Estimating overidentified, non-recursive, time varying coefficients structural VARs

Fabio Canova* Fernando J. Pé rez Forero

March 12, 2014

Abstract

This paper provides a general procedure to estimate structural VARs. 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 non-recursive) 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.

JEL Classification: C11, E51, E52

Key words: Time-varying coefficient structural VAR models, Metropolis algorithm, Identification restrictions, Monetary transmission mechanism.

*EUI and CEPR; and BCRP. Corresponding address: Department of Economics, European University Institute, Via della Piazzuola 43, 50133 Firenze, Italy; Research Department, Banco Central de Reserva del Perú (BCRP), Jr. Miró Quesada 441, Lima, Perú. 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 the 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".

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 Canova et al. (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 DGP, see Canova and Paustian (2011).

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" non-recursive restrictions, using the terminology of Waggoner and Zha (1999), even though these constraints are not theoretically abundant. There exist algorithms to estimate non-recursive structural models (Waggoner and Zha, 2003) and to estimate recursive overidentified models (Kociecki et al., 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 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 non-recursive, 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 non-recursive restrictions. The procedure can be used in systems which are just-identified or overidentified, and allows for both linear and non-linear restrictions on the parameter space. Non-recursive structures have been extensively used to accommodate models which 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 non-recursive scheme is used. In addition, although just-identified systems are easier to construct and estimate, over-identified models have a long history in the literature (see e.g. Leeper et al., 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 general law of motions of the coefficients imply a non-linear state space representation for the structural model, we then nest our basic procedure into Geweke and Tanizaki (2001)'s 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 non-linear simulation methodology.

We use the methodology to identify a monetary policy shock in a overidentified TVC system, whose structure is similar to the one 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 in 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 non-recursive, overidentified TVC-VAR models featuring stochastic volatility. Section 5 extends the algorithm to deal with state space systems which have a non-linear format. Section 6 studies the transmission of monetary policy shocks. Section 7 concludes.

2 A constant coefficients static SVAR

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

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

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

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

Because of $\det\left(A(\alpha)\right)^T$, the Jacobian of the transformation, (2) is non-linear in α. Thus, the posterior of α will be non-standard. 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 which are compatible with the setup.

2.1 The reparameterization and the algorithm

Vectorizing (1) produces $vec\left(A(\alpha) y_t\right) = vec\left(\varepsilon_t\right) = \varepsilon_t$. Amisano and Giannini (1997), assume that $vec\left(A(\alpha)\right) = S_A \alpha + s_A$, where S_A and s_A are matrices of ones and zeros. This reparameterization can be made slightly more general so as to handle certain types of non-linear restrictions. Let $\alpha \in X \subset R^k$, where X is "sufficiently large" - we need X to have this feature because simulations proceed by rejecting draws - and let $f : X \to R^{k'}$ be continuous, where k' may be different from k. Using $vec\left(A(\alpha) y_t\right) = \left(y_t' \otimes I\right)\left(S_A f(\alpha) + s_A\right)$, the model can be expressed as:

$$\widetilde{y}_t = Z_t f(\alpha) + \varepsilon_t \tag{3}$$

where $\widetilde{y}_t \equiv \left(y_t' \otimes I\right) s_A$; $Z_t \equiv -\left(y_t' \otimes I\right) S_A$. The likelihood function is then

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

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

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

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

and

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

where $SSE = \sum_{t=1}^{T} (\widetilde{y}_t - Z_t f^*(\alpha))\,(\widetilde{y}_t - Z_t f^*(\alpha))'$. Set $f(\alpha^0) = f^*(\alpha)$ and, for $i = 1, 2, \ldots, G$:

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

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

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

Note three facts about the algorithm. First, a t-distribution with 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 $f(\alpha)$ vector is jointly sampled and the covariance matrix of $P^*(\alpha)$ is non-diagonal. 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 to draw α since equations (5) and (6) ignore the Jacobian term $D(\alpha)$ appearing in (4).

Kociecki et al. (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 and 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 model (3) is sufficiently general to deal with linear restrictions (both of exclusion and non-exclusion types) and with certain types of non-linear restrictions. We present a few examples for illustration. We focus on over-identified systems because just identified ones only require adjustments of S_A and of s_A.

2.2.1 Short-run linear restrictions

Suppose $A(\alpha)$ features both exclusion and non-exclusion linear restrictions:

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

Then:

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

2.2.2 Short-run non-linear 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

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

If we define $\widetilde{\alpha}_2 \equiv \left(\alpha_2+1\right)^2$, $f\left(\alpha\right) = \left(\alpha_1, \widetilde{\alpha}_2, \alpha_3\right)'$ is still a linear vector-valued function. Given posterior draws for $\widetilde{\alpha}_2$, we can recover $\alpha_2 = \sqrt{\widetilde{\alpha}_2} - 1$, provided $\widetilde{\alpha}_2 \geqslant 0$. Hence, certain non-linear restrictions 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, $f\left(\alpha\right) = \left[\alpha_1, \left(\alpha_2+2\alpha_3\right)^2, \alpha_3\right]'$. Here, we set $\widetilde{\alpha}_2 \equiv \left(\alpha_2+2\alpha_3\right)^2 \geqslant 0$ and use draws of $\widetilde{\alpha}_2 \geqslant 0$ and α_3, to obtain $\alpha_2 = \sqrt{\widetilde{\alpha}_2} - 2\alpha_3$.

A case we can not handle with a reject step is the following:

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

Here

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

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

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

The corresponding VAR is:

$$y_t = B y_{t-1} + [A(\alpha)]^{-1} \varepsilon_t \tag{10}$$

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

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

Let

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

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. $\mathcal{D}_{21} = 0$ 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 non-linear 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 section 5 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 if the first column satisfies the required constraints. Thus, sign restrictions can be dealt with the same accept/reject step we have used for non-linear short run restrictions. Mixture of zero and sign restrictions, as those suggested by Arias et al. (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

$$A(\alpha_t) y_t = \varepsilon_t; \quad \varepsilon_t \sim N(0, I) \tag{13}$$

$$\alpha_t = \alpha_{t-1} + \eta_t; \quad \eta_t \sim N(0, V) \tag{14}$$

where V is a full rank, positive definite matrix, α_0 is given, $A(\alpha_t)$ is assumed to be invertible for almost all α_t, for all t. This model is re-parametrized as:

$$\widetilde{y}_t = Z_t f(\alpha_t) + \varepsilon_t \tag{15}$$

$$f(\alpha_t) = f(\alpha_{t-1}) + \eta_t \tag{16}$$

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

The conditional posterior $p\left(f(\alpha)^T \mid y^T, V\right)$ can not be computed in a standard fashion since Z_t is neither exogenous nor predetermined. Thus, (15)-(16) is not the typical state space model considered in the literature. Still, it is relatively easy to compute the kernel of this conditional posterior. Note that

$$\begin{aligned} p\left(f(\alpha)^T \mid y^T, V\right) &= p\left(f(\alpha_T) \mid y_T, V\right) \prod_{t=1}^{T-1} p\left(f(\alpha_t) \mid f(\alpha_{t+1}), y^t, V\right) \\ &\propto p\left(f(\alpha_T) \mid y_T, V\right) \prod_{t=1}^{T-1} p\left(f(\alpha_t) \mid y^t, V\right) p\left(f(\alpha_{t+1}) \mid f(\alpha_t), V\right) \end{aligned} \tag{17}$$

From (16) $p\left(f(\alpha_{t+1}) \mid f(\alpha_t), V\right)$ is normal. The kernel of other terms in the expression can be constructed noticing that $p(f(\alpha_t)|y^t, V) \propto \tilde{L}(y^t|f(\alpha_t), V)p(f(\alpha_t))$, where $p(f(\alpha_t))$ is given, and

$$\tilde{L}(y^t|f(\alpha_t), V) = (2\pi)^{-M/2} \det\left(D(\alpha_t)\right) exp\{-\frac{1}{2}(\tilde{y}_y - Z_t f(\alpha_t))'(\tilde{y}_t - Z_t f(\alpha_t))\}$$

Thus, given $f(\alpha_{0|0})$ and $P_{0|0}$, and construct Kalman filter updated estimates of $f(\alpha_t)$ and of its covariance matrix for each $t = 1, \ldots, T$ as

$$\begin{aligned} \widehat{f(\alpha_{t|t})} &= \widehat{f(\alpha_{t|t-1})} + K_t \left[\widetilde{y}_t - Z_t \widehat{f(\alpha_{t|t-1})} \right] \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} Z_t' \Omega_t^{-1} Z_t P_{t|t-1}' \end{aligned}$$

where $\widehat{f(\alpha_{t|t-1})} = \widehat{f(\alpha_{t-1|t-1})}$, $P_{t|t-1} = P_{t-1|t-1} + V$, $K_t = P_{t|t-1} Z_t' \Omega_t^{-1}$, $\Omega_t = Z_t' P_{t|t-1} Z_t + I$. Smoothed estimates are $f^*(\alpha_{T|T}) = \widehat{f(\alpha)}_{T|T}$, $P^*_{T|T} = P_{T|T}$ and

$$\begin{aligned} f^*(\alpha_{t|t+1}) &= \widehat{f(\alpha_{t|t})} + P_{t|t} Z_t' P_{t+1|t}^{-1} \left(f^*(\alpha_{t+1|t+2}) - Z_t' \widehat{f(\alpha_{t|t})} \right) \\ P^*_{t|t+1} &= P_{t|t} - P_{t|t} Z_t' P_{t+1|t}^{-1} Z_t P_{t|t-1}' \qquad t = T-1, \ldots, 1 \end{aligned} \tag{18}$$

We use (18) to calibrate the proposal distribution for the Metropolis algorithm.

3.1 The basic algorithm

Set an initial $f(\alpha^0)$ and for $i = 1, 2, \ldots, G$

Step 1: Given $\left(y^T, V^{i-1}\right)$, compute $\{f^*(\alpha^{i-1}_{t|t+1})\}_{t=1}^T, \{P^{*(i-1)}_{t|t+1}\}_{t=1}^T$ using (15)-(16) . Then:

1. For $t = 1, \ldots, T$, draw a candidate $f(\alpha_t^\dagger) \sim p_* \left(f(\alpha_t) \mid f(\alpha_t^{i-1}) \right) = t \left(f^*(\alpha^{i-1}_{t|t+1}), r P^{*(i-1)}_{t|t+1}, \nu \right)$, $r > 0$, $\nu \geq 4$. Set $f(\alpha^\dagger)^T = \{f(\alpha_t^\dagger)\}_{t=1}^T$; $p_*(f(\alpha)^T \mid f(\alpha^{i-1})^T) = \prod_{t=1}^T p_*(f(\alpha_t) \mid f(\alpha_t^{i-1}))$.

2. Compute $\theta = \frac{\widetilde{p}(f(\alpha^\dagger)^T) \cdot p_*(f(\alpha^{i-1})^T | f(\alpha)^T)}{\widetilde{p}(f(\alpha^{i-1})^T) \cdot p_*(f(\alpha)^T | f(\alpha^{i-1})^T)}$, where $\widetilde{p}(f(\alpha^\dagger)^T) = \widetilde{L}(y^T | f(\alpha^\dagger)^T, V) \cdot [p(f(\alpha^\dagger)^T) \mathcal{I}(f(\alpha))]$ and $\mathcal{I}(f(\alpha))$ is an indicator restricting the prior distribution.

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

Step 2: Given $(y^T, f(\alpha^i)^T)$, draw $(V^i)^{-1} \sim p((V^i)^{-1} \mid f(\alpha^i)^T, y^T) = W \left(\overline{v}_V, \overline{V}^{-1} \right)$, where

$$\overline{v}_V = T + \underline{v}_V$$

$$\overline{V}^{-1} = \left[\underline{V} + \sum_{t=1}^T \left(f(\alpha_t) - f(\alpha_{t-1}) \right) \left(f(\alpha_t) - f(\alpha_{t-1}) \right)' \right]^{-1}$$

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

Given the structure of the problem, if coefficients are constant, $f^*(\alpha_t) = f^*(\alpha)$, $P_t^* = P^*$,for all $t = 1, \ldots, T$, and the algorithm collapses to the one described in section 2.1.

4 A standard time-varying coefficients SVAR

Assume that a $M \times 1$ vector of non-stationary variables y_t, $t = 1, \ldots, T$ can be represented with a finite order autoregression of the form:

$$y_t = B_{0,t}C_t + B_{1,t}y_{t-1} + ... + B_{p,t}y_{t-p} + u_t \tag{19}$$

where $B_{0,t}$ is a matrix of coefficients on a $\bar{M} \times 1$ vector of deterministic variables C_t; $B_{j,t}; j = 1, \ldots, p$ 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 $\varepsilon_t \sim N(0, I)$, 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 $\Sigma_t = diag\{\ \sigma_{m,t}\}$ contains the standard deviations of the structural shocks at t in the main diagonal. The SVAR is:

$$y_t = X_t'B_t + A_t^{-1}\Sigma_t\varepsilon_t \tag{20}$$

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

$$B_t = B_{t-1} + \upsilon_t \tag{21}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{22}$$

$$\log(\sigma_{m,t}) = \log(\sigma_{m,t-1}) + \eta_{m,t} \tag{23}$$

and letting $\eta_t = [\eta_{1t}, \ldots, \eta_{Mt}]$, set:

$$\mathcal{V} = Var\left(\begin{bmatrix} \varepsilon_t \\ \upsilon_t \\ \zeta_t \\ \eta_t \end{bmatrix}\right) = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & W \end{bmatrix} \tag{24}$$

where Q, V, W are full rank matrices.

Thus, the setup captures time variations in i) the lag structure (see (21)), ii) the contemporaneous reaction parameters (see (22)) and iii) the structural variances (see (23)). Common patterns of time variations within blocks are possible if the rank of either Q, V or W is reduced. Models with breaks at a specific date can be accommodated by adding restrictions on $(21) - (23)$, see Canova et al. (2012).

4.1 Relaxing standard assumptions

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

$$A_t\left(y_t - X_t'\widehat{B}_t\right) \equiv A_t\widehat{y}_t = \Sigma_t\varepsilon_t \tag{25}$$

Let $vec\,(A_t) = S_A f(\alpha_t) + s_A$, where S_A and s_A are matrices with ones and zeros of dimensions $M^2 \times \dim(f(\alpha))$ and $M^2 \times 1$, respectively. The concentrated model can be reparametrized as

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

and the state space is composed of

$$\begin{aligned} \widetilde{y}_t &= Z_t f(\alpha_t) + \Sigma_t\varepsilon_t \\ f(\alpha_t) &= f(\alpha_{t-1}) + \zeta_t \end{aligned}$$

and of equation (23), where $\widetilde{y}_t \equiv (\widehat{y}_t' \otimes I_M)\,s_A$, $Z_t \equiv -(\widehat{y}_t' \otimes I_M)\,S_A$. Given $(\widehat{B^T}, \Sigma^T, \mathcal{V})$, we need to draw $f(\alpha)^T \equiv \{f(\alpha_t)\}_{t=1}^T$, from $p\left(f(\alpha)^T \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T}\right)$.

Standard algorithms (see Primiceri, 2005) partition $f(\alpha_t)$ into blocks associated with each equation, say $f(\alpha_t) = \left[f(\alpha_t)^{1\prime}, f(\alpha_t)^{2\prime}, \ldots, f(\alpha_t)^{M\prime}\right]'$, and assume that these blocks are independent, so that $\mathcal{V} = diag\,(\mathcal{V}_1, \ldots, \mathcal{V}_M)$. Then

$$p(f(\alpha)^T \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T}) = \prod_{m=2}^{M} p\left(f(\alpha^m)^T \mid f(\alpha^{m-1})^T, \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T}\right) \times p\left(f(\alpha^1)^T \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T}\right) \tag{26}$$

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 $f(\alpha_t)$ 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 non-policy parameters, it will be hard to assume that non-policy 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 $f(\alpha_t)$ is jointly drawn and $\mathcal{V}$ is not necessarily block diagonal. This modification allows us to deal with recursive, non-recursive, just-identified or overidentified structural models in a unified framework. There are, however, computational costs, since system-wide estimation methods are now needed.

4.2 The general algorithm

Set initial values for $((B^0)^T, f(\alpha^0)^T, (s_0)^T, (\Sigma^0)^T, \mathcal{V}^0)$, where s is J-dimensional vector of discrete indicator variables described below. Then:

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

2. Draw $f(\alpha^i)^T$ from

$$\begin{aligned} p\left(f(\alpha^i)^T \mid \widehat{y}^T, (s^{i-1})^T, (\Sigma^{i-1})^T, \mathcal{V}^{i-1}, (B^i)^T)\right) \quad \propto \quad & p\left(f(\alpha_T^i) \mid \widehat{y}^T, s_T^{i-1}, \Sigma_T^{i-1}, \mathcal{V}^{i-1}, B_T^i\right) \times \\ & \prod_{t=1}^{T-1} p\left(f(\alpha_t^i) \mid \widehat{y}^T, s_t^{i-1}, \Sigma_t^{i-1}, \mathcal{V}^{i-1}, B_t^i\right) \times \\ & p_{t+1}\left(f(\alpha_{t+1}^i) \mid f(\alpha_t^i), \widehat{y}^T, s_t^{i-1}, \Sigma_t^{i-1}, \mathcal{V}^{i-1}, B_T^i\right) \end{aligned}$$

   using the approach described in section 3.1.

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

$$\widehat{A}_t \widehat{y}_t \equiv y_t^{**} = \Sigma_t \varepsilon_t$$

   and (23), but the error is not normal. For the $m-th$ equation we have

$$y_{m,t}^* = \log\left[\left(y_{m,t}^{**}\right)^2 + \overline{c}\right] \approx 2\log\left(\sigma_{m,t}\right) + \log \varepsilon_{m,t}^2 \tag{27}$$

   where $\overline{c}$ is a small constant. Since $\varepsilon_{m,t}$ is Gaussian, $\log \varepsilon_{m,t}^2$ is $\log\left(\chi^2\right)$ distributed and can be approximated by a mixture of normals. Conditional on s_t, the indicator for the mixture of normals, the model is linear and Gaussian. Thus, as in Del Negro and Primiceri (2013):

   (a) Draw $(s^i)^T$, given $\left((y^{**})^T, (B^i)^T, f(\alpha^i)^T, (\Sigma^{i-1})^T\right)$ and compute

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

   where $j = 1, \ldots, J; \phi(x)$ is the normal density, q_j 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 $P\left(s_{m,t} \leq j-1 \mid y_{m,t}^{**}, \log\left(\sigma_{m,t}\right)\right) < u \leq P\left(s_{m,t} \leq j \mid y_{m,t}^{**}, \log\left(\sigma_{m,t}\right)\right)$.

(b) Given $(\widehat{y}^T, (B^i)^T, f(\alpha^i)^T, (s^i)^T)$, use standard Kalman smoother recursions to draw $\{\Sigma_t\}_{t=1}^T$ from (27) − (23), given s^T obtained in step (a). To ensure independence of the structural variances, each $\sigma_{m,t}$ is sampled assuming a diagonal W.

4. Draw $\mathcal{V}^i$ from $p\left(\mathcal{V}^i \mid \widehat{y}^T, f(\alpha^i)^T, (s^i)^T, (\Sigma^i)^T, (B^i)^T\right)$. $\mathcal{V}^i$ is sampled assuming that each block follows an independent inverted Wishart distribution.

Then use $(B^i)^T, f(\alpha^i)^T, (s^i)^T, (\Sigma^i)^T, \mathcal{V}^i$ as initial values and repeat the sampling for $i = 1, \ldots, G$.

5 Extensions

In the setup we have used so far, we are constrained about the identification restrictions we can employ; for example, long run restrictions produce a non-linear model for α_t and β_t. Recent identification procedures which restrict certain 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, 2012), also generate a non-linear model for (α_t, β_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 GARCH or Markov switching specifications, which also generate a non-linear or non-normal laws of motion of the coefficients.

In all these cases the sequential Monte Carlo methods discussed in Creel (2012) and Herbst and Schorfheide (2013) are the natural candidates to estimate the structural non-linear model. These methods however are computationally intensive and there are still a number of theoretical and practical issues that are unsolved. Thus, we prefer to take an intermediate step, which still allows us to deal with all these cases, but is much less computationally demanding. Clearly, there are simplifications involved. For example, we will be assuming that the posteriors can be approximated by normals. Nevertheless, we believe it is important to have a tool that can cover these situations without having to pay the costs of a fully non-linear simulation methodology.

We describe next how the setup so far analyzed needs to be modified to deal with the cases of interest. In doing so, we extend Geweke and Tanizaki's (2001) algorithm for estimating non-linear, non-Gaussian state spaces to TVC models.

5.1 The setup

Consider the general non-linear state space model:

$$\begin{aligned} y_t &= z_t(\beta_t, \alpha_t) + u_t(\sigma_t, \xi_{1t}) && (28)\\ \beta_t &= w_t(\beta_{t-1}) + s_t(\beta_{t-1}, \xi_{2t}) && (29)\\ \alpha_t &= t_t(\alpha_{t-1}) + r_t(\alpha_{t-1}, \xi_{3t}) && (30)\\ f_t(\sigma_t) &= h_t(\sigma_{t-1}) + k_t(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})) && (31) \end{aligned}$$

where y_t, ξ_{1t} are $M \times 1$ vectors; β_t and ξ_{2t} are $K_\beta \times 1$ vectors; α_t and ξ_{3t} are $K_\alpha \times 1$ vectors; $\xi_{1t} \sim N(0, Q_{1t})$, $\xi_{2t} \sim N(0, Q_{2t})$, $\xi_{3t} \sim N(0, Q_{3t})$. Assume that $z_t(.)$, $u_t(.)$, $w_t(.)$, $s_t(.)$, $t_t(.)$, $r_t(.)$, $f_t(.)$, $h_t(.)$ and $k_t(.)$ are continuous and differentiable vector-valued functions. To estimate this system, we can linearize it around the previous forecast of the state vector, so that

$$\begin{aligned} z_t(\beta_t, \alpha_t) &\simeq z_t(\widehat{b}_{t|t-1}, \widehat{a}_{t|t-1}) + \widehat{Z}_{1t}(\beta_t - \widehat{b}_{t|t-1}) + \widehat{Z}_{2t}(\alpha_t - \widehat{a}_{t|t-1})\\ u_t(\sigma_t, \xi_{1t}) &\simeq u_t(\widehat{\sigma}_{t|t-1}, 0) + \widehat{u}_{\sigma,t}(\sigma_t - \widehat{\sigma}_{t|t-1}) + \widehat{u}_{\xi_1,t}\xi_{1,t}\\ w_t(\beta_{t-1}) &\simeq w_t(\widehat{b}_{t-1|t-1}) + \widehat{w}_t(\beta_{t-1} - \widehat{b}_{t-1|t-1})\\ s_t(\beta_{t-1}, \xi_{2t}) &\simeq s_t(\widehat{\beta}_{t-1|t-1}, 0) + \widehat{s}_{\beta,t}(\beta_{t-1} - \widehat{b}_{t-1|t-1}) + \widehat{s}_{\xi_2,t}\xi_{2,t}\\ t_t(\alpha_{t-1}) &\simeq t_t(\widehat{a}_{t-1|t-1}) + \widehat{T}_t(\alpha_{t-1} - \widehat{a}_{t-1|t-1})\\ r_t(\alpha_{t-1}, \xi_{3t}) &\simeq r_t(\widehat{\alpha}_{t-1|t-1}, 0) + \widehat{r}_{\alpha,t}(\alpha_{t-1} - \widehat{a}_{t-1|t-1}) + \widehat{r}_{\xi_3,t}\xi_{3,t}\\ f_t(\sigma_t) &\simeq f_t(\widehat{\sigma}_{t|t-1}) + \widehat{f}_t(\sigma_t - \widehat{\sigma}_{t|t-1})\\ h_t(\sigma_{t-1}) &\simeq h_t(\widehat{\sigma}_{t-1|t-1}) + \widehat{h}_t(\sigma_{t-1} - \widehat{\sigma}_{t-1|t-1})\\ k_t(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})) &\simeq k_t(\widehat{u}_{\xi_1,t-1}\xi_{1,t-1}) \end{aligned}$$

where $\widehat{Z}_{i,t}, i = 1, 2; \widehat{u}_{\sigma,t}$, $\widehat{u}_{\xi_1,t}$, $\widehat{w}_t$, $\widehat{T}_t$, $\widehat{s}_{\beta,t}$, $\widehat{s}_{\xi_2,t}$ $\widehat{r}_{\alpha,t}$, $\widehat{r}_{\xi_3,t}$ are matrices corresponding to the Jacobian of $z_t(.)$, $u_t(.)$,$w_t(.)$, $t_t(.)$, $s_t(.)$, $r_t(.)$, evaluated at $\beta_t = \widehat{b}_{t|t-1}, \alpha_t = \widehat{a}_{t|t-1}, \sigma_t = \widehat{\sigma}_{t|t-1}$, $\xi_{1,t} = \xi_{2,t} = \xi_{3,t} = 0$. Thus, the approximated model is

$$\begin{aligned} \widehat{y}_t &\simeq \widehat{Z}_{1t}\beta_t + \widehat{Z}_{2t}\alpha_t + \widehat{d}_t + \widehat{u}_{\xi_1,t}\xi_{1,t} && (32)\\ \beta_t &\simeq \widehat{w}_t\beta_{t-1} + \widehat{g}_t + \widehat{s}_{\xi_2,t}\xi_{2,t} && (33)\\ \alpha_t &\simeq \widehat{T}_t\alpha_{t-1} + \widehat{c}_t + \widehat{r}_{\xi_3,t}\xi_{3,t} && (34)\\ \widehat{f}_t\sigma_t &= \widehat{h}_t\sigma_{t-1} + k_t(\widehat{u}_{\xi_1,t-1}\xi_{1,t-1}) && (35) \end{aligned}$$

where

$$\widehat{d}_t = z_{1t}\left(\widehat{b}_{t|t-1}\right) - \widehat{Z}_{1t}\widehat{b}_{t|t-1} + z_{2t}\left(\widehat{a}_{t|t-1}\right) - \widehat{Z}_{2t}\widehat{a}_{t|t-1} + u(\widehat{\sigma}_{t|t-1}, 0) - \widehat{u}_{\sigma,t}(\widehat{\sigma}_{t|t-1} - \sigma_t) \quad (36)$$

$$\widehat{c}_t = t_t\left(\widehat{a}_{t-1|t-1}\right) - \widehat{T}_t\widehat{a}_{t-1|t-1} + r_t(\widehat{\alpha}_{t|t-1}, 0) - \widehat{r}_{\alpha,t}(\widehat{\alpha}_{t|t-1} - a_{t-1}) \tag{37}$$

$$\widehat{g}_t = w_t\left(\widehat{b}_{t-1|t-1}\right) - \widehat{W}_t\widehat{b}_{t-1|t-1} + s_t(\widehat{\beta}_{t|t-1}, 0) - \widehat{s}_{\beta,t}(\widehat{\beta}_{t|t-1} - b_{t-1}) \tag{38}$$

When i) $z_t(.)$, $w_t(.)$, $t_t(.)$, $u_t(.)$ are linear, ii) $s_t(.)$ is independent of β_t, iii) $r_t(.)$ is independent of α_t and iv) $u_t(.)$ is independent of σ_t, $\widehat{d}_t = \mathbf{0}$, $\widehat{c}_t = \mathbf{0}$, $\widehat{g}_t = \mathbf{0}$. In one of the cases considered by Rubio Ramírez et al. (2010) $\widehat{d}_t \neq \mathbf{0}$, while if the law of motion of the structural coefficient is non-linear or there are non-linear identification restrictions, $\widehat{c}_t \neq \mathbf{0}$ or $\widehat{g}_t \neq \mathbf{0}$.

5.2 Estimation

Since (32)-(35) are linear, the algorithm described in section 4 can now be applied. The only difference is that we now draw from distributions or proposals which are centered at the Extended Kalman Smoother estimates. For example, given $f(\alpha_0, y^T, \Sigma^T)$, we construct updated estimates according to

$$\widehat{f(a_{t|t})} = \widehat{f(a_{t|t-1})} + K_t\left[y_t - z_t\left(\widehat{f(a_{t|t-1})}\right)\right]$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}\widehat{Z}_t'\Gamma_t^{-1}\widehat{Z}_t P_{t|t-1}'$$

where $\widehat{f(a_{t|t-1})} = t_t(\widehat{f(a_{t-1|t-1})})$, $P_{t|t-1} = \widehat{T}_t P_{t-1|t-1}\widehat{T}_t' + \widehat{r}_{\xi_2,t}Q_{2t}\widehat{r}_{\xi_2,t}'$ $K_t = P_{t|t-1}\widehat{Z}_t'\Gamma_t^{-1}$, and $\Gamma_t = \widehat{Z}_t' P_{t|t-1}\widehat{Z}_t + \widehat{u}_{\xi_1,t}Q_{2t}\widehat{u}_{\xi_1,t}'$.

Smoothed estimates are $f^*(\alpha_{T|T}) = \widehat{f(a_{T|T})}$, $P^*_{T|T} = P_{T|T}$ and

$$f(\alpha_{t|t+1})^* = \widehat{f(a_{t|t})} + P_{t|t}\widehat{Z}_t' P_{t+1|t}^{-1}\left(f(\alpha_{t+1|t+2})^* - t_t(\widehat{f(a_{t|t})}\right) \tag{39}$$

$$P^*_{t|t+1} = P_{t|t} - P_{t|t}\widehat{Z}_t'\left[P_{t+1|t} + \widehat{r}_{\xi_2,t}Q_{2t}\widehat{r}_{\xi_2,t}'\right]^{-1}\widehat{Z}_t P_{t|t-1}' \tag{40}$$

for $t = T-1, \ldots, 1,$. Hence, when $f(\alpha)$ is nonlinear, we draw $f(\alpha)^T$ from a proposal centered at (39)-(40). Notice that the approximate model is used only in predicting and updating the mean squared error of $f(\alpha_t)$.

Depending on the exact specification of the non-linear model, one or more steps in the algorithm may require some adjustments. We describe the modifications needed for the application of section 6 in the on-line appendix.

6 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.

6.1 The SVAR model

The vector of endogenous variables is $y_t = (GDP_t, P_t, U_t, R_t, M_t, Pcom_t)'$, where GDP_t is a measure of aggregate output, P_t a measure of aggregate prices, U_t the unemployment rate, R_t the nominal interest rate, M_t a monetary aggregate and $Pcom_t$ represents a commodity price index. The structure of $A(\alpha_t)$ is as in table 1, where X indicates a non-zero coefficient.

Reduced form \ Structural	GDP_t	P_t	U_t	R_t	M_t	$Pcom_t$
Non-policy 1	1	0	0	0	0	0
Non-policy 2	X	1	0	0	0	0
Non-policy 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

Table 1: Identification restrictions

The structural model is identified via exclusion restrictions as follows:

1. *Information equation*: Commodity prices ($Pcom_t$) 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. *Non-policy block*: Following Bernanke and Blinder (1992), the non-policy 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.

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

$$\underbrace{\left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{1,t} & 1 & 0 & 0 & 0 & 0 \\ \alpha_{2,t} & \alpha_{5,t} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11,t} & 0 \\ \alpha_{3,t} & \alpha_{6,t} & 0 & \alpha_{9,t} & 1 & 0 \\ \alpha_{4,t} & \alpha_{7,t} & \alpha_8 & \alpha_{10,t} & \alpha_{12,t} & 1 \end{array} \right]}_{A(\alpha_t)} \times \left[\begin{array}{c} GDP_t \\ P_t \\ U_t \\ R_t \\ M_t \\ Pcom_t \end{array} \right] = A_t^+(L) \left[\begin{array}{c} GDP_{t-1} \\ P_{t-1} \\ U_{t-1} \\ R_{t-1} \\ M_{t-1} \\ Pcom_{t-1} \end{array} \right] + \Sigma_t \left[\begin{array}{c} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \varepsilon_t^3 \\ \varepsilon_t^{mp} \\ \varepsilon_t^{md} \\ \varepsilon_t^i \end{array} \right] \tag{41}$$

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

$$\Sigma_t = \left[\begin{array}{cccccc} \sigma_t^1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_t^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_t^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_t^{mp} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_t^{md} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_t^i \end{array} \right]$$

is the matrix of standard deviations of the structural shocks.

The structural model (41) is non-recursive and overidentified by 3 restrictions. Overidentification obtains because the policy equation is different from the Taylor rule generally employed in the literature. It is easy to check (see on-line appendix A) that the (constant coefficient version of the) system is globally identified and therefore suitable for interesting policy experiments.

6.2 The prior and computation details

The VAR is estimated with 2 lags; this is what the BIC criteria selects for the constant coefficient version of the model. The priors are proper, conjugate for computa-

tional convenience, and given by $B_0^{prior} \sim N\left(\overline{\mathbf{B}}, 4 \cdot \overline{\mathbf{VB}}\right)$, $Q^{prior} \sim IW\left(k_Q^2 \cdot \overline{\mathbf{VB}}, (1+K)\right)$, $\alpha_0^{prior} \sim N\left(\overline{\boldsymbol{\alpha}}, diag\left(abs\left(\overline{\boldsymbol{\alpha}}\right)\right)\right)$, $S^{prior} \sim IW\left(k_S^2 \cdot diag\left(abs\left(\overline{\boldsymbol{\alpha}}\right)\right), (1+\dim\alpha)\right)$, $\log\left(\sigma_0\right)^{prior} \sim N\left(\overline{\boldsymbol{\sigma}}, 10 \cdot I_M\right)$, $W_i^{prior} \sim IW\left(k_W^2, 1+1\right), i = 1, \ldots, M$.

To calibrate the hyperparameters, we use the first 40 observations as a training sample: $\overline{\mathbf{B}}$ and $\overline{\mathbf{VB}}$ are estimated with OLS and $\overline{\boldsymbol{\alpha}}$ and $\overline{\boldsymbol{\sigma}}$ 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 one 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 $I_\alpha(.)$, 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 percent.

Since the structural model has $M = 6$, $\dim(\alpha) = 12$, and S_A and s_A are

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

$$s_A = [e_1', e_2', e_3', e_4', e_5', e_6']'$$

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

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

6.3 The Data

The data comes from the *International Financial Statistics (IFS)* database 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, respectively). 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 GDP deflator, the unemployment rate, the aggregate Gross Domestic Product index (Volume, base 2005=100), the commodity prices index, and M2 are 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 standardized, that is, we use $(y_t^* - E(y_t^*))/std(y_t^*)$, to have all the variables on the same scale.

6.4 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 non-explosive 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 $t = 1, \ldots, T$. 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 the on-line 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 function of a much lower dimensional vector of factors θ_t, who independently move as a random walk, and this can reduce the inefficiency of the algorithm. We describe how this algorithm works in the on-line 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 non-standardized. In fact the acceptance rate is only 0.46% when year-on-year growth rates are used and 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 displays lower persistence. When the data is 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 averages acceptance rates for B^T when the data is standardized are 97% and 91% for year-on-year and quarterly growth rates, respectively. However, the efficiency gains are more than compensated by the higher 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), one factor for each variable (6). Ξ is a 72×15 matrix loading the factors on the required elements of the B_t vector. When Ξ is composed of zeros and ones, we needed about 10 hours to estimate the model and the acceptance rate is 78% when the data is standardized. When the elements of Ξ are also estimated the computational time increases to about 24 hours and the acceptance rate for B^T drops to 24%, when data is standardized.

In sum, both the multi-move and the shrinkage algorithms have reasonable computational costs but the latter has better acceptance rates. 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 this more than compensates for the efficient gains.

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

6.5 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 percent posterior tunnel for the variability of the monetary policy shock and in figure 2 the highest 68 percent posterior tunnel for the non-zero contemporaneous structural parameters 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, which 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 non-policy parameters $[\alpha_{1,t}, \alpha_{2,t}, \alpha_{5,t}]$ exhibit considerable time variations which 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.

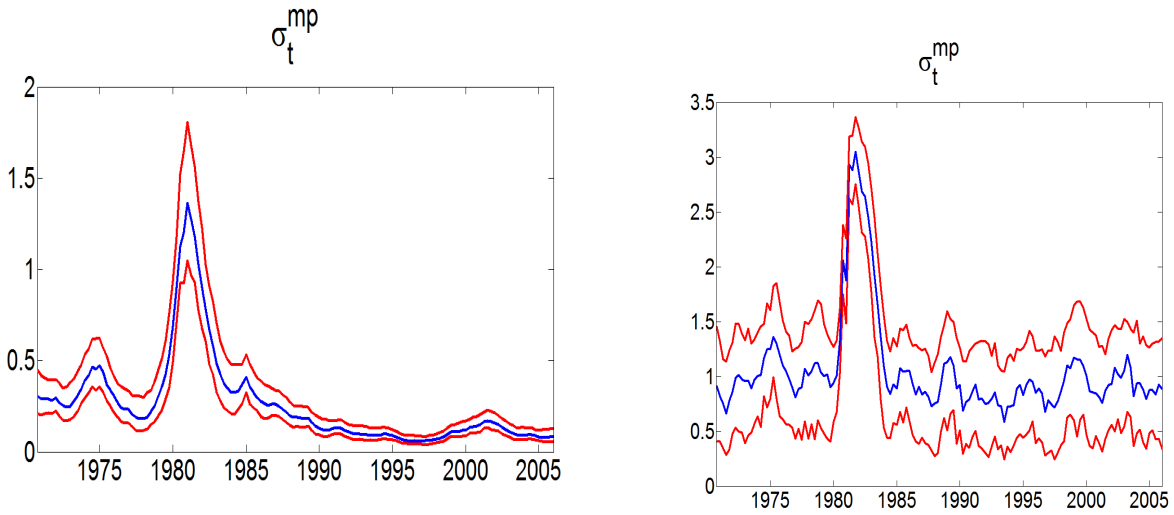

Figure 1: Median and posterior 68 percent tunnels, volatility of monetary policy shock. Left panel: stochastic volatility; right panel GARCH(1,1).

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 in 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 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 one of money in prices does not hold for a large portion of our sample. Finally, 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}). Thus our setup captures the idea that policy and private sector parameters move together.

Thus, in agreement with the 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 non-policy coefficients.

Figure 2: Estimates of α_t

6.6 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 one 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 one and one with the structural shock set to zero.

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.

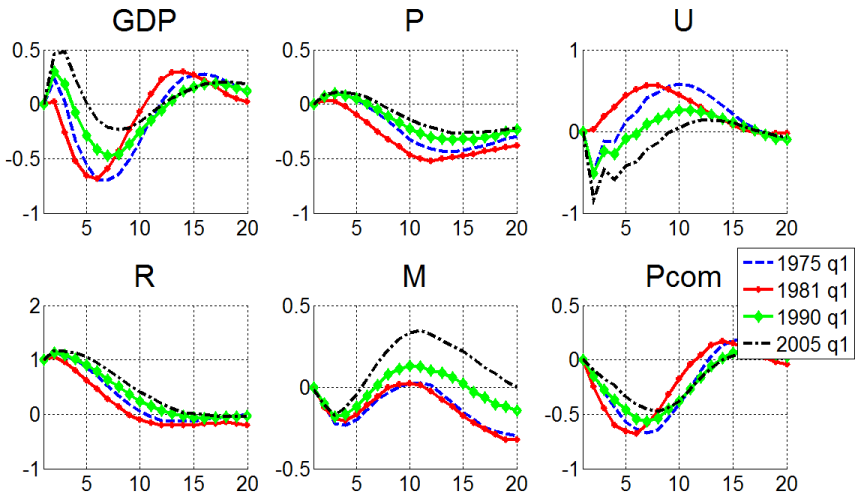

Figure 3: Dynamics following a monetary policy shock, different dates.

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 percent of the forecast error variance of inflation at all dates and about 15-20 percent of the variability of output growth and the unemployment rate, with a maximum of about 25 percent 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 Canova et al. (2008), even though they use sign restrictions to extract structural shocks, and with those of

Boivin and Giannoni (2006), who use sub-sample 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 et al. (2010), who allow for stochastic volatility and time variations only in the coefficients of the policy rule.

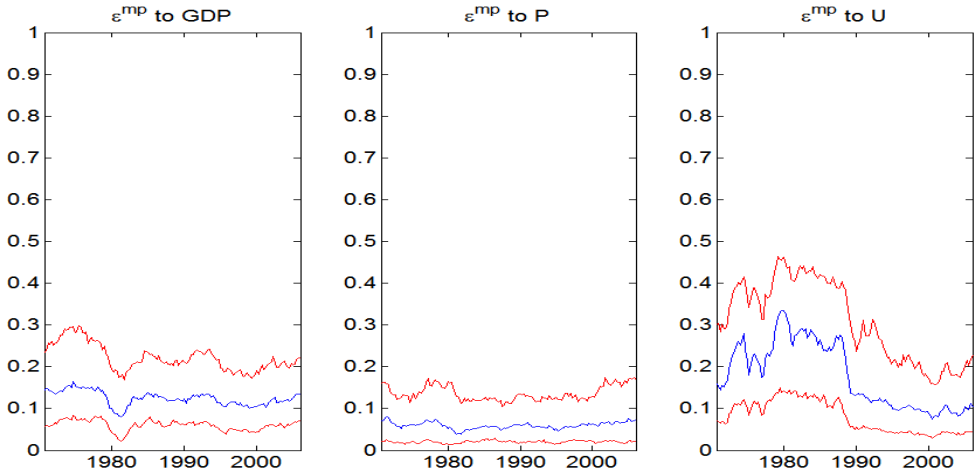

Figure 4: Percentage of the forecast error variance due to monetary policy shock in the long run, different dates.

6.7 A time invariant over-identified 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.

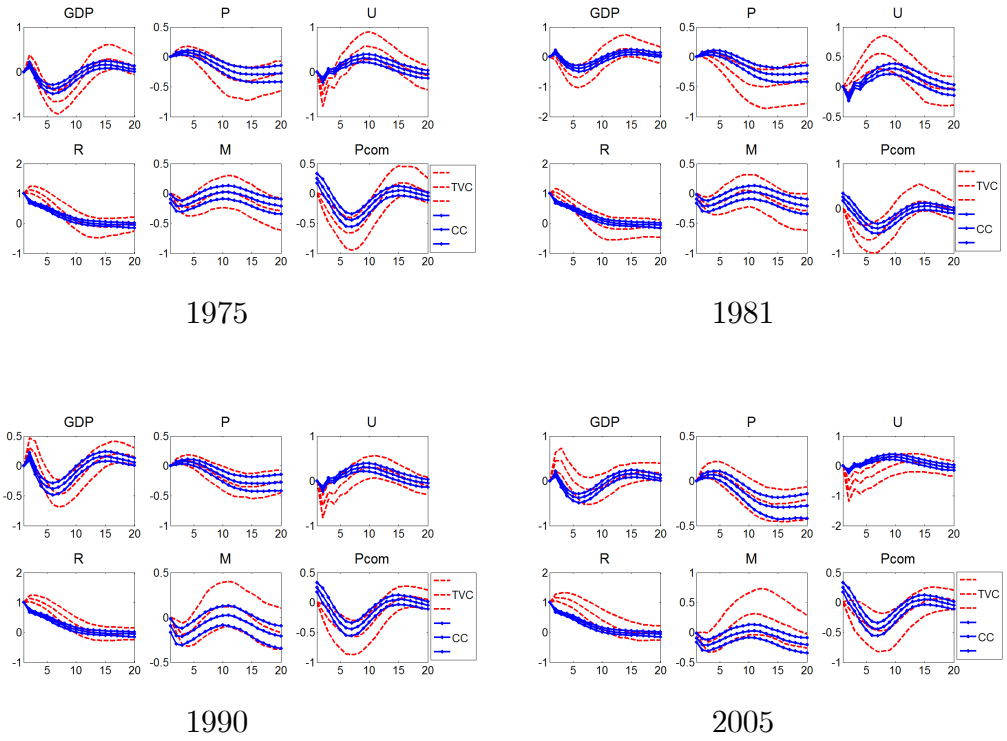

Figure 5: Time varying and time invariant responses.

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.

6.8 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 (23) we use

$$\sigma_{m,t}^2 = (1-\delta) + \delta\sigma_{m,t-1}^2 + \delta\left(y_{m,t-1}^{**}\right)^2 + \eta_{m,t} \tag{42}$$

Since with this GARCH(1,1), the resulting model is non-linear, we employ the setup of section 5 to draw sequences for the parameters. 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)

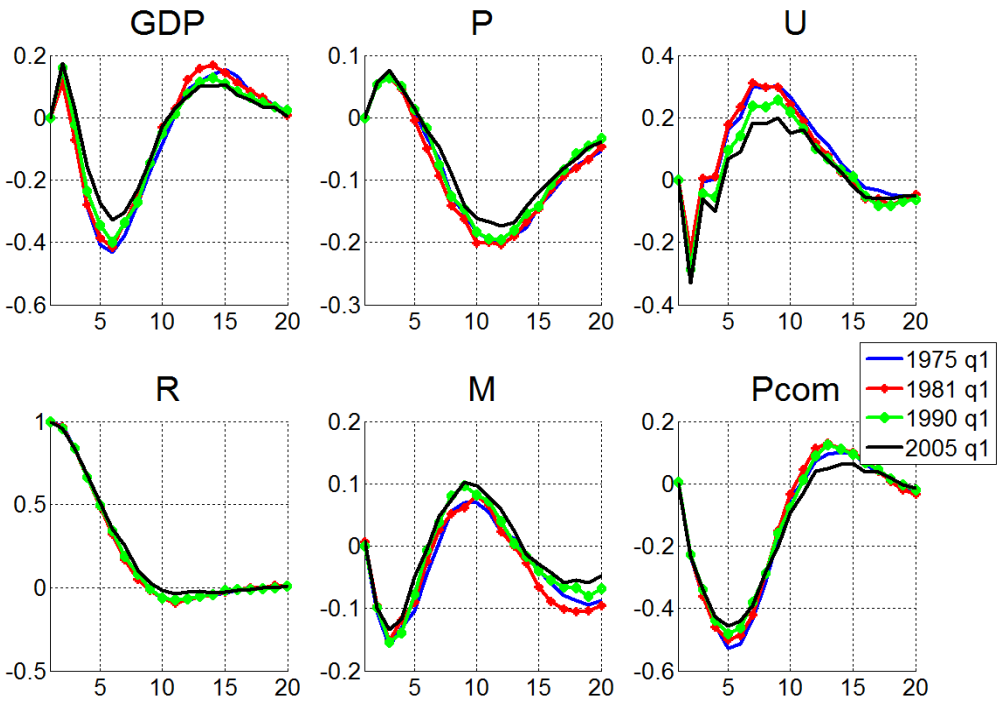

Figure 6: Transmission of monetary policy shocks, GARCH(1,1) specification

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 prize 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 is fifty percent taller. Because a larger portion of the dynamics of the endogenous variables is now capture 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.

6.9 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 which allow long run movements of the unemployment rate in response to monetary policy shocks, see e.g. Farmer and Benhabib (2000). Also in this case, the model becomes nonlinear and the algorithm described in section 5 is used to estimate the model. Details on the modifications needed are in the on-line appendix F.

Figure 7 presents the responses of the variables to a monetary policy shock. The basic qualitative feature 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 stronger. In general, when long run restrictions are imposed, time variations in the transmission of monetary policy shocks are increased.

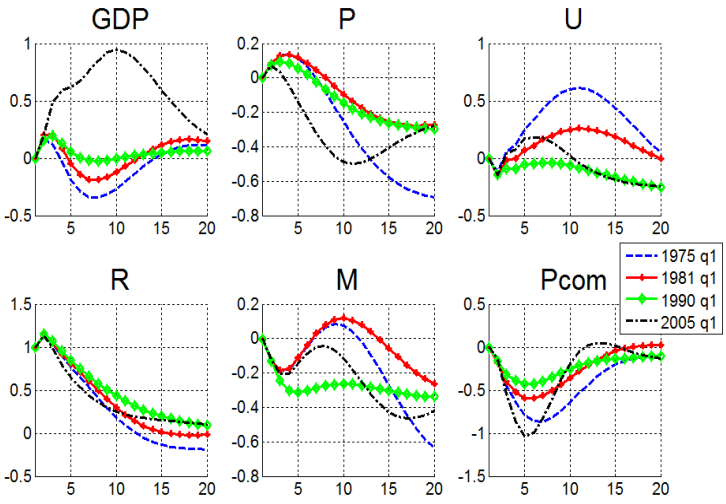

Figure 7: Transmission of monetary policy shocks, long run restrictions.

7 Conclusions

This paper proposes a unified framework to estimate structural VARs. The methodology can handle time varying coefficient or time invariant models identified with

recursive or non-recursive constraints that can be linear or non-linear and that can produce just identified or overidentified systems. Our algorithm adds a Metropolis step to a standard Gibbs sampling routine but nests the structural model into a general non-linear state space system. 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 non-recursive overidentified TVC model similar to the one used by Robertson and Tallman (2001), Waggoner and Zha (2003) with fixed coefficients. We examine the merits of multi-move vs. single-move routines and find that once data are standardized, the computational costs of using a single-move routine are larger than the efficiency gains. We show that there are time variations in the variance of the monetary policy shock and in the estimated contemporaneous coefficients. These variations translate in 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 parametrizes. One could also use the same methodology to identify shocks imposing magnitude restrictions on impulse responses, as in Rubio Ramírez et al. (2010) or variance decomposition restrictions, as in Barsky and Sims (2012). The estimation complexity is important but not overwhelming and all the computations can be performed on a standard PC with sufficient RAM memory.

References

[1] Amisano, G. and C. Giannini (1997). Topics in Structural VAR Econometrics. Springer and Verlag.

[2] Arias, J., D. Waggoner and J. Rubio Ramírez (2014). Inference based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications. Federal Reserve Bank of Atlanta, Working Paper 2014-1.

[3] Barsky, R. and Sims, E (2012). News shocks and the business cycle. Journal of Monetary Economics, 58, 273-289.

[4] Baumeister, C. and J. Hamilton (2013) Sign Restrictions, Structural Vector Autoregressions and Useful Prior Information, manuscript.

[5] Benhabib, J. and R. Farmer (2000) The Monetary Transmission, Review of Economic Dynamics, 3, 523-550.

[6] Bernanke, B. S. and A. S. Blinder (1992). The federal funds rate and the channels of monetary transmission. American Economic Review, 82, 901–921.

[7] Bernanke, B. S. and I. Mihov (1998). Measuring monetary policy. Quarterly Journal of Economics, 113, 869–902.

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

[9] Canova, F. and M. Ciccarelli (2009). Estimating multicountry VAR models. International Economic Review, 50, 929-961.

[10] 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.

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

[12] Canova, F. and F. Ferroni (2012). The dynamics of US inflation: Can monetary policy explain the changes? Journal of Econometrics, 167, 47–60.

[13] Canova, F., L. Gambetti, and E. Pappa (2008). The structural dynamics of U.S. output and inflation: What explains the changes? Journal of Money, Credit and Banking, 40, 369–388.

[14] Canova, F. and M. Paustian (2011). Business cycle measurement with some theory. Journal of Monetary Economics, 58, 345–361.

[15] Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. Biometrika, 81, 541–553.

[16] Chib, S. and E. Greenberg (1995). Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. Journal of Econometrics, 68, iss.2, 339-360.

[17] 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.

[18] Creel, D. (2012) Survey of Sequential Monte Carlo methods for Economics and Finance. Econometric Reviews, 31, 245-296.

[19] Del Negro, M. and G. Primiceri (2013). Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. Federal Reserve Bank of New York Staff Reports May 2013 Number 619.

[20] 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. CEPR discussion paper 7813.

[21] 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.

[22] 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.

[23] Herbst, E. and Schorfheide, F. (2013). Sequential Monte Carlo Sampling for DSGE models, University of Pennsylvania, manuscript.

[24] Justiniano, A. and G. E. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. American Economic Review, 98, 604–641.

[25] Kim, C.-J. and C. R. Nelson (1999). State-Space Models with Regime-Switching: Classical and Gibbs Sampling Approaches with Applications. MIT Press.

[26] Kociecki, A. Rubaszek, M. and Ca' Zorzi, M. (2013) Bayesian Analysis of recursive SVAR models with overidentifying restrictions, Bank of Poland, manuscript.

[27] Koop, G. and Potter, S. (2011).Time varying VARs with inequality restrictions. Journal of Economic Dynamics and Control, 35, 1126-1138.

[28] Lakdawala, A. (2012). Changes in Federal Reserve preferences. Manuscript, Michigan State University.

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

[30] Leeper, E., C. Sims, and T. Zha (1996). What does monetary policy do? Brookings Papers on Economic Activity, 2, 2-78.

[31] Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72, 821–852.

[32] Robertson, J. C. and E. W. Tallman (2001). Improving federal-funds rate forecasts in VAR models used for policy analysis. Journal of Business and Economic Statistics, 19, 324–330.

[33] 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.

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

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

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

[37] 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.

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

[39] Waggoner, D. and T. Zha (2003). A Gibbs sampler for structural vector autoregressions. Journal of Economic Dynamics and Control, 28, 349–366.

On-line Appendix

A: Global Identification of the constant coefficient SVAR of section 6

Consider the constant coefficients version of the SVAR model used in section 6:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11} & 0 \\ \alpha_3 & \alpha_6 & 0 & \alpha_9 & 1 & 0 \\ \alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1 \end{bmatrix}}_{A(\alpha)} \times \begin{bmatrix} GDP_t \\ P_t \\ U_t \\ R_t \\ M_t \\ Pcom_t \end{bmatrix} = A^+(L) \begin{bmatrix} GDP_{t-1} \\ P_{t-1} \\ U_{t-1} \\ R_{t-1} \\ M_{t-1} \\ P_com_{t-1} \end{bmatrix} + \Sigma \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^p \\ \varepsilon_t^u \\ \varepsilon_t^{mp} \\ \varepsilon_t^{md} \\ \varepsilon_t^i \end{bmatrix}$$

with

$$\Sigma = \begin{bmatrix} \sigma^i & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^{md} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^{mp} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^y & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^p & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^u \end{bmatrix}$$

To verify that the system is globally identified, we rewrite the model using the notation of Rubio Ramirez et. al. (2010). Let $y_t \equiv (GDP_t, P_t, U_t, R_t, M_t, Pcom_t)'$ and $\varepsilon_t \equiv \left(\varepsilon_t^y, \varepsilon_t^p, \varepsilon_t^u, \varepsilon_t^{mp}, \varepsilon_t^{md}, \varepsilon_t^i\right)'$. Pre-multiplying by Σ^{-1}, we obtain

$$\Sigma^{-1} A(\alpha) y_t = \Sigma^{-1} A^+(L) y_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim N(0, I_6)$. Define $\mathbf{A}_0' \equiv \Sigma^{-1} A(\alpha)$ and $\mathbf{A}'(L) \equiv \Sigma^{-1} A^+(L)$. Then:

$$y_t' \mathbf{A}_0 = \sum_{L=1}^{p} y_{t-L}' \mathbf{A}_L + \varepsilon_t'$$

where

$$\mathbf{A}_0' = \begin{bmatrix} \sigma^y & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^p & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^u & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^{mp} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^{md} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^i \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11} & 0 \\ \alpha_3 & \alpha_6 & 0 & \alpha_9 & 1 & 0 \\ \alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^y} & 0 & 0 & 0 & 0 & 0 \\ \frac{\alpha_1}{\sigma^p} & \frac{1}{\sigma^p} & 0 & 0 & 0 & 0 \\ \frac{\alpha_2}{\sigma^u} & \frac{\alpha_5}{\sigma^u} & \frac{1}{\sigma^u} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma^{mp}} & \frac{\alpha_{11}}{\sigma^{mp}} & 0 \\ \frac{\alpha_3}{\sigma^{md}} & \frac{\alpha_6}{\sigma^{md}} & 0 & \frac{\alpha_9}{\sigma^{md}} & \frac{1}{\sigma^{md}} & 0 \\ \frac{\alpha_4}{\sigma^i} & \frac{\alpha_7}{\sigma^i} & \frac{\alpha_8}{\sigma^i} & \frac{\alpha_{10}}{\sigma^i} & \frac{\alpha_{12}}{\sigma^i} & \frac{1}{\sigma^i} \end{bmatrix}$$

Denoting $\mathbf{A}_0 = [a_{kj}]$ we have

$$\mathbf{A}_0 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\ 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \end{bmatrix}$$

The matrices $\mathbf{Q}_j$, $j = 1, \ldots, 6$, present in Theorem 1 of Rubio Ramirez et al. (2010) are:

$$\mathbf{Q}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Q}_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{Q}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Q}_5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{Q}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Define the matrices

$$\mathbf{M}_j(\mathbf{A}_0) = \begin{bmatrix} & \mathbf{Q}_j\mathbf{A}_0 \\ \begin{bmatrix} \mathbf{I}_j \end{bmatrix} & \mathbf{0}_{j\times(M-j)} \end{bmatrix}; \quad j = 1, \ldots, M \tag{43}$$

so that

$$\mathbf{M}_1 = \begin{bmatrix} 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}_3 = \begin{bmatrix} 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}; \mathbf{M}_4 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\ 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}_5 = \begin{bmatrix} 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \mathbf{M}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Since all $\mathbf{M}_j$ have full column rank, the model is globally identified.

B: Single-move Metropolis for drawing B_t

The Koop and Potter's (2011) approach for drawing the elements of the B^T sequence separately works as follows. Given $f(\alpha_{i-1})^T, (\Sigma^{i-1})^T, Q^{i-1}, V^{i-1}, W^{i-1}$, the measurement equation is

$$y_t = X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t$$

and the transition equation is

$$B_t = B_{t-1} + \upsilon_t$$

with $\upsilon_t \sim N(0, Q)$, B_0 given, and $A_t^{-1}\Sigma_t\varepsilon_t = u_t \sim N(0, \Omega_t)$. To sample the individual elements of B^T, all $t \geq 1$:

1. Draw a candidate $B_t^{\dagger} \sim N(\mu_t, \Psi_t)$ where

$$\mu_t = \begin{cases} \frac{B_{t-1}^i + B_{t+1}^{i-1}}{2} + G_t \left[y_t - X_t' \left(\frac{B_{t-1}^i + B_{t+1}^{i-1}}{2} \right) \right] & , t < T \\ B_{t-1}^i + G_t \left[y_t - X_t' \left(B_{t-1}^i \right) \right] & , t = T \end{cases}$$

$$G_t = \begin{cases} \frac{1}{2} Q^{i-1} X_t \left(X_t' Q^{i-1} X_t + \Omega_t \right)^{-1} & , t < T \\ Q^{i-1} X_t \left(X_t' Q^{i-1} X_t + \Omega_t \right)^{-1} & , t = T \end{cases}$$

$$\Psi_t = \begin{cases} \frac{1}{2} \left(I_K - G_t X_t' \right) Q^{i-1} & , t < T \\ \left(I_K - G_t X_t' \right) Q^{i-1} & , t = T \end{cases}$$

2. Construct the companion form matrix $\overline{\mathbf{B}}_t^{\dagger}$ and evaluate $\mathcal{I}\left(\max \left| eig\left(\overline{\mathbf{B}}_t^{\dagger} \right) \right| < 1 \right)$, where $\mathcal{I}(.)$ is an indicator function taking the value of 1 if the condition within the parenthesis is satisfied.

3. The acceptance rate of $B_t^{\dagger}$ is

$$\omega_{B,t} = \min\left\{\frac{\frac{\mathcal{I}\left(\max\left|eig\left(\overline{\mathbf{B}}_t^{\dagger}\right)\right|<1\right)}{\lambda\left(B_t^{\dagger},Q^{i-1}\right)}}{\frac{1}{\lambda\left(B_t^{i-1},Q^{i-1}\right)}},1\right\} = \min\left\{\frac{\mathcal{I}\left(\max\left|eig\left(\overline{\mathbf{B}}_t^{\dagger}\right)\right|<1\right)\lambda\left(B_t^{i-1},Q^{i-1}\right)}{\lambda\left(B_t^{\dagger},Q^{i-1}\right)},1\right\}$$

where $\lambda(.)$ is an integrating constant, measuring the proportion of draws that satisfy the inequality constraint. To compute $\lambda(.)$ one first draws $B_t^{\dagger,l} \sim N\left(B_t^{\dagger},Q^{i-1}\right)$, for $l=1,\ldots,\overline{L}$, constructs the companion form matrix $\overline{\mathbf{B}}_t^{\dagger,l}$ and evaluates $\lambda_l = \mathcal{I}\left(\max\left|eig\left(\overline{\mathbf{B}}_t^{\dagger,l}\right)\right|<1\right)$. Second, one evaluates $\lambda\left(B_t^{\dagger},Q^{i-1}\right) = \frac{\sum_{l=1}^{\overline{L}}\lambda_l}{\overline{L}}$ and $\lambda\left(B_{t,i-1},Q^{i-1}\right)$ and compute the acceptance probability. When $t=T$, this probability is

$$\omega_{B,T} = \mathcal{I}\left(\max\left|eig\left(\overline{\mathbf{B}}_t^{\dagger}\right)\right|<1\right)$$

4. Draw a $v\sim U(0,1)$. Set $B_t^i = B_t^c$ if $v<\omega_{B,t}$ and set $B_t^i = B_t^{i-1}$ otherwise.

Since Q depends on B_t, we need to change the sampling scheme also for this matrix. Assume that $Q^{-1}\sim W\left(\underline{v},\underline{Q}^{-1}\right)$ so that the unrestricted posterior is $Q^{-1}\sim W\left(\overline{v},\overline{Q}^{-1}\right)$ with $\overline{v}=\underline{v}+T$ and $\overline{Q}^{-1} = \left[\underline{Q}+\sum_{t=1}^{T}\left(B_{t,i}-B_{t-1,i}\right)\left(B_{t,i}-B_{t-1,i}\right)'\right]^{-1}$. Then draw a candidate $\left(Q^{\dagger}\right)^{-1}\sim W\left(\overline{v},\overline{Q}^{-1}\right)$ and for $t=1,\ldots,T$, evaluate $\lambda\left(B_t^i,Q^{\dagger}\right)$ and $\lambda\left(B_t^i,Q^{i-1}\right)$, for a fixed $\overline{L}$, and calculate

$$\omega_Q = \min\left\{\prod_{t=1}^{T}\frac{\lambda\left(B_t^i,Q^{i-1}\right)}{\lambda\left(B_t^i,Q^{\dagger}\right)},1\right\}$$

Finally, we draw a $v\sim U(0,1)$, set $Q^i=Q^c$ if $v<\omega_Q$ and $Q^i=Q^{i-1}$. In the exercise we conduct in section 6, we set $\overline{L}=25$, when evaluate the integrating constants $\lambda(.)$ at each t.

Note that in a multi-move approach $\lambda(.)=1$, when sampling both B^T and Q. Therefore, Koop and Potter's approach generalizes the multi-move procedure at the cost of making convergence to the posterior, in general, much slower and, because $\lambda(.)$ needs to be simulated at each t, of adding considerable computational time.

C: A shrinkage approach to draw B^T when Ξ is known

The model is still consists of (20), (22) and (23) but now (21) is substituted by

$$\begin{aligned} B_t &= \Xi\theta_t + \upsilon_t \qquad \upsilon_t \sim N(0, I) \quad (44) \\ \theta_t &= \theta_{t-1} + \rho_t \qquad \rho_t \sim N(0, Q) \quad (45) \end{aligned}$$

where $\dim(\theta_t) \ll \dim(B_t)$ and where the matrix Ξ is known, as in Canova and Ciccarelli (2009). Using (45) into (44) we have

$$y_t = X_t'\Xi\theta_t + A_t^{-1}\Sigma_t\varepsilon_t + X_t'\upsilon_t \equiv X_t'\Xi\theta_t + \psi_t \quad (46)$$

where $\psi_t \sim N(0, H_t)$ with $H_t \equiv A_t^{-1}\Sigma_t\Sigma_t'\left(A_t^{-1}\right)' + X_t'X_t$.

To estimate the unknowns we do the following:

1. Sample θ^T with a multi-move routine using (46) and (45) .

2. Given θ^T, we compute $\widehat{y}_t = y_t - X_t'\Xi\theta_t$. Pre-multiplying by A_t, we get the concentrated structural model

$$A_t\widehat{y}_t = A_t\xi_t = \Sigma_t\varepsilon_t + A_tX_t'\upsilon_t$$

As before

$$\left(\widehat{y}_t' \otimes I_M\right)\left(S_A f(\alpha_t) + s_A\right) = \Sigma_t\varepsilon_t + A_tX_t'\upsilon_t$$

so that the second state-space system is

$$\begin{aligned} \widetilde{y}_t &= Z_t f(\alpha_t) + \Sigma_t\varepsilon_t + A_tX_t'\upsilon_t \quad (47) \\ f(\alpha_t) &= f(\alpha_{t-1}) + \zeta_t \quad (48) \end{aligned}$$

and we draw $f(\alpha)^T$ using our proposed Metropolis step. The variance of the measurement error is $\Sigma_t\Sigma_t' + A_t(\alpha_t)X_t'X_tA_t'(\alpha_t)$ and it is evaluated at $f(\alpha_{t|t-1})$.

3. Given $(\theta^T,\ f(\alpha)^T)$:

$$\widehat{A}_t\widehat{y}_t = \Sigma_t\varepsilon_t + \widehat{A}_tX_t'\upsilon_t$$

Since $\widehat{A}_tX_t'$ is known, let the lower-triangular P_t satisfy $P_t\left(\widehat{A}_tX_t'X_t\widehat{A}_t'\right)P_t' = I$. Then

$$P_t\widehat{A}_t\widehat{y}_t = y_t^{**} = P_t\Sigma_t\varepsilon_t + P_t\widehat{A}_tX_t'\upsilon_t$$

with $var\left(P_t\widehat{A}_tX_t'\upsilon_t\right) = I$ and where $P_t\Sigma_t\Sigma_t'P_t' + P_t\left(\widehat{A}_tX_t'X_t\widehat{A}_t'\right)P_t'$ is a diagonal matrix. This transformation is similar to Cogley and Sargent (2005); however, since $\widehat{A}_tX_t'$ is known, we only need to sample the variances of $\epsilon_{m,t}$. We do this using the $\log(\chi^2)$ approximation of a mixture of J normals.

4. Given $(\theta^T, f(\alpha)^T, \Sigma^T)$, sample Q, V, W from independent inverted Wishart distributions.

5. Given new values of $\sigma_{m,t}$, we construct $A_t^{-1}\Sigma_t\Sigma_t'\left(A_t^{-1}\right)' + X_t'X_t$ and go back to step 1.

D: A Shrinkage approach to draw B^T when Ξ is unknown

When the $\Xi's$ are known, the algorithm needs to be modified as follows.

The TVC-SVAR model is:

$$y_t = X_t'B_t + A_t^{-1}\Sigma_t\varepsilon_t$$

where $X_t' = I_M \otimes \left[D_t', y_{t-1}', \ldots, y_{t-k}'\right]$, with

$$B_t = \Xi\theta_t + \omega_t$$

$$\theta_t = \theta_{t-1} + \upsilon_t$$

$$f(\alpha_t) = f(\alpha_{t-1}) + \zeta_t$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t$$

$$Var\left(\begin{bmatrix} \varepsilon_t \\ \omega_t \\ \upsilon_t \\ \zeta_t \\ \eta_t \end{bmatrix}\right) = \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 & 0 \\ 0 & 0 & R & 0 & 0 \\ 0 & 0 & 0 & V & 0 \\ 0 & 0 & 0 & 0 & W \end{bmatrix}$$

where Q and R are diagonal matrices. We exploit the hierarchical structure of the model to simulate the posterior distribution, as in Chib and Greenberg (1995):

1. Given $(\alpha_t,\ \sigma_t\ Q)$, sample B_t using:

$$y_t = X_t'B_t + A_t^{-1}\Sigma_t\varepsilon_t$$

with $A_t^{-1}\Sigma_t\varepsilon_t = u_t \sim N(0, H_t)$. That is, for each $t = 1, \ldots, T$ draw:

$$B_t \sim N\left(\overline{B}_t, \overline{VB}_t\right)$$

where

$$\begin{aligned} \overline{VB}_t &= \left(\underline{VB}^{-1} + X_tH_t^{-1}X_t'\right)^{-1} \\ \overline{B}_t &= \overline{VB}_t\left(\underline{VB}^{-1}\underline{B}_t + X_tH_t^{-1}y_t\right) \end{aligned}$$

and priors

$$\underline{VB} = Q;\ \ \underline{B}_t = \Xi\theta_t$$

2. Given (B_t, θ_t), compute the residuals $(B_t - \Xi\theta_t)$ and sample Q using an inverse Wishart distribution.

3. Given B_t, sample θ_t using the state space form:

$$B_t = \Xi\theta_t + \omega_t$$

$$\theta_t = \theta_{t-1} + \upsilon_t$$

4. Given θ_t, sample R using an inverse Wishart distribution.

5. Given (B_t, θ_t, Q) draw Ξ using:

$$B_t = \Xi\theta_t + \omega_t; \;\; t = 1, \ldots, T$$

where, in order to achieve identification, we normalize the first upper block of Ξ to be an identity matrix, as in Koop and Korobilis (2010). That is, denote $\mathcal{F} = \dim(\theta_t)$ and $K = \dim(B_t)$, then Ξ is a $K \times \mathcal{F}$ matrix. The first $\mathcal{F}$ rows of Ξ are:

$$\Xi_{(1:\mathcal{F})\times(1:\mathcal{F})} = I_{\mathcal{F}}$$

Moreover, since $\omega_t \sim N(0, Q)$, and we have assumed that Q is diagonal, we draw the loadings row by row for each element of B_t. That is, for each $f = \mathcal{F} + 1, \ldots, K$ draw:

$$\Xi_{f\times(1:\mathcal{F})} \sim N\left(\overline{\Xi}_f, \overline{V\Xi}_f\right)$$

with

$$\begin{aligned} \overline{V\Xi}_f &= \left(\underline{V\Xi}^{-1} + Q^{-1}_{(f,f)}\left(\theta^T\right)\left(\theta^T\right)'\right)^{-1} \\ \overline{\Xi}_f &= \overline{V\Xi}_f\left(\underline{V\Xi}^{-1}\underline{\Xi}_f + Q^{-1}_{(f,f)}\theta^T B_f^T\right) \end{aligned}$$

where θ^T is a $\mathcal{F} \times T$ matrix of explanatory variables, B_f^T is a $T \times 1$ vector that contains the dependent variable and $Q_{(f,f)}$ is the corresponding element of matrix Q drawn previously. The priors are $\underline{\Xi}_f = \mathbf{0}_{\mathcal{F}\times 1}$; $\underline{V\Xi} = k_\Xi^2 I_{\mathcal{F}}$ with the hyperparameter $k_\Xi^2 = 0.01$.

6. Given (B_t, Q), sample (f(α_t), V, σ_t, W) as before. Then go back to 1.

E: Sampling the GARCH model

To sample volatilities when their law of motion is assumed to be a GARCH(1,1), we need to modify the transition and the measurement equations used in step 3 of the algorithm of section 4. The $m-th$ equation of the model is:

$$y_{m,t}^{**} = \sigma_{m,t}\varepsilon_{m,t} \tag{49}$$

where $\sigma_{m,t}$ is the $m-$th diagonal element of Σ_t. Assume

$$\sigma_{m,t}^2 = (1-\delta+\delta\sigma_{m,t-1}^2+\delta\left(y_{m,t-1}^{**}\right)^2)+\eta_{m,t} \tag{50}$$

with $\eta_t \sim N(0,W)$, where δ and W are known parameters.

The system $(49)-(50)$ is now non-linear. Equation (49) can be written as:

$$y_{m,t}^{**} = z(\sigma_{m,t}) + u_t(\sigma_{m,t},\varepsilon_{m,t})$$

Since $z(\sigma_{m,t})=0$, the linear approximation is:

$$\sigma_{m,t}\varepsilon_{m,t} \simeq u_t(\widehat{\sigma}_{m,t|t-1},0)+\widehat{u}_{\sigma,t}(\sigma_{m,t}-\widehat{\sigma}_{m,t|t-1})+\widehat{u}_{\varepsilon_{m,t}}\varepsilon_{m,t} = \widehat{\sigma}_{m,t|t-1}\varepsilon_{m,t}$$

because:

- $u_t(\widehat{\sigma}_{m,t|t-1},0)=\widehat{\sigma}_{m,t|t-1}\times 0=0$
- $\widehat{u}_{\sigma,t} = \left.\frac{\partial u_t(\sigma_{m,t},\varepsilon_{m,t})}{\partial\sigma_{m,t}}\right|_{\left(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0\right)} = \varepsilon_{m,t}|_{\left(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0\right)}=0$
- $\widehat{u}_{\varepsilon_{m,t}} = \left.\frac{\partial u_t(\sigma_{m,t},\varepsilon_{m,t})}{\partial\varepsilon_{m,t}}\right|_{\left(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0\right)} = \sigma_{m,t}|_{\left(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0\right)}=\widehat{\sigma}_{m,t|t-1}$

The transition equation (50) can be written as:

$$\sigma_{m,t}^2 \equiv f_t(\sigma_{m,t}) = h_t(\sigma_{m,t-1})+k_t(\sigma_{m,t-1},\eta_{m,t}) \equiv \left(1-\delta+\delta\sigma_{m,t-1}^2+\delta\left(y_{m,t-1}^{**}\right)^2\right)+\eta_{m,t}$$

Linearizing the two sides of the equation we have:

$$f_t(\sigma_{m,t}) \simeq f_t(\widehat{\sigma}_{m,t|t-1})+\widehat{f}_t(\widehat{\sigma}_{m,t-1|t-1})(\sigma_{m,t-1}-\widehat{\sigma}_{m,t-1|t-1})$$

$$h_t(\sigma_{m,t-1}) \simeq h_t(\widehat{\sigma}_{m,t-1|t-1})+\widehat{h}_t(\widehat{\sigma}_{m,t-1|t-1})(\sigma_{m,t-1}-\widehat{\sigma}_{m,t-1|t-1})$$

where $\widehat{f}_t(\widehat{\sigma}_{m,t|t-1},0) = 2\sigma_{m,t}|_{(\widehat{\sigma}_{m,t|t-1},0)}$ and $\widehat{h}_t(\widehat{\sigma}_{m,t-1|t-1},0) = 2\delta\sigma_{m,t-1}|_{(\widehat{\sigma}_{m,t|t-1},0)}$.

F: Long-run restrictions

Long run restrictions are non-linear in the SVAR coefficients, but linear in the impulse responses. For the sake of presentation we omit the intercept $B_{0,t}$. Let

$$y_t = B_{1,t}y_{t-1} + \ldots + B_{p,t}y_{t-p} + \left[A\left(\alpha_t\right)\right]^{-1}\Sigma_t\varepsilon_t$$

Then, we only need to modify how draws for the B_t block are made. In particular,

1. At iteration i, given $A\left(\alpha_t\right)^{i-1}, \Sigma_t^{i-1}$ sample $\{B_t^i\}_{t=1}^T$ using Carter and Kohn's routine or one of the other routines described in section 6. With the sampled vector, compute the companion matrix

$$\mathbf{B}_t^i = \begin{bmatrix} B_{1,t}^i & \cdots & B_{p-1,t}^i & B_{p,t}^i \\ I_M & \cdots & \mathbf{0}_{M\times M} & \mathbf{0}_{M\times M} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{M\times M} & \cdots & I_M & \mathbf{0}_{M\times M} \end{bmatrix}$$

   where $B_t^i = \left[vec\left(B_{1,t}^i\right)', \ldots, vec\left(B_{p,t}^i\right)'\right]'$.

2. Given B_t^i, $A\left(\alpha_t\right)^{i-1}, \Sigma_t^{i-1}$compute the long run matrix for each t

$$\begin{aligned} \mathbf{D}_t^i &= \mathbf{J}\left(I_{Mp} - \mathbf{B}_t^i\right)^{-1}\mathbf{J}'\left[A\left(\alpha_t\right)^{i-1}\right]^{-1}\Sigma_t^{i-1} \\ &= \left(I_M - B_{1t}^i - \ldots - B_{p,t}^i\right)^{-1}\left[A\left(\alpha_t\right)^{i-1}\right]^{-1}\Sigma_t^{i-1} \end{aligned} \tag{51}$$

   where $\mathbf{J} = \begin{bmatrix} I_M & \mathbf{0}_{M\times M} & \cdots & \mathbf{0}_{M\times M} \end{bmatrix}$ is a selection matrix.

3. Impose long run restrictions i.e. construct $\widetilde{\mathbf{D}}_t^i = R\mathbf{D}_t^i$ where R is matrix restricting the entries of $\mathbf{D}_t^i$.

4. Given $\widetilde{\mathbf{D}}_t^i$, $A\left(\alpha_t\right)^{i-1}$, Σ_t^{i-1} and $B_{j,t}^i$, $j = 1, \ldots, p-1$, solve for $\widetilde{B}_{p,t}^i$ using (51), so that

$$\widetilde{B}_{p,t}^i = I_M - B_{1,t}^i - \ldots - B_{p-1,t}^i - \left[A\left(\alpha_t\right)^{i-1}\right]^{-1}\Sigma_t^{i-1}\left[\widetilde{\mathbf{D}}_t^i\right]^{-1}$$

   and with this construct the restricted draw $\widetilde{B}_t^i = \left[vec\left(B_{1,t}^i\right)', \ldots, vec\left(\widetilde{B}_{p,t}^i\right)'\right]'$.

5. Evaluate whether

$$\widetilde{\mathbf{B}}_t^i = \begin{bmatrix} B_{1,t}^i & \cdots & B_{p-1,t}^i & \widetilde{B}_{p,t}^i \\ I_M & \cdots & \mathbf{0}_{M\times M} & \mathbf{0}_{M\times M} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{M\times M} & \cdots & I_M & \mathbf{0}_{M\times M} \end{bmatrix}$$

has all its eigenvalues inside the unit circle. If so, we accept $\widetilde{B}_t^i$; otherwise discard it.

Given a draw for $\widetilde{B}_t$, the sampling of the remaining blocks $(A(\alpha_t), \Sigma_t, s, \mathcal{V})$ is unchanged.