t \sim N(0, I),$$

$$\theta_t = \theta_{t-1} + \rho_t, \quad \rho_t \sim N(0, Q),$$

where Ξ is a 72×15 matrix loading the factors on the required elements of the B_t vector. When Ξ is composed of 0s and 1s, we needed about 10 hours to estimate the model, the acceptance rate is 78% when the data are standardized, and the serial correlation of the draws is 0.55. When the elements of Ξ are also estimated, the computational time increases to about 24 hours, the acceptance rate for B^T drops to 24%, and the first order serial correlation of the draws is 0.43, when data are standardized.

In sum, both the multi-move and the shrinkage algorithms have reasonable computational costs and have comparable efficiency properties. The single-move algorithm is computationally much more demanding—we need to compute a constant of integration at each t and at each step of the Gibbs sampler—and its efficiency seems lower.

In the next subsections, we comment on the results obtained using standardized year-on-year growth rates and the multi-move algorithm.

5.5 Time variations in structural parameters

We first describe the time variations that our model delivers. In the left panel of Figure 1, we report the highest 68% posterior tunnel for the variability of the monetary policy shock, and in Figure 2, we report the highest 68% posterior tunnel for the nonzero contemporaneous structural parameters α_t .

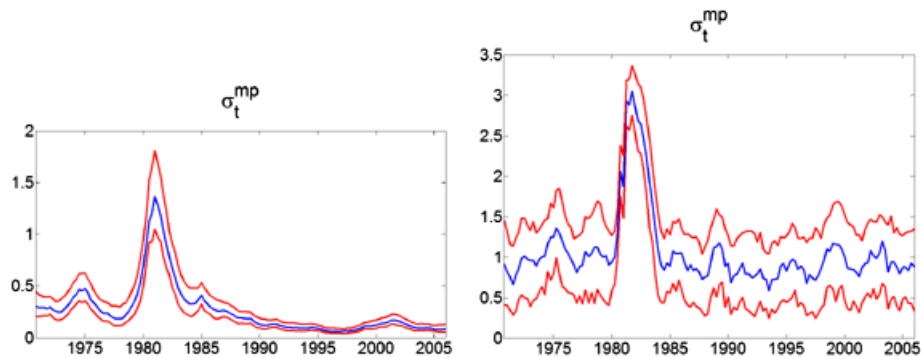
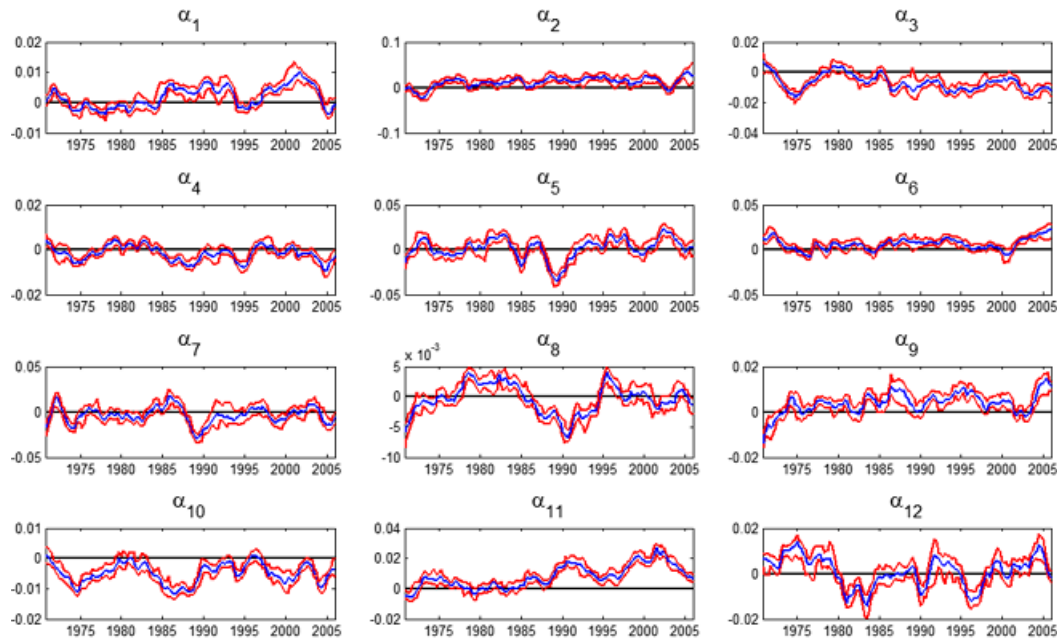


FIGURE 1. Median and posterior 68% tunnels: volatility of monetary policy shock. Left panel: stochastic volatility model; right panel: GARCH(1, 1) model.

FIGURE 2. Estimates of α_t .

There are significant changes in the standard deviation of the policy shocks and a large swing in the late 1970s–early 1980s is visible. Given the identification restrictions, this increase in volatility must be attributed to some unusual and unexpected policy action that made the typical relationship between interest rates and money growth different. This pattern is consistent with the arguments of Strongin (1995) and Bernanke and Mihov (1998), who claim that monetary policy in the 1980s was run differently, and agrees with the results of Sims and Zha (2006).

A few of the nonpolicy parameters $[\alpha_{1,t}, \alpha_{2,t}, \alpha_{5,t}]$ exhibit considerable time variations that are a posteriori significant. Note that it is not only the magnitude that changes; the sign of the posterior tunnel is also affected. Also worth noting is the fact that both the GDP coefficient in the inflation equation ($\alpha_{1,t}$) and the inflation coefficient in the unemployment equation ($\alpha_{5,t}$) change sign, suggesting a generic sign switch in the slope of the Phillips curve.

The parameter $\alpha_{11,t}$, which controls the reaction of the nominal interest rates to money growth, also displays considerable changes. In particular, while in the 1970s and the first half of the 1980s the coefficient was generally small and at times insignificant, it became much stronger in the rest of the sample (1986–2005). Interestingly, this time period coincides with the Greenspan era, where official statements claimed that monetary policy was conducted using the interest rate as an instrument and money aggregates were endogenous.

The coefficients of the money demand equation, $[\alpha_{3,t}, \alpha_{6,t}, \alpha_{9,t}]'$ are also unstable. For example, the elasticity of money demand to the nominal interest rate ($\alpha_{9,t}$) is negative at the beginning of the sample and turns positive since the middle of the 1970s, with

some episodes when it is not significantly different from zero. The elasticity of money (growth) demand to inflation is low and sometimes insignificant, but increasing in the last decade. Thus, homogeneity of degree 1 of money in prices does not hold for a large portion of our sample. Since time variations in elements of α_t are correlated (see, in particular, $\alpha_{5,t}$ and $\alpha_{8,t}$ or α_{1t} and α_{11t}), our setup captures the idea that policy and private sector parameters move together.

Thus, in agreement with the dynamic stochastic general equilibrium (DSGE) evidence of Justiniano and Primiceri (2008) and Canova and Ferroni (2012), time variations appear in the variance of the monetary policy shock, and in contemporaneous policy and nonpolicy coefficients.

5.6 The transmission of monetary policy shocks

We next study how the observed time variations affect the transmission of monetary policy shocks. Since σ_t^{mp} is time-varying, we normalize the impulse to be 1 at all t . Thus, the time variations we describe are due to changes in the propagation but not in the size of the shocks. We compute responses as the difference between two conditional projections: one with the structural shock set to 1 and one with the structural shock set to 0.

In theory, a surprise increase in the monetary policy instrument, should make money growth, output growth, and inflation fall, while unemployment should go up. Such a pattern is present in the early part of the sample, but disappears as time goes by. As Figure 3 indicates, monetary policy shocks have the largest effects in 1981; the pattern is similar but weaker in 1975 and 1990. In 2005, responses are somewhat perverse (inflation and output growth significantly increase and unemployment significantly falls after an interest rate increase). Differences in the responses of output and unemployment between, say, 1981 and 2005 are a posteriori significant. Thus, the ability of monetary policy to affect the real economy has considerably changed over time and policy surprises are interpreted in different ways across decades.

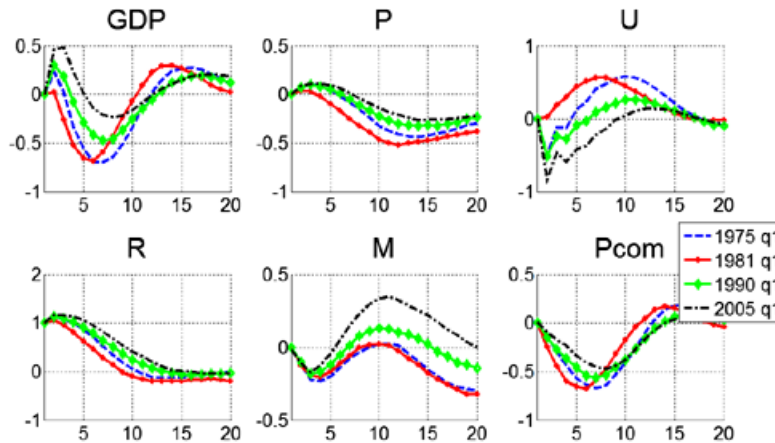


FIGURE 3. Dynamics following a monetary policy shock: various dates.

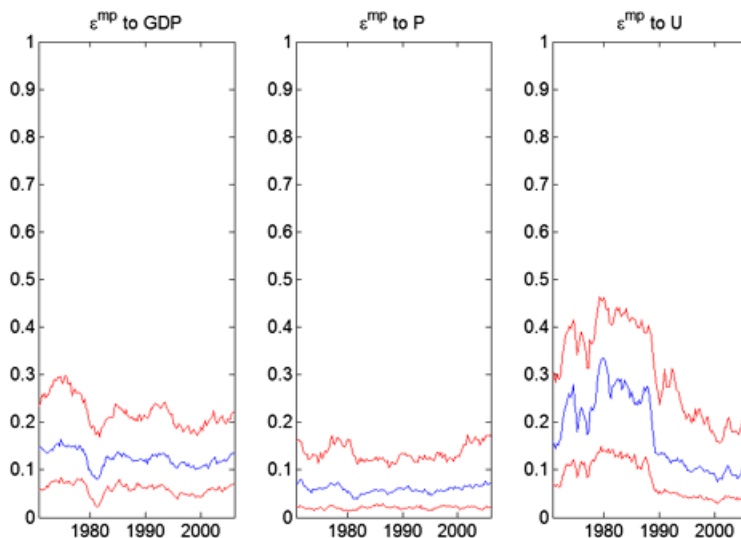


FIGURE 4. Long-run share of the forecast error variance due to monetary policy shock: various dates.

Despite these noticeable variations, the proportion of the forecast error variance of output, prices, and unemployment due to policy shocks is consistently small (see Figure 4). Monetary policy shocks explain 10% of the forecast error variance of inflation at all dates and about 15–20% of the variability of output growth and the unemployment rate, with a maximum of about 25% in the early 1980s. Thus, as in Uhlig (2005) or Sims and Zha (2006), monetary policy has modest real effects.

These results are very much in line with those of Gambetti, Pappa, and Canova (2008), even though they use sign restrictions to extract structural shocks, and with those of Boivin and Giannoni (2006), who use subsample analysis to make their points. They differ somewhat from those reported in Sims and Zha (2006), primarily because they do not allow for time variations in the instantaneous coefficients, and from those in Fernández Villaverde, Guerron Quintana, and Rubio Ramírez (2010), who allow for stochastic volatility and time variations only in the coefficients of the policy rule.

5.7 A time-invariant overidentified model

We compare these results with those obtained in a constant coefficient overidentified structural model. Given that time variations seem relevant, we would like to know how the interpretation of the evidence would change if one estimates a model with constant coefficients.

To illustrate the differences, we report in Figure 5 the responses of the variables to a unexpected monetary policy impulse at four dates (1975, 1981, 1990, 2005) in the two systems. Clearly, there is more uncertainty regarding the liquidity effect in the time-varying SVAR model at some dates. Furthermore, the responses of output growth, inflation, and unemployment in the constant coefficients model are different and the dynamics prevailing in the 1970s seem to dominate.

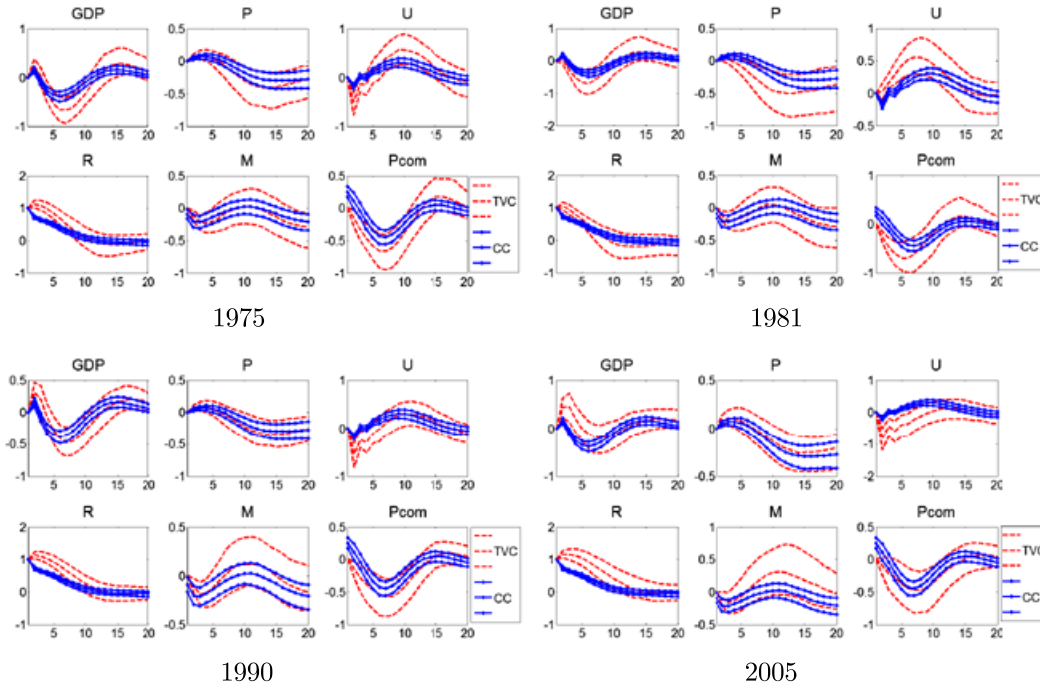


FIGURE 5. Time-varying and time-invariant responses.

Overall, differences between TVC and time-invariant models are generally smaller than previously reported. The reason is that we standardize the data prior to estimation. If this transformation is not performed, differences in the two systems become substantially larger.

5.8 Altering the law of motion of the volatility

To check whether our results depend on the specification of the law of motion of the volatilities, we now assume that instead of (26) we use

$$\sigma_{m,t}^2 = (1 - \delta) + \delta \sigma_{m,t-1}^2 + \delta (y_{m,t-1}^{**})^2 + \eta_{m,t}. \quad (33)$$

Since with this GARCH(1, 1), the resulting model is nonlinear, we calibrate the proposal using an extended Kalman filter algorithm and a linearized version of the model. Details on how the algorithm is modified in this case are in Appendix E. For comparison purposes, we report the time profile of the posterior distribution of the standard deviation of the monetary policy shock (in the second panel of Figure 1) and the responses to a monetary policy shock in 1975:1, 1981:1, 1990:1, and 2005:1 (see Figure 6).

A few interesting conclusions emerge from the figures. The qualitative features of the results are broadly unaltered. For example, there is a peak in the volatility of the monetary policy shock in the late 1970s–early 1980s and a standard price puzzle in response to a policy shock. Quantitatively, however, important changes occur. The volatility of the monetary policy shock is estimated to be generally larger and the peak in the early 1980s

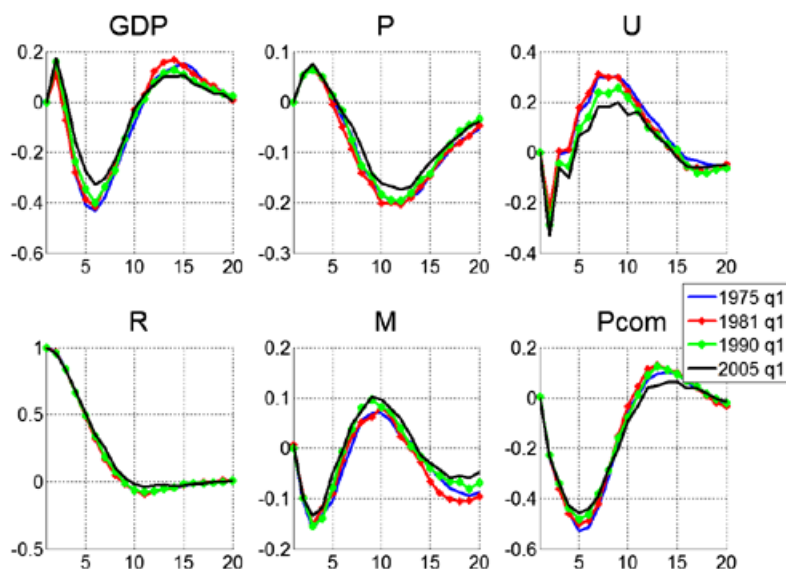


FIGURE 6. Transmission of monetary policy shocks: GARCH(1, 1) specification.

is 50% taller. Because a larger portion of the dynamics of the endogenous variables is now captured by volatility changes, the responses to policy shocks are generally smaller and less significant than in the baseline case. For instance, contrary to what we had in Figure 3, the responses of prices and money are never significant, and those of the unemployment rate are significant only in the very short run. In addition, time variations in the transmission of monetary policy shocks are smaller: if we exclude the medium term response of the unemployment rate, the responses of the other five variables are roughly constant at all horizons. Thus, inference about the effects of policy shocks and the changes in the transmission mechanism may depend on nuisance features.

5.9 Using short- and long-run restrictions

As a final robustness check, together with the restrictions we have imposed in Table 1, we also impose the restriction that monetary policy shocks have no long-run effect on output. We do not restrict the long-run behavior of the unemployment rate since there are theories that allow long-run movements of the unemployment rate in response to monetary policy shocks; see, e.g., Benhabib and Farmer (2000). Also in this case, the model becomes nonlinear. Details on the modifications needed in the sampling algorithm are in Appendix E.

Figure 7 presents the responses of the variables to a monetary policy shock. The basic qualitative features of the responses are unchanged: it takes some time to output and the unemployment rate to react; prices are sluggish in response to a surprise increase in interest rates. Quantitatively some differences emerge. In 2005, the response of output is perverse: output increases in response to a monetary policy contraction for at least 20 quarters. In 1975, the response of prices is much more persistent than without long-run restrictions and the peak response of the unemployment rate is stronger. In general,

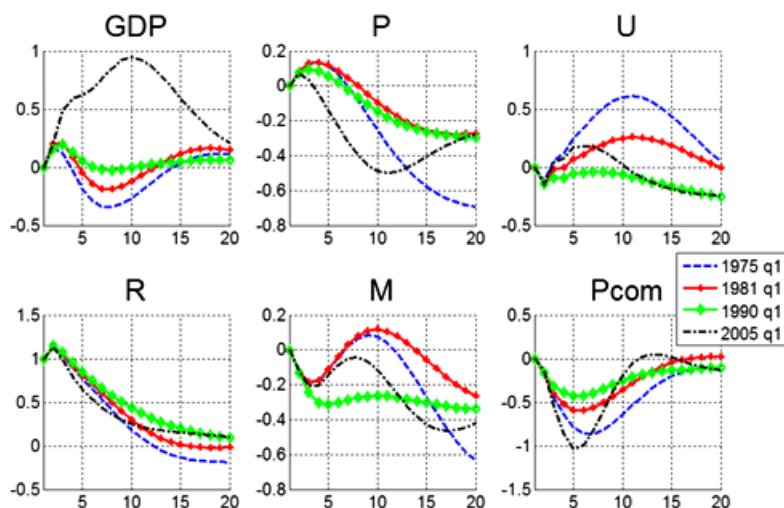


FIGURE 7. Transmission of monetary policy shocks: long-run restrictions.

when long-run restrictions are imposed, time variations in the transmission of monetary policy shocks are increased.

6. CONCLUSIONS

We propose a unified framework to estimate structural VARs. The methodology can handle time-varying coefficients or time-invariant models identified with recursive or nonrecursive constraints that can be linear or nonlinear and that can produce just identified or overidentified systems. Our algorithm adds a Metropolis step to a standard Gibbs sampling routine. With minor modifications, it can also deal with nonlinear structural state space systems. Thus, we greatly expand the set of structural VARs that researchers can deal with within the same estimation framework.

We apply the methodology to study the transmission of monetary policy shocks in a nonrecursive overidentified TVC model similar to that used by [Robertson and Tallman \(2001\)](#) and [Waggoner and Zha \(2003\)](#) with fixed coefficients. We examine the merits of multi-move versus single-move routines and find that once data is demeaned and expressed in the same scale, the computational costs of using a single-move routine are larger than the efficiency gains. We also show that there are time variations in the variance of the monetary policy shock and in the estimated contemporaneous coefficients. These variations translate into significant changes in the transmission of monetary policy shocks. The time variations are considerably reduced when an alternative law of motion for the standard deviation of the shocks is used. We show that when a mixture of long- and short-run restrictions are employed, the transmission of monetary policy shocks in the 2000s is affected.

The range of potential applications of the methodology is large. For example, one could use the same setup to identify fiscal shocks or externally generated shocks in models that theory tightly parameterizes. One could also use the same methodology to identify shocks imposing magnitude restrictions on impulse responses as in [Rubio Ramírez,](#)

Waggoner, and Zha (2010) or variance decomposition restrictions as in Barsky and Sims (2011). The estimation complexity is important but not overwhelming and all the computations can be performed on a standard personal computer with sufficient random access memory.

REFERENCES

- Amisano, G. and C. Giannini (1997), *Topics in Structural VAR Econometrics*, second edition. Springer-Verlag, Berlin. [362]
- Arias, J., D. Waggoner, and J. Rubio Ramírez (2014), “Inference based on SVARs identified with sign and zero restrictions: Theory and applications.” Working Paper 2014-1, Federal Reserve Bank of Atlanta. [365]
- Barsky, R. and E. Sims (2011), “News shocks and the business cycle.” *Journal of Monetary Economics*, 58, 273–289. [370, 382]
- Baumeister, C. and J. Hamilton (2013), “Sign restrictions, structural vector autoregressions and useful prior information.” Manuscript. [363]
- Benhabib, J. and R. Farmer (2000), “The monetary transmission mechanism.” *Review of Economic Dynamics*, 3, 523–550. [380]
- Bernanke, B. S. and A. S. Blinder (1992), “The federal funds rate and the channels of monetary transmission.” *American Economic Review*, 82, 901–921. [371]
- Bernanke, B. S. and I. Mihov (1998), “Measuring monetary policy.” *Quarterly Journal of Economics*, 113, 869–902. [376]
- Boivin, J. and M. P. Giannoni (2006), “Has monetary policy become more effective?” *Review of Economics and Statistics*, 88, 445–462. [378]
- Canova, F. and M. Ciccarelli (2009), “Estimating multicountry VAR models.” *International Economic Review*, 50, 929–959. [374]
- Canova, F., M. Ciccarelli, and E. Ortega (2012), “Do institutional changes affect business cycles? Evidence from Europe.” *Journal of Economic Dynamics and Control*, 36, 1520–1533. [368]
- Canova, F. and G. De Nicoló (2002), “Monetary disturbances matter for business fluctuations in the G-7.” *Journal of Monetary Economics*, 49, 1131–1159. [359]
- Canova, F. and F. Ferroni (2012), “The dynamics of US inflation: Can monetary policy explain the changes?” *Journal of Econometrics*, 167, 47–60. [377]
- Canova, F. and M. Paustian (2011), “Business cycle measurement with some theory.” *Journal of Monetary Economics*, 58, 345–361. [359]
- Carter, C. K. and R. Kohn (1994), “On Gibbs sampling for state space models.” *Biometrika*, 81, 541–553. [369, 374]
- Cogley, T. and T. J. Sargent (2005), “Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.” *Review of Economic Dynamics*, 8, 262–302. [359, 374]

- Creel, D. (2012), “Survey of sequential Monte Carlo methods for economics and finance.” *Econometric Reviews*, 31, 245–296. [370]
- Del Negro, M. and G. Primiceri (2013), “Time-varying structural vector autoregressions and monetary policy: A corrigendum.” Federal Reserve Bank of New York Staff Report 619. [370]
- Fernández Villaverde, J., P. A. Guerron Quintana, and J. F. Rubio Ramírez (2010), “Fortune or virtue: Time-variant volatilities versus parameter drifting in U.S. data.” Discussion Paper 7813, CEPR. [378]
- Gambetti, L., E. Pappa, and F. Canova (2008), “The structural dynamics of U.S. output and inflation: What explains the changes?” *Journal of Money, Credit and Banking*, 40, 369–388. [359, 378]
- Geweke, J. and H. Tanizaki (2001), “Bayesian estimation of state–space models using the Metropolis–Hastings algorithm within Gibbs sampling.” *Computational Statistics and Data Analysis*, 37, 151–170. [360, 370]
- Gordon, D. B. and E. M. Leeper (1994), “The dynamic impacts of monetary policy: An exercise in tentative identification.” *Journal of Political Economy*, 102, 1228–1247. [360]
- Herbst, E. and F. Schorfheide (2014), “Sequential Monte Carlo sampling for DSGE models.” *Journal of Applied Econometrics*, 29, 1073–1098. [370]
- Justiniano, A. and G. E. Primiceri (2008), “The time-varying volatility of macroeconomic fluctuations.” *American Economic Review*, 98, 604–641. [377]
- Kociecki, A., M. Rubaszek, and M. Ca’ Zorzi (2013), “Bayesian analysis of recursive SVAR models with overidentifying restrictions.” Manuscript, Bank of Poland. [360, 363]
- Koop, G. and S. Potter (2011), “Time varying VARs with inequality restrictions.” *Journal of Economic Dynamics and Control*, 35, 1126–1138. [369, 374]
- Lakdawala, A. (2012), “Changes in Federal Reserve preferences.” Manuscript, Michigan State University. [369]
- Lanne, M. and H. Lütkepohl (2008), “Identifying monetary policy shocks via changes in volatility.” *Journal of Money, Credit and Banking*, 40, 1131–1149. [359]
- Leeper, E., C. Sims, and T. Zha (1996), “What does monetary policy do?” *Brookings Papers on Economic Activity*, 2, 2–78. [360]
- Primiceri, G. (2005), “Time varying structural vector autoregressions and monetary policy.” *Review of Economic Studies*, 72, 821–852. [359, 360, 368]
- Robertson, J. C. and E. W. Tallman (2001), “Improving federal-funds rate forecasts in VAR models used for policy analysis.” *Journal of Business & Economic Statistics*, 19, 324–330. [361, 381]
- Rubio Ramírez, J. F., D. F. Waggoner, and T. Zha (2010), “Structural vector autoregressions: Theory of identification and algorithms for inference.” *Review of Economic Studies*, 77, 665–696. [381, 382]

Sims, C. A. and T. Zha (1998), “Bayesian methods for dynamic multivariate models.” *International Economic Review*, 39, 949–968. [360, 361]

Sims, C. A. and T. Zha (2006), “Were there regime switches in U.S. monetary policy?” *American Economic Review*, 96, 54–81. [361, 374, 376, 378]

Strongin, S. (1995), “The identification of monetary policy disturbances: Explaining the liquidity puzzle.” *Journal of Monetary Economics*, 35, 463–497. [376]

Uhlig, H. (2005), “What are the effects of monetary policy on output? Results from an agnostic identification procedure.” *Journal of Monetary Economics*, 52, 381–419. [359, 378]

Waggoner, D. F. and T. Zha (1999), “Conditional forecasts in dynamic multivariate models.” *Review of Economics and Statistics*, 81, 639–651. [360]

Waggoner, D. F. and T. Zha (2003), “A Gibbs sampler for structural vector autoregressions.” *Journal of Economic Dynamics and Control*, 28, 349–366. [360, 361, 381]

Submitted September, 2012. Final version accepted July, 2014.