Essays in Structural Macroeconometrics

Fernando José Pérez Forero

TESI DOCTORAL UPF / ANY 2013

DIRECTOR DE LA TESI
Prof. Fabio Canova (EUI) and Prof. Kristoffer Nimark
Departament Economia i Empresa

UNIVERSITAT POMPEU FABRA

Contents

---

[1]We would like to thank F. Schorfheide, G. Primiceri, R. Casarin, H. Van Dijk and three anonymous referees for comments and suggestions. A previous version of the paper has circulated with the title "Estimating overidentified, non-recursive, time varying coefficients structural VARs".

List of Figures

List of Tables

To my family,
to Sarelita

Acknowledgements

In first place, I would like to thank Fabio Canova and Kristoffer Nimark for their continuous support and guidance during the development of this thesis. I also thank them for giving me the opportunity to undertake different research projects as co-author and therefore allowing me to gain experience as a researcher.

I also would like to thank Vasco Carvalho, Jordi Galí, Christian Matthes, Barbara Rossi and all the faculty members and PhD Students of Pompeu Fabra who actively participate in the CREI Macroeconomics Breakfast. Their very valuable comments and suggestions at every stage of my research gave me the possibility of completing it.

In addition, I would like to thank Marta Araque and Laura Agustí for being there every time I needed help with many administrative procedures at UPF. Thanks also to Mariona Novoa for her help in scheduling my presentations. Thanks to them for their invaluable efficiency.

During these years I have made a lot of friends. I thank to all of them for the opportunity to share tons of experiences. Thanks to Miguel and Mapi for all the moments we spent together, starting from the problem sets in 2008 until being flatmates in the last year 2012-2013. Thanks to the 'Peruvian Community', starting with Miguel (again), Cynthia, Sofía and Silvio. Thanks to Marc and Jorg for having the opportunity to share our opinions about research, football, history and religion. I would like to also mention the people with whom I had great times during these years: Rodrigo, Benjamín, Mauro, Giorgio and Paula, Mapi (again) and Davide, Oriol, Sergio and Johanna, Elisa, José, Alicia, José, José Carlos, Jagdish, Tom and Michael. Please forgive me if I forgot anyone. Thanks to my office mates Ciccio, Tanya, Vicky and Bruno. Thanks to the musicians Ciccio (again), Kiz, Miguel Karlo and Gene for giving me the opportunity to play the guitar with them, you rock!

My family deserves a special mention. I am indebted with them for all the support and love that they given to me during these years. All the Skype phone calls, plus the good times we spent together every time I was there were extremely important. Finally, thanks to Sarelita (with Lolita and Renata). I cannot describe how important her role was. This thesis is for all of you.

Abstract

This thesis is concerned with the structural estimation of macroeconomic models via Bayesian methods and the economic implications derived from its empirical output. The first chapter provides a general method for estimating structural VAR models. The second chapter applies the method previously developed and provides a measure of the monetary stance of the Federal Reserve for the last forty years. It uses a pool of instruments and taking into account recent practices named Unconventional Monetary Policies. Then it is shown how the monetary transmission mechanism has changed over time, focusing the attention in the period after the Great Recession. The third chapter develops a model of exchange rate determination with dispersed information and regime switches. It has the purpose of fitting the observed disagreement in survey data of Japan. The model does a good job in terms of fitting the observed data.

Resumen

Esta tesis trata sobre la estimación estructural de modelos macroeconómicos a través de métodos Bayesianos y las implicancias económicas derivadas de sus resultados. El primer capítulo proporciona un método general para la estimación de modelos VAR estructurales. El segundo capítulo aplica dicho método y proporciona una medida de la posición de política monetaria de la Reserva Federal para los últimos cuarenta años. Se utiliza una variedad de instrumentos y se tienen en cuenta las prácticas recientes denominadas políticas no convencionales. Se muestra cómo el mecanismo de transmisión de la política monetaria ha cambiado a través del tiempo, centrando la atención en el período posterior a la gran recesión. El tercer capítulo desarrolla un modelo de determinación del tipo de cambio con información dispersa y cambios de régimen, y tiene el propósito de capturar la dispersión observada en datos de encuestas de expectativas de Japón. El modelo realiza un buen trabajo en términos de ajuste de los datos.

Foreword

This thesis is concerned with the structural estimation of macroeconomic models via Bayesian methods and the economic implications derived from its output. It is mainly developed within the context of structural vector autoregressive (SVAR) models and general state space models.

The first chapter," A general algorithm for estimating structural VARs", is a joint work with the professor Fabio Canova. It provides the method for estimating structural VAR models, which are non-recursive and potentially overidentified, with both constant and time varying coefficients. The procedure allows for linear and non-linear restrictions on the parameters, maintains the multi-move structure of standard algorithms and can be used to estimate structural models with different identification restrictions. The transmission of monetary policy shocks is studied with the proposed approach and results are compared with those obtained with traditional methods.

The second chapter, "Measuring the Stance of Monetary Policy in a Time-Varying world", applies the method previously developed and focuses its attention in measuring the monetary policy stance. The stance of monetary policy is of general interest for macroeconomists and the private sector. But it is not necessarily observable, since a Central Bank can use different instruments at different points in time. This chapter provides a measure of this stance for the last forty years using a pool of instruments. Different operating procedures are quantified by computing the time varying weights of these instruments and taking into account recent practices named Unconventional Monetary Policies. The measure describes how tight/loose was monetary policy conduction over time and takes into account the uncertainty related with posterior estimates of the parameters. Then it is shown how the monetary transmission mechanism has changed over time, focusing the attention in the period after the Great Recession.

The third chapter of this thesis, "Heterogeneous Information and Regime Switches in a Structural Exchange Rate model: Evidence from Survey Data", develops a model of exchange rate determination in the context of dispersed information and Higher Order Expectations. Exchange Rates Survey Data exhibits a considerable amount of disagreement across participants. Moreover, the mentioned disagreement is not constant over time, exhibiting substantial and persistent variation across time. We introduce regime switches to a model of exchange rate determination with disparately informed agents, and we provide an empirical exercise using actual survey data, with the purpose purpose of fitting the observed disagreement. We assume that the information structure is such that high-volatility regimes are associated with the appearance of a very noisy public signal about fundamentals. Given that this signal is very imprecise, and because of the higher volatility of shocks, disagreement increases. The model-implied dispersion closely follows the observed disagreement, which means the proposed model does a good job in

terms of fitting. We confirm the latter when comparing the model fit with respect to a restricted model without regimes switches and with no informational frictions at all. Furthermore, the model solution using the captured regime switches from the data is interpreted as evidence in favor of parameter instability in exchange rate models. The latter is, together with rational confusion, an additional explanation for the disconnection between Exchange Rates and future fundamentals.

Chapter 1

A GENERAL ALGORITHM FOR ESTIMATING STRUCTURAL VARS (JOINT WITH F. CANOVA)

1.1   Introduction[1]

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 among the firsts 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 theoret-

ically abundant. Algorithms to estimate non-recursive structural models exist, see e.g. Waggoner and Zha (2003) or Kociecki and Ca' Zorzi (2013). However, their extension to overidentified or TVC models is problematic.

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.

TVC-VAR models are typically estimated using a Bayesian Gibbs sampling routine. In this routine, 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 predetermined draws for the parameters belonging to previous equations. However, when the system is non-recursive, such an approach disregards the restrictions existing across equations. Hence, 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 blocks of the conditional distributions used in the Gibbs sampling do not necessarily have a known format. Primiceri (2005) suggests to use a Metropolis-step to deal with this problem. We follow his lead and nest the step into Geweke and Tanizaki (2001)'s approach to estimate general nonlinear state space models. This setup is convenient since it can accommodate general non-linear identification restrictions. Thus, many structural systems can be dealt with in a compact and unified way.

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 compare the results with those obtained in an overidentified, but fixed coefficient model. We show that there are important time variations in the variance of the monetary policy shock and in the estimated contemporaneous coefficients. These

time variations translate in important changes in the transmission of monetary policy shocks which are consistent with the idea that the ability of monetary policy to influence the real economy has waned, especially in the 2000s. We also show that the characterization of the dynamics in response to monetary policy shocks one obtains in an overidentified but fixed coefficient VAR is different.

The paper is organized as follows, Section 2 builds up intuition, shows how to apply the algorithm to estimate a simple SVAR with time invariant coefficients, and the identification restrictions that are allowed for. Section 3 extends the setup to a time varying coefficients static SVAR. Section 4 presents the general algorithm that is applicable to non-recursive, overidentified TVC-VAR models with stochastic volatility and quite general identification restrictions. Section 5 studies the transmission of monetary policy shocks. Section 6 summarizes the conclusions.

1.2 Constant coefficients static SVAR

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

$$A\left(\alpha\right)y_t = \varepsilon_t; \quad \varepsilon_t \sim N\left(0, I_M\right) \tag{1.1}$$

where $t = 1, \ldots, T$; y_t and ε_t are $M \times 1$ vectors, $A\left(\alpha\right)$ is a non-singular $M \times M$ matrix and α a vector of structural parameters. The likelihood function of (1.1) is

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(A\left(\alpha\right)y_t\right)'\left(A\left(\alpha\right)y_t\right)\right\} \tag{1.2}$$

Because of the Jacobian, $\det\left(A\left(\alpha\right)\right)^T$, the likelihood 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, draws of the reduced-form covariance matrix $\Omega\left(\alpha\right)^{-1} = A\left(\alpha\right)A\left(\alpha\right)'$ can be used to derive draws for α. However, when the system is overidentified $\Omega\left(\alpha\right)^{-1}$ will be 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 we proceed in four steps. First, we reparametrize the model. Second, we suggest a proposal distribution whose parameters can be estimated using the reparametrized model. Third, we provide a numerical example to highlight the properties of our algorithm. Fourth, we indicate the type of identification restrictions that are compatible with the setup.

3

1.2.1 Reparameterization of the SVAR

There are a number of ways to reparametrize the SVAR. Here we show that they are equivalent in terms of the likelihood.

Amisano and Giannini's setup

In Amisano and Giannini (1997), the matrix $A(\alpha)$ is re-parametrized as

$$vec(A(\alpha)) = S_A \alpha + s_A$$

Since

$$
\begin{aligned}
(A(\alpha)y_t)'(A(\alpha)y_t) &= tr\left[(A(\alpha)y_t)'(A(\alpha)y_t)\right] \\
tr\left[(A(\alpha)y_t)'(A(\alpha)y_t)\right] &= [vec(A(\alpha)y_t)]'[vec(A(\alpha)y_t)] \\
vec(A(\alpha)y_t) &= (y_t' \otimes I_M)(S_A\alpha + s_A)
\end{aligned}
\tag{1.3}
$$

after a number of manipulation (see on-line appendix), the likelihood for the re-parametrized model can be written as

$$
L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det(A(\alpha))^T \exp\left\{ -\frac{1}{2}\sum_{t=1}^{T}\left[\begin{array}{c} (3\alpha'S_A' + s_A')(I_M \otimes y_t'y_t) \\ \times (S_A\alpha + 2s_A) \end{array}\right]\right\}
\tag{1.4}
$$

Waggoner and Zha's setup

Waggoner and Zha (2003) rewrite the $A(\alpha)$ matrix as

$$
\begin{aligned}
A(\alpha) &= \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_M \end{array}\right] \\
&= \left[\begin{array}{cccc} \mathcal{U}_1\overline{\alpha}_1 + \mathcal{R}_1 & \mathcal{U}_2\overline{\alpha}_2 + \mathcal{R}_2 & \cdots & \mathcal{U}_M\overline{\alpha}_M + \mathcal{R}_M \end{array}\right]
\end{aligned}
$$

such that $\alpha = \left[\begin{array}{cccc} \overline{\alpha}_1' & \overline{\alpha}_2' & \cdots & \overline{\alpha}_M' \end{array}\right]'$ is the original column vector. That is, they perform a linear transformation of each of the columns of $A(\alpha)$. This reparameterization allows them to develop a sampling routine where each $\overline{\alpha}_i$, $i = 1, \ldots, M$ is drawn from a mixture of normal and gamma distributions. For the sake of concreteness, suppose that:

$$
A(\alpha) = \left[\begin{array}{ccc} 1 & 0 & \alpha_3 \\ \alpha_1 & 1 & 0 \\ 0 & \alpha_2 & 1 \end{array}\right]
\tag{1.5}
$$

so that the system is non-recursive and overidentified (we require that the variances of the shocks are unity). The Amisano and Giannini's reparameterization is

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

The Waggoner and Zha reparameterization is

$$a_1 = \begin{bmatrix} 1 \\ \alpha_1 \\ 0 \end{bmatrix}; \ \mathcal{U}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \ \mathcal{R}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \ \overline{\alpha}_1 = \alpha_1$$

$$a_2 = \begin{bmatrix} 0 \\ 1 \\ \alpha_2 \end{bmatrix}; \ \mathcal{U}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \ \mathcal{R}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \ \overline{\alpha}_2 = \alpha_2$$

$$a_3 = \begin{bmatrix} \alpha_3 \\ 0 \\ 1 \end{bmatrix}; \ \mathcal{U}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \ \mathcal{R}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \ \overline{\alpha}_3 = \alpha_3$$

Clearly

$$vec\left(A\left(\alpha\right)\right) = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

so that

$$S_A = diag\left(\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3\right); \ s_A = \begin{bmatrix} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \mathcal{R}_3 \end{bmatrix}$$

where $diag\left(.\right)$ indicates a block-diagonal matrix. Hence, Waggoner and Zha reparameterization also delivers the likelihood (1.4).

Alternative re-parametrization

Vectorizing (1.1) produces

$$vec\left(A\left(\alpha\right)y_t\right) = vec\left(\varepsilon_t\right)$$

5

Using (1.3) and the fact that $vec(\varepsilon_t) = \varepsilon_t$, the model can be expressed as:

$$\widetilde{y}_t = Z_t \alpha + \varepsilon_t \tag{1.6}$$

where $\widetilde{y}_t \equiv (y_t' \otimes I_M) s_A$; $\quad Z_t \equiv -(y_t' \otimes I_M) S_A$. The likelihood function of (1.6) is (see Appendix A.1 for details)

$$\widetilde{L}\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} (\det D)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}[\widetilde{y}_t - Z_t\alpha]'[\widetilde{y}_t - Z_t\alpha]\right\} \tag{1.7}$$

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

$$vec(D) = vec(A(\alpha))$$

and the likelihood in (1.7) is equal to the likelihood in (1.4). Note that (1.6) tells us that estimates of α can be obtained using data correlations. For the example in equation (1.5), (1.6) is equivalent to the following three linear regressions:

$$
\begin{aligned}
y_{1t} &= -y_{2t}\alpha_1 + \epsilon_{1t} \\
y_{2t} &= -y_{3t}\alpha_2 + \epsilon_{2t} \\
y_{3t} &= -y_{1t}\alpha_3 + \epsilon_{3t}
\end{aligned} \tag{1.8}
$$

1.2.2 The proposal distribution and the MH algorithm

The advantage of the reparameterization in (1.6) is that it allows us to easily design a proposal distribution to be used in a Metropolis routine. First, get estimates of the α parameters in (1.6)

$$\alpha^* = \left[\sum_{t=1}^{T} Z_t'Z_t\right]^{-1}\left[\sum_{t=1}^{T} Z_t'\widetilde{y}_t\right] \tag{1.9}$$

and of the covariance matrix

$$P^*(\alpha^*) = \left[\sum_{t=1}^{T} Z_t'(SSE)^{-1}Z_t\right]^{-1} \tag{1.10}$$

where $SSE = \sum_{t=1}^{T}(\widetilde{y}_t - Z_t\alpha^*)(\widetilde{y}_t - Z_t\alpha^*)'$. Then, the algorithm to draw α is as follows. Set $\alpha_0 = \alpha^*$ and for $i = 1, 2, \ldots, G$:

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

2. Compute $\theta = \frac{\widetilde{p}(\alpha\dagger|y^T) \cdot p_{*\alpha}(\alpha_{i-1}|\alpha_i)}{\widetilde{p}(\alpha_{i-1}|y^T) \cdot p_{*\alpha}(\alpha_i|\alpha_{i-1})}$, where $\widetilde{p}(.|y^T) = \widetilde{L}(y^T|.)p(.)$ is the posterior kernel of $(\alpha\dagger, \alpha_{i-1})$. Draw a $v \sim U(0,1)$. Set $\alpha_i = \alpha\dagger$ if $v < \omega$ and $\alpha_i = \alpha_{i-1}$ otherwise, where

$$\omega \equiv \left\{ \begin{array}{cc} \min\{\theta, 1\}, & \text{if } I_\alpha(\alpha\dagger) = 1 \\ 0, & \text{if } I_\alpha(\alpha\dagger) = 0 \end{array} \right.$$

Here $I_\alpha(.)$ is a truncation indicator and G is the total numbers of draws. Note that since $P^*(.)$, depends on α^*, the algorithm can be easily nested into a Gibbs sampling scheme. A t-distribution with small number of degrees of freedom is chosen to account for possible deviations from normality: when ν is large the proposal resembles a normal distribution.

Notice two facts about this algorithm. First, the α vector is jointly sampled. Second, the covariance matrix of $P^*(\alpha^*)$ is generally non-diagonal. As we explain later, these features distinguish our algorithm from those in the literature and provides the flexibility needed to accommodate a variety of structural models.

Kociecki and Ca' Zorzi (2013) have derived a closed form solution for the posterior of α under the assumption that $\det(A) = 1$. Interestingly, their posterior collapses to our proposal when the prior for α is diffuse.

1.2.3 A Numerical example

We illustrate the properties of our Metropolis approximation in the example of equation (1.5), when $\alpha = \left(\begin{array}{ccc} 0.8 & 0.5 & 0.5 \end{array}\right)'$. We simulate data according to (1.1) for $t = 1, \ldots, 500$, re-parametrize the model as in (1.6) and estimate α^* and P^* using (1.9) and (1.10). We use flat priors, i.e., $p(\alpha_i) \propto 1$, $i = 1, 2, 3$. We set $G = 150,000$, discard the first $100,000$, and keep 1 every 100 from the remaining. The acceptance rate is 24%.

Figure 1.1 indicates that the simulator does a good job in reproducing the DGP (the vertical lines indicate true values).

1.2.4 Identification restrictions

The framework can deal with linear restrictions (both of exclusion and non-exclusion type) and with certain non-linear restrictions. To show the type of constraints that are allowed, we present a few examples. While the focus is on over-identified systems, just identified ones only require appropriate adjustments of the matrices S_A and s_A.
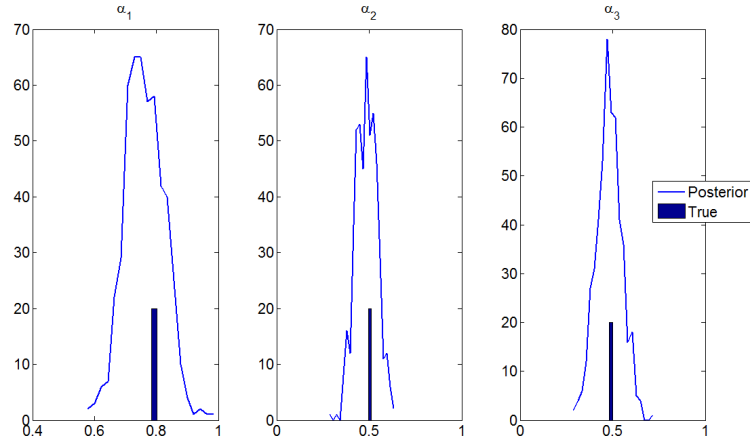
Figure 1.1: Posterior estimates of α

Short-run linear restrictions

Suppose

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

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

Since the restrictions are linear, the setup fits the above framework.

Short-run non-linear restrictions

Suppose now

$$A\left(\alpha\right) = \begin{bmatrix} 1 & 0 & \alpha_3 \\ \alpha_1 & 1 & 0 \\ 0 & \left(\alpha_2+1\right)^2 & 1 \end{bmatrix} \tag{1.11}$$

8

The model is re-parametrized as

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

where $F\left(\alpha\right)$ is a non-linear vector-valued function. Linearity is lost here, but if we define $\widetilde{\alpha}_2 \equiv (\alpha_2+1)^2$ as a new parameter, the procedure applies to the vector $\widetilde{\alpha} = (\alpha_1, \widetilde{\alpha}_2, \alpha_3)$. In fact, given posterior draws for $\widetilde{\alpha}_2$, we can recover $\alpha_2 = \sqrt{\widetilde{\alpha}_2} - 1$ once we impose the extra restriction $\widetilde{\alpha}_2 \geqslant 0$. Adding this restriction avoids us to deal with the fact that $F(\alpha)$ is non-linear.

Consider now:

$$
A\left(\alpha\right) = \begin{bmatrix} 1 & 0 & \alpha_3 \\ \alpha_1 & 1 & 0 \\ 0 & (\alpha_2+2\alpha_3)^2 & 1 \end{bmatrix} \tag{1.12}
$$

Here

$$
F\left(\alpha\right) = \begin{bmatrix} \alpha_1 \\ (\alpha_2+2\alpha_3)^2 \\ \alpha_3 \end{bmatrix}
$$

Also in this case the procedure can be employed, if we define $\widetilde{\alpha}_2 \equiv (\alpha_2 + 2\alpha_3)^2$ as a new parameter. In fact, with draws from the posterior of $\widetilde{\alpha}_2$ and α_3, we can recover $\alpha_2 = \sqrt{\widetilde{\alpha}_2} - 2\alpha_3$, provided that $\widetilde{\alpha}_2 \geqslant 0$.

Consider a final example:

$$
A\left(\alpha\right) = \begin{bmatrix} 1 & 0 & \alpha_1\alpha_2 - 1 \\ \alpha_1 & 1 & 0 \\ 0 & \alpha_2 & 1 \end{bmatrix} \tag{1.13}
$$

9

The reparametrized model is

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

where $\alpha = [\alpha_1, \alpha_2]'$. Here linearity is lost and adding an inequality constraint does not help since the third component of $F\left(\alpha\right)$ depends on the other two. Letting $z_t\left(\alpha\right) \equiv Z_t F\left(\alpha\right)$, the model is:

$$
\widetilde{y}_t = z_t\left(\alpha\right) + \varepsilon_t
$$

In general, if a closed-form solution for combinations of the parameters is available, the procedure can deal with short run non-linear restrictions. However, when a closed-form is not available, we need to treat the model as a non-linear system of equations, and the tools we describe in section 4 are useful.

Long-run restrictions

Long run restrictions are non-linear, but can be dealt in our Metropolis algorithm with a accept/reject step. To see this consider the more general SVAR model:

$$
A\left(\alpha\right) y_t = A_+ y_{t-1} + \varepsilon_t; \;\; \varepsilon_t \sim N\left(0, I_M\right) \tag{1.14}
$$

Letting $B \equiv \left[A\left(\alpha\right)\right]^{-1} A_+$, the VAR is:

$$
y_t = B y_{t-1} + \left[A\left(\alpha\right)\right]^{-1} \varepsilon_t \tag{1.15}
$$

The (long run) cumulative matrix is:

$$
\mathbf{D} \equiv \left(I_M - B\right)^{-1} \left[A\left(\alpha\right)\right]^{-1} \tag{1.16}
$$

Given draws of B and α, one can immediately construct $\mathbf{D}$ using (1.16) and check whether the required restrictions are satisfied. For example, suppose the cumulative impact matrix is restricted as

$$
\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & 0 \\ 0 & D_{22} & D_{23} \\ D_{13} & 0 & D_{33} \end{bmatrix}
$$

10

This set of restrictions can be summarized as

$$R'vec\left(\mathbf{D}\right) = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right] \tag{1.17}$$

with

$$R' = \left[\begin{array}{ccccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

From (1.15) setting $\widehat{y}_t \equiv y_t - By_{t-1}$, we have that

$$A\left(\alpha\right)\widehat{y}_t = \varepsilon_t$$

where

$$A\left(\alpha\right) = \left[\begin{array}{ccc} 1 & \alpha_3 & \alpha_5 \\ \alpha_1 & 1 & \alpha_6 \\ \alpha_2 & \alpha_4 & 1 \end{array}\right] \tag{1.18}$$

To estimate the structural parameters, we need first to draw B, then draw candidate α's using the suggested reparameterization and for each draw use an accept-reject step to make sure the long run restrictions (1.17) are satisfied. Seen through these lenses, long run and non-linear short run restrictions are similar. Clearly, if partial multipliers or the structural lagged coefficients A_+ are restricted with zero constraints, the same acceptance/rejection framework can be used.

A situation that leads to a non-linear model is one where there are both long and short run restrictions (see e.g. Gali (1991)). For example, suppose

$$\mathbf{D} = \left[\begin{array}{ccc} D_{11} & D_{12} & D_{13} \\ 0 & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{array}\right]$$

and in (1.18) $\alpha_4 = \alpha_5 = 0$. Let

$$\left(I_M - B\right)^{-1} \equiv \left[\begin{array}{ccc} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{array}\right]$$

Then

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

11

Thus, $D_{21} = 0$ implies $-b_{21}\left(\alpha_4\alpha_6 - 1\right) - b_{23}\left(\alpha_2 - \alpha_1\alpha_4\right) - b_{22}\left(\alpha_1 - \alpha_2\alpha_6\right) = 0$ and, using $\alpha_4 = 0$ and $\alpha_5 = 0$, we have $-b_{22}\left(\alpha_1 - \alpha_2\alpha_6\right) = 0$. Hence, long run restrictions require $\alpha_1 = \alpha_2\alpha_6$ and the impact matrix is

$$A\left(\alpha\right) = \left[\begin{array}{ccc} 1 & \alpha_3 & 0 \\ \alpha_2\alpha_6 & 1 & \alpha_6 \\ \alpha_2 & 0 & 1 \end{array}\right]$$

Therefore

$$vec\left(A\left(\alpha\right)\right) = \left[\begin{array}{c} 1 \\ \alpha_2\alpha_6 \\ \alpha_2 \\ \alpha_3 \\ 1 \\ 0 \\ 0 \\ \alpha_6 \\ 1 \end{array}\right] = \underbrace{\left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]}_{S_A} \underbrace{\left[\begin{array}{c} \alpha_2\alpha_6 \\ \alpha_2 \\ \alpha_3 \\ \alpha_6 \end{array}\right]}_{F(\alpha)} + \underbrace{\left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{array}\right]}_{s_A}$$

Sign restrictions

Although sign restrictions are not the focus of this paper, it is straightforward to show that the algorithm can be applied also to VARs identified this way. Let $A(\alpha)$ be a general matrix with no zero elements and impose inequality constraints on, say, the first column. Then, one can draw α's as in section 2.2 and check if the first column satisfies the required inequality restrictions. Thus, sign restrictions can be dealt with in the same way as long run restrictions.

1.3   Time-varying coefficients static SVAR

Before we move to a full fledged TVC-SVAR model, it is useful to study the intermediate step of a static TVC-SVAR. The model is

$$A\left(\alpha_t\right)y_t = \varepsilon_t; \quad \varepsilon_t \sim N\left(0, I_M\right) \tag{1.19}$$

$$\alpha_t = \alpha_{t-1} + \eta_t; \quad \eta_t \sim N\left(0, V\right) \tag{1.20}$$

where V is positive definite and α_0 is given. This model is re-parametrized as:

$$\widetilde{y}_t = Z_t\alpha_t + \varepsilon_t \tag{1.21}$$

$$\alpha_t = \alpha_{t-1} + \eta_t \tag{1.22}$$

where, as before, $\widetilde{y}_t \equiv (y_t' \otimes I_M) s_A$ and $Z_t \equiv -(y_t' \otimes I_M) S_A$. We wish to compute $p\left(\alpha^T \mid y^T, V\right)$ and $p\left(V \mid y^T, \alpha^T\right)$ to be used in the Gibbs sampler. Given the assumptions the latter is inverted Wishart and its parameters are easy to compute.

Using the Markovian structure of the model, the conditional posterior $p\left(\alpha^T \mid y^T, V\right)$ can be factorized as

$$
\begin{aligned}
p\left(\alpha^T \mid y^T, V\right) &= p\left(\alpha_T \mid y_T, V\right) \prod_{t=1}^{T-1} p\left(\alpha_t \mid \alpha_{t+1}, y^t, V\right) \\
&\propto p\left(\alpha_T \mid y_T, V\right) \prod_{t=1}^{T-1} p\left(\alpha_t \mid y^t, V\right) p\left(\alpha_{t+1} \mid \alpha_t, V\right) \tag{1.23}
\end{aligned}
$$

Since each term in the last expression is normal, to sample α^T from (1.23) we just need the mean and the variance of each of the terms.

Thus, set initial values α_0 and $P_{0|0}$ and for each $t = 1, \ldots, T$ construct

$$\widehat{\alpha}_{t|t-1} = \widehat{\alpha}_{t-1|t-1}$$

$$P_{t|t-1} = P_{t-1|t-1} + V$$

and the Kalman gain $K_t = P_{t|t-1} Z_t' \Omega_t^{-1}$, where $\Omega_t = Z_t' P_{t|t-1} Z_t + I_M$. Estimates of α_t and of its variance are updated according to

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

To smooth the estimates set $\alpha_{T|T}^* = \widehat{\alpha}_{T|T}$, $P_{T|T}^* = P_{T|T}$ and, for $t = T-1, \ldots, 1,$, compute

$$\alpha_{t|t+1}^* = \widehat{\alpha}_{t|t} + P_{t|t} Z_t' P_{t+1|t}^{-1} \left(\alpha_{t+1|t+2}^* - Z_t' \widehat{\alpha}_{t|t}\right)$$

$$P_{t|t+1}^* = P_{t|t} - P_{t|t} Z_t' P_{t+1|t}^{-1} Z_t P_{t|t-1}'$$

1.3.1 The basic algorithm

Step 1: Given $\left(y^T, V_{i-1}\right)$, we take an initial value $\alpha_0^T = \{\alpha_{0,t}\}_{t=1}^T$ and:

1. Compute $\left\{\alpha_{t|t+1}^{*(i-1)}\right\}_{t=1}^T$ and $\left\{P_{t|t+1}^{*(i-1)}\right\}_{t=1}^T$.

2. At each $t = 1, \ldots, T$, draw a candidate $\alpha\dagger_t \sim p_{*\alpha}\left(\alpha_t \mid \alpha_{i-1,t}\right) = t\left(\alpha_{i-1,t}, r P_{t|t+1}^{*(i-1)}, \nu\right)$, $r > 0$, $\nu \geq 4$. Set $p_{*\alpha}\left(\alpha^T \mid \alpha_{i-1}^T\right) = \prod_{t=1}^T p_{*\alpha}\left(\alpha_t \mid \alpha_{i-1,t}\right)$.

13

3. Compute $\theta = \frac{p\left((\alpha\dagger)^T\right)\cdot p_{*\alpha}\left(\alpha^T_{i-1}|\alpha^T\right)}{p\left(\alpha^T_{i-1}\right)\cdot p_{*\alpha}\left(\alpha^T|\alpha^T_{i-1}\right)}$ where $p\left(.\right)$ is the posterior kernel (1.23). Draw a $v \sim U\left(0,1\right)$. Set $\alpha^T_i = (\alpha\dagger)^T$ if $v < \omega$ and set $\alpha^T_i = \alpha^T_{i-1}$ otherwise, where

$$\omega \equiv \begin{cases} \min\left\{\theta,1\right\}, & \text{if } I_\alpha\left((\alpha\dagger)^T\right) = 1 \\ 0, & \text{if } I_\alpha\left((\alpha\dagger)^T\right) = 0 \end{cases}$$

and $I_\alpha\left(.\right)$ is a truncation indicator function.

Step 2: Given (α^T_i, y^t), draw V_i from $(V^{-1}_i \mid \alpha^T_i, y^t) \sim W\left(\overline{v}_V, \overline{V}^{-1}\right)$, where

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

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

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

We then use α^T_i, V_i as initial values and repeat steps 1 and 2 for $i = 1, \ldots, G$.

Given the structure of the problem, if α is constant, V is the null matrix. Thus, Kalman smoother and OLS estimates α^* and P^* will coincide and the algorithm collapses to the one described in section 2.2.

1.4   A 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:

$$y_t = B_{0,t}C_t + B_{1,t}y_{t-1} + \ldots + B_{p,t}y_{t-p} + u_t \tag{1.24}$$

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 $u_t \sim N\left(0, \Omega_t\right)$, where Ω_t is symmetric, positive definite, and full rank for every t. For the sake of presentation, exogenous variables are excluded, but the setup can be easily extended to account for them. Since (1.24) is a reduced form, u_t does not have an economic interpretation. Denote the structural shocks by $\varepsilon_t \sim N\left(0, I_M\right)$ and let

$$u_t = A^{-1}_t \Sigma_t \varepsilon_t \tag{1.25}$$

where $A_t \equiv A(\alpha_t)$ is the contemporaneous coefficients matrix and $\Sigma_t = diag\{\sigma_{i,t}\}$ contains the standard deviations of the structural shocks at t. The SVAR is:

$$y_t = X'_t B_t + A^{-1}_t \Sigma_t \varepsilon_t \tag{1.26}$$

14

where $X'_t = I_M \otimes [C'_t, y'_{t-1}, \ldots, y'_{t-p}]$ and $B_t = [vec(B_{0,t})', vec(B_{1,t})', \ldots, vec(B_{p,t})']'$ are a $M \times K$ matrix and a $K \times 1$ vector, $K = \bar{M} \times M + pM^2$. It is typical to assume that (B_t, A_t, Σ_t) evolve as independent random-walks:

$$B_t = B_{t-1} + \upsilon_t \tag{1.27}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{1.28}$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t \tag{1.29}$$

where α_t denotes the vector of free parameters of A_t, and let:

$$\mathcal{V} = Var \left(\begin{bmatrix} \varepsilon_t \\ \upsilon_t \\ \zeta_t \\ \eta_t \end{bmatrix} \right) = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & W \end{bmatrix} \tag{1.30}$$

where Q, V, W are full rank matrices. Common patterns of time variations are possible if the rank of some of these matrices is reduced.

Thus, the setup captures time variations in i) the lag structure (see (1.27)), ii) the contemporaneous reaction parameters (see (1.28)) and iii) the structural variances (see (1.29)). As shown in Canova et al. (2012), models with breaks at a specific date can be accommodated by adding restrictions on the law of motions $(1.27) - (1.29)$.

1.4.1   Relaxing standard assumptions

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

$$A_t \left(y_t - X'_t \widehat{B}_t \right) \equiv A_t \widehat{y}_t = \Sigma_t \varepsilon_t \tag{1.31}$$

As before, let

$$vec(A_t) = S_A \alpha_t + s_A \tag{1.32}$$

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

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

and the state space is composed of

$$\widetilde{y}_t = Z_t \alpha_t + \Sigma_t \varepsilon_t \tag{1.33}$$

of (1.28) and (1.29), 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 draws for $(\hat{B}_t, \sigma_t)$, we need to draw $\alpha^T \equiv \{\alpha_t\}_{t=1}^T$, from $p\left(\alpha^T \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T}\right)$.

15

The standard approach is to partition α_t into blocks associated with each equation, say $\alpha_t = \left[\alpha_t^{1\prime}, \alpha_t^{2\prime}, \ldots, \alpha_t^{M\prime} \right]^{\prime}$, and assume that these blocks are independent, so that $V = diag\left(V_1, \ldots, V_M \right)$. Under these assumptions

$$
p\left(\alpha^T \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T} \right) = \prod_{m=2}^{M} p\left(\alpha^{m,T} \mid \alpha^{m-1,T}, \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T} \right) \quad (1.34)
$$
$$
\times p\left(\alpha^{1,T} \mid \widetilde{y}^T, \Sigma^T, \mathcal{V}, \widehat{B^T} \right)
$$

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 α_t blocks are uncorrelated. However, if we insist that each element of α_t has some economic meaning, the diagonality of 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 strictly invariant to changes in the policy parameters (see e.g. Lakdawala (2011)).

The algorithm to draw α we have described relaxes both assumptions, that is, the vector α_t is jointly drawn and 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.

In a constant coefficient SVAR one identifies shocks imposing short run, long run, or heteroschedasticity restrictions. In TVC-VARs identification restrictions are typically employed only on A_t but, as we have seen, certain type of restrictions produce non-linear state space models. In some situations one may want to identify shocks imposing shape restrictions on certain medium term multipliers (the maximum effect of a monetary shock on output occurs x-months after the disturbances) or on the variance decomposition, as it is done in the news shock literature (see e.g. Barsky and Sims (2012)), and these may also generate non-linear structural VARs. Furthermore, while it is standard to employ a log linear setup for the time variations in $\log(\sigma_t)$, one may want to use GARCH or Markov switching specifications, which also generate a non-linear or non-normal law of motion for some of the coefficients.

To be able to deal with all these cases, we embed the Metropolis algorithm to draw α^T into a modified version of Geweke and Tanizaki (2001)'s routine for estimating non-linear, non-Gaussian state space models. This greatly expands the type of structural models one can consider within the same estimation framework.

1.4.2 Estimation

Consider the general state space model:

$$\widehat{y}_t = z_t(\alpha_t) + u_t(\sigma_t, \xi_{1t}) \tag{1.35}$$

$$\alpha_t = t_t(\alpha_{t-1}) + r_t(\alpha_{t-1}, \xi_{2t}) \tag{1.36}$$

$$h_t(\sigma_t) = k_t(\sigma_{t-1}) + \xi_{3t} \tag{1.37}$$

where $\widehat{y}_t, \xi_{1t}, \xi_{3t}$ are $M \times 1$ vectors; α_t and ξ_{2t} are $K \times 1$ vectors; $\xi_{1t} \sim N(0, Q_{1t})$, $\xi_{2t} \sim N(0, Q_{2t}), \xi_{3t} \sim N(0, Q_{3t})$. Assume that $z_t(.), t_t(.), r_t(.), u_t(.), h_t(.), k_t(.)$ are vector-valued functions.

To estimate this system, it is typical to linearize it around the previous forecast of the state vector, so that

$$z_t(\alpha_t) \simeq z_t(\widehat{a}_{t|t-1}) + \widehat{Z}_t(\alpha_t - \widehat{a}_{t|t-1})$$

$$u_t(\sigma_t, \xi_{1t}) \simeq u_t(\widehat{\sigma}_{t|t-1}, 0) + \widehat{u}_{\sigma,t}(\sigma_t - \widehat{\sigma}_{t|t-1}) + \widehat{u}_{\xi_1,t}\xi_{1,t}$$

$$t_t(\alpha_{t-1}) \simeq t_t(\widehat{a}_{t-1|t-1}) + \widehat{T}_t(\alpha_{t-1} - \widehat{a}_{t-1|t-1})$$

$$r_t(\alpha_{t-1}, \xi_{2t}) \simeq r_t(\widehat{\alpha}_{t-1|t-1}, 0) + \widehat{r}_{\alpha,t}(\alpha_{t-1} - \widehat{a}_{t-1|t-1}) + \widehat{r}_{\xi_2,t}\xi_{2,t}$$

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

$$k_t(\sigma_t) \simeq k_t(\widehat{\sigma}_{t|t-1}) + \widehat{k}_t(\sigma_t - \widehat{\sigma}_{t|t-1})$$

where $\widehat{Z}_t, \widehat{u}_{\sigma,t}, \widehat{u}_{\xi_1,t}, \widehat{T}_t, \widehat{r}_{\alpha,t}, \widehat{r}_{\xi_2,t}, \widehat{h}_t, \widehat{k}_t$ are matrices corresponding to the Jacobian of $z_t(.), u_t(.), t_t(.), r_t(.), h_t(.), k_t(.)$, evaluated at $\alpha_t = \widehat{a}_{t|t-1}, \sigma_t = \widehat{\sigma}_{t|t-1}, \xi_{1,t} = 0, \xi_{2,t} = 0$. Thus, the approximated model is

$$\widehat{y}_t \simeq \widehat{Z}_t\alpha_t + \widehat{d}_t + \widehat{u}_{\xi_1,t}\xi_{1,t} \tag{1.38}$$

$$\alpha_t \simeq \widehat{T}_t\alpha_{t-1} + \widehat{c}_t + \widehat{r}_{\xi_2,t}\xi_{2,t} \tag{1.39}$$

$$\widehat{h}_t\sigma_t \simeq \widehat{k}_t\sigma_{t-1} + \widehat{f}_t + \xi_{3t} \tag{1.40}$$

where

$$\widehat{d}_t = z_t(\widehat{a}_{t|t-1}) - \widehat{Z}_t\widehat{a}_{t|t-1} + u(\widehat{\sigma}_{t|t-1}, 0) - \widehat{u}_{\sigma,t}(\widehat{\sigma}_{t|t-1} - \sigma_t) \tag{1.41}$$

$$\widehat{c}_t = t_t(\widehat{a}_{t-1|t-1}) - \widehat{T}_t\widehat{a}_{t-1|t-1} + r_t(\widehat{\alpha}_{t|t-1}, 0) - \widehat{r}_{\alpha,t}(\widehat{\alpha}_{t|t-1} - a_{t-1}) \tag{1.42}$$

$$\widehat{f}_t = k_t(\widehat{\sigma}_{t|t-1}) - \widehat{k}_t\widehat{\sigma}_{t|t-1} - h_t(\widehat{\sigma}_{t|t-1}) + \widehat{h}_t\widehat{\sigma}_{t|t-1} \tag{1.43}$$

Equations $(1.38), (1.39), (1.40)$ are similar to equations (1.33) and $(1.28), (1.29)$. When $z_t(.), t_t(.),$

$k_t(.), h_t(.), u_t(.)$ are linear, r_t is independent of α_t and u_t is independent of σ_t. $\widehat{d}_t = \mathbf{0}, \widehat{c}_t = \mathbf{0}, \widehat{f}_t = \mathbf{0}$. In one of the cases considered by Rubio Ramírez et al. (2010) or in some of those of section 2.4 $\widehat{d}_t \neq \mathbf{0}$, while if the law of motion of the structural coefficient is non-linear $\widehat{c}_t \neq \mathbf{0}$ or $\widehat{f}_t \neq \mathbf{0}$ or both.

17

The algorithm

Set initial values for $(B_0^T, \alpha_0^T, \Sigma_0^T, s_0^T, \mathcal{V}_0)$, where s^T is J-dimensional vector of discrete indicator variables described below. Then:

1. Draw B_i^T from from $p\left(B_i^T \mid \alpha_{i-1}^T, \Sigma_{i-1}^T, s_{i-1}^T, \mathcal{V}_{i-1}\right) \cdot I_B\left(B_i^T\right)$, where $I_B\left(.\right)$ truncates the posterior to insure stationarity of impulse responses.

2. Draw α_i^T from

$$
\begin{aligned}
p\left(\alpha_i^T \mid \widehat{y}_i^T, \Sigma_{i-1}^T, s_{i-1}^T, \mathcal{V}_{i-1}\right) \quad \propto \quad & p\left(\alpha_{i,T} \mid \widehat{y}_i^T, \Sigma_{i-1,T}, s_{i-1,T}, \mathcal{V}_{i-1}\right) \times \\
& \prod_{t=1}^{T-1} p\left(\alpha_{i,t} \mid \widehat{y}_i^t, \Sigma_{i-1,t}, s_{i-1,t}, \mathcal{V}_i\right) \times \\
& p_{t+1}\left(\alpha_{i,t+1} \mid \alpha_{i,t}, \Sigma_{i-1,t}, s_{i-1,t}, \mathcal{V}_{i-1}\right)
\end{aligned}
$$

   using the Metropolis approach described in section 3.1, where $\alpha_{i|t+1}, P_{i|t+1}$ are estimated with the extended Kalman smoother (EKS) described below.

3. Draw Σ_i^T using a log-normal approximation as in Kim et al. (1998). Given $\left(B_i^T, \alpha_i^T\right)$, the model is linear and composed of

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

   and (1.29), but the error is not normal. The $m - th$ equation is $y_{m,t}^{**} = \sigma_{m,t}\varepsilon_{m,t}$, where $\sigma_{m,t}$ is the $m-$th diagonal element of Σ_t. Then

$$
y_t^* = \log\left[\left(y_{m,t}^{**}\right)^2 + \overline{c}\right] \approx 2\log\left(\sigma_{m,t}\right) + \log\varepsilon_{m,t}^2 \qquad (1.44)
$$

   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. Such a distribution 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. Hence, standard Kalman smoother recursions can be used to draw $\{\Sigma_t\}_{t=1}^T$ from $(1.44) - (1.29)$. To ensure independence of the structural variances, each element of $\{\sigma_{m,t}\}_{m=1}^M$ is sampled assuming a diagonal W.

4. To draw s_i^T, given (Σ_i^T, y_t^*), draw $u \sim U\left(0, 1\right)$ and compare it to

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

   where $j = 1, \ldots, J$; $\phi\left(.\right)$ is the normal density function, q_j a set of weights, the term inside the parenthesis 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 component. Then assign $s_{m,t} = j$ iff $P\left(s_{m,t} \le j-1 \mid y^*_{m,t}, \log\left(\sigma_{m,t}\right)\right) < u \le P\left(s_{m,t} \le j \mid y^*_{m,t}, \log\left(\sigma_{m,t}\right)\right).$

5. Draw $\mathcal{V}_i$ from $p\left(\mathcal{V}_i \mid \alpha^T_i, \widehat{y}^T_i, \Sigma^T_{i-1}, s^T_{i-1}\right)$. The matrix $\mathcal{V}_i$ is sampled assuming that each block follows an independent inverted Wishart distribution.

Then one uses $B^T_i, \alpha^T_i, \Sigma^T_i, s^T_i, \mathcal{V}_i$ as initial values and repeat the sampling for the five blocks for $i = 1, \dots, G$.

The details of step 2

Given (y^T, Σ^T), we predict the mean and mean square error of α_t for $t = 1, \dots, T$:

$$\widehat{a}_{t|t-1} = t_t(\widehat{a}_{t-1|t-1})$$

$$P_{t|t-1} = \widehat{T}_t P_{t-1|t-1} \widehat{T}'_t + \widehat{r}_{\xi_2,t} Q_{2t} \widehat{r}'_{\xi_2,t}$$

and compute the Kalman gain $K_t = P_{t|t-1} \widehat{Z}'_t \Gamma^{-1}_t$, where $\Gamma_t = \widehat{Z}'_t P_{t|t-1} \widehat{Z}_t + \widehat{u}_{\xi_1,t} Q_{2t} \widehat{u}'_{\xi_1,t}$.
As new information arrives, estimates are updated according to

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

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} \widehat{Z}'_t \Gamma^{-1}_t \widehat{Z}_t P'_{t|t-1}$$

To smooth the estimates, set $\alpha^*_{T|T} = \widehat{a}_{T|T}$, $P^*_{T|T} = P_{T|T}$ and compute

$$\alpha^*_{t|t+1} = \widehat{a}_{t|t} + P_{t|t} \widehat{Z}'_t P^{-1}_{t+1|t} \left(\alpha^*_{t+1|t+2} - t_t(\widehat{a}_{t|t})\right)$$

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

for $t = T-1, \dots, 1,$. To start the iterations, we use $\widehat{a}_{1|0} = \mathbf{0}_{K \times 1}$ and $P_{0|0} = I_K * \kappa, \kappa >> 1$. Notice that the approximate model is used only in predicting and updating the mean square error of α.

1.4.3 Discussion

The advantage of nesting our setup into Geweke and Tanizaki's framework should be clear. However, there is no free lunch and costs are involved. For example, we are assuming that the posteriors can be approximated by normals and that no asymmetries exist. While normality may be appropriate in large samples, it is unclear that it is when the data is short, it includes financial or other fast moving

variables, and rare (and large) shocks hit the economy. The alternative would be to use recently developed sequential Montecarlo methods, see e.g., Creel (2012) and Herbst and Schorfheide (2013), to compute the posterior of the unknown of the non-linear state space model. While such an approach is feasible, it complicates computations quite a lot. As it will be clear in the application section, our approach allows us to estimate medium scale VARs in reasonable amount of time. Furthermore, in most applications identification restrictions imply a linear state space. Thus, it is important to have a tool that can extensively cover that situation and can deal with certain non-linear restrictions used in the literature without having to pay the full costs of having a complete non-linear methodology.

1.4.4 Single-move Metropolis for drawing B_t

To draw B^T in step 1 of the algorithm one can employ a standard 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) have 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. The assumption of non-explosive impulse responses is appealing in many macroeconomic applications and since Cogley and Sargent (2005), it is common to assume that all the eigenvalues of the companion form matrix associated with B_t lie within the unit circle for $t = 1, \ldots, T$. Thus, draws that do not satisfy the restrictions are discarded. When the Carter and Kohn (1994) a multi-move logic is used, if one element of the sequences violates the restrictions, the entire sequence is discarded, making the algorithm inefficient.

To solve this problem, Koop and Potter suggest to evaluate the elements of the B^T sequence separately using a single-move algorithm and use an accept/reject step. The approach works as follows. Given draws of $B_{i-1}^T, \alpha_{i-1}^T, \Sigma_{i-1}^T, Q_{i-1}, V_{i-1}, W_{i-1}$, the measurement equation is

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

and the transition equation for B_t 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 $B_t^c \sim N(\mu_t, \Psi_t)$ where

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

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

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

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

3. The acceptance rate of B_t^c is

$$\begin{aligned} \omega_{B,t} &= \min \left\{ \frac{\frac{\mathbf{1}\left(\max \left| eig \left(\overline{\mathbf{B}}_t^c \right) \right| < 1 \right)}{\lambda \left(B_t^c, Q_{i-1} \right)}}{\frac{1}{\lambda \left(B_{t,i-1}, Q_{i-1} \right)}}, 1 \right\} \\ &= \min \left\{ \frac{\mathbf{1} \left(\max \left| eig \left(\overline{\mathbf{B}}_t^c \right) \right| < 1 \right) \lambda \left(B_{t,i-1}, Q_{i-1} \right)}{\lambda \left(B_t^c, Q_{i-1} \right)}, 1 \right\} \end{aligned}$$

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

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

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 a-priori 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

$$\overline{Q}^{-1} = \left[\underline{Q} + \sum_{t=1}^{T} \left(B_{t,i} - B_{t-1,i} \right) \left(B_{t,i} - B_{t-1,i} \right)' \right]^{-1}$$

To draw Q we need to draw a candidate $\left(Q^c \right)^{-1} \sim W \left(\overline{v}, \overline{Q}^{-1} \right)$ and take the inverse Q^c. Then, for $t = 1, \dots, T$, we evaluate $\lambda \left(B_{t,i}, Q^c \right)$ and $\lambda \left(B_{t,i}, Q_{i-1} \right)$,

21

for a fixed $\overline{L}$, and calculate the acceptance probability

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

Finally, we draw a $v \sim U\left(0, 1\right)$, set $Q_i = Q^c$ if $v < \omega_Q$ and $Q_i = Q_{i-1}$ otherwise.

In a standard multi-move approach $\lambda\left(.\right) = 1$, 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, as we will see later on, of adding considerable computational time.

1.4.5 A shrinkage approach

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. The main problem with the standard setup is that 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 at each t. By making B_t function of a lower dimensional vector of factors θ_t, who independently move as a random walk, the approach can reduce the computational costs and the inefficiency of the algorithm.

The model is still consists of (1.26), (1.28) and (1.29) but now (1.27) is substituted by

$$
\begin{aligned}
B_t &= \Xi\theta_t + \upsilon_t & \upsilon_t &\sim N(0, I) & (1.45) \\
\theta_t &= \theta_{t-1} + \rho_t & \rho_t &\sim N(0, Q) & (1.46)
\end{aligned}
$$

where $\dim(\theta_t) \ll \dim(B_t)$ and where the matrix Ξ is known and composed of ones and zeros as in Canova and Ciccarelli (2009). The setup where the matrix Ξ is unknown and estimated along the other unknown quantities is presented in the on-line appendix. Using (1.46) into (1.45) we have

$$y_t = X'_t\Xi\theta_t + A_t^{-1}\Sigma_t\varepsilon_t + X'_t\upsilon_t \equiv X'_t\Xi\theta_t + \psi_t \qquad (1.47)$$

where $\psi_t \sim N\left(0_{M\times 1}, H_t\right)$ with $H_t \equiv A_t^{-1}\Sigma_t\Sigma'_t\left(A_t^{-1}\right)' + X'_tX_t$.

To estimate the unknowns we do the following:

1. Sample θ^T using a multi-move routine using (1.47) and (1.46).

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

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

22

As before
$$(\widehat{y'_t} \otimes I_M)(S_A \alpha_t + s_A) = \Sigma_t \varepsilon_t + A_t X'_t \upsilon_t$$

so that the second state-space system is

$$\widetilde{y}_t = Z_t \alpha_t + \Sigma_t \varepsilon_t + A_t X'_t \upsilon_t \tag{1.48}$$
$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{1.49}$$

and we draw α^T using our proposed Metropolis step. Here, the variance of the measurement error is $\Sigma_t \Sigma'_t + A_t(\alpha_t) X'_t X_t A'_t(\alpha_t)$ and it is evaluated at the current prediction $\alpha_{t|t-1}$. α^T is sampled using the extended Kalman smoother previously described.

3. Given (θ^T, α^T):
$$\widehat{A}_t \widehat{y}_t = \Sigma_t \varepsilon_t + \widehat{A}_t X'_t \upsilon_t$$

Since $\widehat{A}_t X'_t$ is known, let the lower-triangular P_t satisfy $P_t \left(\widehat{A}_t X'_t X_t \widehat{A}'_t \right) P'_t = I$. Then
$$P_t \widehat{A}_t \widehat{y}_t = y_t^{**} = P_t \Sigma_t \varepsilon_t + P_t \widehat{A}_t X'_t \upsilon_t$$

with $var\left(P_t \widehat{A}_t X'_t \upsilon_t \right) = I$ and where $P_t \Sigma_t \Sigma'_t P'_t + P_t \left(\widehat{A}_t X'_t X_t \widehat{A}'_t \right) P'_t$ is a diagonal matrix. This transformation is similar to Cogley and Sargent (2005); however, since $\widehat{A}_t X'_t$ is known, we only need to sample the variances of $\epsilon_{m,t}$. As in algorithm 4.2.1, we do this using the $\log(\chi^2)$ approximation that consists on a mixture of 7 normals (see (1.44)).

4. Given $(\theta^T, \alpha^T, \Sigma^T)$, we sample Q, V, W from independent inverted Wishart distributions as in algorithm 4.2.1.

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.

We evaluate the relative merits of different approaches in the specific example discussed in the next section.

1.5 An Application

We apply our procedures to study the transmission of monetary policy shocks in an overidentified structural TVC-VAR. We are interested in knowing whether the propagation of policy shocks has changed over time and in identifying the sources of variation in certain macroeconomic variables. For comparison, we will also examine the conclusions obtained estimating a constant coefficient overidentified SVAR.

1.5.1 The SVAR

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. Since researchers working with this set of variables are typically interested in the dynamic response to monetary policy shocks, see e.g. Sims and Zha (2006), the structure of A_t is restricted 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 1.1: Identification restrictions

The structural form 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 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 non-policy 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 $\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 & \varepsilon_t^2 & \varepsilon_t^3 & \varepsilon_t^{mp} & \varepsilon_t^{md} & \varepsilon_t^i \end{bmatrix}'$ be the vector of structural innovations. The structural model is

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

$$(1.50)$$

where $A_t^+ (L)$ is a function of A_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,

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

is the matrix of standard deviations of the structural shocks.

The structural model (1.50) 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) that the (constant coefficient version of the) system is globally identified and therefore suitable for interesting policy experiments. While the structural model we consider is conditionally linear, more general non-linear models are easy to generate using the restrictions described in section 2.4 or considering a non-linear law of motion for the parameters. Our setup can accommodate for all these possibilities.

1.5.2 The Data

The data we use 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, $x_t = (y_t^* - E(y_t^*)) / std(y_t^*)$, to have all the variables on the same scale.

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

To calibrate the parameters of the prior, 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 with 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. Convergence was checked using standard statistics - see on-line appendix. Draws for B_t are monitored and discarded if the stability condition fails. The indicator 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 $M = 6$, and $\dim(\alpha) = 12$, then

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

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

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

and also

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

Finally, computations were performed on an Intel (R) CORE(TM) i5-2400 CPU @ 3.1GHz machine with 16GB of RAM.

1.5.4 Multi-move, single move, shrinkage algorithms

The Carter and Kohn (1994) routine is fairly popular in the literature. However, when one needs to impose stationarity of the impulse responses, it becomes very inefficient, in particular in situations like ours, when the SVAR has more than three variables and year-on-year growth rates of the variables are used. The problem can be somewhat diminished if quarterly growth rates are used, since they tend to be less persistent [2] and considerably reduced if data is standardized. We show these facts in table 1.2, which reports the acceptance rates in the four possible options.

In the Koop and Potter's algorithm, we set $\overline{L} = 25$, when evaluate the integrating constants $\lambda(.)$ in each period. The priors, the number of draws and the skipping scheme are the same as in the multi-move algorithm. The averages acceptance rates for B^T are 97% and 91% for year-on-year and quarterly growth

---

[2] We thank an anonymous referee for suggesting us this possibility.

Detrending \ Scales	Non-standardized	Standardized
Year-on-year growth rates	0.46%	10.2%
Quarterly growth rates	0.60%	11.5%

Table 1.2: Acceptance Rates from multi-move routine

Year-on-year, non-standardized   Quarterly, non-standardized
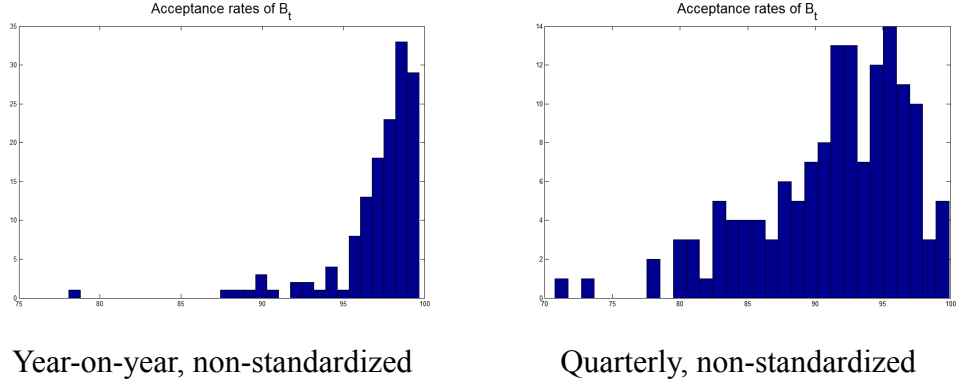
Figure 1.2: Acceptance rates of the single-move algorithm

rates, respectively, much higher than in the multi-move algorithm. The acceptance rate greatly differs in different time periods (see Figure 1.2) - this is consistent with the fact that in the multi-move algorithm the whole B^T is rejected quite often. However, the higher acceptance rate comes at the cost of higher computational time: we need about 12 hours to estimate the model with the multi-move routine but about 96 hours with the single move routine. Thus, the computational costs offset the efficiency gains.

Apart from the constants, the vector B_t has 72 components. Since we want to maintain as much as possible the covariance structure of the data unchanged, we estimate the shrinkage model using 15 factors. There is one common factor, one factor for each equation (6), one factor for each lag (2), one factor for each variable (6). Ξ_t simply loads the factors on the required elements of the B_t vector. Thus, it is a 72x15 matrix with zeros and ones. To maintain comparability, we use the same training sample and the same hyperparameters $\left(k_Q^2, k_S^2, k_W^2\right)$ as in the benchmark case. The computational time for the algorithm was about 12 hours and the acceptance rate for B^T was around 78% when the 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 shrinkage algorithm has less problems with explosiveness by construction, but requires important restrictions on the law of motion of B^T and θ^T, making the trades-off

28

roughly similar. The single move algorithm is instead computationally much more demanding because we need to compute the constant of integration $\lambda\left(.\right)$ at each t and this cancels out the advantages of having more efficient draws for B^T.

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

1.5.5   Time variations in structural parameters

We first describe the time variations that our model delivers. In figure 1.3 we report the highest 68 percent posterior tunnel for the variability of the monetary policy shock and in figure 1.4 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 (1998b), who claim that monetary policy in the 1980s was run differently, and agrees with the results of Sims and Zha (2006).

Figure 1.4 indicates that 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 in the Phillips curve.

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

$$\sigma_t^{mp}$$
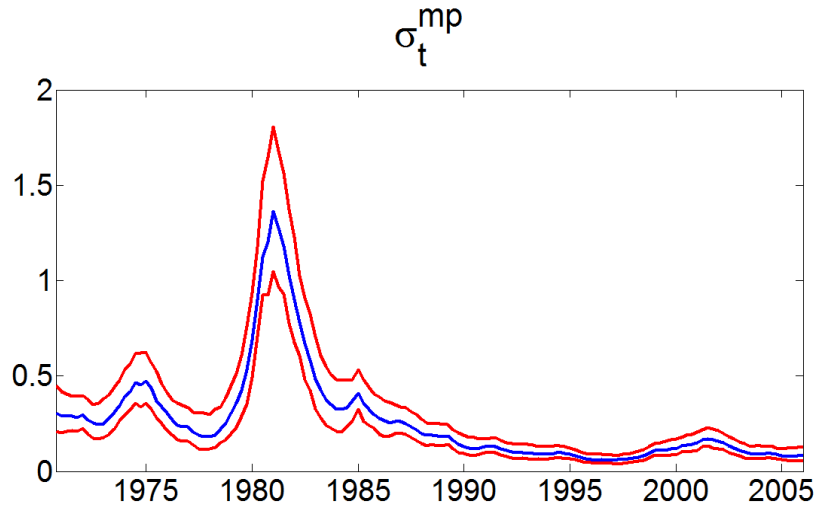
Figure 1.3: Median and posterior 68 percent tunnel, volatility of monetary policy shock.

Also interesting is the fact that 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.

One additional features of figure 1.4 needs to be mentioned. 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.

In sum, 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 the contemporaneous policy and non-policy coefficients.

1.5.6   The transmission of monetary policy shocks

We study how the time variations we have described 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 normalized to one and one with the structural shock normalized to zero.

Figure 1.4: Estimates of α

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 data in the early part of the sample, but disappears as time goes by. As figure 1.5 indicates, monetary policy shocks have the largest effects in 1981; the pattern is similar but weaker in 1975 and 1990. In 2005, prices, output and unemployment effects are perverse (inflation and output growth significantly increase and unemployment significantly falls after an interest rate increase). Note that the differences in the responses of output and unemployment between, say, 1981 and 2005 are a-posteriori significant. Thus, it appears that the ability of monetary policy to affect the real economy has considerably weakened over time and policy surprises are interpreted in different ways across decades.

Despite these noticeable variations, the proportion of the forecast error variance of output, prices and unemployment due to policy shocks is consistently small. Monetary policy shocks explain little of the forecast error variance of inflation at all times 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.

31
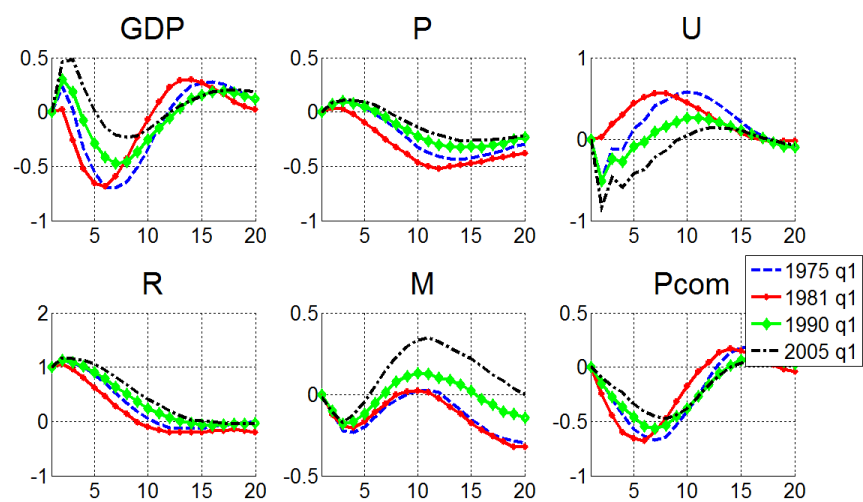
Figure 1.5: Dynamics following a monetary policy shock, different dates.
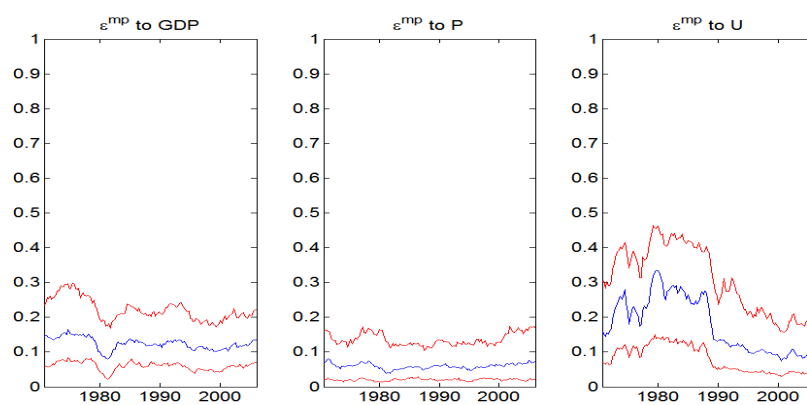
Figure 1.6: Long-run effects of monetary policy shocks
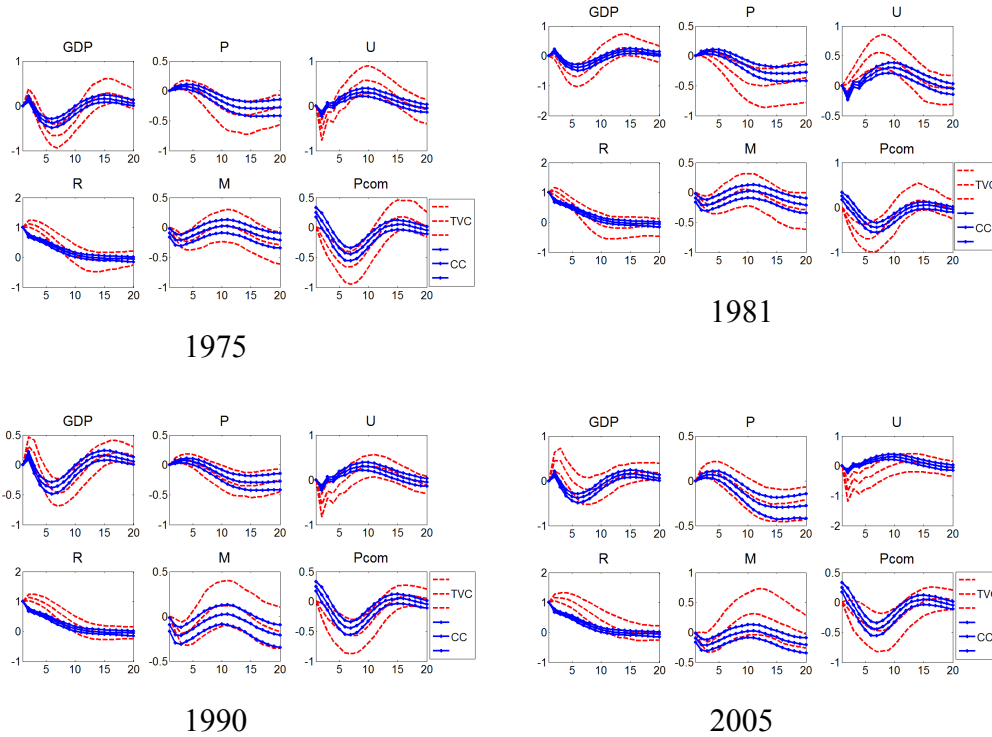
Figure 1.7: Time varying and time invariant responses.

Our results are very much in line with those of Canova et al. (2008), even though they use sign restrictions to extract structural shocks, and 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.

1.5.7   A time invariant over-identified model

We compare our 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 fixed coefficients.

To illustrate the difference that the two systems produce, we report the responses of the variables to a unexpected monetary policy impulse at four dates

(1975, 1981, 1990, 2005). 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. Thus, the two systems give quite a different interpretation of the transmission of monetary policy shocks.

1.6   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 restrictions, that are just identified or overidentified, and where the restrictions are of linear or non-linear type. Our algorithm adds a Metropolis step to a standard Gibbs sampling routine but nests the model into a general non-linear state space. Thus, we greatly expand the set of structural VAR models that researchers can deal with within the same estimation framework.

We apply the methodology to the estimation of a monetary policy shock in a non-recursive overidentified TVC model similar to the one used by Robertson and Tallman (2001), Waggoner and Zha (2003) with fixed coefficients. In the context of this example, we examine the merits of multi-move vs. a 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 important time variations in the variance of the monetary policy shock and in the estimated non-zero contemporaneous relationships. These time variations translate in important changes in the transmission of monetary policy shocks to the variables in the economy. We also show that a different characterization of the dynamics in response to monetary policy shocks would emerge in an overidentified but fixed coefficient VAR.

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 which 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), long run restrictions as in Galí and Gambetti (2009) or variance decomposition restrictions. The computational complexity is important but it is not overwhelming and all the computations can be easily performed on a standard PC with sufficient RAM memory.

Chapter 2

MEASURING THE STANCE OF MONETARY POLICY IN A TIME-VARYING WORLD

2.1   Introduction

The stance of monetary policy is of general interest for macroeconomists and the private sector. It provides an important input to understand the current state of the economy and contributes to the expectations formation of future states. Despite its importance, it has been difficult to have an exact measure of this stance, given the lack of consensus on what were the instruments of monetary policy and operating procedures at each point in time. Currently, this task has turned even more difficult after the introduction of so-called Unconventional Monetary Policies (UMP) and the achievement of the Zero-Lower-Bound (ZLB) of the Federal Funds Rate (FFR), given that the latter used to be considered the core instrument at least for the last two decades. The purpose of this paper is to provide a measure of the policy stance which takes into account changes in the operating procedures of the Fed.

Monetary Policy is implemented through intervention in the reserves market. In this market, each participant has to meet a reserve requirement set by the Federal Reserve in advance. To do that, market participants use the interbank loans market so that banks that have a deficit in reserves can borrow from the ones that have excessive reserves. These loans are granted only if the borrowers have an amount of collateral equivalent to the asked loan. The equilibrium price of this market is the FFR. The Fed performs open market operations (OMO) in order to set the supply of reserves and a thereby affect the equilibrium outcome of this market. If a bank can not meet its reserve requirement, it has the option of borrow

reserves from the Fed at the Discount Window (DW), these are called Borrowed Reserves (BR). Given the Total Reserves (TR) that market participants have at the end of the period, the difference is called Non-Borrowed Reserves (NBR=TR-BR). Finally, the way these operations (OMO and DW) are implemented is what we call operating procedures.

In the recent monetary history of the United States we can find evidence of different episodes of operating procedures implementation which depend on what the Fed targets at each point in time, e. g. targeting borrowed reserves, non-borrowed reserves, total teserves or the federal funds rate (see, Cosimano and Jansen (1988), Cosimano and Sheehan (1994), Bernanke and Mihov (1998b), among others).

Structural Vector Autoregressions (SVARs) have been popular techniques to identify monetary policy shocks and measure the policy stance. In these models the FFR has been considered the core instrument since the seminal work of Bernanke and Blinder (1992)[12]. Christiano and Eichenbaum (1992) and Strongin (1995) on the other hand used reserves of banks as a monetary policy instrument. Bernanke and Mihov (1998b) reconcile the two strands of the literature with an eclectic approach that identifies monetary policy shocks as a linear combination of innovations in different instruments. Christiano et al. (1999) summarize the literature. In the last decade, SVARs have been extended to resolve some problems with the procedure. In particular, larger information sets have been used to solve the price-puzzle (see Bernanke et al. (2005)), instability of parameters have been considered and regime changes explicitly discussed (see Primiceri (2005), Sims and Zha (2006)). Nevertheless, all these extensions still consider the FFR as the core instrument. Clearly, approaches that used the FFR as policy instrument turned out to be unsuitable to discuss UMPs and the achievement of the ZLB. Recent attempts to identify monetary policy where the ZLB binds and UMP is active can be found in Baumeister and Benati (2012) and also Peersman (2011). They use sign-restrictions as in Canova and De Nicoló (2002) and Uhlig (2005), and identify UMP shocks distinct from FFR innovations. UMP has different dimensions, and an innovation in each of them must be associated with monetary policy actions. As a result, considering a different policy shocks for each UMP dimension (as in Baumeister and Benati (2012) and Peersman (2011)) is not a good strategy. We believe, that the strategy of identifying monetary policy shocks as a linear combination of innovations in different instruments as Bernanke and Mihov have suggested is a good one, given the various dimensions of UMPs (e.g. to influence Financial Markets conditions through Large Scaled Asset Purchases, Forward Guidance, Direct Financial Intermediation, Quantitative Easing)[3]. Thus,

---

[1] See also Sims (1986) and Leeper et al. (1996).

[2] See Kilian (2012) for a recent survey about SVARs and identification.

[3] See Williams (2011), Williams (2012b) and Williams (2012a) for a detailed description of UMP. See also Borio and Disyatat (2010) and Cecioni et al. (2011) for a thorough survey of the

in this paper we modify Bernanke and Mihov (1998b)'s approach to take into account the multidimensionality of the UMP and the possibility that different instruments matter at different points in time.

The building blocks of our approach are as follows. Bernanke and Mihov (1998b) characterize Federal Reserve's operating procedures and provide a measure of the stance of monetary policy for the period 1965-1996. Essentially, their model has an interbank market of reserves where monetary policy can be implemented, depending on the parameter values of the model, by setting either interest rates (price of reserves) or the supply of reserves. What makes this model useful is its capability to identify monetary policy shocks for different contexts and instruments (see eq. (12) in the mentioned reference). These authors make explicit their concern about stability of parameters along their sample of analysis because operating procedures might have changed (e.g. Volcker's experiment in early 1980s or the recent UMPs in our case). Instead, when measuring policy stance one should take into account that the weight of each instrument is likely to be time-varying. These weights are nonlinear functions of estimated structural parameters. To study the posterior distribution of the path of the monetary policy stance and the weights taken by each component, we follow Canova and Pérez Forero (2013), who extend the Time-Varying Coefficients (TVC) VAR with Stochastic Volatility of Primiceri (2005) to deal with non-recursive and potentially overidentified SVAR models. We extend the framework used by Bernanke and Mihov (1998b) to account for the monetary policy (UMPs) practices. The model can be used to study the role of Quantitative Easing (QE), since it identifies demand and supply of reserves shocks and discount window operations shocks. However, the model needs to be slightly modified in order to capture Large Scaled Asset Purchases (LSAPs) and Forward Guidance (the announcement of future path for interest rates), actions aimed to affect medium and long-term interest rates. Thus, we specify an informational equation that represents financial markets' dynamics through an indicator of the level of yield-to-maturity spreads with respect to the short term interest rate[4]. One potential limitation of our approach is that we do not make explicit the role of communication (i.e. FOMC meeting Releases, Minutes, Speeches, etc.) and the recently introduced interest paid for holding reserves. Our strategy to chacarterize the Monetary Policy Stance is robust in terms of specification, since we are allowing structural parameters in both policy and non-policy blocks to vary over time. On the other hand, we believe that part of the effect of FOMC communication is captured through changes in the level of spreads in the yield curve, which is

---

different dimensions of UMPs.

[4]Reis (2009), Blinder (2010), Lenza et al. (2010) and Hamilton and Wu (2012) present the main characteristics of UMPs, emphasizing the role of yield curve spreads as a powerful indicator that summarizes both credit policy as well as the expectations of future paths for interest rates (Forward Guidance).

included in our SVAR.

We find that the stance of monetary policy has varied quite a lot over the last 40 years. It was loose for the first half of the 1970s and roughly neutral for the second half, it becomes tighter at the beginning of Volcker's period, i.e. the so-called Volcker's disinflation experiment (1980-1982) and then becomes loose again. Volcker's period ends with a relatively tight stance but showing more uncertainty than before. Greenspan's first ten years (1987-1996) exhibit a tight stance with a short period of loose policy in 1989. A long episode of loose stance (1996-2001) is observed with a subsequent neutral stage (2002-2003). Last Greenspan's years (2003-2005) display a relatively tight stance but shows an upward trend starting in late 2004. The stance turns to be loose when Bernanke's period starts until the outbreak of the Great Recession in 2007:Q4, when the stance turns to be tight again since 2008:Q4. We finally observe a reversal of this pattern after the implementation of UMPs, when the stance turns to be relatively loose in 2011-2012. This result is also in line with Beckworth (2011), who claims that the Monetary Policy Stance was relatively tight in 2008. The relative weights of these instruments are time-varying, where the most important result is the weight of zero for the FFR at the end of the sample, consistent with the binding ZLB. What matters here is the fact that the model is capable of capturing significant changes in operating procedures.

Model estimates allow us to explore time variations in the transmission of policy shocks. Overall, the transmission of monetary policy shocks is stable for a large portion of our sample, but it exhibits significant changes after the outbreak of the Great Financial Crisis and the achievement of the ZLB. We find that the effect of expansionary policy shocks on the spreads is positive before 2007, but turns to be negative afterwards. The latter is consistent with the purpose of UMPs, i.e. since the FFR is constant, the objective is to cut medium and long term interest rates. We also show that the *liquidity effect* vanishes over time and that the volatility of monetary policy shocks is changing.

We explore the sensitivity of our results using alternative specifications. We find that the paths of structural parameters and variances might differ across models. However, the main features of our result are robust.

In sum, the approach this paper presents is capable of capturing changes in monetary policy impementation across different episodes. We present a monetary policy stance index that hope will be useful for both policy makers and researchers. More work is needed for exploring the explicit role of communication in UMPs, the announcement of future paths of interest rates and credibility. We believe that these type of issues should be explored in a richer setup and therefore we leave it for future agenda. In this regard, some structural models that incorporate different dimensions of UMPs can be found in Gertler and Karadi (2011), Cúrdia and Woodford (2011) and Chen et al. (2012).

The paper is organized as follows: section 2.2 presents the Structural VAR model used for the analysis, section 2.3 describes the estimation procedure, section 2.4 presents an estimate of the monetary policy stance, sections 2.5 and 2.6 explore the transmission mechanism and the volatility of monetary policy shocks, respectively, section 2.7 presents the sensitivity analysis and section 2.8 concludes.

2.2 The Model

2.2.1 A Structural Dynamic System

We are interested in specifying a dynamic setting that allow us to identify monetary policy shocks. Therefore, we closely follow the methodology proposed by Bernanke and Blinder (1992) and Bernanke and Mihov (1998b). That is, assume that the structure of the economy is linear and given by

$$
Y_t = c_t^{np}D_t + \sum_{i=0}^{p} R_{i,t}Y_{t-i} + \sum_{i=0}^{p} T_{i,t}P_{t-i} + C_t^{np}v_t^{np}
$$

$$
P_t = c_t^{p}D_t + \sum_{i=0}^{p} S_{i,t}Y_{t-i} + \sum_{i=0}^{p} G_{i,t}P_{t-i} + C_t^{p}v_t^{p}
$$

where Y_t is a vector of macroeconomic variables, P_t is a vector of monetary policy instruments, c_t^{np} and c_t^{p} are matrices of coefficients on a vector of deterministic variables D_t and v_t^{np} and v_t^{p} are vectors of structural shocks that hit the economy at any point in time $t = 1, \ldots, T$ with

$$
v_t^{k} \sim N\left(\mathbf{0}, \Sigma_t^{k}\Sigma_t^{k\prime}\right); \quad k = \{np, p\}
$$

where $\Sigma_t^{k}\Sigma_t^{k\prime}$ is a diagonal positive definite matrix and $Cov\left(v_t^{np}, v_t^{p}\right) = \mathbf{0}$. Notice that here we allow for potential time variation in matrix coefficients and variances and therefore we include the index t for each of them. We assume that the macroeconomic variables Y_t do not react within the same period to innovations in policy instruments, i.e. $T_{0,t} = \mathbf{0} \; \forall t$, so that

$$
Y_t = c_t^{np}D_t + \sum_{i=0}^{p} R_{i,t}Y_{t-i} + \sum_{i=1}^{p} T_{i,t}P_{t-i} + C_t^{np}v_t^{np} \qquad (2.1)
$$

$$
P_t = c_t^{p}D_t + \sum_{i=0}^{p} S_{i,t}Y_{t-i} + \sum_{i=0}^{p} G_{i,t}P_{t-i} + C_t^{p}v_t^{p}
$$

where we assume that a period t is a quarter. For now we can say that the structure of the economy (2.1) takes the form of a system of Vector Autoregressions (VAR)

of order p. Denote the vector of variables $\mathbf{y}_t \equiv [Y_t', P_t']'$, the vector of intercepts $c_t \equiv \left[c_t^{np'}, c_t^{p'} \right]'$ and the matrices

$$
\begin{aligned}
A_t &\equiv \left[\begin{array}{cc} A_{11,t} & A_{12,t} \\ A_{21,t} & A_{22,t} \end{array} \right] = \left[\begin{array}{cc} I - R_{0,t} & \mathbf{0} \\ -S_{0,t} & I - G_{0,t} \end{array} \right] \qquad (2.2) \\
A_{i,t} &\equiv \left[\begin{array}{cc} R_{i,t} & T_{i,t} \\ S_{i,t} & G_{i,t} \end{array} \right]; \ i = 1, \ldots, p \\
C_t &\equiv \left[\begin{array}{cc} C_{11,t} & C_{12,t} \\ C_{21,t} & C_{22,t} \end{array} \right] = \left[\begin{array}{cc} C_t^{np} & \mathbf{0} \\ \mathbf{0} & C_t^{p} \end{array} \right]; \Sigma_t \equiv \left[\begin{array}{cc} \Sigma_t^{np} & \mathbf{0} \\ \mathbf{0} & \Sigma_t^{p} \end{array} \right]
\end{aligned}
$$

so that the model can be re-expressed as a Structural VAR with time-varying coefficients:

$$
A_t \mathbf{y}_t = c_t D_t + A_{1,t} \mathbf{y}_{t-1} + \ldots + A_{p,t} \mathbf{y}_{t-p} + C_t v_t \qquad (2.3)
$$

Without additional assumptions, the economic model expressed in its structural form cannot be directly estimated. In order to identify the vector of structural shocks $\left[v_t^{np'}, v_t^{p'} \right]'$ we need to identify the matrices A_t and C_t. We will describe the structural model in subsection 2.2.3 but first we will describe the basic setup in detail.

2.2.2 Basic setup

Consider a vector of M variables $\mathbf{y}_{t(M \times 1)}$ with data available for T periods. I assume that the data generating process for $\mathbf{y}_t$ is the reduced-form version of the model (2.3), i.e. a VAR(p) process such that:

$$
\mathbf{y}_t = B_{0,t} D_t + B_{1,t} \mathbf{y}_{t-1} + \ldots + B_{p,t} \mathbf{y}_{t-p} + u_t; \ \ t = 1, \ldots, T \qquad (2.4)
$$

where $B_{0,t} \equiv A_t^{-1} c_t$ is a matrix of coefficients on a $\bar{M} \times 1$ vector of deterministic variables D_t and $B_{i,t} \equiv A_t^{-1} A_{i,t}; i = 1, \ldots, p$ are $M \times M$ matrices containing the coefficients on the lags of the endogenous variables and the error term is distributed as $u_{t(M \times 1)} \sim N\left(0, \Omega_t\right)$, where $\Omega_{t(M \times M)}$ is a symmetric, positive definite, full rank matrix for every t. Equation (2.4) is a reduced form and the error terms u_t do not have an economic interpretation. Let the structural shocks be $\varepsilon_t \sim N\left(0, I_M\right)$ and let the mapping between these shocks and their reduced form counterpart be

$$
u_t = A_t^{-1} C_t \Sigma_t \varepsilon_t \qquad (2.5)
$$

where $A_{t(M \times M)}$, $C_{t(M \times M)}$ and $\Sigma_{t(M \times M)}$ are defined in (2.2). In order to be in line with the notation of previous subsection, we should note that $\varepsilon_t \equiv \Sigma_t^{-1} v_t$ is the normalized version of the structural shocks. I substitute (2.5) into (2.4) so that we get structural form of the VAR(p) model:

$$
\mathbf{y}_t = X_t' B_t + A_t^{-1} C_t \Sigma_t \varepsilon_t \qquad (2.6)
$$

The matrix of regressors is $X_t' = I_M \otimes \left[D_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}' \right]$ is a $M \times K$ matrix where D_t potentially includes a constant term, trends, seasonal dummys, etc and $K = M \times \overline{M} + p \times M^2$. Parameter blocks $\left(B_t, A_t, C_t^{-1}, \sigma_t \right)$ are treated as latent variables that evolve as independent random walks:

$$B_t = B_{t-1} + \upsilon_t \tag{2.7}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{2.8}$$

$$\widetilde{c}_t = \widetilde{c}_{t-1} + \varrho_t \tag{2.9}$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t \tag{2.10}$$

where $B_{t(K \times 1)} = \left[vec(B_{0,t})', vec(B_{1,t})', \ldots, vec(B_{p,t})' \right]'$ is a $K \times 1$ vector; α_t and $\widetilde{c}_t$ denote free parameters of matrices A_t and C_t^{-1}, respectively[5]. In addition, $\sigma_{t(M \times 1)}$ is the main diagonal of Σ_t. Finally, the covariance matrix for the error vector is:

$$V = Var \left(\begin{bmatrix} \varepsilon_t \\ \upsilon_t \\ \zeta_t \\ \varrho_t \\ \eta_t \end{bmatrix} \right) = \begin{bmatrix} I_M & 0 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 & 0 \\ 0 & 0 & S_a & 0 & 0 \\ 0 & 0 & 0 & S_{\widetilde{c}} & 0 \\ 0 & 0 & 0 & 0 & W \end{bmatrix} \tag{2.11}$$

The model presented captures time variations of different parameter blocks: i) lag structure (2.7), ii) structural parameters (2.8) and (2.9) and iii) structural variances (2.10). In other words, the model is capable of capturing the sources of potential structural changes, i.e. drifting coefficients $(B_t, \alpha_t, \widetilde{c}_t)$ or stochastic volatility (σ_t) without imposing prior information about specific dates or number of structural breaks. In particular, in the process of identifying parameters that affect the policy stance and represent the operating procedures, a subset of $(\alpha_t, \widetilde{c}_t)$ will have a major relevancy.

2.2.3 A Structural VAR model with an Interbank Market

Bernanke and Mihov (1998b) proposed a *semi-structural* VAR model that characterizes Federal Reserve's operating procedures. The purpose of this section is to present an extension of the framework in order to take into account conventional and unconventional policies.

---

[5]The reason of why we are focused on C_t^{-1} instead of C_t is for computational convenience after the construction of the State-Space model (see Appendix B.3.4 for details). For instance, denote $\overline{c}_t$ as the vector of free parameters of C_t. Then, if C_t is lower-triangular with ones in the main diagonal (which is indeed the case here), then the set of free parameters of C_t^{-1} will be simply the vector $\widetilde{c}_t = -\overline{c}_t$. Thus, recovering the original parameters will be straightforward.

Consider the vector of variables

$$\mathbf{y}_t = [x_t, \pi_t, \Delta Pcom_t, SPR_t, TR_t, FFR_t, NBR_t]'$$

where x_t represents output growth rate, π_t represents the inflation rate, $\Delta Pcom_t$ is the growth rate of an index of commodity prices, SPR_t is an index that summarizes the evolution of financial markets, TR_t is the total amount of reserves that banks hold at the Central Bank, FFR_t is the Federal Funds Rate in annual terms and NBR_t is the total amount of non-borrowed reserves. Regarding the model specification, we re-write equation (2.5) as follows[6]

$$A_t u_t = C_t v_t$$

where $v_t = \Sigma_t \varepsilon_t$ is the re-scaled vector of structural shocks and u_t is the vector of reduced-form innovations. Moreover, recall the system partition described in subsection 2.2.1. That is, there is a *non-policy* block and a *policy* block:

$$\begin{bmatrix} A_{11,t} & A_{12,t} \\ A_{21,t} & A_{22,t} \end{bmatrix} \begin{bmatrix} u_t^{np} \\ u_t^p \end{bmatrix} = \begin{bmatrix} C_{11,t} & C_{12,t} \\ C_{21,t} & C_{22,t} \end{bmatrix} \begin{bmatrix} v_t^{np} \\ v_t^p \end{bmatrix} \qquad (2.12)$$

Within the non-policy block there is output growth, inflation, and commodity prices growth, so that $u_t^{np} = [u_t^x, u_t^\pi, u_t^c]'$ and $v_t^{np} = [v_t^x, v_t^\pi, v_t^c]'$. The system has $M = 7$ variables, where we denote the number of non-policy variables as $M_{np} = \dim(u_t^{np}) = 3$. The policy block contains the remaining variables of the system, i.e. $M_p = M - M_{np} = \dim(u_t^p) = 4$, which will be called policy instruments. We specify a sub-system of equations for the portion of u_t^p that is orthogonal to the non-policy block u_t^{np}. The set of assumptions embedded in the system of equations above can be summarized as follows:

1. *Non-policy block*: First, non-policy variables only react to policy changes with some delay, i.e. according to (2.2), we have $A_{12,t} = \mathbf{0}_{(M_p \times M_{np})}, \forall t$. The intuition behind this assumption is that the private sector considers the lagged stance of policy as a state variable. That is, non-policy variables will not change in the same quarter after an innovation in a particular instrument that belongs to u_t^p. Moreover, following Bernanke and Mihov (1998b), We will keep this non-policy block unmodeled and just assume that $A_{11,t}$ is lower triangular

$$A_{11,t} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{x,t}^\pi & 1 & 0 \\ \alpha_{x,t}^c & \alpha_{\pi,t}^c & 1 \end{bmatrix}$$

---

[6]According to Amisano and Giannini (1997) and Lütkepohl (2005) (ch. 9), the model presented in is one version of the AB model. See the mentioned references for details.

The ordering in this block is an open question, but for the question of interest it does not matter, since the results we present are robust. We also assume that innovations in u_t^{np} will affect the policy block in the same quarter, i.e. $A_{21,t}$ is an unrestricted $M_{np} \times M_p$ matrix of potentially non-zero parameters (see Appendix B.3.4). In addition, according to (2.2), we have $C_{12,t} = \mathbf{0}_{(M_p \times M_{np})}$ and $C_{21,t} = \mathbf{0}_{(M_{np} \times M_p)}, \forall t$. We also assume that $C_{11,t} = \mathbf{I}_{M_p}$ is the identity matrix, which means that structural shocks v_t^x, v_t^π and v_t^c only affect output growth, inflation and commodity growth on impact, so that there are no cross-effects on impact.

Turning to the policy block, the next four equations have the aim to describe the Interbank Market of Reserves. That is, each period t banks have to meet their reserve requirements determined by the Fed. The sum of the level of reserves across banks determines the term "Total Reserves" denoted by TR_t. Moreover, these reserves pay an interest that is closely related to the Federal Funds Rate, FFR_t and as a result the latter is a relevant indicator for the demand of reserves. In order to meet their reserve requierements banks have three alternatives: they could get liquidity from the Discount Window ("Borrowed Reserves", BR_t), or through interbank loans and Open Market Operations ("Non-Borrowed Reserves", NBR_t). Banks have a pool of assets that are used as collaterals in order to get liquidity. It is natural to assume that the pool of assets owned by banks are also traded in the secondary market. As a result, the evolution of the Spreads in the yield curve with respect to the short term interest rate, SPR_t, is also a relevant indicator. If banks use more of these assets to get reserves TR_t, they are going to affect the relative supply of them in the secondary market, so that the final price is going to change. For that reason, we assume that innovations in the terms TR_t and NBR_t potentially affect SPR_t contemporaneously. We also allow that the Federal Funds Rate FFR_t affects SPR_t contemporaneously. To close the model, we assume that the Federal Reserve intervenes in the market by deciding the amount of liquidity that is going to be injected through open market operations. I will now proceed to describe the structural equations that describe the Interbank Market. In equation (2.12):

2. *Financial markets equation*: Here we characterize the dynamics of Financial Markets. To do that, I construct the indicator SPR_t which is essentially the first principal component of a large dataset of spreads w.r.t. the Federal Funds Rate at different maturities[7]. This principal component is associated

---

[7]I define a set of variables x_t that contains the spreads of the Treasury Bonds yield with respect to the Federal Funds Rate FFR_t for every term included (3M,6M,1Y, 2Y, 3Y, 5Y, 10Y, 30Y as well as AAA and BAA bonds). The indicator $Spread_t$ is the first principal component of the

with the level of the spreads. Moreover, according to Bernanke et al. (2005), the inclusion of more information in the VAR model is also crucial to avoid the so-called *price puzzle*. The structural residual v_t^f represents a financial market shock. Moreover, the fact that one dimension of UMP is to directly influence interest rate spreads makes it possible to consider this equation within the policy block[89].

$$u_t^{SPR} = -\alpha_{TR,t}^{SPR} u_t^{TR} - \alpha_{FFR,t}^{SPR} u_t^{FFR} - \alpha_{NBR,t}^{SPR} u_t^{NBR} + v_t^f \qquad (2.13)$$

3. *Demand for Reserves equation:* Represents the total demand for reserves of banks. In particular, the portion of u_t^{TR} which is orthogonal to the nonpolicy block depends negatively on the Federal Funds Rate's innovation u_t^{FFR} and v_t^d represents a shock to reserves' demand.

$$u_t^{TR} = -\alpha_{1,t}^d u_t^{FFR} + v_t^d \qquad (2.14)$$

4. *Demand for discount window operations equation*: Borrowed Reserves (BR) is the portion of reserves obtained through the discount window. They depend positively on the Federal Funds Rate's innovations u_t^{FFR} and v_t^b represents a shock to the discount window operations' demand. This potential source of fluctuation could become relevant in episodes of financial stress or under BR targeting.

$$u_t^{BR} = u_t^{TR} - u_t^{NBR} = \alpha_{1,t}^b u_t^{FFR} + v_t^b \qquad (2.15)$$

5. *Federal Reserve equation*: Represents the money supply process, i.e. liquidity provision through open market operations in order clear the money market. The portion of u_t^{NBR} which is orthogonal to the nonpolicy block responds contemporaneously to shocks in the spreads, the demand for total and borrowed reserves. Every other action unrelated with the mentioned shocks is called an exogenous monetary policy shock, v_t^s.

$$u_t^{NBR} = \phi_t^f v_t^f + \phi_t^d v_t^d + \phi_t^b v_t^b + v_t^s \qquad (2.16)$$

matrix $x_t x_t'$, a vector that is normalized according to Bernanke et al. (2005). We take only the first factor, which is related with the level of spreads. Second and Third factors are related with the slope and curvature, see Mumtaz and Surico (2009).

[8]See also Baumeister and Benati (2012) for a detailed version of a TVP-SVAR identified with sign-restrictions. The document finds a powerful effect of lowering interest rate spreads on the aggregated economy.

[9]Several empirical papers have found that the Fed is able to influence the level of these spreads at different maturities through Large-Scale Asset Purchases and Forward guidance policies. See e.g. Gagnon et al. (2010), Gagnon et al. (2011), Swanson and Williams (2012), Hamilton and Wu (2012), Christensen and Rudebusch (2012), Chung et al. (2012) among others.

Equations (2.14) and (2.15) and a slightly modified equation (2.16) also appear in the benchmark version from Bernanke and Mihov (1998b). It is worth to notice that the signs in equations do not necessarility imply an identification strategy via sign-restrictions. Thus, our main contribution is equation (2.13). As in the above reference, we use the idea that each of the four variables could be considered a monetary policy instrument. In this framework, we abstract for the role of Fed's monetary policy communication (i.e. FOMC meeting Releases, Minutes, Speaches, etc.) as well as for interest paid by reserves. Our strategy is robust in terms of specification, since we are allowing structural parameters from both policy and non-policy blocks as well as structural variances to vary over time. Since the model is an approximation that could potentially be misspecified, it is likely that posterior estimates of structural parameters will vary across subsamples. Therefore, it is even better to allow for continuous drifting parameters[10]. On the other hand, we believe that part of the effect of FOMC communication could be captured through changes in the level of SPR_t included in our SVAR.

Recall (2.12) and consider the sub-system of equations for the portion of u_t^p that is orthogonal to the non-policy block u_t^{np}, i.e. $A_{22,t} u_t^p = C_{22,t} v_t^p$. The system is[11]:

$$\underbrace{\begin{bmatrix} 1 & \alpha_{TR,t}^{SPR} & \alpha_{FFR,t}^{SPR} & \alpha_{NBR,t}^{SPR} \\ 0 & 1 & \alpha_{1,t}^d & 0 \\ 0 & 1 & -\alpha_{1,t}^b & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{A_{22,t}} \times \begin{bmatrix} u_t^{SPR} \\ u_t^{TR} \\ u_t^{FFR} \\ u_t^{NBR} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \phi_t^f & \phi_t^d & \phi_t^b & 1 \end{bmatrix}}_{C_{22,t}} \times \begin{bmatrix} v_t^e \\ v_t^d \\ v_t^b \\ v_t^s \end{bmatrix}$$

The latter system can be solved for structural shocks $v_t^p = C_{22,t}^{-1} A_{22,t} u_t^p$. In particular, the last equation of this system corresponds to the monetary policy shock v_t^s, that is:

$$\begin{aligned} v_t^s = {} & -\phi_t^f u_t^{SPR} \\ & - \left(\phi_t^b + \phi_t^d + \phi_t^f \alpha_{TR,t}^{SPR} \right) u_t^{TR} \\ & + \left(\phi_t^b \alpha_{1,t}^b - \phi_t^d \alpha_{1,t}^d - \phi_t^f \alpha_{FFR,t}^{SPR} \right) u_t^{FFR} \\ & + \left(\phi_t^b - \phi_t^f \alpha_{NBR,t}^{SPR} + 1 \right) u_t^{NBR} \end{aligned} \qquad (2.17)$$

The intuition behind equation (2.17) is that monetary policy actions can be represented as a linear combination of innovations in different instruments. Policy

---

[10]See Cogley and Yagihashi (2010), Chang et al. (2010) and Canova and Pérez Forero (2013) for more details about this issue.

[11]We have considered that now $\sqrt{var\left(v_t^b\right)} = \sigma_t^b / \left|\alpha_{1,t}^b\right|$. See also Favero (2001) for a similar approach using Bernanke and Mihov (1998b)'s model.

actions are characterized not only as interest rate innovations as commonly suggested in the literature and (2.17) can be used to evaluate the Monetary Policy Stance. Relative to Bernanke and Mihov (1998b), we take into account innovations in Yield-Curve Spreads that are orthogonal to the non-policy block, since that captures the new UMP instruments.

Because we are allowing coefficients to vary over time, the weight of each instrument will be time-varying as well. These weights are nonlinear functions of structural parameters that come from the estimated SVAR model. As a result, this characterization of policy actions is robust to changes in the operating procedures during the sample of analysis. For instance, consider the case of monetary policy conducted by interest rate setting. If the ZLB is binding, then the FFR will no longer be the policy instrument, at least temporarily. As a result, the Fed will re-design its operating procedures putting more weight on other instruments and assigning zero weight to the FFR[12].

The sub-system has $M_p = 4$ variables. Therefore the variance covariance matrix of the reduced form error terms u_t^p will have $4 \times (4 + 1)/2 = 10$ parameters. On the other hand, the vector of structural parameters has 12 elements, i.e.

$$\theta_t = \left(\alpha_{TR,t}^{SPR}, \alpha_{FFR,t}^{SPR}, \alpha_{NBR,t}^{SPR}, \alpha_{1,t}^d, \alpha_{1,t}^b, \phi_t^f, \phi_t^d, \phi_t^b, \sigma_t^f, \sigma_t^d, \sigma_t^b, \sigma_t^s \right)'$$

These parameters are treated as latent variables (see Canova and Pérez Forero (2013) for details). Thus, to achieve identification it is necessary to impose 2 additional restrictions. Following Bernanke and Mihov (1998b), we focus our attention on equation (2.17) and assume that monetary policy is associated with a particular instrument, i.e. set restrictions such that all the brackets are equal to zero except the one associated with our instrument of interest. Since our sample covers periods where the Fed had different chairmen and the period of UMP, we have less reasons to restric our attention to a particular instrument. Instead, we want to capture changes in operating procedures, which means that all the brackets in (2.17) could potentially be different from zero. For that reason we assume that

$$\alpha_{TR,t}^{SPR} = \alpha_{NBR,t}^{SPR} = 0 \tag{2.18}$$

Thus, the portion of u_t^{SPR} which is orthogonal to the nonpolicy block does not react within the same quarter to innovations in either Total or Non-Borrowed Reserves. This is not necessarily a strong assumption since what makes spreads react is the announcement of future changes in the money supply, such as QE1, and not necessarily effective changes in reserves. On the other hand, unlike Strongin (1995) and Bernanke and Mihov (1998b), we assume that $\alpha_t^d \neq 0$ and it means that the demand of Total Reserves is not necessarily inelastic with respect to the

---

[12]See for instance Reis (2009) and Blinder (2010).

Federal Funds Rate. We believe that in the context of a binding ZLB the demand for Total Reserves is not inelastic and for that reason we choose this assumption. Our system of equations will be exactly identified if (2.18) is imposed but we are still free to consider different instruments of monetary policy at any given time. We also test the sensitivity and plausibility of these identification restrictions in section 2.7.

The estimation of the path for the structural parameters and of v_t^s in (2.17) will give us a measure of the monetary policy stance that internalizes changes in operating procedures. On the other hand, the estimated path of the variance of policy shocks will shed light on the relative importance of the non-systematic component.

2.3 Bayesian Estimation

The purpose of this section is to describe the procedure used to estimate the parameters of the statistical model described in (2.6). In particular, we are interested in the posterior distribution of the latent variables described in (2.7), (2.8), (2.9) and (2.10). I will adopt a Bayesian perspective and, following Primiceri (2005), I will use a Multi-move Gibbs Sampling procedure. Moreover, we will sample structural coefficients from (2.8) and (2.9) introducing two Metropolis-type steps, as suggested by Canova and Pérez Forero (2013).

2.3.1 Data description

The time series used for the analysis are in quarterly frequency and were taken from the *International Financial Statistics (IFS)* Database of the International Monetary Fund (IMF), from the Federal Reserve Board's website and from S&P Dow Jones Indices[13]. From the former database I took the GDP Deflator and Aggregate Gross Domestic Product index (Volume, base 2005=100). From the second database I took Total Reserves of aggregated depository institutions, Non-borrowed Reserves of aggregated depository institutions (both seasonally adjusted), Federal Funds Rate and the Treasury Bonds yields from maturities 3M, 6M, 1Y, 2Y, 3Y, 5Y, 10Y, 30Y as well as AAA and BAA bonds. Finally, I took the Dow Jones average index from the third database. The sample runs from 1959:Q1 - 2012:Q3 (215 obs.) and includes approximately four years after the outbreak of the Great Recession.

---

[13]http://www.imfstatistics.org/imf/about.asp,
http://www.federalreserve.gov/econresdata/releases/statisticsdata.htm,
http://www.djaverages.com/?view=industrial&page=overview, respectively.

Industrial Production and Consumer Price Index variables are expressed in annual growth rates, i.e. $y^*_{i,t} = 100 * (\log(y_{i,t}) - \log(y_{i,t-4}))$. Federal Funds Rate and the remaining interest rates from the yield curve are expressed in annual terms. To induce stationarity, Total and Nonborrowed Reserves were standardized using the mean and standard deviation of Non-Borrowed Reserves for a window of 16 quarters. Bernanke and Mihov (1998b) also divide Total and Non-Borrowed reserves by the average of Total Reserves using a window of 36 months. However their approach is not useful for inducing stationarity given the recent changes in reserves[14].

2.3.2 Priors and setup

The Priors of the VAR are shown in Table 2.1 and they are chosen to be conjugated. As a result the posterior distribution will be Normal and Inverted-Wishart for each corresponding case. To insure stationarity of impulse responses, the posterior of B^T is truncated. That is, the associated companion form of the VAR (2.4) is computed for each draw of B^T and it is discarded if it does not satisfy the stability condition for $t = 1, \ldots, T$. The latter procedure is captured by the indicator function $I_B(.)$. In addition, the prior for initial states of structural parameters is calibrated using the first $\tau = 40$ observations (1959:Q1 - 1969:Q4) as a training sample. Thus, we estimate $\left(\overline{\mathbf{B}}, \overline{\mathbf{VB}}\right)$ via OLS and $\left(\overline{\boldsymbol{\alpha}}', \overline{\mathbf{c}}', \overline{\boldsymbol{\sigma}}'\right)'$ via Maximum Likelihood[15][16].

Moreover, I set $k^2_Q = 0.5 * 1 \times 10^{-4}, k^2_W = 1 \times 10^{-4}$, following Primiceri (2005), and $k^2_{S_a} = k^2_{S_{\tilde{c}}} = 1 \times 10^{-3}$, following Canova and Pérez Forero (2013). Finally, lag length is set to $p = 2$.

<div align="center">

Table 2.1: Priors

</div>

$B_0 \sim N\left(\overline{\mathbf{B}}, 4 \cdot \overline{\mathbf{VB}}\right)$	$Q \sim IW\left(k^2_Q \cdot \overline{\mathbf{VB}}, 1 + \dim B\right)$
$\alpha_0 \sim N\left(\overline{\boldsymbol{\alpha}}, diag\left(\|\overline{\boldsymbol{\alpha}}\|\right)\right)$	$S_a \sim IW\left(\begin{array}{c} k^2_{S_a} \cdot diag\left(\|\overline{\boldsymbol{\alpha}}\|\right), \\ 1 + \dim \alpha \end{array}\right)$
$\widetilde{c}_0 \sim N\left(-\overline{\mathbf{c}}, diag\left(\|-\overline{\mathbf{c}}\|\right)\right)$	$S_{\widetilde{c}} \sim IW\left(\begin{array}{c} k^2_{S_{\widetilde{c}}} \cdot diag\left(\|-\overline{\mathbf{c}}\|\right), \\ 1 + \dim \widetilde{c} \end{array}\right)$
$\log\left(\sigma_0\right) \sim N\left(\log\left(\overline{\boldsymbol{\sigma}}\right), I_M\right)$	$W_i \sim IG\left(k^2_W, 1/2\right), \; i = 1, \ldots, M$

[14] We have also tried to regress Total and Non-Borrowed Reserves on a constant and linear trend with breaks in the third quarter of 2008. However, the obtained residuals are extremely volatile.

[15] The MATLAB code **csminwel.m** from professor C. Sims website is used (http://sims.princeton.edu/yftp/optimize/mfiles/). I have chosen randomly 100 different starting points in order to find a global maximum.

[16] Alternatively, we could use a Minnesota-style prior for calibrating $\left(\overline{\mathbf{B}}, \overline{\mathbf{VB}}\right)$ (see Del Negro (2003), Canova (2007) (ch.10), among others), but we do not cover this issue here.

Since it is assumed that the blocks $\left(B^T, \alpha^T, \widetilde{c}^T, \sigma^T\right)$ follow random walks (see equations (2.7), (2.8), (2.9) and (2.10)), we use the mean and the variance of the priors of B_0, α_0, $\widetilde{c}_0$, $\log\left(\sigma_0\right)$ to initialize the Kalman Filter at each iteration. The sampling procedure is described below.

2.3.3 Sampling parameter blocks

We have to sample parameter blocks $\left(B^T, \alpha^T, \widetilde{c}^T, \sigma^T, s^T, V\right)$ and we do it sequentially using the logic of Gibbs Sampling (see Chib (2001)). The block s^T is an auxiliar one used as an intermediate step for sampling σ^T, see Kim et al. (1998). The sampling algorithm is as follows:

1. Set an initial value for $\left(B_0^T, \alpha_0^T, \widetilde{c}_0^T, \sigma_0^T, s_0^T, V_0\right)$ and set $i = 1$.

2. Draw reduced-form coefficients B_i^T from $p\left(B_i^T \mid \alpha_{i-1}^T, \widetilde{c}_{i-1}^T, \sigma_{i-1}^T, s_{i-1}^T, V_{i-1}\right)$. $I_B\left(B_i^T\right)$.

3. Draw structural parameters α_i^T from $p\left(\alpha_i^T \mid B_i^T, \widetilde{c}_{i-1}^T, \sigma_{i-1}^T, s_{i-1}^T, V_{i-1}\right)$.

4. Draw structural parameters $\widetilde{c}_i^T$ from $p\left(\widetilde{c}_i^T \mid B_i^T, \alpha_i^T, \sigma_{i-1}^T, s_{i-1}^T, V_{i-1}\right)$.

5. Draw volatilities σ_i^T from $p\left(\sigma_i^T \mid B_i^T, \alpha_i^T, \widetilde{c}_i^T, s_{i-1}^T, V_{i-1}\right)$.

6. Draw the indicator s_i^T from $p\left(s_i^T \mid B_i^T, \alpha_i^T, \widetilde{c}_i^T, \sigma_i^T, V_{i-1}\right)$.

7. Draw hyperparameters V_i from $p\left(V_i \mid \alpha_i^T, \widetilde{c}_i^T, y_i^T, \sigma_i^T, s_i^T\right)$.

8. Set $B_i^T, \alpha_i^T, \widetilde{c}_i^T, \sigma_i^T, s_i^T, V_i$ as the initial value for the next iteration. If $i < N$, set $i = i + 1$ and go back to 2, otherwise stop.

The indicator function $I_B\left(.\right)$ truncates the posterior distribution of B^T for draws that violate stationarity for $t = 1, \ldots, T$. I perform $N = 150,000$ draws discarding the first $100,000$ and I store one every 100 draws for the last $50,000$ to reduce serial correlation. Details about the algorithm are in Appendix B.3.2. The reader is also referred to the Appendix B.4 for the diagnosis of convergence.

2.4 The Stance of Monetary Policy

Using equation (2.17) and the identification restrictions (2.18), replacing the reduced form innovations by the data (standardized) included in the VAR and renaming the resulting index as the monetary policy stance (MPS), the expression

49
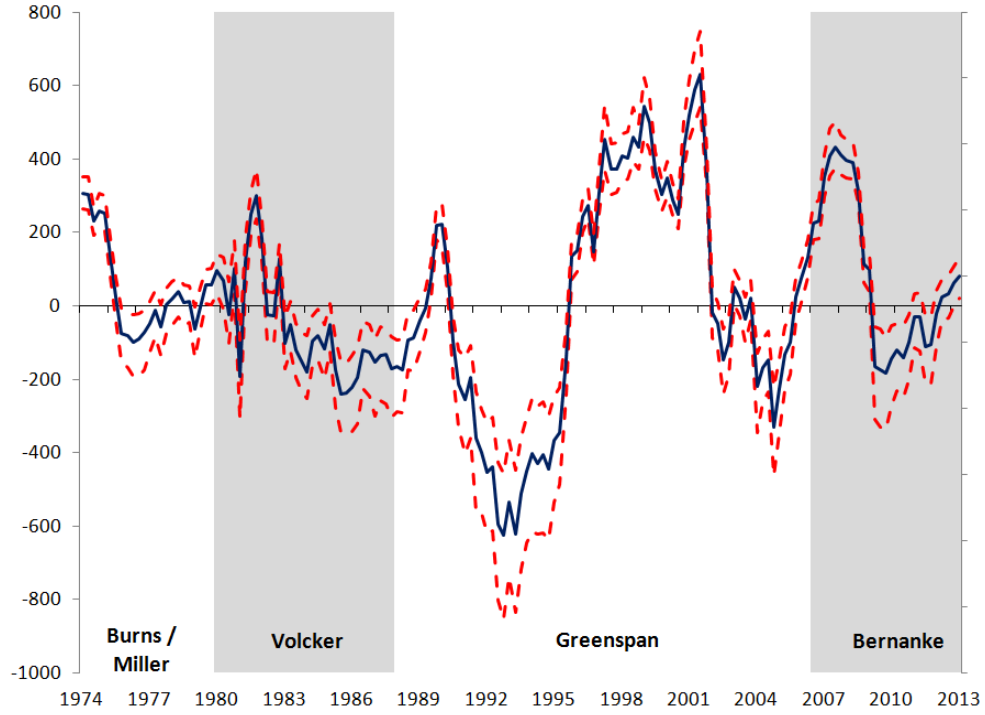
Figure 2.1: Posterior distribution of the Monetary Policy Stance, median value and 90 percent posterior bands

of interest is:

$$
\begin{aligned}
MPS_t \;\; = \;\; & -\phi_t^f SPR_t - \left(\phi_t^b + \phi_t^d\right) TR_t + \qquad\qquad\qquad\qquad (2.19) \\
& + \left(\phi_t^b \alpha_{1,t}^b - \phi_t^d \alpha_{1,t}^d - \phi_t^f \alpha_{FFR,t}^{SPR}\right) FFR_t + \left(\phi_t^b + 1\right) NBR_t
\end{aligned}
$$

We compute the posterior distribution for MPS_t at each point in time using posterior estimates of the parameters and VAR variables divided by their sample standard deviation. Moreover, in oder to normalize this index, we substract the resulting sample mean for each draw of MPS_t. Thus, the index has zero mean and therefore it is easy to associate positive (negative) values with a loose (tight) stance. In addition, since this index is a linear combination of standardized data, we assume that the index is also standardized.

Figure 2.1 depicts the index capturing the Monetary Policy Stance for the period 1974-2012. To the best of our knowledge, this is the first time that such an index is produced taking into account the uncertainty of parameter estimates. Overall, the stance varies quite a lot over the last 40 years. It is loose for a first half of the decade of 1970s and is about neutral for the second half. We can see a tight stance at the beginning of Volcker's period, i.e. during the so-called Volcker's

50
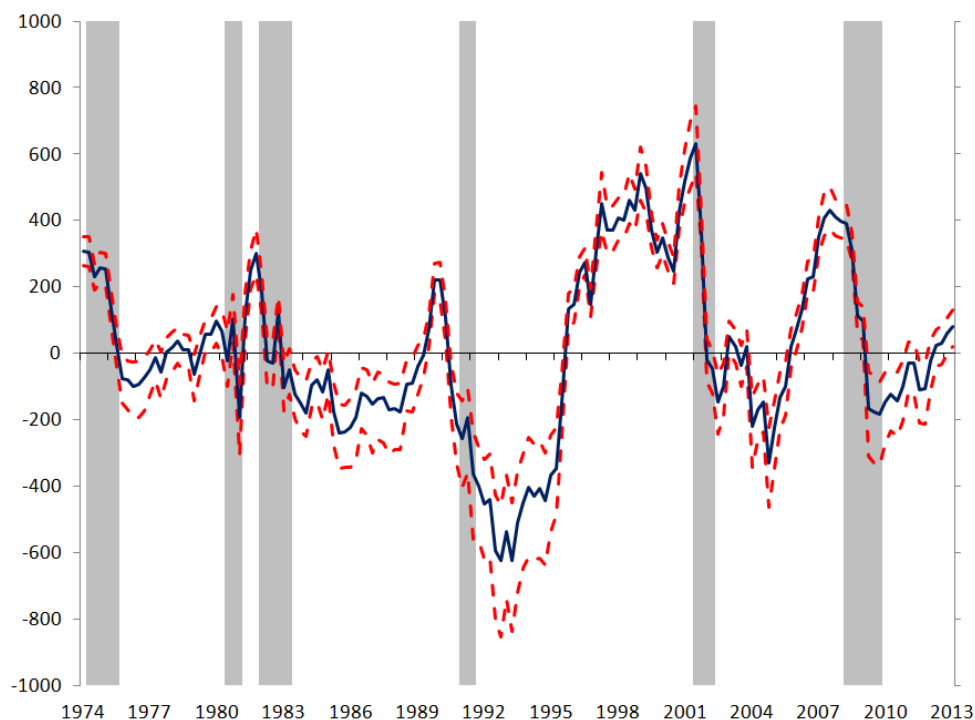
Figure 2.2: Monetary Policy Stance and NBER recession dates (shaded areas)

disinflation experiment (1980-1982). Volcker's period ends with a relatively tight stance but showing more uncertainty than in previous periods. Greenspan's first ten years (1987-1996) were characterized by a tight stance with a short period of loose stance in 1989. A long episode of loose stance is observed in 1996-2001 with a subsequent neutral stage (2002-2003). Last Greenspan's years (2003-2005) exhibit a relatively tight stance but policy was increasingly looser since 2004. The stance turns loose when Bernanke's period starts until the outbreak of the Great Recession in 2007:Q4. Since 2008:Q4 the stance turns to be tight again. Finally we observe a reversion in this pattern after the implementation of UMPs, when the stance turns to be relatively loose in 2011-2012. This result is in line with Beckworth (2011), who claims that the Monetary Policy Stance was relatively tight in 2008.

It is important to associate the monetary policy stance with each appointed chairman. In Figure 2.1 we can differentiate episodes of Burns-Miller (until 1979:Q2), Volcker (1979:Q3-1987:Q2), Greenspan (1987:Q3-2005:Q4) and Bernanke (2006:Q1-present).

In addition, we show the same stance but including the NBER recession dates in shaded areas in Figure 2.2. Note that during Volcker's disinflation period (1980-

51

1982), there is a sharp fall in the index as in late 1980 related with a sharp fall in long term intest rates[17]. In general, we observe a that our MPS index falls during recessions.

The contribution of each of the included instruments into the MPS index is presented in Figure 2.3. As expected, the contribution of FFR, Total and Non-Borrowed Reserves is important. The interesting part is related with the level of spreads (SPR), which also exhibits a substantial historical contribution. The latter is simply to intepret, it indicates that expected future interest rates, inflation expectations and credibility play a key role for determining the policy stance. Thus, managing and anchoring expectations has been always a relevant policy instrument[18]. Regarding the issue of managing expectations and spreads, we can collect evidence from Goodfriend and King (2005) (the Volcker's disinflation episode) and the recent implementation of Unconventional Monetary Policies, with the specific focus on Forward Guidance. In addition, Total Reserves (TR) seem to also explain the recent monetary policy stance. Thus, in line with Cúrdia and Woodford (2011), the size of the balance sheet of the Federal Reserve plays an important role, even when the ZLB is binding.

We have shown how each instrument effectively contributes to the evolution of the MPS index. However, it turns out that their relative weights are time-varying, as it is shown in Figure 2.4. The fact that these weights are time varying reflects the overwhelming evidence about changes in operating procedures. One should notice that interest rate spreads always played a significant role, and a role that actually has been increasing over time (see absolute values in panel (a). On the other hand, we find a significant weight of Total and Non-Borrowed reserves (see panels (b) and (c)) in the index.

In line with a textbook approach, a negative weight for FFR_t (panel (c)) means that monetary policy is loose after lowering this rate. It turns out that the weight of the FFR in the monetary policy stance is relatively higher during the Greenspan period (1987-2005). This weight has been decreasing in absolute value since 2005. Note also that the weight for the FFR is close to zero in early 1980s and turns to be statistically insignificant at the end of the sample. These two effects are consistent with the conventional wisdom and a binding ZLB. All in all, the model seems capable of capturing significant changes in operating procedures and this should increase our trust in what it delivers about the transmission of monetary policy shocks.

---

[17]See Goodfriend and King (2005).

[18]Recall that this is the portion of $Spreads_t$ that is orthogonal to the non-policy block.
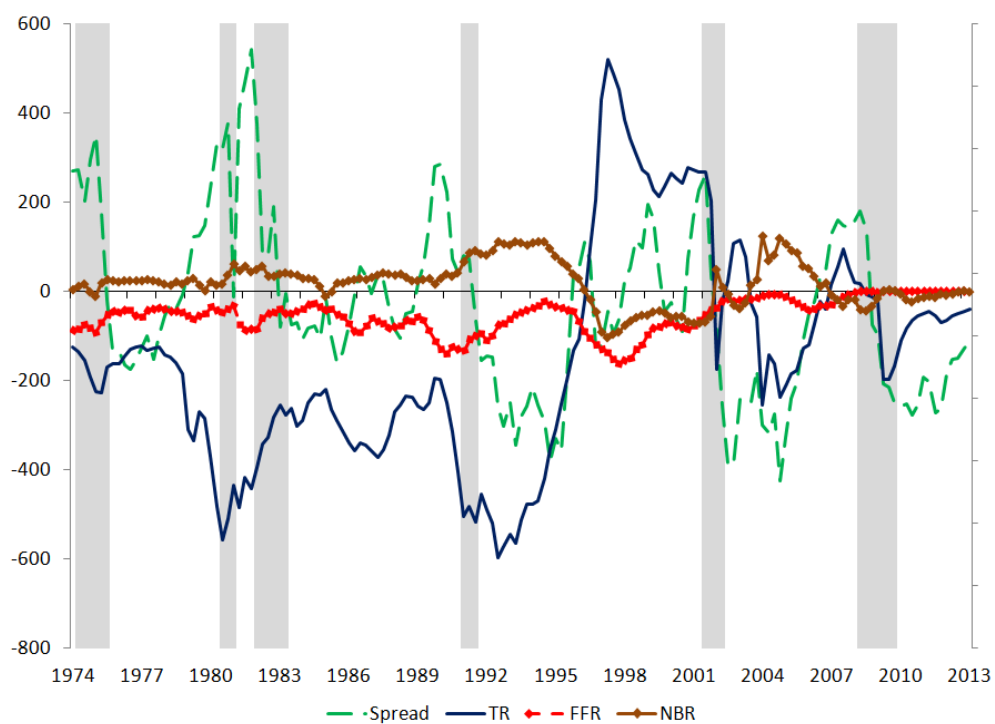
Figure 2.3: Historical Decomposition of Monetary Policy index and NBER recession dates (shaded areas)

(a) SPR

(b) TR

(c) FFR

(d) NBR
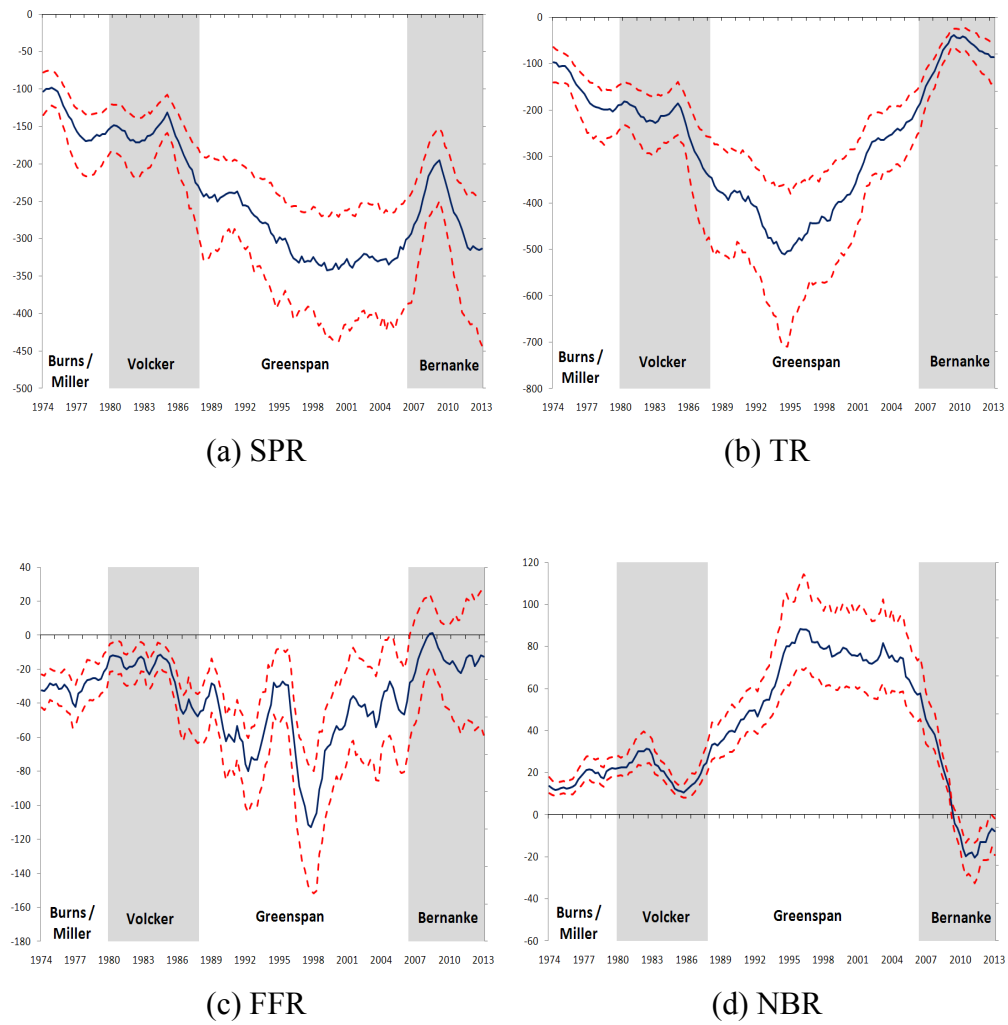
Figure 2.4: Weights of various instruments in the Monetary Policy index, median value and 90 percent confidence bands

2.5 The Transmission Mechanism of Monetary Policy revisited

In this section we explore the transmission of monetary policy shocks (ε_t^s) on the interbank market and the aggregate economy. Since parameters vary continuously, it is possible to trace impulse responses along the time dimension and explore their evolution over time. Let the impulse response function be

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = F_j \left(\{B_i\}_{i=t}^{t+j}, A_t, C_t^{-1}, \Sigma_t \right); \quad j = 0, 1, \ldots \tag{2.20}$$

where $F_j(.)$ depends on the companion form matrix of (2.4) for periods $t, t + 1, \ldots, t + j$ and the blocks A_t, C_t^{-1}, Σ_t and depends on when the shock occurs.In particular:

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = E_t \left[\mathbf{J} \left(\prod_{k=0}^{j-1} \mathbf{A}_{t+j-k}^c \right) \mathbf{J}' H_t \left(\widetilde{\varepsilon}_t - \overline{\varepsilon}_t \right) \right] \tag{2.21}$$

where $H_t = A_t^{-1} C_t \Sigma_t$ and $\mathbf{A}_t^c$ is the companion form matrix of (2.4). Details on the derivation of equation (2.21) are in Appendix B.2.

Figure 2.5 depicts the response of each variable after an expansionary policy shock (ε_t^s) in 1996, a date that we associate with normal times. First, output growth and inflation exhibith a hump-shaped response where the peaks are 6 and 12 quarters after the shock. The sluggish response of Inflation suggests some form of price rigidity. The expansionary policy produces an increase in Total Reserves (TR). Non-Borrowed Reserves (NBR) and Federal Funds Rate (FFR) move in opposite directions after the shock occurs, and this effect disappears in the long run. Finally, this shock produces a positive response in our Spreads indicator, meaning that short term interest rate (FFR) moves faster than long-term ones.

Figure 2.6 depicts the response of each variable after an expansionary policy shock (ε_t^s) in 2012. In comparison to Figure 2.5, it is noticeable that the transmission mechanism of monetary policy has been altered. In particular, the response of inflation is stronger and more persistent and the peak is higher and it is achieved in less than 12 quarters. The response of the Spreads index is negative, meaning that now long-term interest rates move faster than short term interest rate (FFR). As a matter of fact, the response of the FFR is small and almost insignificant, since the target rate is close to zero[19]. The latter result is also in line with Baumeister and Benati (2012), where we can observe that an expansionary policy shock is associated with a compression in spreads and a positive response in output growth and inflation. However, we do not use additional restrictions such as zeroing the coefficients of the interest rate equation for a certain amount of periods. As a matter

---

[19]Responses of other dates are avaiable in Appendix B.1.
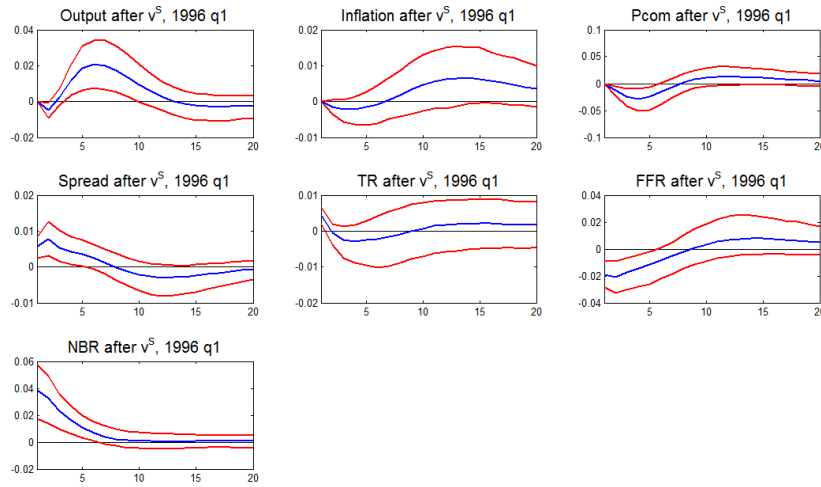
Figure 2.5: Responses to Monetary Policy shocks in 1996, 90 percent confidence interval

of fact, unlike Baumeister and Benati (2012), our identification strategy is more flexible, in the sense that it identifies a pure policy shock, regardless of which is the current instrument used for conducting monetary policy.

Since the monetary transmission mechanism has changed over time, we want to dig into a particular issue of interest, named the *liquidity effect*. According to the literature[20], a temporary increase in money supply introduces a negative reaction of interest rates which vanishes in the long run. This pattern is clear, in particular, when rather than M1 or M2, Non-Borrowed Reserves are included in the SVAR model[21]. Our model possesses this feature and suggests that this pattern is changing over time. Figure 2.7 depicts the evolution of the response of the Federal Funds Rate after a policy shock in Non-Borrowed Reserves of the same size for each date. It is clear that the response is strong up to late 1980s, but it decreases afterwards until the interest rates hits the Zero Lower Bound (ZLB) in 2009.

In sum, there are changes in the transmission of monetary policy shocks which can be associated with changes in the conduct of monetary policy. Note also that our result controls for changes in the private sector behavior, since the parameters of non-policy block can also vary over time, and these changes are correlated with

---

[20]See. e.g. Christiano and Eichenbaum (1991), Strongin (1995), Bernanke and Mihov (1998b), Christiano et al. (1999), among others.

[21]In general, see Bernanke and Mihov (1998a) for a deep analysis of the Liquidity Effect using the model of Bernanke and Mihov (1998b), but using long run restrictions instead.
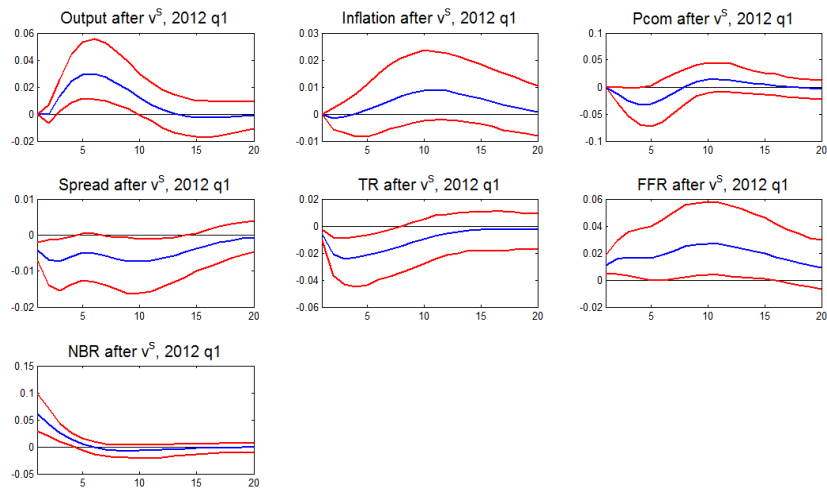
Figure 2.6: Responses to Monetary Policy shocks in 2012, 90 percent confidence interval
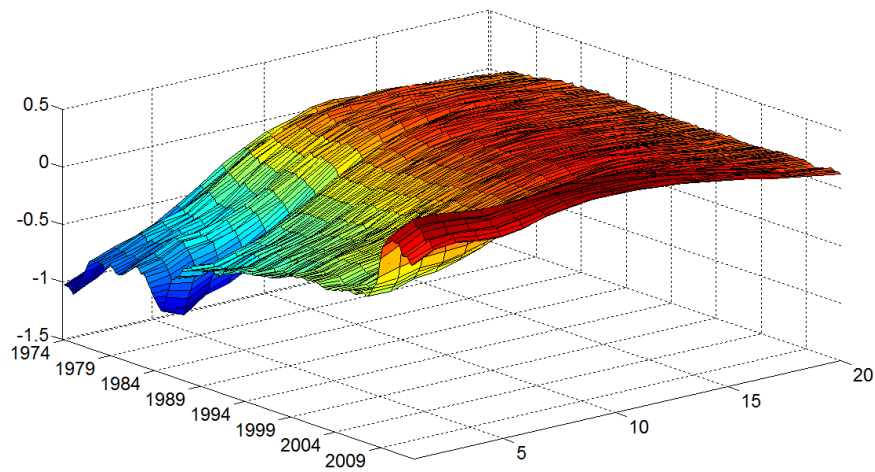
Figure 2.7: Response of the Federal Funds Rate to an expansionary NBR policy shock

57

changes in the policy design.

2.6 The evolution of the Systematic and Non-systematic components of Monetary Policy

The recent literature suggests that monetary policy shocks were more volatile in early 1980s, a date which is associated with changes in the conduct of monetary policy and with the use of non-standard instruments (see Primiceri (2005), Sims and Zha (2006), Justiniano and Primiceri (2008), Canova et al. (2008), Canova and Gambetti (2009) among others). In part this result could be driven by the fact that most of these models only allow for a single monetary policy instrument, the short term interest rate. Since, as we pointed out, we identify monetary policy shocks allowing for different instruments and controlling for changes in the systematic component, we have the chance of controlling for this possibility. The policy rule equation (2.16) is

$$u_t^{NBR} = \phi_t^f v_t^f + \phi_t^d v_t^d + \phi_t^b v_t^b + v_t^s$$

The systematic component is captured by the policy coefficients $\phi_t = \left(\phi_t^f, \phi_t^d, \phi_t^b\right)' = -\widetilde{c}_t$ and the non-systematic one is governed by the shock $v_t^s \sim N\left(0, (\sigma_t^s)^2\right)$. We will capture changes in the systematic component with the evolution of ϕ_t and changes in the non-systematic one with the evolution of σ_t^s.

Figure 2.8 depicts the evolution of vector ϕ_t. There is an increase in the absolute value of ϕ_t^f and ϕ_t^b over time, which can be associated with a stronger reaction of the money suppy to Spreads and Borrowing Reserves shocks, especially during the last decade. On the other hand, ϕ_t^d rises and falls over time. Interestingly, there is an increase in this value starting from 2008, which could be interpreted as an increasing importance of the demand of reserves for implementing monetary policy.

Figure 2.9 shows the evolution of the non-systematic component of monetary policy. In line with the literature, we observe a high level of volatility in 1970s and early 1980s. However, we also observe a second episode of high volatility starting in 2007, an episode related with the Great Financial Crisis. This jump is capturing other dimensions of Unconventional Monetary Policy not related with Reserves and affecting future interest rates, for example Direct Financial Intermediation as in Gertler and Karadi (2011). For a thorough description of different dimensions of UMPs see also Reis (2009), Borio and Disyatat (2010), Cúrdia and Woodford (2011), Cecioni et al. (2011), Williams (2012a).

Since the setup of the model is such that the variances of structural shocks evolve independently, i.e. matrix W is diagonal, we cannot attribute the latter
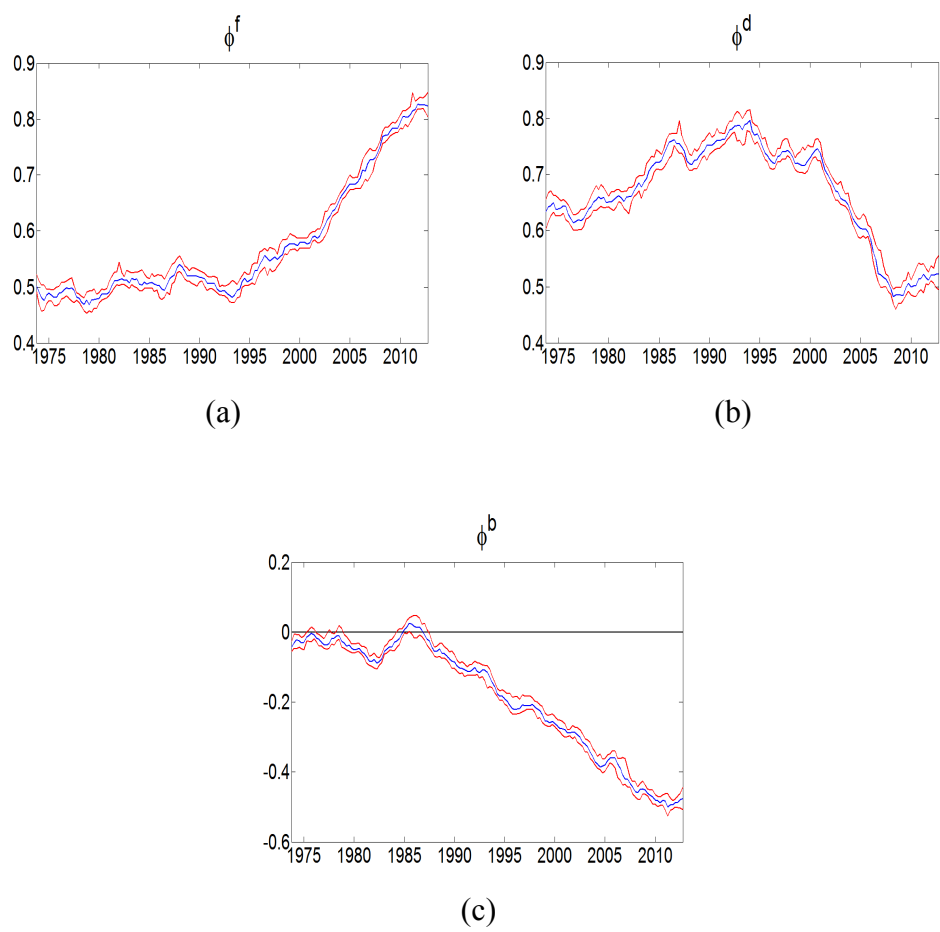
Figure 2.8: Policy rule coefficients, median value and 90 percent bands
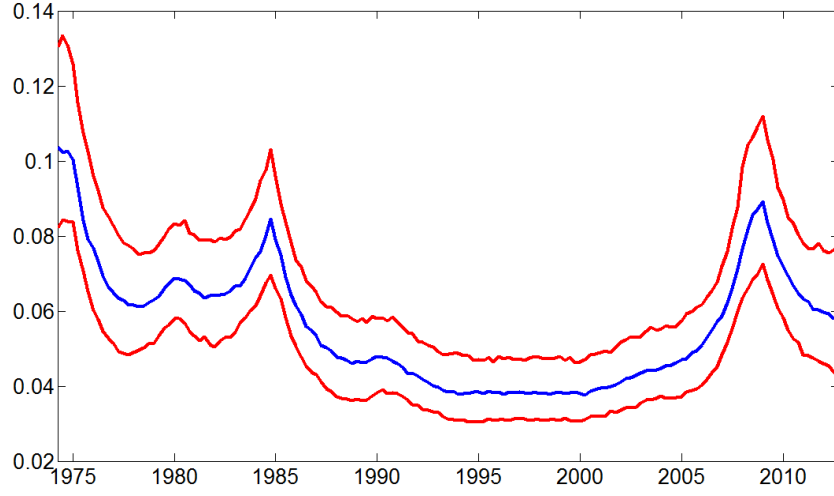
Figure 2.9: Standard Deviation of Policy shock σ_t^s

result to the fact that these shocks were correlated. Furthermore, we are capturing the portion of structural change that can be considered as non-systematic, i.e. different than changes in the operating procedures represented by changes in parameters of matrices A_t and C_t.

2.7 Sensitivity analysis

The results presented so far have important implications for monetary policy. Here we are interested in showing that they are robust to changes in the specification of the empirical model. For comparison, we consider two additional models:

A *Federal Funds Rate* model (*FFR*): In this model we achieve identification by setting coefficient restrictions according to equation (2.17).

Recall the policy stance equation (2.17):

$$
\begin{aligned}
v_t^s \;=\; & -\phi_t^f u_t^{SPR} \\
& - \left(\phi_t^b + \phi_t^d + \phi_t^f \alpha_{TR,t}^{SPR} \right) u_t^{TR} \\
& + \left(\phi_t^b \alpha_{1,t}^b - \phi_t^d \alpha_{1,t}^d - \phi_t^f \alpha_{FFR,t}^{SPR} \right) u_t^{FFR} \\
& + \left(\phi_t^b - \phi_t^f \alpha_{NBR,t}^{SPR} + 1 \right) u_t^{NBR}
\end{aligned}
$$

Here we set parameter restrictions such that the weights associated with instruments different than the Federal Funds Rate are equal to zero. Thus, our identific-

ation restrictions are as follows:

$$\phi_t^f = 0 \tag{2.22}$$

which implies

$$\phi_t^b = -1 \tag{2.23}$$

$$\phi_t^d = -\phi_t^b = 1 \tag{2.24}$$

Thus, if we assume that the Fed uses only FFR for when implementing monetary policy, we will have that policy coefficients in matrix C_t will be constant. As a result, the monetary policy stance will be

$$v_t^s = -\left(\alpha_{1,t}^b + \alpha_{1,t}^d\right) u_t^{FFR}$$

The model is over-identified, since we have imposed 3 restrictions $(2.22)-(2.23)-$ (2.24). As a result, we have the following SVAR matrices:

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{\pi} & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{Pc} & \alpha_{\pi,t}^{Pcom} & 1 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{SPR} & \alpha_{\pi,t}^{SPR} & \alpha_{Pc,t}^{SPR} & 1 & \alpha_{TR,t}^{SPR} & \alpha_{FFR,t}^{SPR} & \alpha_{NBR,t}^{SPR} \\ \alpha_{x,t}^{TR} & \alpha_{\pi,t}^{TR} & \alpha_{Pc,t}^{TR} & 0 & 1 & \alpha_{1,t}^{d} & 0 \\ \alpha_{x,t}^{NBR} & \alpha_{\pi,t}^{NBR} & \alpha_{Pc,t}^{NBR} & 0 & 1 & -\alpha_{1,t}^{b} & -1 \\ \alpha_{x,t}^{FFR} & \alpha_{\pi,t}^{FFR} & \alpha_{Pc,t}^{FFR} & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.25}$$

$$C_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix} \tag{2.26}$$

Clearly, in this particular case vectors α_t and c_t enter linearly in matrices $(2.25) -$ (2.26). Therefore, we can sample the vector α_t as in section 2.3 and as it is clear, it will not be necessary to sample c_t.

A *Bernanke and Mihov* model (*BM*): This is a model without the spread indicator, so that:

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{\pi} & 1 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{c} & \alpha_{\pi,t}^{c} & 1 & 0 & 0 & 0 \\ \alpha_{x,t}^{TR} & \alpha_{\pi,t}^{TR} & \alpha_{c,t}^{TR} & 1 & \alpha_{1,t}^{d} & 0 \\ \alpha_{x,t}^{FFR} & \alpha_{\pi,t}^{FFR} & \alpha_{c,t}^{FFR} & 1 & -\alpha_{1,t}^{b} & -1 \\ \alpha_{x,t}^{NBR} & \alpha_{\pi,t}^{NBR} & \alpha_{c,t}^{NBR} & 0 & 0 & 1 \end{bmatrix} \tag{2.27}$$

61

$$C_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \phi_t^d & \phi_t^b & 1 \end{bmatrix} \qquad (2.28)$$

As a result, the monetary policy stance turns to be:

$$v_t^s = - \left(\phi_t^b + \phi_t^d \right) u_t^{TR} + \left(\phi_t^b \alpha_{1,t}^b - \phi_t^d \alpha_{1,t}^d \right) u_t^{FFR} + \left(\phi_t^b + 1 \right) u_t^{NBR}$$

Here the sub-system of policy equations needs one restriction to be identified. Following the original model, we set $\alpha_{1,t}^d = 0$, i.e. the so-called *just-identified* model, so that:

$$v_t^s = - \left(\phi_t^b + \phi_t^d \right) u_t^{TR} + \left(\phi_t^b \alpha_{1,t}^b \right) u_t^{FFR} + \left(\phi_t^b + 1 \right) u_t^{NBR}$$

We also label our specification, for comparison purposes, as *Baseline* and we now proceed to the model comparison in next subsection.

Bernanke and Mihov (1998b) analyze the possibility of regime switches in the conduct of policy. More specifically, they allow policy coefficients under their baseline (just-identified) specification to vary according to a hidden discrete latent variable that takes two values (see Hamilton (1989) and Kim and Nelson (1999)). They find strong evidence of switches for early eighties (see Table I in pp. 883 and Figure I in pp. 891). However, they do not allow shock variances or non-policy parameters to change, and this may bias the conclusions. Using the flexible approach of this paper, it is possible to pin down the evolution of these coefficients over time and, at the same time, control for stochastic volatility and policy changes. Indeed, we find strong evidence of changes in policy rule coefficients even in this case. However, as shown in Figures 2.10 and 2.11, the resulting path could be potentially different if we consider a richer specification such as *Baseline*. Regarding Figure 2.10, Baseline model (panel a) exhibits larger fluctuations relative to panel (b). In particular, the changes are stronger for the last decade, in particular starting by 2005.

Regarding Figure 2.11, we observe even a change in the sign of the coefficient, but dates of changes are very different across models.

Furthermore, the extent of regime switches depends also on how the model is specified and identified. In particular, our *Baseline* has one more instrument in the policy rule and also has one more variable in the SVAR, and thus a larger information set.

Finally, we explore whether the results concerning the variance of the policy shocks depend on the specification of the model.
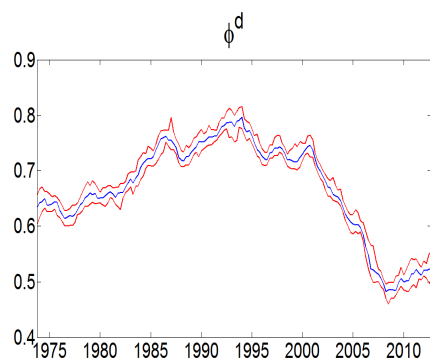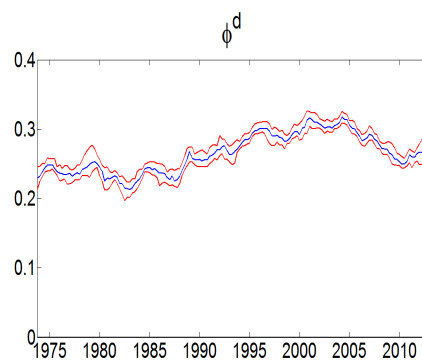
Figure 2.10: Sensitivity to Demand shocks ϕ_t^d
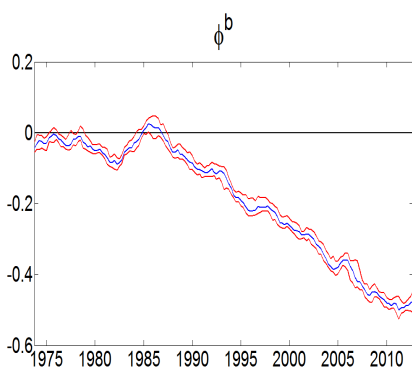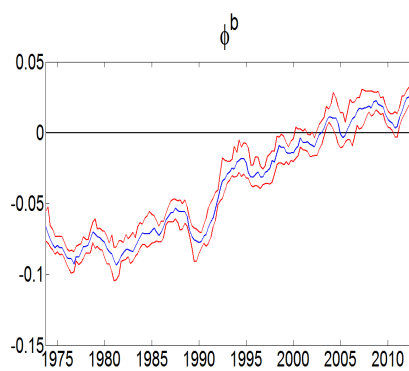
Figure 2.11: Sensitivity to Discount window shocks ϕ_t^b

(a) Baseline

(b) FFR

(c) BM
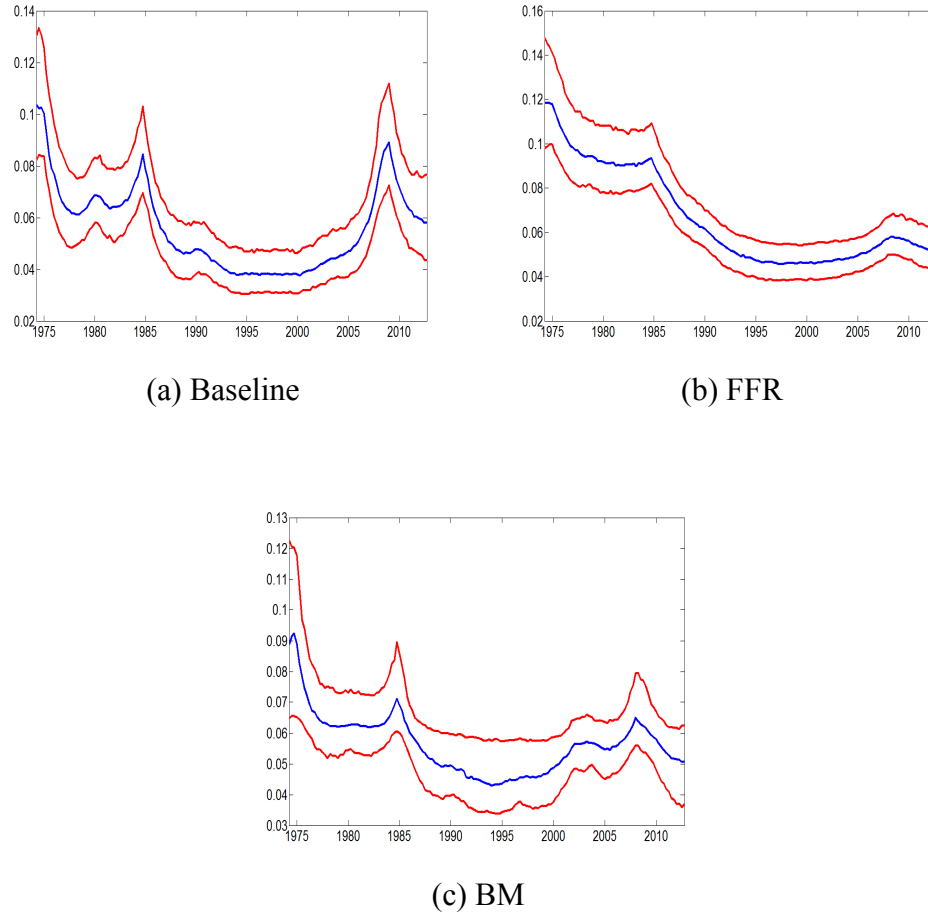
Figure 2.12: Comparison of Standard Deviation of Policy shock σ_t^s

The volatility of the monetary policy shock is higher in 1970s and 1980s and again during the recent Financial Crisis episode. However, as it is depicted in Figure 2.12, the pattern is similar with the other specifications, but the magnitude of the changes is different. In particular, in panels (b) and (c) the recent financial crisis seem to have induced a much smaller increase in the volatility of policy shocks.

2.8 Concluding Remarks

We have estimated the monetary policy stance index for the U.S. economy for the period 1974-2012 and taking into account the time-varying operating procedures.

To the best of our knowledge, this is the first paper that presents a policy stance index with bayesian error bands. Moreover, we identify periods of loose and tight monetary policy, in line with previous literature, and we also quantify the relative importance of instruments in this process.

We find evidence of changes in the monetary transmission mechanism after the financial crisis, and we consider this results important for future research. We also document that the *Vanishing Liquidity Effect* has been broken after the last financial crisis. Furthermore, the described results are a consequence of changes in both systematic and non-systematic components.

Our results are robust to alternative specifications. We find that the paths of structural parameters and variances might differ across models. However, the main features of our result do not change. Overall, this paper presents a powerful approach that is capable of capturing different episodes regarding monetary policy design, especially the so-called Unconventional Monetary Policies. Even more important is the fact that we present here a monetary policy stance index with bayesian error bands, which we expect to be useful for both policy makers and researchers. It remains to be explored the role of communication in Unconventional Monetary Policies, the announcement of future paths of interest rates and monetary authority reputation. We believe that these type of issues should be explored in a richer setup and we leave it for future agenda.

Chapter 3

HETEROGENEOUS INFORMATION AND REGIME SWITCHES IN A STRUCTURAL EXCHANGE RATE MODEL: EVIDENCE FROM SURVEY DATA

3.1   Introduction

Survey Data on Exchange Rates forecasts exhibits a considerable amount of disagreement among the poll's participants.This mentioned dispersion is not constant over time, exhibiting substantial and persistent variation across time as shown in Figure 3.1. What explains this time varying dispersion? More concretely, what determines the fluctuation of this disagreement that is inherent to heterogeneous expectations over time? To answer these questions, we need to first dig deeply into how is the expectation formation in the context of heterogeneous agents and, how are these expectations connected to fundamentals and actual exchange rates. To do so, we need to model the decision of economic agents based on what they observe and how they process information. We present an extension of the theoretical model of exchange rate determination with rational but disparately informed agents by Bacchetta and Van Wincoop (2006), and we perform an empirical exercise using actual Survey Data, with the purpose of fitting the observed disagreement.

Disagreement suggests that poll's participants have access to different pieces of information about future fundamentals, henceforth signals. Consequently, each rational agent solves a signal extraction problem, and it is clear that signals' pre-
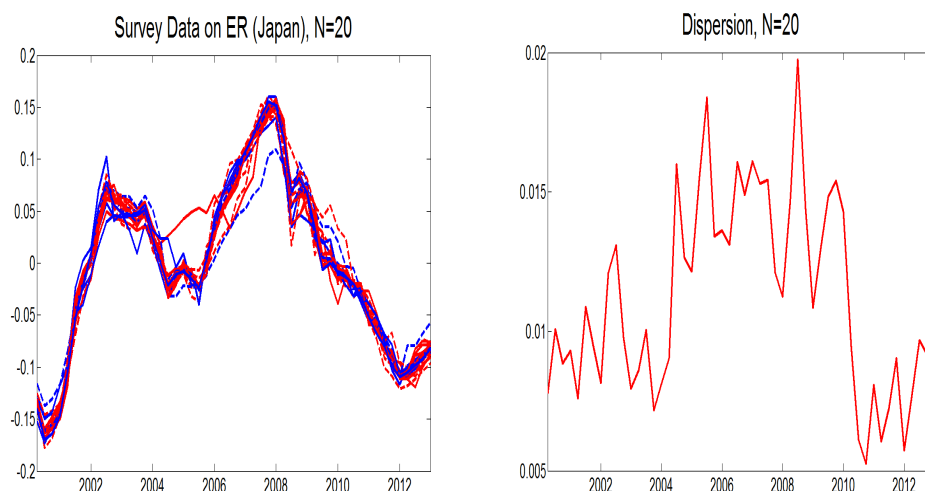
Figure 3.1: TANKAN Survey (BoJ): Log-Predicted Exchange Rates by sectors (next quarter)

cision play an important role. However, this is not enough for explaining the observed pattern in the data. Therefore, to allow for time varying dispersion we need to consider time-varying information sets. We do so considering a flexible approach that accounts for regime switches in volatilities of fundamentals and in the amount of information available for agents.

Bacchetta and Van Wincoop (2006) argues that the *Exchange Rate determination puzzle* (Meese and Rogoff (1983)) can be resolved by allowing for the presence of heterogeneously informed agents, forcing agents to form the so-called Higher-order expectations (HOE). Bacchetta and van Wincoop's argument is as follows: Information heterogeneity matters, since each trader has a portion of private information about aggregate fundamentals, and the equilibrium exchange rate is determined by future values of these aggregate fundamentals, which are not observable. As a result, agents need to form expectations about aggregate fundamentals, which means that they need to forecast the forecasts of the others. Because of information heterogeneity, expectations about future fundamentals (HOE) also react to aggregate shocks. But agents may confound one shock with another, so expectations of other fundamentals also react, generating endogenous persistence in Exchange Rates reflected in hump-shaped responses to shocks. This type of mechanism was first called Rational confusion by Bacchetta and Van Wincoop (2006). In this context of dispersed information, we introduce regime switches in the volatility of fundamentals and in information sets as in Nimark (2012a) and estimate the model with bayesian methods following Barillas et al. (2013). We use the TANKAN Survey from the Bank of Japan, which in

68

particular has heterogeneous forecasts of exchange rates across industries.

The model in this paper is connected with a large strand of the empirical literature of Exchange Rate determination. Starting from Meese and Rogoff (1983), today researchers often model exchange rates as an asset price, e.g. Engel and West (2005), Rossi (2005), Engel et al. (2010). That type of modelling allows us to consider the case where expectations of future fundamentals matter for the actual exchange rate. Our framework is also connected with the parameter instability and regime-switching in Exchange Rate models. In these papers it is argued that the relationship between exchange rates and fundamentals is unstable, starting with Wolff (1987), and also Engel and Hamilton (1990), Canova (1993), Kaminsky (1993), Engel (1994), Evans and Lewis (1995), Rossi (2006), Bacchetta and Van Wincoop (2009), Bacchetta et al. (2010), among others. Regarding previous work using Survey Data, disagreement in Survey Data at different horizons has been covered by Patton and Timmermann (2010) for the specific case of Term-Structure Interest Rates. Bacchetta et al. (2009) use survey of expectations for studying the excess return predictability in Financial Markets.

We fit a structural model with regime switches to Exchange Rates Survey Data. The assumed information structure is such that high-volatility regimes are associated with the appearance of a very noisy public signal about fundamentals. Given that this signal is very imprecise, and because of the higher volatility of shocks, disagreement increases. The model-implied dispersion closely follows the observed disagreement, meaning that the presented model does a good job in describing our dataset. We confirm the latter when comparing the model fit with respect to a restricted model without regimes switches and without informational frictions. Given the presence of regime switches, the model solution implies a time varying relationship between exchange rates and fundamentals. This result can be linked with the empirical literature of parameter instability in exchange rate models. The latter is, together with Rational Confusion, an additional explanation for the disconnection between Exchange Rates and future fundamentals. We leave as future agenda the idea to explore several surveys of Exchange Rate expectations about different currencies, such as Consensus Economics or Bloomberg.

The paper is organized as follows: Section 3.2 explains the extent of Heterogeneous Information in the context of Asset Pricing models, Section 3.3 describes the Exchange Rate model used for the analysis, Section 3.4 covers the empirical analysis, Section 3.5 discusses the main results and Section 3.6 concludes.

3.2   Asset Prices and Heterogeneous Information

According to Engel and West (2005), a good starting point for studying exchange rates dynamics is to consider an asset pricing model. In this section we describe

the extent of heterogeneous information in the context of asset pricing models. First, let the price euler equation be:

$$p_t = a \int E_t^i p_{t+1} di + f_t \qquad (3.1)$$

where p_t is the price at time t, f_t is the fundamental that determines the price and a is a positive scalar. The operator E_t^i indicates that each agent makes forecasts conditional on his available information set $\Omega_t(i)$, so that

$$E_t^i = E\left[. \mid \Omega_t(i)\right] \qquad (3.2)$$

Under full or common information, an asset price with common information is solved by exploiting the 'Law of Iterated Expectations' and ruling out bubbles, since $a \in (0,1)$:

$$p_t = a \int E_t^i \left(a \left(\int E_{t+1}^i p_{t+2} di + f_{t+1} \right) \right) di + f_t$$

$$p_t = a^2 \int E_t^i p_{t+2} di + a E_t f_{t+1} di + f_t$$

$$\dots$$

$$p_t = \sum_{k=0}^{\infty} a^k E_t f_{t+k} = \frac{1}{1-a} f_t \qquad (3.3)$$

where E_t is the average expectation operator and where the equilibrium price is a result of the fact that agents have nested information sets.

Turning to heterogeneous information, according to Admati (1985), the equilibrium price (p_t) is a result of aggregation of information. That is, each market participant $i \in (0,1)$ has a different piece of noisy information that is private (I_i). When each market participant observes the equilibrium price (p_t), he can learn more about the information of the others (I_{-i}), but since this learning is not perfect, disagreement in forecasts will be present. Moreover, if we take heterogeneous information seriously, then we need to reconsider the future expectations of f_t, and find an expression equivalent to (3.3). First, if agents have different information sets, then their forecasts will be different so that:

$$E_t^i f_{t+1} \neq E_t^{i'} f_{t+1}$$

where $(i \neq i')$ but there exists an average forecast $f_{t+1}^{(1)} \equiv \int E_t^i f_{t+1} di$. Going back to euler equation (3.1), then

$$p_t = a^2 \int E_t^i \left(\int E_{t+1}^i p_{t+2} di \right) di + a \int E_t^i f_{t+1} di + f_t$$

$$\cdots$$

$$p_t = \sum_{k=0}^{\infty} a^k f_{t+k|\cdots|t}^{(k)} \tag{3.4}$$

In general, we define the higher-order expectations as:

$$f_{t+k}^{(k)} \equiv \int E_t^i f_{t+k-1}^{(k-1)} di \tag{3.5}$$

The latter implies that, in order to form the k-th order of expectations, agents need to 'Forecast the Forecasts of the others' (as in Townsend (1983)) up to a level $k-1$. By observing (3.4) and (3.5), it is clear that higher-order expectations of the fundamental f_t are an important determinant of the equilibrium price and, as a consequence, they constitute additional state variables. See Allen et al. (2006) for an extensive description of the role of higher-order expectations in asset prices.

The richer setup comes at the cost of additional technical complications in order to find a model solution. In particular, as pointed out by Townsend (1983) and Sargent (1991), the dimension of the state vector can be infinite, and therefore the model turns to be difficult to solve without assuming lagged revelation, as in Singleton (1987) or Bacchetta and Van Wincoop (2006). On the other hand, Kasa (2000) and Kasa et al. (2007) use frequency domain techniques in order to solve this type of models. Furthermore, Nimark (2011) shows how to solve these type of models in a dynamic setting without assuming lagged shock revelation and, in addition Nimark (2012a) and Barillas et al. (2013) extend Nimark (2011)'s procedure to allow for regime switches. In this paper we will follow these last references in order to solve and estimate an exchange rates model and fit the observed dispersion in the data.

3.3 The model

3.3.1 Benchmark setup

The model is taken from Bacchetta and Van Wincoop (2006). This is a two-country economy where the Purchasing Power Parity (PPP) holds:

$$p_t = p_t^* + s_t \tag{3.6}$$

where p_t is the log of price index of home economy, p_t^* is the same for the foreign economy and s_t is the log of exchange rate. Each economy is populated by a continuum of investors $i \in [0,1]$, they live two periods but can only invest in the first one, i.e. there are overlapping generations. When each investor i dies, he passes his information to his offspring. Each investor i has the possibility

to invest in three types of assets: (i) domestic currency (m_t), (ii) foreign bonds (b^i_{Ft}), (iii) technology fixed real return r (inifinite supply). Domestic and foreign money supplies (m_t, m^*_t) are deterministic and stochastic, respectively. On the other hand, wealth (w^i_t) is a fixed endowment. Production depends on exchange rate (s_t) and real money holdings $(\widetilde{m}_t = m_t - p_t)$, so that:

$$y^i_{t+1} = b^i_t s_{t+1} - \frac{\widetilde{m}^i_t \left[\log\left(\widetilde{m}^i_t\right) - 1\right]}{\alpha} \tag{3.7}$$

where $\alpha > 0$ and b^i_t is the privately observed exchange rate exposure of nonasset income of investor i. Following Bacchetta and Van Wincoop (2006), the fact that $\widetilde{m}^i_t$ is in the production function and not in the utility means that we do not have to specify a money demand.

Each investor i has a Constant Absolute Risk-Aversion (CARA) concave utility and solves the problem:

$$\max_{\left\{\widetilde{m}^i_t, b^i_{Ft}\right\}} - E\left[\exp\left\{-\gamma c^i_{t+1}\right\} \mid \Omega_t\left(i\right)\right]$$

subject to

$$c^i_{t+1} = (1 + i_t) w^i_t + (s_{t+1} - s_t + i^*_t - i_t) b^i_{Ft} - i_t \widetilde{m}^i_t + y^i_{t+1} \tag{3.8}$$

where $\Omega_t\left(i\right)$ is the information set of inverstor i at time t and γ is the risk aversion coefficient. We assume that future consumption c^i_{t+1} is normally distributed conditional on the available information, so that

$$c^i_{t+1|\Omega_t(i)} \sim N\left(\mu, \sigma^2_c\right)$$

where

$$\mu \equiv E\left(c^i_{t+1} \mid \Omega_t\left(i\right)\right) \tag{3.9}$$

$$\sigma^2_c \equiv var\left(c^i_{t+1} \mid \Omega_t\left(i\right)\right) \tag{3.10}$$

Exploiting the CARA utility and the normality assumption, we rewrite the utility function:

$$-E^i_t \exp\left\{-\gamma c^i_{t+1}\right\} = -\left[\exp\left\{-\gamma\mu + \frac{\gamma^2}{2}\sigma^2_c\right\}\right] \tag{3.11}$$

where:

$$\mu = (1 + i_t) w^i_t + \left(E^i_t s_{t+1} - s_t + i^*_t - i_t\right) b^i_{Ft} - i_t \widetilde{m}^i_t + E^i_t y^i_{t+1}$$

with

$$E^i_t y^i_{t+1} = b^i_t E^i_t s_{t+1} - \frac{\widetilde{m}^i_t \left[\log\left(\widetilde{m}^i_t\right) - 1\right]}{\alpha}$$

72

and

$$\sigma_c^2 = \left(b_{Ft}^i + b_t^i\right)^2 \sigma_t^2$$

because b_t^i is observed and $\sigma_t^2 \equiv var\left(s_{t+1}^i \mid \Omega_t\left(i\right)\right)$. The latter result comes from the fact that the stochastic components of c_{t+1}^i in (3.8) is the future exchange rate s_{t+1}.

The first order conditions of this problem are:

$$\widetilde{m}_t^i : \quad -\alpha i_t = \log\left(\widetilde{m}_t^i\right) \tag{3.12}$$

which on aggregate means

$$m_t - p_t = -\alpha i_t$$

and it is also symmetric for foreign investors, so that:

$$m_t^* - p_t^* + \sigma_{mp}\varepsilon_t^{mp} = -\alpha i_t^*$$

We introduce the shock ε_t^{mp} to rule out a perfect revealing equilibrium.

The optimal condition for bonds is:

$$b_{Ft}^i : \quad b_{Ft}^i = \frac{E_t^i s_{t+1} - s_t + i_t^* - i_t}{\gamma \sigma_t^2} - b_t^i$$

Intuitively, since this is a risky asset, both a higher risk aversion coefficient γ and higher volatility in forecast errors σ_t^2 can reduce the demand, but a higher expected value of returns $E_t^i s_{t+1}$ increases it. Finally, the demand is negatively related with the individual nonasset income exposure to foreign currency b_t^i.

The complete monetary model of Exchange Rates as in Bacchetta and Van Wincoop (2006) is given by:

$$p_t = p_t^* + s_t \tag{3.13}$$

$$m_t - p_t = -\alpha i_t \tag{3.14}$$

$$m_t^* - p_t^* + \sigma_{mp}\varepsilon_t^{mp} = -\alpha i_t^* \tag{3.15}$$

$$b_{Ft}^i = \frac{E_t^i\left(s_{t+1}\right) - s_t + i_t^* - i_t}{\gamma \sigma_t^2} - b_t^i \tag{3.16}$$

The market clearing condition is

$$\int_0^1 b_{Ft}^i di = \sigma_\eta \eta_t$$

where $\eta_t \sim N\left(0, 1\right)$. The fact that there is a noisy component in the supply of bonds also prevents a perfectly revealing rational expectations equilibrium; see also Admati (1985) and Vives (2010). Aggregating the equation (3.16):

$$\sigma_\eta \eta_t = \frac{\overline{E}_t\left(s_{t+1}\right) - s_t + i_t^* - i_t}{\gamma \sigma_t^2} - b_t$$

we get an expression similar to the Uncovered Interest Rate Parity (UIRP)

$$s_t = \overline{E}_t \left(s_{t+1} \right) - \left(i_t - i_t^* \right) - \gamma \sigma_t^2 \left(b_t + \sigma_\eta \eta_t \right) \tag{3.17}$$

where $\overline{E}_t \left(s_{t+1} \right) = \int_0^1 E \left(s_{t+1} \mid \Omega_t \left(i \right) \right) di$ and where the aggregate exposure component $\int_0^1 b_t^i di = b_t$ is not observable as is going to be a relevant aggregate fundamental. As it is clear, the latter expression has an explicit time-varying risk-premium component $-\gamma \sigma_t^2 \left(b_t + \sigma_\eta \eta_t \right)$, which will be important for our empirical analysis. Unlike Bacchetta and Van Wincoop (2006), we allow the component σ_t^2 to vary over time.

The equilibrium condition can be simplified. Working with the conditions $(3.14) - (3.15)$, we get:

$$i_t - i_t^* = \frac{s_t - f_t + \sigma_{mp} \varepsilon_t^{mp}}{\alpha} \tag{3.18}$$

where we define the monetary fundamental

$$f_t \equiv m_t - m_t^*$$

as the log of the ratio of money supplies. Plugging the interest rate differential in (3.17)

$$s_t = \frac{\alpha}{1+\alpha} \overline{E}_t \left(s_{t+1} \right) + \frac{f_t - \alpha \gamma \sigma_t^2 b_t}{1+\alpha} - \frac{\alpha \gamma \sigma_t^2 \sigma_\eta \eta_t + \sigma_{mp} \varepsilon_t^{mp}}{1+\alpha} \tag{3.19}$$

we get the equilibrium condition of the model. It states that the log of the exchange rate is a linear function of its average expectation, a linear combination of aggregate fundamentals and a linear combination of shocks. As it is clear from this expression, the presence of the shocks η_t and ε_t^{mp} rules out the possibility of a perfect revealing equilibrium. Moreover, the model in this form needs to be solved. The equilibrium condition can be re-written similar to equation (3.4) using the definition in (3.5):

$$s_t = \frac{1}{1+\alpha} \sum_{k=0}^{\infty} \left[\frac{\alpha}{1+\alpha} \right]^k \left(f_{t+k|\cdots|t}^{(k)} - \alpha \gamma \sigma_t^2 b_{t+k|\cdots|t}^{(k)} \right) - \frac{1}{1+\alpha} \left(\sigma_{mp} \varepsilon_t^{mp} + \alpha \gamma \sigma_t^2 \sigma_\eta \eta_t \right) \tag{3.20}$$

where we assume that there is no bubble solution, i.e.

$$\lim_{k \to \infty} \left[\frac{\alpha}{1+\alpha} \right]^k s_{t+k|\cdots|t}^{(k)} = 0 \tag{3.21}$$

The equation (3.20) makes explicit that the equilibrium exchange rate s_t has to be expressed as a function of the state variables of the model, i.e. fundamentals plus

higher-order expectations $\left\{ f_{t+k}^{(k)}, b_{t+k}^{(k)} \right\}_{k=0}^{\infty}$ and shocks $(\varepsilon_t^{mp}, \eta_t)$, i.e. a particular case of equation (3.4). In what follows we will use the strategy of Nimark (2011) to solve this model. In the mentioned paper (see section 8), the model-implied cross sectional dispersion will be a constant non-linear function of parameters. This feature is not desirable given our dataset in Figure 3.1. Therefore, with the purpose of getting time variation in this cross sectional dispersion, we introduce regime switches in information sets $\Omega_t(i)$ in the next subsection.

3.3.2 Introducing Regime Switches

Assume now that the economy will have two regimes and that these regimes are governed by an underlying dichotomic latent variable τ_t. Then assume that the information available for each individual is related with the current regime of the economy, so that $\Omega_t(i)$ can take two values accordingly:

$$\Omega_t(i) = \left\{ \begin{array}{ll} \Omega_t^0(i) & \text{, if } \tau_t = 0 \\ \Omega_t^1(i) & \text{, if } \tau_t = 1 \end{array} \right.$$

where

$$\Pr(\tau_t = 1) = \omega; \ \ t = 1, 2, \ldots, T$$

is the probability of being in regime $\tau_t = 1$. This is a simplified version of the widely used Regime-Switching models based on a hidden markov chain, see Hamilton (1989). As a result, probabilities of each regime are constant over time, i.e.

$$\Pr(\tau_t = 1 \mid \tau_{t-1} = 0) = \Pr(\tau_t = 1 \mid \tau_{t-1} = 1) = \omega$$

meaning that the event $\tau_t = 1$ does not depend on past values of τ. The implication of time-varying information sets is that it is likely that it will impact on the model implied dispersion. Therefore, we consider that this is the minimal ingredient that the presented structural model needs for that purpose, and turning to a markov-switching setup will be a generalization with a consequent additional persistence in regimes. On the other hand, one important reason for keeping our approach is the fact that, given the pattern observed in Figure 3.1, jumps in dispersion do not seem to be persistent.

Turning to the model implications of the introduced regime-switches, recall that before c_{t+1}^i was normally distributed with a well defined mean μ and variance σ_c^2 given by (3.9) and (3.10). The latter moments were constructed conditional on the information set $\Omega_t(i)$. Thus, the resulting expression in (3.11) was straightforward. With regime switches the result is different. Denote the conditional

75

moments for $j = 0, 1$:

$$\mu_j = E_{t,j}^i c_{t+1}^i = E\left(c_{t+1}^i \mid \Omega_t^j\left(i\right)\right)$$
$$\sigma_{c,j}^2 = var_{t,j}^i c_{t+1}^i = var\left(c_{t+1}^i \mid \Omega_t^j\left(i\right)\right)$$

Strictly speaking, c_{t+1}^i is now following a mixture of normals:

$$c_{t+1}^i \sim \omega N\left(\mu_1, \sigma_{c,1}^2\right) + \left(1 - \omega\right) N\left(\mu_0, \sigma_{c,0}^2\right)$$

with mean

$$\mu = \omega\mu_1 + \left(1 - \omega\right)\mu_0 \tag{3.22}$$

and variance

$$\sigma_c^2 = \omega\sigma_{c,1}^2 + \left(1 - \omega\right)\sigma_{c,0}^2 + \omega\left(1 - \omega\right)\left(\mu_1 - \mu_0\right)^2 \tag{3.23}$$

The expected utility will be given by:

$$E\left[-\exp\left(-\gamma c_{t+1}^i\right)\right] = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right]$$
$$+ \left(1 - \omega\right)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$

That is, the technical complication comes from the fact that individuals take into account the possibility of regime switches for next period. For tractability, following Yang (2011), we use a second-order approximation of the utility function[1]:

$$E\left[-\exp\left(-\gamma c_{t+1}^i\right)\right] \approx -\left[\exp\left\{-\gamma\mu + \frac{\gamma^2}{2}\sigma_c^2\right\}\right] \tag{3.24}$$

with μ and σ_c^2 defined in $(3.22) - (3.23)$. This approximation resembles equation (3.11) but with the additional feature of the mixture of normals. As a result, the optimality conditions using (3.24) are now:

$$m_t - p_t = -\alpha i_t$$

$$m_t^* - p_t^* + \sigma_{mp}\varepsilon_t^{mp} = -\alpha i_t^*$$

$$b_{Ft}^i = \frac{E_t^i\left(s_{t+1}\right) - s_t + i_t^* - i_t}{\gamma\sigma_t^2} - b_t^i$$

with

$$E_t^i\left(s_{t+1}\right) = \omega E_t^i\left(s_{t+1}^1\right) + \left(1 - \omega\right)E_t^i\left(s_{t+1}^0\right) \tag{3.25}$$

$$\sigma_t^2 = \omega\sigma_{t,1}^2 + \left(1 - \omega\right)\sigma_{t,0}^2 + \omega\left(1 - \omega\right)\left[E_t^i\left(s_{t+1}^1\right) - E_t^i\left(s_{t+1}^0\right)\right]^2 \tag{3.26}$$

[1]See details in Appendix C.4

so that we get similar expressions to (3.19):

$$s_t = \frac{\alpha}{1+\alpha} \overline{E}_t (s_{t+1}) + \frac{f_t - \alpha\gamma\sigma_t^2 b_t}{1+\alpha} - \frac{\alpha\gamma\sigma_t^2\sigma_\eta\eta_t + \sigma_{mp}\varepsilon_t^{mp}}{1+\alpha}$$

and (3.20):

$$s_t = \frac{1}{1+\alpha} \sum_{k=0}^{\infty} \left[\frac{\alpha}{1+\alpha}\right]^k \left(f_{t+k|\cdots|t}^{(k)} - \alpha\gamma\sigma_t^2 b_{t+k|\cdots|t}^{(k)} \right) - \frac{1}{1+\alpha} \left(\sigma_{mp}\varepsilon_t^{mp} + \alpha\gamma\sigma_t^2\sigma_\eta\eta_t \right)$$

Notice that now it is clear that σ_t^2 is time varying depending on the regimes τ_t. Having these regimes explicit, we turn to the information structure and model solution.

3.3.3 Information Structure

In order to pin down a solution for equation (3.20), it is necessary to make additional assumptions regarding the information sets $\Omega_t(i)$. First, denote the vector $\mathbf{x}_t$ as the fundamentals and assume that each one follows an autoregressive process AR(1):

$$\mathbf{x}_t \equiv \left[\begin{array}{c} f_t \\ b_t \end{array} \right] = \left[\begin{array}{cc} \rho_f & 0 \\ 0 & \rho_b \end{array} \right] \left[\begin{array}{c} f_{t-1} \\ b_{t-1} \end{array} \right] + \left[\begin{array}{cc} \sigma_{f,t} & 0 \\ 0 & \sigma_b \end{array} \right] \left[\begin{array}{c} \varepsilon_t^f \\ \varepsilon_t^b \end{array} \right] \tag{3.27}$$

This structure is known for every agent and moreover we have that

$$\sigma_{f,t} = (1 - \tau_t)\sigma_f + \tau_t\sigma_{f_2}, \sigma_{f_2} > \sigma_f$$

That is, we link regimes with states of low ($\tau_t = 0$) and high ($\tau_t = 1$) volatility of fundamental f_t. In addition, whenever the economy is in regime ($\tau_t = 1$), all investors observe the public signal

$$\mathbf{z}_t^p = f_t + \sigma_p v_t^p$$

The availability of a public signal plus being in a state of high volatility will impact directly on the model implied dispersion, and this will be time varying. Intuitively, in more unstable regimes e.g. a financial turbulence episode, we observe more sources of information such as news, media releases, special reports, etc. These additional sources of information about fundamentals can potentially be very imprecise. See Nimark (2012a) for the specific case of macroeconomic news.

Dispersion comes from the fact that each agent i observes a private signal at time t:

$$\mathbf{z}_t(i) = \mathbf{x}_t + \left[\begin{array}{cc} \widetilde{\sigma}_f & 0 \\ 0 & \widetilde{\sigma}_b \end{array} \right] \left[\begin{array}{c} v_t^f(i) \\ v_t^b(i) \end{array} \right] \tag{3.28}$$

and finally, all investors also observe the realized price s_t.

As a result, the information set of agent i exhibits time variation:

$$\Omega_t\left(i\right) = \begin{cases} \Omega_t^0\left(i\right) = \left\{\mathbf{z}_t\left(i\right), s_t, \tau_t, \Omega_{t-1}\left(i\right)\right\} & \text{, if } \tau_t = 0 \\ \Omega_t^1\left(i\right) = \left\{\mathbf{z}_t\left(i\right), \mathbf{z}_t^p, s_t, \tau_t, \Omega_{t-1}\left(i\right)\right\} & \text{, if } \tau_t = 1 \end{cases} \qquad (3.29)$$

3.3.4 Solving the model

In this section we closely follow the reference of Barillas et al. (2013). See further details in Appendix C.5. Recall the exchange rate equilibrium equation (3.19):

$$s_t = \frac{\alpha}{1+\alpha}\overline{E}_t\left(s_{t+1}\right) + \frac{f_t - \alpha\gamma\sigma_t^2 b_t}{1+\alpha} - \frac{\alpha\gamma\sigma_t^2\sigma_\eta\eta_t + \sigma_{mp}\varepsilon_t^{mp}}{1+\alpha}$$

and the law of motion (3.27):

$$\mathbf{x}_t \equiv \begin{bmatrix} f_t \\ b_t \end{bmatrix} = \begin{bmatrix} \rho_f & 0 \\ 0 & \rho_b \end{bmatrix} \begin{bmatrix} f_{t-1} \\ b_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma_{f,t} & 0 \\ 0 & \sigma_b \end{bmatrix} \begin{bmatrix} \varepsilon_t^f \\ \varepsilon_t^b \end{bmatrix}$$

These two equations can be nested into a large class of models.

That is, consider Noisy Rational Expectations (NRE) model:

$$s_t = A\left(\tau_t\right)\int E\left(s_{t+1} \mid \Omega_t\left(i\right)\right)di + B\left(\tau_t\right)\mathbf{x}_t + F\left(\tau_t\right)\mathbf{u}_t \qquad (3.30)$$

with state variables:
$$\mathbf{x}_t = \boldsymbol{\rho}\left(\tau_t\right)\mathbf{x}_{t-1} + \nu\left(\tau_t\right)\mathbf{u}_t \qquad (3.31)$$

with shocks $\mathbf{u}_t \sim N(0, I)$, regimes $\tau_t \in \{0, 1\}$ and information sets $\Omega_t\left(i\right)$.

Because of the presence of rational agents in a dispersed information context, the full state of the model consists of the extended state vector $\mathbf{X}_t$ defined as

$$\mathbf{X}_t \equiv \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k})} \end{bmatrix} \qquad (3.32)$$

where

$$\mathbf{x}_t^{(k+1)} \equiv \int E\left[\mathbf{x}_t^{(k)} \mid \Omega_t\left(i\right)\right]di \qquad (3.33)$$

is similar to (3.5). As in the mentioned references, since the state vector $\mathbf{X}_t$ can potentially have infinite dimension, we assume that the hierachy of expectations is truncated up to an order $\overline{k} < \infty$.

The procedure for solving the model is of the type of guess-and-verify. Thus, we conjecture a solution of the form:

$$s_t = G(\tau^t)\mathbf{X}_t + F(\sigma(\tau^t))\mathbf{u}_t \tag{3.34}$$

$$\mathbf{X}_t = M(\tau^t)\mathbf{X}_{t-1} + N(\tau^t)\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \tag{3.35}$$

Solve the model means to find the undetermined matrices $G(\tau^t)$, $F(\tau^t)$, $\sigma(\tau^t)$, $M(\tau^t)$, $N(\tau^t)$. To do that it is necessary to track the history of regimes τ^t. As in the cited reference, we proceed by specifying a maximum lag of τ_t, say $\tau_{t-\overline{r}}$, that matters for current dynamics and then check whether increasing the lag $\overline{r}$ changes the dynamics sufficiently to motivate the increased computational burden.

The matrix H is defined so that:

$$\int E\left[\mathbf{X}_t \mid \Omega_t\left(i\right)\right] di = H\mathbf{X}_t \tag{3.36}$$

The Euler equation for s_t can then be written as

$$
\begin{aligned}
s_t = \ & \omega A(\tau^t) G\left(\tau_1^{t+1}\right) M\left(\tau_1^{t+1}\right) H\mathbf{X}_t \\
& + (1 - \omega) A G\left(\tau_0^{t+1}\right) M\left(\tau_0^{t+1}\right) H\mathbf{X}_t \\
& + \left[\begin{array}{cc} B(\tau^t) & \mathbf{0} \end{array}\right] \mathbf{X}_t + F\left(\tau^t\right) \mathbf{u}_t
\end{aligned}
$$

Using the conjectured form (3.34) and equating coefficients implies:

$$
\begin{aligned}
G\left(\tau^t\right) = \ & \omega A(\tau^t) G\left(\tau_1^{t+1}\right) M\left(\tau_1^{t+1}\right) H \\
& + (1 - \omega) A G\left(\tau_0^{t+1}\right) M\left(\tau_0^{t+1}\right) H \\
& + \left[\begin{array}{cc} B(\tau^t) & \mathbf{0} \end{array}\right]
\end{aligned} \tag{3.37}
$$

Define the conditional variance $\sigma\left(\tau^t\right)$ as

$$\sigma\left(\tau^t\right) \equiv E\left(s_{t+1} - E\left[s_{t+1} \mid \Omega_t\left(i\right)\right]\right)\left(s_{t+1} - E\left[s_{t+1} \mid \Omega_t\left(i\right)\right]\right)' \tag{3.38}$$

It can then be computed as:

$$
\begin{aligned}
\sigma^2\left(\tau^t\right) = \ & \omega\left\{\widehat{\mathbf{G}}_{\overline{k}}\left(\tau_1^{t+1}\right) \widehat{P}\left(\tau_1^{t+1}\right) \widehat{\mathbf{G}}_{\overline{k}}\left(\tau_1^{t+1}\right)'\right\} \\
& + (1 - \omega)\left\{\widehat{\mathbf{G}}_{\overline{k}}\left(\tau_0^{t+1}\right) \widehat{P}\left(\tau_0^{t+1}\right) \widehat{\mathbf{G}}_{\overline{k}}\left(\tau_0^{t+1}\right)'\right\}
\end{aligned} \tag{3.39}
$$

where $\widehat{\mathbf{G}}_{\overline{k}}\left(.\right)$ and $\widehat{P}\left(.\right)$ are the loadings from the solution and the posterior variance from the Kalman updating, considering the augmented state vector $\widehat{\mathbf{X}}_t \equiv [\mathbf{X}_t', \eta_t, \varepsilon_t^{mp}]'$ as in Appendix A of Nimark (2011).

Algorithm for model solution - Barillas et al. (2013):

79

1. Set the values for the dimension of the numerical approximation $\overline{k}$ and $\overline{r}$.

2. Start by making initial guesses for the $|\mathcal{T}|$ different versions of the matrices $G\left(\tau_{\overline{r}}^{t}\right)$, $M(\tau_{\overline{r}}^{t})$, $N(\tau_{\overline{r}}^{t})$ and, if applicable, $\sigma(\tau_{\overline{r}}^{t})$. A good initial guess is to set them such that the dynamics are equivalent to the model solution without regime switching along the lines of Nimark (2011).

3. For given matrices $G\left(\tau_{\overline{r}}^{t}\right)$, $M(\tau_{\overline{r}}^{t})$, $N(\tau_{\overline{r}}^{t})$ and $\sigma(\tau_{\overline{r}}^{t})$ compute $|\mathcal{T}|$ "new" $G\left(\tau_{\overline{r}}^{t}\right)$ using (3.37). (That is, one need to loop through the $|\mathcal{T}|$ different matrices $G\left(\tau_{\overline{r}}^{t}\right)$).

4. If applicable, for given matrices $M(\tau_{\overline{r}}^{t})$, $N(\tau_{\overline{r}}^{t})$ and the "new" $G\left(\tau_{\overline{r}}^{t}\right)$ compute $|\mathcal{T}|$ "new" $\sigma(\tau_{\overline{r}}^{t})$ using (3.39).

5. For given matrices $M\left(\tau_{\overline{r}}^{t}\right)$ and $N\left(\tau_{\overline{r}}^{t}\right)$ and the "new" $G\left(\tau_{\overline{r}}^{t}\right)$ and $\sigma(\tau_{\overline{r}}^{t})$ compute $|\mathcal{T}|$ "new" $M(\tau_{\overline{r}}^{t})$ and $N(\tau_{\overline{r}}^{t})$ using $(C.26)$ and $(C.27)$.

6. Iterate on steps 3 to 5 until convergence.

See specific details regarding this setup can also be found in Nimark (2011) and Nimark (2012a).

3.4 Empirical analysis

3.4.1 Data

In order to estimate the model, data of exchange rates, interest rates differentials and expectations is needed. For our specific case, we use the Japanesse Yen/US Dollar (in logs, detrended), Interest rate differentials $(i_t - i_t^*)$ being the Call Overnight Rate (Japan) and Federal Funds Rate (in percentages). In addition, we use the data shown in Figure 3.1, i.e. the Predicted Exchange Rates from TANKAN Survey (BoJ) for 20 sectors (in logs, detrended) as a proxy for Exchange Rates expectations $(E_t s_{t+1}(i))$. Data is in quarterly frequency and covers the episode 2000:Q1-2012:Q4, see details in Appendix C.1.

3.4.2 Bayesian Estimation

Log Likelihood evaluation

The model solution can be expressed as a State Space System. Including the observables we get the measurement equations:

$$s_t = G\left(\tau^t\right)\mathbf{X}_t + F\left(\sigma\left(\tau^t\right)\right)\mathbf{u}_t \qquad (3.40)$$

80

$$i_t - i_t^* = \frac{1}{\alpha} \left[G\left(\tau^t\right) - e_1' \right] \mathbf{X}_t + \frac{1}{\alpha} \left[F\left(\sigma\left(\tau^t\right)\right) + \sigma_{mp} e_4' \right] \mathbf{u}_t \qquad (3.41)$$

$$E_t s_{t+1}(i) = \left[\begin{array}{c} \omega \mathbf{G}_{\overline{k}}\left(\tau_1^{t+1}\right) M\left(\tau_1^{t+1}\right) \\ + (1-\omega)\, \mathbf{G}_{\overline{k}}\left(\tau_0^{t+1}\right) M\left(\tau_0^{t+1}\right) \end{array} \right] H\mathbf{X}_t + \Sigma_t^s(i)\,\varepsilon_t^e \qquad (3.42)$$

and the transition equation:

$$\mathbf{X}_t = M(\tau^t)\mathbf{X}_{t-1} + N(\tau^t)\mathbf{u}_t$$

Equation (3.40) comes directly from the model solution. Plugging this solution in (3.18) we get equation (3.41). Finally, for each of the expectations in the survey we specify equation (3.42), which is basically the one-step ahead forecast using again the solution (3.40). In order to avoid stochastic singularity, we include the measurement error ε_t^e, and we assume that has a standard error equal to the model implied cross-section dispersion at time t.

It is clear that the latter system is linear and gaussian conditional on a history of regimes τ^t. As a result, the log-likelihood function can be evaluated, so that:

$$\log L\left(\Theta, \tau^T \mid Z^T\right) = -\frac{1}{2} \sum_{t=1}^{T} \left(\begin{array}{c} \ln 2\pi + \ln |\Omega_t\left(\tau^t\right)| \\ + \widetilde{Z}_t'\left(\tau^t\right) \left[\Omega_t\left(\tau^t\right)\right]^{-1} \widetilde{Z}_t\left(\tau^t\right) \end{array} \right)$$

with $Z_t = \left(s_t, i_t - i_t^*, \{E_t s_{t+1}(i)\}_{i=1}^{\overline{N}} \right)'$.

A Multiple-Block Metropolis Hastings procedure

We have identified the parameter set $\Psi = \left\{\Theta, \tau^T\right\}$, being Θ the vector of structural parameters. In order to sample the posterior distribution of Ψ we employ, a Multiple-Block Metropolis Hastings procedure (see also Chib (2001)). The latter strategy allows us to exploit the conditional distribution of parameter blocks. That is, consider the joint posterior distribution and apply the basic Bayes' rule:

$$
\begin{aligned}
P\left(\Theta, \tau^T \mid Z^T\right) &= \frac{P\left(Z^T \mid \Theta, \tau^T\right) \times P\left(\Theta, \tau^T\right)}{P\left(Z^T\right)} \\
&\propto P\left(Z^T \mid \Theta, \tau^T\right) \times P\left(\Theta, \tau^T\right)
\end{aligned}
$$

That is, the posterior distribution is proportional to the kernel specified in the right hand side of this expression. Moreover, the term $P\left(\Theta, \tau^T\right)$ is a joint prior for the two blocks, which can be factorized such that

$$P\left(\Theta, \tau^T \mid Z^T\right) \propto P\left(Z^T \mid \Theta, \tau^T\right) \times P\left(\tau^T \mid \Theta\right) \times P\left(\Theta\right)$$

In other words, we adopt a hierarchical structure for the prior distribution.

The distributions of τ_t are independent across time:

$$P\left(\tau^T \mid \Theta\right) = \prod_{t=1}^{T} P\left(\tau_t \mid \Theta\right)$$

This is not a restrictive assumption, given the specification of the model in which the probability of $(\tau_t = 1)$ does not depend on the history of regimes. Assume also independent priors for structural parameters:

$$P\left(\Theta\right) = \prod_{k=1}^{\dim\Theta} P\left(\Theta_k\right)$$

The Algorithm is as follows:

Parameter set $\Psi = \left\{\Theta, \tau^T\right\}$

1. Specify an initial value $\Psi_0 = \left\{\Theta_0, \tau_0^T\right\}$.

2. Repeat for $j = 1, 2, \ldots, J$
   (a) **Block 1:** Draw Θ_j from $p\left(\Theta_j \mid \Theta_{j-1}, \tau_{j-1}^T, Z^T\right)$:
   (i) Draw Θ^* from $q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right)$
   (ii) Set $\alpha_j^\Theta = \min\left\{1, \frac{L\left(Z^T|\Theta^*,\tau_{j-1}^T\right)p\left(\tau_{j-1}^T|\Theta^*\right)p(\Theta^*)q_\Theta(\Theta_{j-1}|\Theta^*)}{L\left(Z^T|\Theta_{j-1},\tau_{j-1}^T\right)p\left(\tau_{j-1}^T|\Theta_{j-1}\right)p(\Theta_{j-1})q_\Theta(\Theta^*|\Theta_{j-1})}\right\}$
   (iii) Set $\Theta_j = \Theta^*$ if $U\left(0,1\right) \le \alpha_j^\Theta$ and $\Theta_j = \Theta_{j-1}$ otherwise.
   (b) **Block 2:** Draw τ_j^T from $p\left(\tau_j^T \mid \tau_{j-1}^T, \Theta_j, Z^T\right)$:
   (i) Draw $\left(\tau^T\right)^* = \{\tau_t^*\}_{t=1}^T$ from $q_\tau\left(\left(\tau^T\right)^* \mid \tau_{j-1}^T, \Theta_j\right)$
   (ii) Set $\alpha_j^\tau = \min\left\{1, \frac{L\left(Z^T|\Theta_j,(\tau^T)^*\right)p\left((\tau^T)^*|\Theta_j\right)q_\tau\left(\tau_{j-1}^T|(\tau^T)^*,\Theta_j\right)}{L\left(Z^T|\Theta_j,\tau_{j-1}^T\right)p\left(\tau_{j-1}^T|\Theta_j\right)q_\tau\left((\tau^T)^*|\tau_{j-1}^T,\Theta_j\right)}\right\}$
   (iii) Set $\tau_j^T = \left(\tau^T\right)^*$ if $U\left(0,1\right) \le \alpha_j^\tau$ and $\tau_j^T = \tau_{j-1}^T$ otherwise.

3. Return values $\left\{\Theta_0, \Theta_1, \ldots, \Theta_J\right\}$ and $\left\{\tau_0^T, \tau_1^T, \ldots, \tau_J^T\right\}$.

Adaptive Proposal Distributions

In models with potentially multiple peaks and local modes, it is important to invest resources in improving mixing properties of the posterior simulator. We adopt the adaptive approach pionereed by Haario et al. (2001) and include mixtures of proposals.

1. **Block 1 - Structural Parameters** (Θ)**:** Draw a candidate

   $$\Theta^* \sim \delta_\theta N\left(\Theta_{j-1}, C_j\right) + (1 - \delta_\theta) N\left(\Theta_{j-1}, \kappa C_j\right) \equiv q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right)$$

   with $\delta_\theta \le 1$ and $\kappa > 1$. The covariance matrix depends on the history of draws, so

   $$C_j = \left\{ \begin{array}{ll} C_0, & j \le j_0 \\ c_d cov\left(\Theta_0, \ldots, \Theta_{j-1}\right) + c_d \varepsilon I_d, & j > j_0 \end{array} \right.$$

82

where $d = \dim(\Theta)$ and (c_d, ε) are constants that are calibrated in order to get an optimal acceptance rate.

The Mixture of Gaussian distributions is very useful for generating candidates with occasionaly large variances, meaning that it is possible to escape from local modes. Following Strid et al. (2010) , we set $\delta_\theta = 0.95$ and $\kappa = 9$.

Since proposal is symmetric:

$$q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right) = q_\Theta\left(\Theta_{j-1} \mid \Theta^*\right)$$

we have as a result that the acceptance probability is

$$\alpha_j^\Theta = \min\left\{1, \frac{L\left(Z^T \mid \Theta^*, \tau_{j-1}^T\right) p\left(\tau_{j-1}^T \mid \Theta^*\right) p\left(\Theta^*\right)}{L\left(Z^T \mid \Theta_{j-1}, \tau_{j-1}^T\right) p\left(\tau_{j-1}^T \mid \Theta_{j-1}\right) p\left(\Theta_{j-1}\right)}\right\}$$

2. **Block 2** $\left(\tau^T\right)$**:** Sampling the posterior of parameters that follow continuous distributions has been widely covered, especially in DSGE estimation literature, as in An and Schorfheide (2007). On the other hand, sampling the posterior of discrete variables is a more complex issue. We adopt the approach of Fiorentini et al. (2012) and Giordani and Kohn (2010) and use adaptive proposals. However, this time we cannot use the covariance matrix structure proposed by Haario et al. (2001).

   Draw a candidate $\left(\tau^T\right)^* = \{\tau_t^*\}_{t=1}^T$ from $q_\tau\left(\left(\tau^T\right)^* \mid \tau_{j-1}^T, \Theta_j, c_\tau^j\right)$. That is, for $t = 1, 2, \ldots, T$ draw

$$\tau_t^* = \begin{cases} 1, & U\left(0,1\right) \le c_\tau^j(t) \\ 0, & \text{otherwise} \end{cases}$$

   where for $t = 1, 2, \ldots, T$ we compute

$$c_\tau^j(t) = \begin{cases} c_\tau^0(t), & j \le j_0 \\ \delta c_\tau^0(t) + (1-\delta)\frac{\sum_{k=1}^{j-1}\tau_t^k}{j-1}, & j > j_0 \end{cases}$$

   $c_\tau^0(t)$ and $\delta \in (0,1)$ are constants that are calibrated in order to get an optimal acceptance rate.

   Since values τ_t for each period $t = 1, 2, \ldots, T$ are independent, the proposal density is the product of T Bernoulli-type distributions:

$$q_\tau\left(\left(\tau^T\right)^* \mid \tau_{j-1}^T, \Theta_j, c_\tau^j\right) = \prod_{t=1}^T \left[c_\tau^j(t)\right]^{\tau_t^*}\left[1 - c_\tau^j(t)\right]^{(1-\tau_t^*)}$$

The use of the two proposal distributions, together with the hierarchical prior specification, turns to be a consistent procedure for detecting regimes in Survey Data. In particular, Barillas et al. (2013) provide a Monte Carlo Experiment using this routine.

3.5 Results

3.5.1 Posterior distribution of parameters

After running the posterior simulator for $200,000$ repetitions, we get the following results. It is worth to notice that informative rather than flat priors for Θ are included in the estimation procedure. These priors are not very restrictive, since their standard deviations are fairly large, but are included because the sample horizon is relatively small (about 52 quarters of observations).

Structural parameter estimates are displayed in Table 3.1. First important feature is to see that aggregate fundamentals exhibit a considerable persistence reflected in the value of parameters $\left(\rho_f, \rho_b\right)$, especially for f_t. This result is in line with the empirical literature of exchange rates, e.g. Rossi (2005). The second important aspect is the fact that the volatility of the noise in f_t is much higher in regime $\tau_t = 1$, with $(\sigma_{f_2}/\sigma_f) \approx 2.83$. In addition, the volatility of the public signal is much higher than the private one, with $(\sigma_p/\widetilde{\sigma}_f) \approx 2$. In other words, in regime $\tau_t = 1$ agents observe an additional public signal with a relative precision of 25% with respect to the private signal. This very imprecise public signal, together withthe fact that $(\sigma_{f_2}/\sigma_f) > 1$, are consistent elements with our story of an increase in disagreement.

Θ	$\left\|\text{Mode}\left(\widetilde{\Theta}\right)\right\|$	Prior		
		Distribution	Mean	S.d.
α	0.75	$N\,(10,4)$	10.00	2.00
γ	1.79	$Gamma\,(1.6, 1.25)$	2.00	1.01
ρ_f	0.98	$Beta\,(2.6, 2.6)$	0.50	0.20
ρ_b	0.53	$Beta\,(2.6, 2.6)$	0.50	0.20
$\widetilde{\sigma}_f$	1.67	$I-Gamma\,(6, 1/4)$	0.80	0.40
$\widetilde{\sigma}_b$	2.52	$I-Gamma\,(6, 1/4)$	0.80	0.40
σ_f	0.24	$I-Gamma\,(9, 1/4)$	0.50	0.19
σ_b	0.77	$I-Gamma\,(9, 1/4)$	0.50	0.19
σ_η	0.29	$I-Gamma\,(3, 1/2)$	1.00	1.00
σ_{mp}	0.62	$I-Gamma\,(3, 1/2)$	1.00	1.00
σ_{f_2}	0.68	$I-Gamma\,(9, 1/4)$	0.50	0.19
σ_p	3.35	$I-Gamma\,(10, 1/20)$	2.22	0.78
ω	0.18	$Beta\,(0.5, 5)$	0.09	0.11
$\log P\left(\widetilde{\Theta}, \widetilde{\tau}^T \mid Z^T\right)$	400.88			

Table 3.1: Posterior estimates for 2000-2012
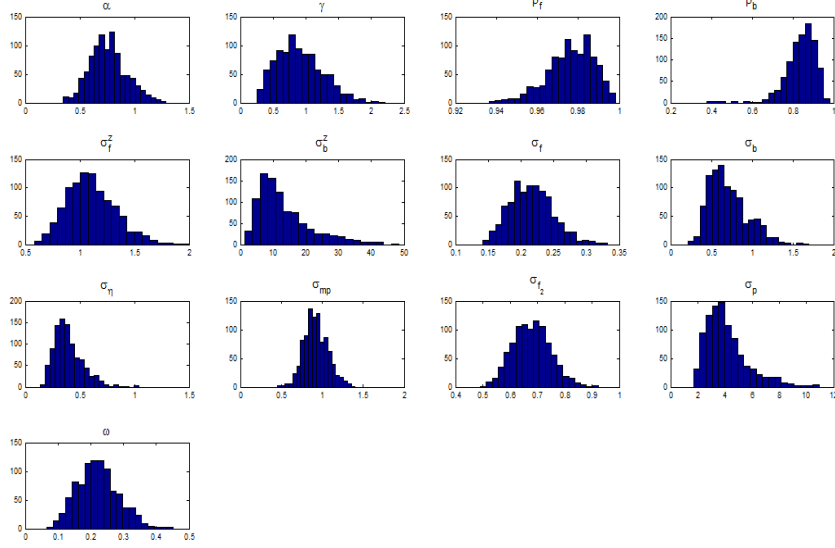
Figure 3.2: Marginal Posterior Distributions for Θ

As a matter of fact, if we inspect the signal-to-noise ratios (see e.g. Melosi (2012)), they are:

$$
\begin{aligned}
\left(1/\widetilde{\sigma}_f\right)/\left(1/\sigma_f\right) &\approx 0.14 \quad (\tau_t = 0) \\
\left(1/\widetilde{\sigma}_f\right)/\left(1/\sigma_{f_2}\right) &\approx 0.41 \quad (\tau_t = 1) \\
\left(1/\widetilde{\sigma}_b\right)/\left(1/\sigma_b\right) &\approx 0.31
\end{aligned}
$$

Clearly, because of the regime switches, the signal-to-noise ratio of f_t changes across regimes and it is higher in $\tau_t = 1$ (0.41) than in $\tau_t = 0$ (0.14). This means that agents pay more attention to the private signal under $\tau_t = 1$ than under $\tau_t = 0$. The signal-to-noise ratio also serves as a device to differentiate which fundamental is relatively more important for private agents. Interestingly, in regime $\tau_t = 1$ agents pay more attention to fundamental f_t than to b_t (0.41 > 0.31), but in regime $\tau_t = 0$ agents pay more attention to fundamental b_t than to f_t (0.31 > 0.14).

Finally, we observe at the mode that $\Pr(\tau_t = 1) = \omega = 0.18$, which means that the regime of higher volatility, and therefore lower disagreement, is quite infrequent with respect to the regime with lower volatility and higher disagreement, with posterior odds $\omega/(1-\omega) \approx 0.22$. Furthermore, histograms from the Regime-Switching model are depicted in Figure 3.2.

Having observed a first evidence of regime switches through the posterior distribution of Θ, we now turn to study the posterior distribution of regimes τ^t. As
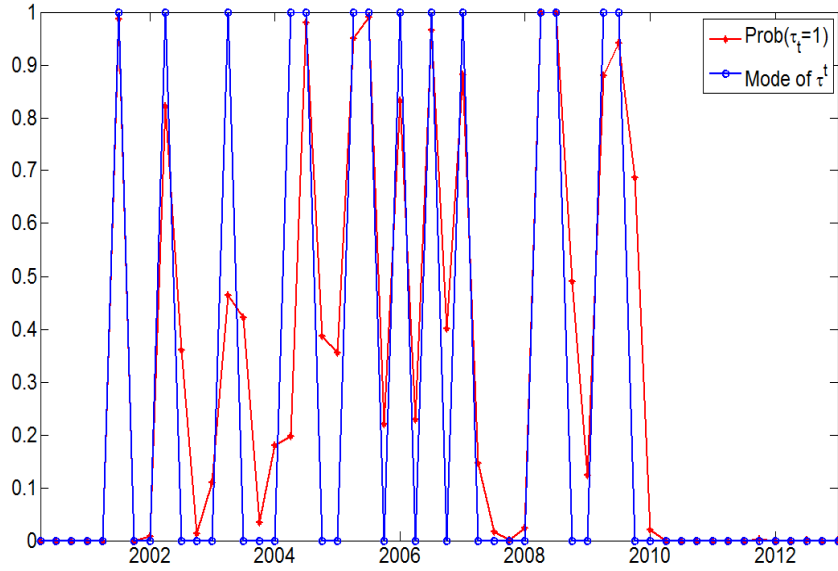
Figure 3.3: Posterior estimates of τ_t

it is depicted in Figure 3.3, the posterior probability of being in regime $\tau_t = 1$ is not the same for each point in the data. The latter implies that our dataset exhibits evidence of regime switches at particular dates related with higher volatility of fundamentals and the observation of a very imprecise public signal. We also include the history of regimes τ^t associated with the posterior mode. An interesting result is that it shows some persistence, i.e. at least two consecutive regimes with $\tau_t = 1$, even though we have imposed that these regimes are independent each other. This result does not necessarily implies that a markov-switching type model should be implemented. It simply means that in some specific cases the regime $\tau_t = 1$ continues ongoing, but it is still the case that these episodes occur eventually. The observed persistence, if any, is only for two periods, which is different than standard markov switching results. We can focus our attention to the period of the zero interest rates in Japan (until 2006), with the specific year of 2001, or the last financial crisis episode, starting in 2008. These episodes of turbulence imply a higher volatility in the monetary fundamentals, and also more imprecise information about fundamentals that can generate more disagreement among poll participants in the TANKAN Survey of Bank of Japan.
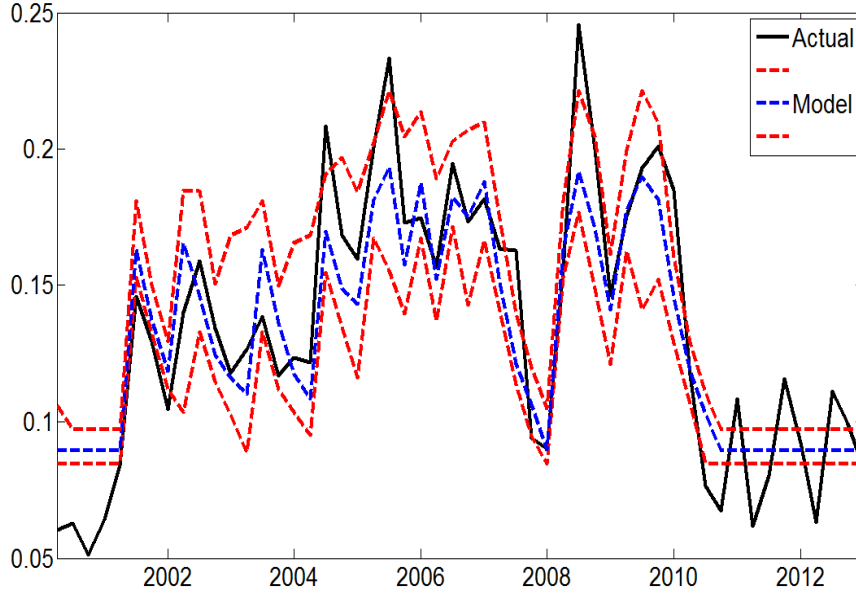
Figure 3.4: Model implied dispersion $\Sigma_{i,t}$

3.5.2 Model Implied Dispersion

We want to see, given posterior estimates of parameters $\Psi = \left\{ \Theta, \tau^T \right\}$, to what extent the cross-sectional dispersion implied by the model matches with the actual data, as in Figure 3.1. In this regard, Figure 3.4 depicts this result using a 90 percent confidence interval and compares it with the actual observables. It can be inferred that the model does a good job in describing the dynamics of disagreement. In other words, the information frictions plus the regime-switching approach are good ingredients for this purpose. Of course, the match is not perfect, it gives us evidence pointing into the right direction.

3.5.3 Rational Confusion and Impulse responses

Once we have identified the parameter values, we proceed to explore the transmission of shocks in the model. Clearly, one can expect that differences in responses will be reflected across regimes, and more importantly, the presence of informational frictions and higher-order expectations provokes the Rational Confusion, as pointed out by Bacchetta and Van Wincoop (2006). That is, whenever a shock hits the economy, since it is a shock to unobservable fundamentals, agent cannot distinguish the true source of the innovation in the short run. As a result, higher-
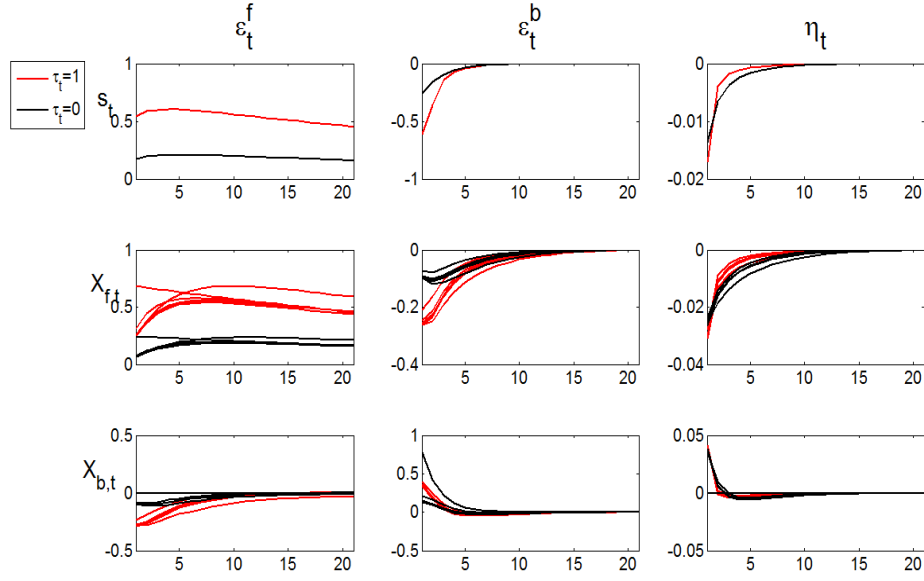
Figure 3.5: Responses of s_t and $\mathbf{X}_t$ to ε_t^f, ε_t^b and η_t

order expectations for both fundamentals will react to an innovation in only one fundamental, generating additional persistence in the response of exchange rates.

Concretely, we can see that the response of the exchange rate s_t to each of the fundamentals as well as all the responses of the different orders of expectations $\left(\mathbf{X}_t^f, \mathbf{X}_t^b\right)$ in Figure 3.5. As it is clear, responses across regimes differ because of the higher volatility, $\sigma_{f_2} > \sigma_f$ and because of the time varying information sets $\Omega_t(i)$. We explore the impulse responses with respect to the initial regime $\tau_t = 1$ ot $\tau_t = 0$, i.e. the current regime when the shock occurs. These differences are reflected in both amplification and propagation. The fact that the hierarchy of expectations about fundamentals exhibit hump-shaped responses, and given that the exchange rate is a linear function of this hierarchy, we can see a persistent reaction of exchange rates s_t. For the specific case of monetary fundamentals, the hump-shaped response of s_t can also be linked with the so-called *delayed overshooting puzzle* (see Dornbusch (1976), Eichenbaum and Evans (1995), Kim (2005)), but recalling that our model has flexible prices. Furthermore, rational confusion is even more clear when we observe the reaction of the system to the transitory shock η_t, the stochastic component of bond supply. This shock is not persistent at all, however we observe the reaction of st as well as the hierachies for f_t and b_t. As a result, under the presence of informational frictions and rational confusion, even a transitory i.i.d. shock can produce considerable persistent effects in exchange rates.

3.5.4 Model comparison

The purpose of this section is to provide a measure of comparison between the estimated regime-switching model with respect to two competing ones. Following a large literature of Bayesian statistics, our measure of model comparison is the Marginal Data Density. In our case we have the set of models:

$$M = \{\text{Benchmark}, \text{ Bacchetta and van Wincoop}, \text{Common Information}\} \quad (3.43)$$

The model labeled as 'Benchmark' is the one described in this paper. The model labeled as 'Bacchetta and van Wincoop' is a restricted version of 'Benchmark' that rules out regime switches. That is, if we set $\tau_t = 0$, $\forall t$, we go back to the benchmark model of Bacchetta and Van Wincoop (2006). Solving and estimating the model under this restriction is much simpler, it is an asset pricing model as in Nimark (2011) and it is estimated as in e.g. Nimark (2012b) or Barillas and Nimark (2012). In addition, the model labeled as 'Common Information' is a restricted version of 'Bacchetta and van Wincoop' that rules out informational frictions. As a result, this is a simpler model solved through the 'Law of Iterated Expectations'. Survey data are treated using the particular case of equation (3.42) with no regime changes and no higher-order expectations, i.e. $E_t s_{t+1}(i) = \mathbf{G}M\mathbf{x}_t$. We want to test to what extent our regime-switching model fits the data better than the two competing ones.

Denote the Marginal Likelihood for model M_j as:

$$P\left(Z^T \mid M_j\right) = \int_{\Theta^j} L\left(Z^T \mid \Theta_s^j, M_j\right) P\left(\Theta_s^j \mid M_j\right) d\Theta_s^j$$

where $\Theta^j = \{\Theta_s^j\}_{s=1}^{\overline{S}}$ is the set of posterior estimates, $L\left(.\right)$ is the Likelihood Function and $P\left(.\right)$ is the prior density. For each of the competing models M_j we calculate the empirical counterpart of this measure using the harmonic-mean estimator of Geweke (1999):

$$MDD_j = \left[\frac{1}{\overline{S}}\sum_{s=1}^{\overline{S}} \frac{f\left(\Theta_s^j \mid M_j\right)}{L\left(Z^T \mid \Theta_s^j, M_j\right) P\left(\Theta_s^j \mid M_j\right)}\right]^{-1} \quad (3.44)$$

where

$$\begin{aligned}
f\left(\Theta_s^j\right) &= p^{-1}\left(2\pi\right)^{-d_j/2} \exp\left[-1/2\left(\Theta_s^j - \overline{\Theta}^j\right)' \Sigma^j \left(\Theta_s^j - \overline{\Theta}^j\right)\right] \times \\
&\quad I\left(\left(\Theta_s^j - \overline{\Theta}^j\right)' \Sigma^j \left(\Theta_s^j - \overline{\Theta}^j\right) \leq \chi_p^2\left(d_j\right)\right)
\end{aligned}$$

and $\overline{\Theta}^j$ and Σ^j are respectively the mean and covariance matrix of the posterior distribution Θ^j and $p \in (0,1)$ is the percentile of the χ^2 distribution. In other words, the expression (3.44) is the integral the Posterior Distribution with respect to the MCMC draws Θ^j. The latter represents the marginal data density (MDD) or Marginal Likelihood of model j. Having estimated the three models using $200,000$ draws, we set $\overline{S} = 100$ and evaluate the expression (3.44).

p	Benchmark	Bacchetta and van Wincoop	Common Information
0.1	420.966	327.536	323.625
0.3	419.867	328.187	323.743
0.5	421.022	328.869	323.952
0.7	421.445	329.020	323.809
0.9	420.935	329.147	323.986

Table 3.2: Log-Marginal Likelihood: Harmonic-Mean estimator

Table 3.2 shows that the model with Regime Switches fits the data better than the two other competing models. It is also interesting to see that the Benchmark model with information frictions does a better job than the model with common information. It is worth to remark that, for comparison purposes, for each these three models we use the same dataset with the same detrending procedure. Seen through these lenses, our model does a good job in describing Survey Data that exhibits disagreement.

3.6 Concluding Remarks

We fit a structural model with regime switches to Exchange Rates Survey Data. The assumed information structure is such that high-volatility regimes are associated with the appearance of a very noisy public signal about fundamentals. Given that this signal is very imprecise, and the higher volatility of shocks, disagreement increases. The model-implied dispersion closely follows the observed disagreement, which means the proposed model does a good job in terms of fitting. We confirm the latter when comparing the model fit with respect to a restricted model without regimes switches and with no informational frictions at all.

Furthermore, the model solution using the captured regime switches from the data is interpreted as evidence in favor of parameter instability in exchange rate models. The latter is, together with rational confusion, an additional explanation for the disconnection between Exchange Rates and future fundamentals.

We leave as future agenda the idea of exploring several surveys of Exchange Rate expectations about different currencies, such as Consensus Economics or

Bloomberg.

Appendix A

APPENDIX TO CHAPTER 1

A.1 Equivalent reparameterizations of a SVAR

Consider the SVAR model:

$$A\left(\alpha\right)y_t = \varepsilon_t; \ \ \varepsilon_t \sim N\left(0, I_M\right) \tag{A.1}$$

$t = 1, \ldots, T$. y_t and ε_t are $M \times 1$ vectors, $A\left(\alpha\right)$ is a non-singular $M \times M$ matrix and α is the vector of structural parameters. The Likelihood function of $(A.1)$ is:

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

Following Hamilton (1994) and Amisano and Giannini (1997), the matrix $A\left(\alpha\right)$ is re-parametrized as

$$vec\left(A\left(\alpha\right)\right) = S_A\alpha + s_A$$

We know that

$$
\begin{aligned}
vec\left(A\left(\alpha\right)y_t\right) &= vec\left(I_M A\left(\alpha\right)y_t\right) \tag{A.3}\\
&= \left(y_t' \otimes I_M\right) vec\left(A\left(\alpha\right)\right)\\
&= \left(y_t' \otimes I_M\right)\left(S_A\alpha + s_A\right)
\end{aligned}
$$

Since $A\left(\alpha\right)y_t'\left(A\left(\alpha\right)y_t\right)$ is a scalar

$$\left(A\left(\alpha\right)y_t\right)'\left(A\left(\alpha\right)y_t\right) = tr\left(\left(A\left(\alpha\right)y_t\right)'\left(A\left(\alpha\right)y_t\right)\right)$$

Also, since:

$$tr\left(\left(A\left(\alpha\right)y_t\right)'\left(A\left(\alpha\right)y_t\right)\right) = \left[vec\left(A\left(\alpha\right)y_t\right)\right]'\left[vec\left(A\left(\alpha\right)y_t\right)\right]$$

we have:

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

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{\begin{array}{c}-\frac{1}{2}\sum_{t=1}^{T}\left(S_A\alpha + s_A\right)'\left(y_t' \otimes I_M\right)' \\ \times\left(y_t' \otimes I_M\right)\left(S_A\alpha + s_A\right)\end{array}\right\}$$

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{\begin{array}{c}-\frac{1}{2}\sum_{t=1}^{T}\left(S_A\alpha + s_A\right)'\left(y_t \otimes I_M\right) \\ \times\left(y_t' \otimes I_M\right)\left(S_A\alpha + s_A\right)\end{array}\right\}$$

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

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

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{\begin{array}{c}-\frac{1}{2}\sum_{t=1}^{T}\left(\alpha'S_A'\left(y_t y_t' \otimes I_M\right) + s_A'\left(y_t y_t' \otimes I_M\right)\right) \\ \times\left(S_A\alpha + s_A\right)\end{array}\right\}$$

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\begin{array}{c}\alpha'S_A'\left(y_t y_t' \otimes I_M\right)S_A\alpha \\ +\alpha'S_A'\left(y_t y_t' \otimes I_M\right)s_A \\ +s_A'\left(y_t y_t' \otimes I_M\right)S_A\alpha \\ +s_A'\left(y_t y_t' \otimes I_M\right)s_A\end{array}\right]\right\}$$

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\begin{array}{c}\alpha'S_A'\left(y_t y_t' \otimes I_M\right)S_A\alpha \\ +2\alpha'S_A'\left(y_t y_t' \otimes I_M\right)s_A \\ +s_A'\left(y_t y_t' \otimes I_M\right)s_A\end{array}\right]\right\}$$

$$L\left(y^T \mid \alpha\right) = (2\pi)^{-MT/2} \det\left(A\left(\alpha\right)\right)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\left(3\alpha'S_A' + s_A'\right)\left(I_M \otimes y_t'y_t\right)\left(S_A\alpha + 2s_A\right)\right]\right\}$$

(A.4)

where $y_t y_t' \otimes I_M = \left(y_t y_t' \otimes I_M\right)' = I_M \otimes y_t'y_t$, because it is a symmetric matrix.

The repameterization used Waggoner and Zha (2003) delivers the same likelihood function because as we have shown in the text

$$vec\left(A\left(\alpha\right)\right) = \left[\begin{array}{c}a_1 \\ a_2 \\ a_3\end{array}\right]$$

so that

$$S_A = diag\left(U_1, U_2, U_3\right); \; s_A = \left[\begin{array}{c}R_1 \\ R_2 \\ R_3\end{array}\right]$$

94

where $diag\,(.)$ indicates a block-diagonal matrix.

To show that the reparameterization employed in section 2 of the paper also delivers the same likelihood, we apply the vec operator to $(A.1)$:

$$vec\,(A\,(\alpha)\,y_t) = vec\,(\varepsilon_t)$$

Using $(A.3)$ and the fact that $vec\,(\varepsilon_t) = \varepsilon_t$, the model can be re-expressed as:

$$\widetilde{y}_t = Z_t \alpha + \varepsilon_t \qquad\qquad (A.5)$$

with

$$\widetilde{y}_t \equiv (y_t' \otimes I_M)\,s_A; \;\; Z_t \equiv - (y_t' \otimes I_M)\,S_A$$

Hence, the likelihood function is

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

where T is the number of observations and where the **Jacobian** D such that

$$
\begin{aligned}
D &= \frac{\partial\left[vec\,(A\,(\alpha)\,y_t)\right]}{\partial y_t'} \\
&= \frac{\partial\left[\widetilde{y}_t - Z_t\alpha\right]}{\partial y_t'} \\
&= \frac{\partial\,(y_t' \otimes I_M)\,s_A}{\partial y_t'} + \frac{\partial\,(y_t' \otimes I_M)\,S_A\alpha}{\partial y_t'} \\
&= \frac{\partial I_M D_y y_t}{\partial y_t'} + \frac{\partial I_M D_z y_t}{\partial y_t'} \\
&= \frac{\partial D_y y_t}{\partial y_t'} + \frac{\partial D_z y_t}{\partial y_t'} \\
&= D_y + D_z
\end{aligned}
$$

where $vec\,(D_y) = s_A$ and $vec\,(D_z) = S_A\alpha$. Applying the vec operator we get

$$
\begin{aligned}
vec\,(D) &= vec\,(D_y + D_z) \\
&= vec\,(D_y) + vec\,(D_z) \\
&= s_A + S_A\alpha \\
&= vec\,(A\,(\alpha))
\end{aligned}
$$

Hence

$$D = A\,(\alpha) \qquad\qquad (A.6)$$

95

The Likelihood function of $(A.5)$:

$$
\begin{aligned}
\widetilde{L}\left(y^T \mid \alpha\right) \ & \propto \ \det D^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\widetilde{y}_t' - \alpha' Z_t'\right]\left[\widetilde{y}_t - Z_t\alpha\right]\right\} \\
& \propto \ \det D^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\widetilde{y}_t'\widetilde{y}_t - \widetilde{y}_t' Z_t\alpha - \alpha' Z_t'\widetilde{y}_t + \alpha' Z_t' Z_t\alpha\right]\right\} \\
& \propto \ \det D^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\widetilde{y}_t'\widetilde{y}_t - 2\alpha' Z_t'\widetilde{y}_t + \alpha' Z_t' Z_t\alpha\right]\right\}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
\widetilde{y}_t'\widetilde{y}_t \ & = \ s_A\left(y_t' \otimes I_M\right)'\left(y_t' \otimes I_M\right)s_A \\
& = \ s_A\left(y_t \otimes I_M\right)\left(y_t' \otimes I_M\right)s_A \\
& = \ s_A'\left(y_t y_t' \otimes I_M\right)s_A \\
& = \ s_A'\left(I_M \otimes y_t' y_t\right)s_A
\end{aligned}
$$

$$
\begin{aligned}
Z_t' Z_t \ & = \ S_A'\left(y_t' \otimes I_M\right)'\left(y_t' \otimes I_M\right)S_A \\
& = \ S_A'\left(y_t \otimes I_M\right)\left(y_t' \otimes I_M\right)S_A \\
& = \ S_A'\left(y_t y_t' \otimes I_M\right)S_A \\
& = \ S_A'\left(I_M \otimes y_t' y_t\right)S_A
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{y}_t' Z_t \ & = \ -s_A'\left(y_t' \otimes I_M\right)'\left(y_t' \otimes I_M\right)S_A \\
& = \ s_A'\left(y_t \otimes I_M\right)\left(y_t' \otimes I_M\right)S_A \\
& = \ s_A'\left(y_t y_t' \otimes I_M\right)S_A \\
& = \ -s_A'\left(I_M \otimes y_t' y_t\right)S_A
\end{aligned}
$$

so that

$$
Z_t'\widetilde{y}_t = -S_A'\left(I_M \otimes y_t' y_t\right)s_A
$$

Therefore

$$
\widetilde{L}\left(y^T \mid \alpha\right) \propto \det D^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\begin{array}{c} s_A'\left(I_M \otimes y_t' y_t\right)s_A \\ +2\alpha' S_A'\left(I_M \otimes y_t' y_t\right)s_A \\ +\alpha' S_A'\left(I_M \otimes y_t' y_t\right)S_A\alpha \end{array}\right]\right\}
$$

and using $(A.6)$ we get

$$
\widetilde{L}\left(y^T \mid \alpha\right) \propto \det A\left(\alpha\right)^T \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\left(s_A' + 3\alpha' S_A'\right)\left(I_M \otimes y_t' y_t\right)\left(2s_A + S_A\alpha\right)\right]\right\}
$$
$$\tag{A.7}$$

96

A.2   Global Identification

Consider the constant coefficients version of SVAR model used in section 5:

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

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 whether the system is globally identified, we first need to re-express this model using the notation of Rubio-Ramírez 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 the model is re-expressed as[1]

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

---

[1] See equation (1) in Rubio-Ramírez et al. (2010).

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

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

an expression similar to equation (9) in Rubio-Ramírez et al. (2010). The matrices $\mathbf{Q}_j$ for $j = 1, \ldots, 6.$ corresponding to Theorem 1 are:

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

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

98

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

Define the matrices

$$
\mathbf{M}_j\left(\mathbf{A}_0\right) =
\begin{bmatrix}
\mathbf{Q}_j \mathbf{A}_0 \\
\begin{bmatrix} \mathbf{I}_j & \mathbf{0}_{j\times(M-j)} \end{bmatrix}
\end{bmatrix}
; \; j = 1, \ldots, M \tag{A.8}
$$

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.

A.3   Lower-dimensional systems

We have described in the text the reparametrized SUR model that results from using a shrinkage prior on the B_t under the assumption that the Ξ are known. Here we describe a modified algorithm which can be used when the Ξ are unknown.

Consider the TVC-SVAR model:

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

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

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

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

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

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

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

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) and Koop and Korobilis (2010):

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

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

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

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

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 the equation:

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

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. 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

$$\overline{V\Xi}_f = \left(\underline{V\Xi}^{-1} + Q_{(f,f)}^{-1} \left(\theta^T \right) \left(\theta^T \right)' \right)^{-1}$$

$$\overline{\Xi}_f = \overline{V\Xi}_f \left(\underline{V\Xi}^{-1} \underline{\Xi}_f + Q_{(f,f)}^{-1} \theta^T B_f^T \right)$$

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:

$$\underline{\Xi}_f = \mathbf{0}_{\mathcal{F} \times 1}; \quad \underline{V\Xi} = k_{\Xi}^2 I_{\mathcal{F}}$$

with the hyperparameter $k_{\Xi}^2 = 0.01$.

6. Given (B_t, Q), sample α_t, V, σ_t and W as in the text. Once these are obtained, go back to 1.

A.4 Convergence diagnostics

Following Geweke (1992), Primiceri (2005) and Baumeister and Benati (2012) among others, we check for the autocorrelation properties of the different blocks of the Markov Chain via the inefficiency factor. Let the Relative Numerical Efficiency (RNE) be:

$$RNE = \frac{1}{2\pi} \frac{1}{S(0)} \int_{-\pi}^{\pi} S(\omega) \, d\omega \tag{A.9}$$

where $S(.)$ is the spectral density of an element of the Markov Chain. The inefficiency factor $IF = 1/RNE$, can be interpreted as an estimate of $(1 + 2\sum_{k=1}^{\infty} \rho_k)$, where ρ_k denotes the autocorrelation of k-th order. Thus, high values of IF indicate strong serial correlation across draws. We use the MATLAB file **coda.m** from James P. Lesage toolbox to calculate the IF. We set a 4% tapered window for the estimation of the spectral density at frequency zero and take values around or below 20 as cut-off point (and values below are considered as satisfactory). Figure A.1 presents the IF statistic. Overall, serial correlation across does not seem to be an issue.

A.4.1 Markov Chain plots

Figures A.2 and A.3 depict the evolution of the Markov chain for the non-discarded draws for selected parameters. Recall that we have discarded the first $100,000$ draws and kept 1 for every 100 draws for the remaining $50,000$ draws.
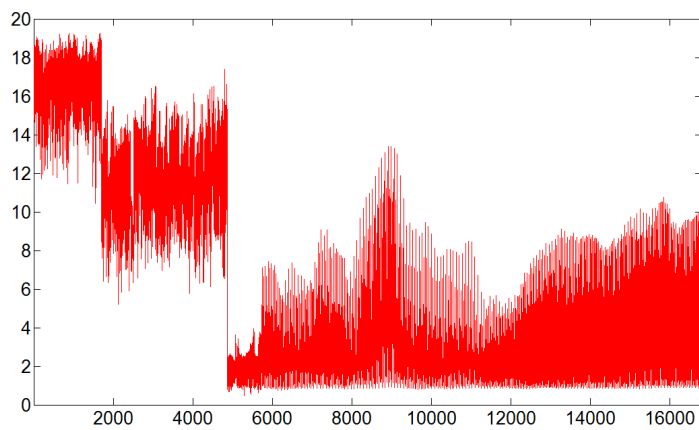
Figure A.1: Inefficiency factor for each parameter of the model
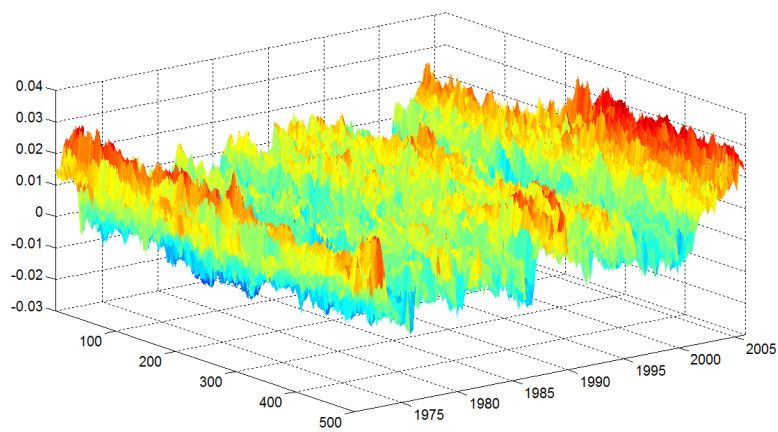
Figure A.2: Plot of $\alpha_{6,t}$
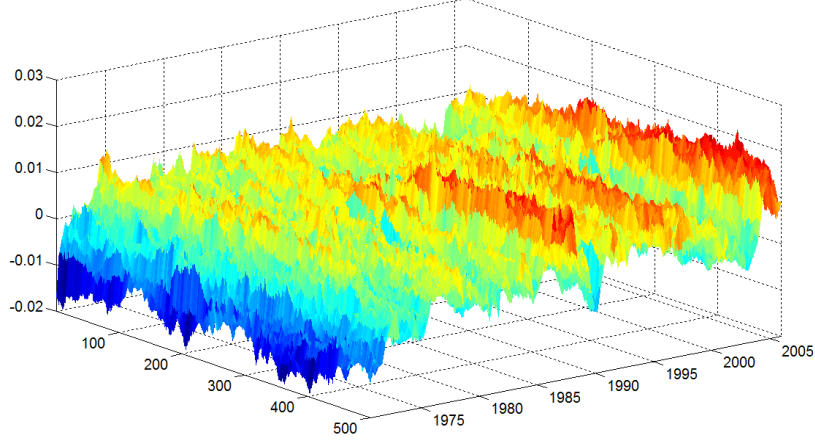
Figure A.3: Plot of $\alpha_{9,t}$

In all figures parameter fluctuations are small indicating that convergence to the ergodic distribution has already occurred.

Figure A.4 presents rolling estimates of the diagonal elements of the covariance matrix, where estimates at each point are obtained with 10 draws - we do this for every t. we For space reasons, we only show the results for $t = 25$. In almost all the cases estimates of the variances are stable or do not fluctuate too much (values are most of them in the order of 1×10^{-5}), which suggest that the chain converged to the ergodic distributions (see Casella and Robert (2004)).

A.4.2   Histograms

Figures A.5, A.6 and A.7 present histograms for coefficients $\alpha_{11,t}$, $\alpha_{6,t}$ and $\alpha_{2,t}$ at selected dates. The empirical distribution look broadly unimodal, which is also a good indication of convergence.

A.5   Dynamics in the single-move algorithm

This section reports the figures presented in the text when a single-move algorithm is used to draw sequences for B^T and all 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. We set the same priors as in Table 1
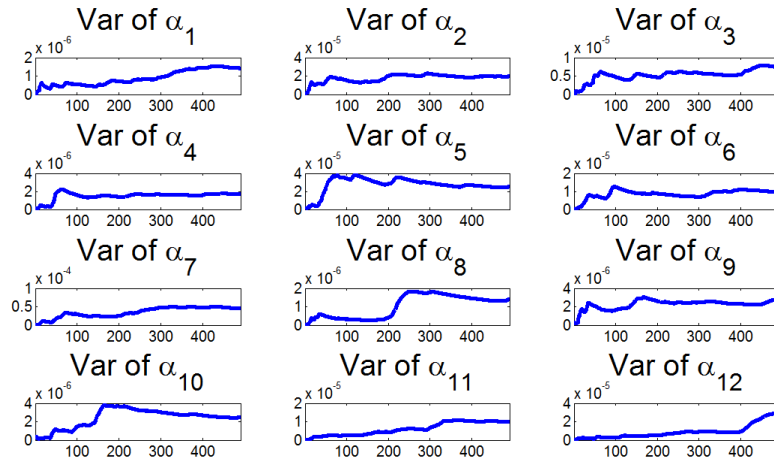
Figure A.4: Rolling Covariance Matrix of MCMC draws
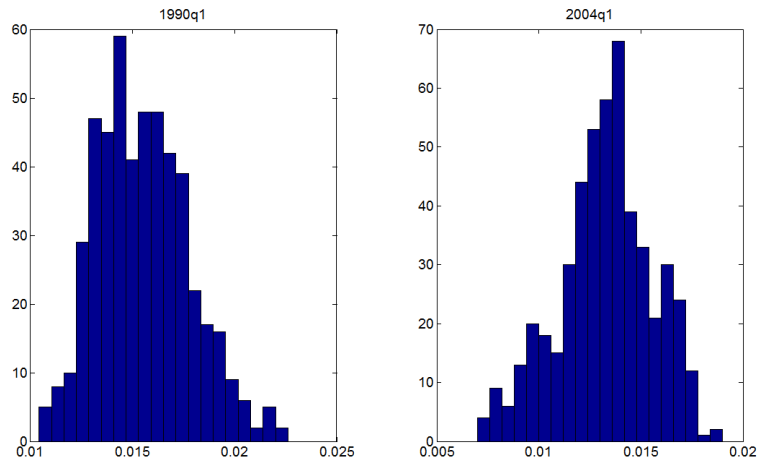
Figure A.5: Histograms for $\alpha_{11,t}$, selected dates.
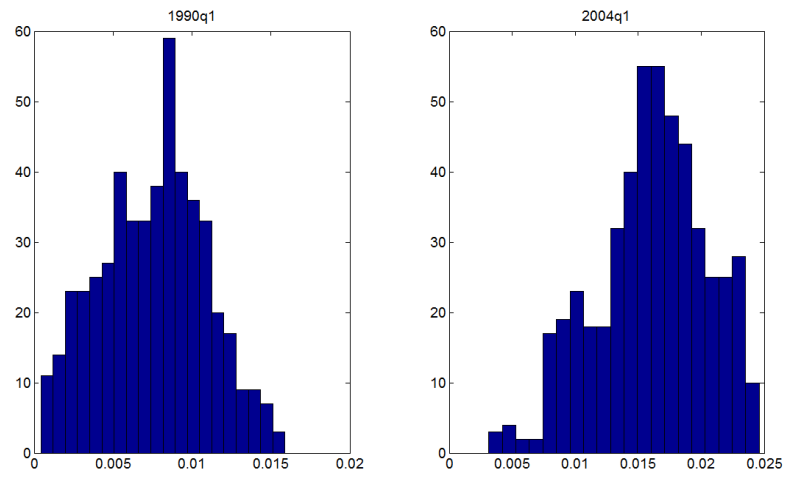
Figure A.6: Histograms for $\alpha_{6,t}$, selected dates.
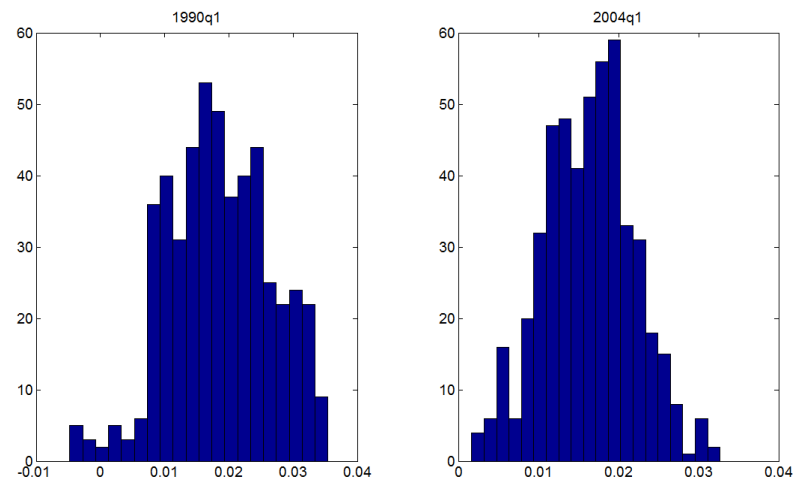
Figure A.7: Histograms for $\alpha_{2,t}$, selected dates.

with hyperparameters $k_Q^2 = 0.5 \times 10^{-4}$, $k_S^2 = 1 \times 10^{-3}$, $k_W^2 = 1 \times 10^{-4}$ and, for comparison purposes, we use the same scaling for the variables as in the multi-move case.

Figures A.8, A.9 and A.10 slightly differ from those in the text (see Figures 1.3, 1.5, and 1.4). We still observe the same in the evolution of volatility of monetary policy shocks as in the benchmark case.
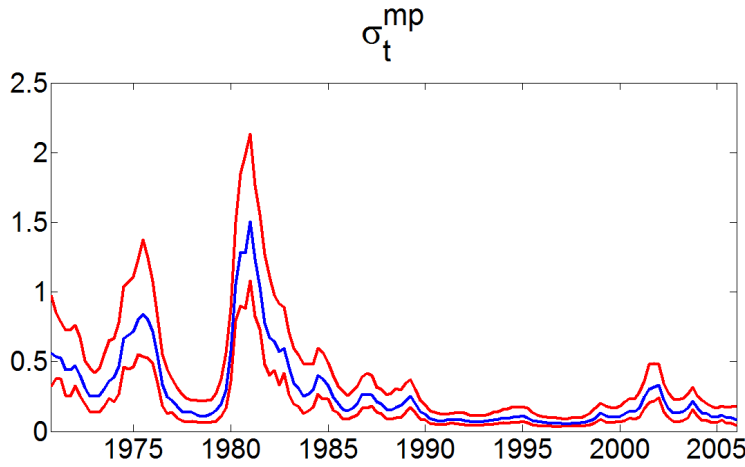
Figure A.8: Volatility of monetary policy shock (single-move)

Since the hyper-parameters $\left(k_Q^2, k_S^2, k_W^2\right)$ are fixed and since these parameters control our prior beliefs about the extent of time variation in each parameter block $\left(B^T, \alpha^T, \sigma^T\right)$, some difference are also present over time. In particular, the same parameterization implies a tighter prior for B^T relative to the multi-move algorithm and, as a result, we will have less variation impulse responses across dates. Clearly, a tighter prior for B^T leads to an increase in the acceptance rates for stable impulse responses.

The changes in structural parameters are also different as can be seem in Figure A.10.
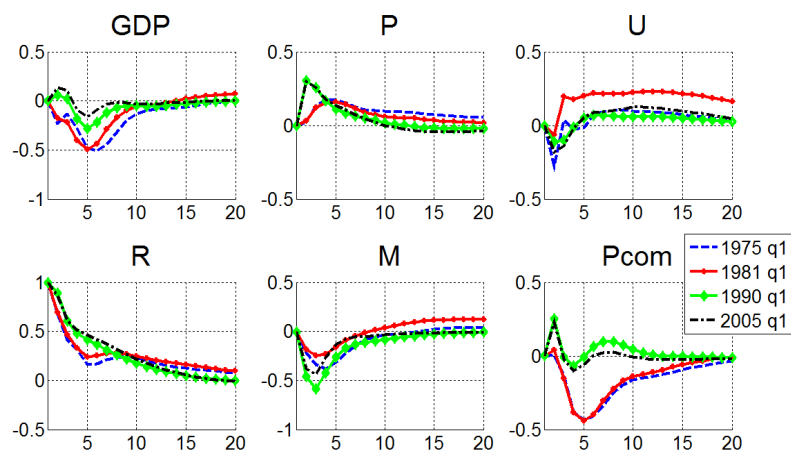
Figure A.9: Impulse responses to monetary shocks (single-move)
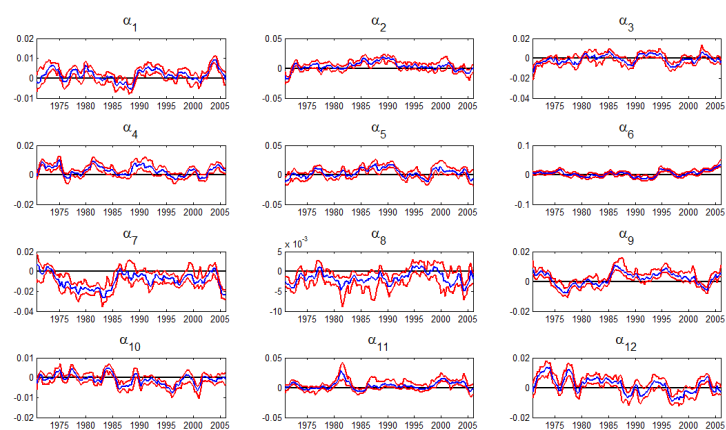
Figure A.10: Estimates for α (single-move)

Appendix B

APPENDIX TO CHAPTER 2

B.1 Impulse responses at selected dates

Figure B.1 depicts responses after a policy shock for years 1975, 1981, 1996 and 2004, dates related with normal times. On the other hand, Figure B.2 depicts responses after a policy shock for years 2009, 2010, 2011 and 2012, i.e. after the Great Financial crisis.For comparison purposes, responses have been normalized such that here is a shock of $\varepsilon_t^s = 1$ in period t. Details about the computation of these responses can be found in Appendix B.2.

B.2 Computation of Impulse Responses in a TVC-SVAR

B.2.1 Setup

Impulse response analysis can be performed using the posterior distribution of parameter blocks and structural shocks. Following Canova (2007) and Canova and Ciccarelli (2009), the Impulse Response Function in a Time-Varying-Coefficients setup can be interpreted as the difference between two conditional expectations,

$$IR\left(t, t+j\right) = E\left(\mathbf{Y}_{t+j} \mid \mathbf{I}_1\right) - E\left(\mathbf{Y}_{t+j} \mid \mathbf{I}_2\right), \quad j = 0, 1, 2, \ldots \qquad \text{(B.1)}$$

where the information sets are

$$\mathbf{I}_1 = \left\{\mathbf{Y}^{t-1}, B^t, V, \alpha_t, \widetilde{c}_t, \sigma_t, \varepsilon_{i,t} = \widetilde{\varepsilon}_{i,t}, \varepsilon_{-i,t}\right\}$$

$$\mathbf{I}_2 = \left\{\mathbf{Y}^{t-1}, B^t, V, \alpha_t, \widetilde{c}_t, \sigma_t, \varepsilon_{i,t} = \overline{\varepsilon}_{i,t}, \varepsilon_{-i,t}\right\}$$

Figure B.1: Responses after a Monetary Policy shocks and before the Great Financial Crisis, 90 percent bands
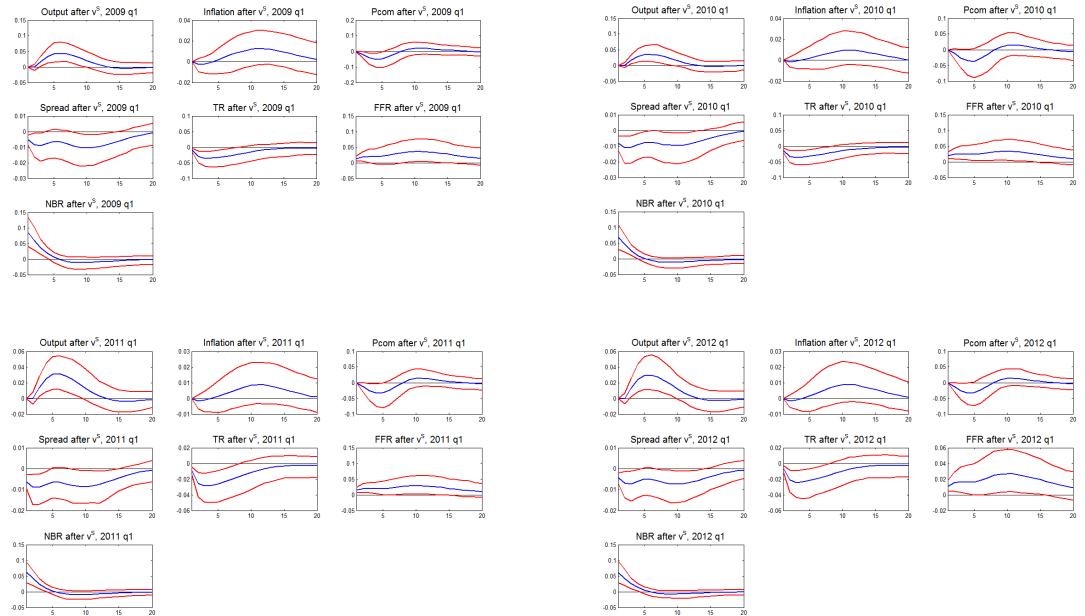
Figure B.2: Responses after a Monetary Policy shocks after the Great Financial Crisis, 90 percent bands

110

That is, given a particular realization of the $i-$th structural shock at time t, $\varepsilon_{i,t} = \widetilde{\varepsilon}_{i,t}$, we compare the forecast using this information with respect to a different forecast using the posterior mean $\varepsilon_{i,t} = \overline{\varepsilon}_{i,t}$.

We now proceed to explain the details. Denote the companion form of (2.4) as

$$\mathbf{Y}_{t+j} = \mathbf{A}_t^c \mathbf{Y}_{t+j-1} + \mathbf{U}_{t+j} \tag{B.2}$$

where

$$\mathbf{A}_t^c = \begin{bmatrix} B_{1,t} & B_{2,t} & \cdots & B_{p-1,t} & B_{p,t} \\ I_M & 0 & \cdots & 0 & 0 \\ 0 & I_M & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_M & 0 \end{bmatrix} \tag{B.3}$$

is a $(Mp \times Mp)$ matrix, $\mathbf{Y}_{t+j}$ is a demeaned version of $\left[y'_{t+j}, y'_{t+j-1}, \ldots, y'_{t+j-p+1} \right]'$ and $\mathbf{U}_{t+j} = \left[u'_{t+j}, \mathbf{0}', \ldots, \mathbf{0}' \right]'$. Solving backwards equation $(B.2)$ yields
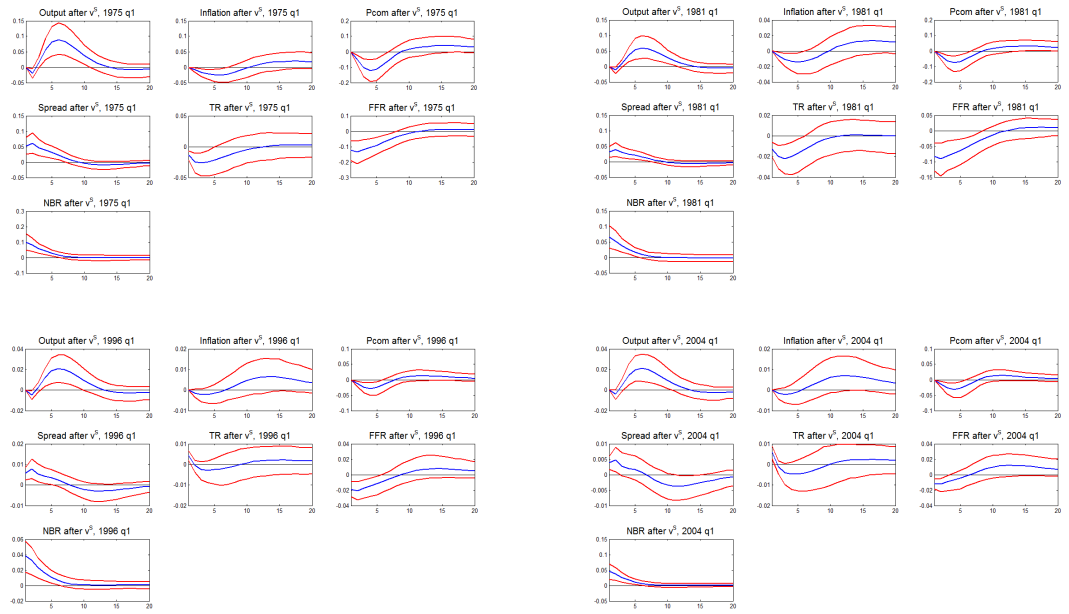
$$\mathbf{Y}_{t+j} = \mathbf{A}_t^c \left(\mathbf{A}_{t-1}^c \mathbf{Y}_{t+j-2} + \mathbf{U}_{t+j-1} \right) + \mathbf{U}_{t+j}$$

$$\mathbf{Y}_{t+j} = \mathbf{A}_t^c \mathbf{A}_{t-1}^c \left(\mathbf{A}_{t-2}^c \mathbf{Y}_{t+j-3} + \mathbf{U}_{t+j-2} \right) + \mathbf{A}_t^c \mathbf{U}_{t+j-1} + \mathbf{U}_{t+j}$$

$$\mathbf{Y}_{t+j} = \mathbf{A}_t^c \mathbf{A}_{t-1}^c \mathbf{A}_{t-2}^c \mathbf{Y}_{t+j-3} + \mathbf{A}_t^c \mathbf{A}_{t-1}^c \mathbf{U}_{t+j-2} + \mathbf{A}_t^c \mathbf{U}_{t+j-1} + \mathbf{U}_{t+j}$$

$$\cdots$$

$$\mathbf{Y}_{t+j} = \left(\prod_{k=0}^{j} \mathbf{A}_{t+j-k}^c \right) \mathbf{Y}_{t-1} + \sum_{h=1}^{j} \left(\prod_{k=0}^{h-1} \mathbf{A}_{t+j-k}^c \right) \mathbf{U}_{t+j-h} \tag{B.4}$$

Recall also that

$$\mathbf{U}_{t+j} = \left[u'_{t+j}, \mathbf{0}', \ldots, \mathbf{0}' \right]'$$

which can be written as

$$\mathbf{U}_{t+j} = \mathbf{H}_{t+j} \mathbf{E}_{t+j}$$

where

$$\mathbf{H}_{t+j} = \begin{bmatrix} H_{t+j} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

$\mathbf{E}_{t+j} = \left[\varepsilon'_{t+j}, \mathbf{0}', \ldots, \mathbf{0}' \right]'$ and $H_{t+j} = A_{t+j}^{-1} C_{t+j} \Sigma_{t+j}$. As a result, equation $(B.4)$ is re-written as

$$\mathbf{Y}_{t+j} = \left(\prod_{k=0}^{j} \mathbf{A}_{t+j-k}^c \right) \mathbf{Y}_{t-1} + \sum_{h=1}^{j} \left(\prod_{k=0}^{h-1} \mathbf{A}_{t+j-k}^c \right) \mathbf{H}_{t+j-h} \mathbf{E}_{t+j-h} \tag{B.5}$$

111

Next step is to compute the conditional expectations and take differences to pin down impulse responses as in $(B.1)$. Two aspects are worth to remark, i) the first term in $(B.5)$ will be common for both conditional expectations given the information sets $\mathbf{I}_1$ and $\mathbf{I}_2$, therefore it will be cancelled, ii) past realizations of $\mathbf{E}_{t+j-h}$ are also common to $\mathbf{I}_1$ and $\mathbf{I}_2$, thus the only term that survives in the summation is the one indexed with t (where $h = j$). As a result $(B.1)$ can be written as

$$IR\left(t, t+j\right) = E_t\left[\left(\prod_{k=0}^{j-1}\mathbf{A}^c_{t+j-k}\right)\mathbf{H}_t\left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right)\right] \tag{B.6}$$

Since the latter expression $(B.6)$ is based on the companion form vector $\mathbf{Y}_{t+j}$, it is useful to re-write it in terms of its original elements. First, define the $(M \times Mp)$ selection matrix $\mathbf{J} = \begin{bmatrix} I_M & 0 & \cdots & 0 \end{bmatrix}$ and notice that $y_{t+j} = \mathbf{J}\mathbf{Y}_{t+j}$, $\mathbf{J}\mathbf{J}' = I_M$ and $\mathbf{J}'\mathbf{J} = I_{Mp}$. As a result, $H_t = \mathbf{J}\mathbf{H}_t\mathbf{J}'$, $\mathbf{J}'\mathbf{J}\left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right) = \left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right)$ and $\mathbf{J}\left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right) = \left(\widetilde{\varepsilon}_t - \overline{\varepsilon}_t\right)$, where $\widetilde{\varepsilon}_t = [\widetilde{\varepsilon}_{i,t}, \varepsilon_{-i,t}]$ and $\overline{\varepsilon}_t = [\overline{\varepsilon}_{i,t}, \varepsilon_{-i,t}]$. Premultiplying $(B.6)$ by $\mathbf{J}$ yields

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = E_t\left[\mathbf{J}\left(\prod_{k=0}^{j-1}\mathbf{A}^c_{t+j-k}\right)\mathbf{J}'\mathbf{J}\mathbf{H}_t\mathbf{J}'\mathbf{J}\left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right)\right]$$

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = E_t\left[\mathbf{J}\left(\prod_{k=0}^{j-1}\mathbf{A}^c_{t+j-k}\right)\mathbf{J}'\left(\mathbf{J}\mathbf{H}_t\mathbf{J}'\right)\left\{\mathbf{J}\left(\widetilde{\mathbf{E}}_t - \overline{\mathbf{E}}_t\right)\right\}\right]$$

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = E_t\left[\mathbf{J}\left(\prod_{k=0}^{j-1}\mathbf{A}^c_{t+j-k}\right)\mathbf{J}'H_t\left(\widetilde{\varepsilon}_t - \overline{\varepsilon}_t\right)\right] \tag{B.7}$$

The latter expression is equivalent to equation (2.21).

B.2.2   Algorithm for computing impulse responses

Denote the set of posterior estimates as

$$PE = \left\{B_i^T, \alpha_i^T, \widetilde{c}_i^T, \sigma_i^T, s_i^T, Q_i, Sa_i, Sc_i, W_i\right\}_{i=1}^{Nd}$$

Then use the following algorithm to compute responses to shock in variable $m \in \{1, \ldots, M\}$ of size δ:

1. Set a date $t \in \{1, \ldots, T\}$ and set $k = 1$.

2. For $l = 1, \ldots, L$ do the following:

   a) Draw a random number $\overline{R} \sim U(1, Nd)$, take the nearest integer R.

   b) Pick the elements $\left(B_R^T, \alpha_R^T, \widetilde{c}_R^T, \sigma_R^T, s_R^T, Q_R, S_{a_R}, S_{c_R}, W_R\right)$ and set $B_t^R, \alpha_t^R, \widetilde{c}_t^R, \log\left(\sigma_t^R\right)$ and y_t as initial values.

   c) Construct the matrix of impacts $H_t^R = \left(A_t^R\right)^{-1} C_t^R \Sigma_t^R$. Then draw $\varepsilon_{1,t}^R \sim N(0, I_M)$, set $u_{1,t}^l = H_t^R \varepsilon_{1,t}^l$ and $u_{2,t}^l = H_t^R \varepsilon_{2,t}^l$, where $\varepsilon_{2,t}^l = \varepsilon_{1,t}^l$ but set the entry m of $\varepsilon_{2,t}^l$ equal to δ.

   d) For $j = 1, \ldots, h - 1$ do the following:

   i) Forecast $(B_{t+j}, \alpha_{t+j}, \widetilde{c}_{t+j}, \log(\sigma_{t+j}))$ using $(Q_R, S_{a_R}, S_{c_R}, W_R)$ and the equations (2.7), (2.8), (2.9) and (2.10).

   ii) Construct the matrix of impacts $H_{t+j} = (A_{t+j})^{-1} C_{t+j} \Sigma_{t+j}$. Then draw $\varepsilon_{1,t+j}^l \sim N(0, I_M)$, set $u_{1,t+j}^l = H_{t+j} \varepsilon_{1,t+j}^l$ and $u_{2,t+j}^l = u_{1,t+j}^l$.

   e) For $j = 0, \ldots, h - 1$ do the following:

   i) Forecast $y_{1,t+j}^l$ and $y_{2,t+j}^l$ by using the companion form $(B.2)$ and using B_{t+j} and $u_{1,t+j}^l$ and $u_{2,t+j}^l$.

   ii) Compute $IR(t + j, k, l) = y_{1,t+j}^l - y_{2,t+j}^l$

   f) Define $IR(t : t + h, k, l) = \{IR(t + j, k, l)\}_{j=0}^{h-1}$.

3. Take averages

$$IR(t : t + h, k) = \frac{1}{L} \sum_{l=1}^{L} IR(t : t + h, k, l)$$

4. If $k < N$, set $k = k + 1$ and go back to 2. Otherwise stop.

  Finally, take percentiles over $\{IR(t : t + h, k)\}_{k=1}^{N}$.

B.3   Sampling Parameter blocks

This section takes an extended version of the algorithm described in chapter 1. We describe the sampling procedure for parameter blocks $\left(B^T, \alpha^T, \widetilde{c}^T, \Sigma^T, s^T, V\right)$ and we do it sequentially using the logic of Gibbs Sampling (see Chib (2001)). We emphasize how to sample blocks $\left(\alpha^T, \widetilde{c}^T\right)$ and we refer the reader to Primiceri (2005)'s Appendix A for specific details regarding sampling blocks $\left(B^T, \Sigma^T, s^T, V\right)$.

B.3.1   Setting the State Space form for matrices A_t and C_t^{-1}

Consider the state space model generated after sampling the reduced-form coefficients $\widehat{B}_t$. From (1.26) let

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

Then the state-space form can be written as

$$\widetilde{y}_t = Z_{\alpha,t}\alpha_t + C_t\Sigma_t\varepsilon_t \tag{B.8}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{B.9}$$

where $\widetilde{y}_t$ and $Z_{\alpha,t}$ are defined in subsection B.3.4, α_t are the free elements in A_t and $Var\left(\zeta_t\right) = S_a$.

Similarly, consider the following state space model generated after sampling the vector α_t. From (1.26) let

$$C_t^{-1}A_t\left(y_t - X_t'\widehat{B}_t\right) = C_t^{-1}\widetilde{y}_t = \overrightarrow{y}_t = \Sigma_t\varepsilon_t$$

Then the state-space form can be written as

$$\overrightarrow{y}_t = Z_{\widetilde{c},t}\widetilde{c}_t + \Sigma_t\varepsilon_t \tag{B.10}$$

$$\widetilde{c}_t = \widetilde{c}_{t-1} + \varrho_t \tag{B.11}$$

where $Z_{\widetilde{c},t}$ is also defined in subsection B.3.4, $\widetilde{c}_t$ are the free elements in C_t^{-1} and $Var\left(\varrho_t\right) = S_{\widetilde{c}}$.

B.3.2   The algorithm

Let $\left\{\{s_{l,t}\}_{t=1}^T\right\}_{l=1}^M$ be a discrete indicator variable which takes $j = 1,\ldots,k$ possible values. The procedure has 8 steps and 6 sampling blocks:

1. Set an initial value for $(B_0^T, \alpha_0^T, \widetilde{c}_0^T, \Sigma_0^T, s_0^T, V_0)$ and set $i = 1$.

2. Draw B_i^T from from $p\left(B_i^T \mid \alpha_{i-1}^T, \widetilde{c}_{i-1}^T, \Sigma_{i-1}^T, s_{i-1}^T, V_{i-1}\right) \cdot I_B\left(B_i^T\right)$ using kalman smoothed estimates $B_{t|T}$ obtained from the system $(1.26) - (2.7)$ and compute $\widehat{y}_i^T$, where $I_B\left(.\right)$ truncates the posterior distribution to insure stationarity of companion form.

3. Draw α_i^T from

$$
\begin{aligned}
p\left(\alpha_i^T \mid \widehat{y}_i^T, \widetilde{c}_{i-1}^T, \Sigma_{i-1}^T, s_{i-1}^T, V_{i-1}\right) &= p\left(\alpha_{i,T} \mid \widehat{y}_i^T, \widetilde{c}_{i-1,T}, \Sigma_{i-1,T}, s_{i-1,T}, V_{i-1}\right) \times \tag{B.12} \\
&\qquad \prod_{t=1}^{T-1} p\left(\alpha_{i,t} \mid \alpha_{i,t+1}, \widehat{y}_i^t, \widetilde{c}_{i-1,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_{i-1}\right) \\
&\propto p\left(\alpha_{i,T} \mid \widehat{y}_i^T, \widetilde{c}_{i-1,T}, \Sigma_{i-1,T}, s_{i-1,T}, V_{i-1}\right) \times \\
&\qquad \prod_{t=1}^{T-1} p\left(\alpha_{i,t} \mid \widehat{y}_i^t, \widetilde{c}_{i-1,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_i\right) \times \\
&\qquad f_{t+1}\left(\alpha_{i,t+1} \mid \alpha_{i,t}, \widetilde{c}_{i-1,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_{i-1}\right)
\end{aligned}
$$

114

4. Draw $\widetilde{c}_i^T$ from

$$
\begin{aligned}
p\left(\widetilde{c}_i^T \mid \overrightarrow{y}_i^T, \alpha_i^T, \Sigma_{i-1}^T, V_{i-1}\right) &= p\left(\widetilde{c}_{i,T} \mid \overrightarrow{y}_i^T, \alpha_{i,T}, \Sigma_{i-1,T}, s_{i-1,T}, V_{i-1}\right) \quad \text{(B.13)} \\
&\quad \times \prod_{t=1}^{T-1} p\left(\widetilde{c}_{i,t} \mid \widetilde{c}_{i,t+1}, \overrightarrow{y}_i^t, \alpha_{i,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_{i-1}\right) \\
&\propto p\left(\widetilde{c}_{i,T} \mid \overrightarrow{y}_i^T, \alpha_{i,T}, \Sigma_{i-1,T}, s_{i-1,T}, V_{i-1}\right) \\
&\quad \times \prod_{t=1}^{T-1} p\left(\widetilde{c}_{i,t} \mid \overrightarrow{y}_i^t, \alpha_{i,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_{i-1}\right) \\
&\quad \times f_{t+1}\left(\widetilde{c}_{i,t+1} \mid \widetilde{c}_{i,t}, \alpha_{i,t}, \Sigma_{i-1,t}, s_{i-1,t}, V_{i-1}\right)
\end{aligned}
$$

5. Draw Σ_i^T using a log-normal approximation to their distribution as in Kim et al. (1998). After sampling $\left(B_i^T, \alpha_i^T, \widetilde{c}_i^T\right)$, the state space is linear but the error term is not normally distributed. To see this, note that given $\left(B_i^T, \alpha_i^T, \widetilde{c}_i^T\right)$, the state space system is

$$
\widehat{C}_t^{-1}\widehat{A}_t\widetilde{y}_t = y_t^{**} = \Sigma_t\varepsilon_t
$$

and (2.10). Consider the $i-th$ equation $y_{i,t}^{**} = \sigma_{i,t}\varepsilon_{i,t}$, where $i = 1, \ldots, M$, $\sigma_{i,t}$ is the $i-$th element in the main diagonal of Σ_t and $\varepsilon_{i,t}$ is the $i-$th element of ε_t.

$$
y_t^* = \log\left[\left(y_{i,t}^{**}\right)^2 + \overline{c}\right] \approx 2\log\left(\sigma_{i,t}\right) + \log\varepsilon_{i,t}^2 \quad \text{(B.14)}
$$

Then where $\overline{c}$ is a small constant. Since $\varepsilon_{i,t}$ is Gaussian, $\log\varepsilon_{i,t}^2$ is $\log\left(\chi^2\right)$ distributed and we approximate this distribution by a mixture of normals. Since conditional on s_t, the model is linear and Gaussian, standard Kalman Filter recursions can be used to draw $\{\Sigma_t\}_{t=1}^T$ from the system $(B.14) - (2.10)$. To ensure independence across structural variances, each element of the sequence $\{\sigma_{i,t}\}_{i=1}^M$ is sampled assuming that the covariance matrix W is diagonal.

6. Draw the indicator of mixture of normals s_i^T. Conditional on Σ_i^T, y_t^*, and given l and t, we draw $u \sim U(0,1)$ and compare it with the discrete distribution of $s_{l,t}$ which is given by

$$
\begin{aligned}
P\left(s_{l,t} = j \mid y_{l,t}^*, \log\left(\sigma_{l,t}\right)\right) &\propto q_j \times \phi\left(\frac{y_{l,t}^* - 2\log\left(\sigma_{l,t}\right) - m_j + 1.2704}{v_j}\right); \\
j &= 1, \ldots, k; l = 1, \ldots, M
\end{aligned}
$$

where $\phi\left(.\right)$ is the probability density function of a normal distribution, and the argument of this function is the standarized error term $\log\varepsilon_{l,t}^2$ (see Kim

115

et al. (1998)). Then we assign $s_{l,t} = j$ iff $P\left(s_{l,t} \leq j - 1 \mid y_{l,t}^*, \log\left(\sigma_{l,t}\right)\right) < u \leq P\left(s_{l,t} \leq j \mid y_{l,t}^*, \log\left(\sigma_{l,t}\right)\right)$.

7. Draw V_i from $P\left(V_i \mid \alpha_i^T, \widetilde{c}_i^T, y_i^T, \Sigma_{i-1}^T, s_{i-1}^T\right)$ using definitions $(2.7) - (2.10)$. The covariance matrix V_i is sampled assuming that each block follows an independent Wishart distribution.

8. Set $B_i^T, \alpha_i^T, \widetilde{c}_i^T, \Sigma_i^T, s_i^T, V_i$ as the initial value for the next iteration and set $i = i + 1$. Repeat 2 to 7 if $i < N$, otherwise stop.

The complete cycle of draws is repeated $N = Nb + Nd$ times and the first Nb draws are discarded to ensure convergence in distribution. Because the draws are generally serially correlated, one every $nthin$ of the last Nd draws is used for inference.

B.3.3 The details in steps 3 and 4

This subsection follows chaper 1. For steps 3 and 4 we use a Metropolis step to decide whether a draw from a proposal distribution is retained or not.

We only illustrate the case of sampling vector α_t, since sampling vector $\widetilde{c}_t$ will be completely symmetric following the Multi-move Gibbs Sampling logic. The densities $p\left(\alpha_t \mid \widehat{y}^t, \widetilde{c}_t, \Sigma_t, s, V\right)$ are obtained applying the Extended Kalman Smoother to the original system $(1.33) - (B.9)$. To draw α_i^T given $\widehat{y}_i^T, \widetilde{c}_{i-1}^T, \Sigma_{i-1}^T, s_{i-1}^T, V_{i-1}$, we proceed as follows:

1. If $i = 0$, take an initial value $\alpha_0^T = \{\alpha_{0,t}\}_{t=1}^T$. If not,

2. Given α_{i-1}^T, compute $\left\{\alpha_{t|t+1}^{*(i-1)}\right\}_{t=1}^T$ and $\left\{P_{t|t+1}^{*(i-1)}\right\}_{t=1}^T$ using the EKS where $\left\{P_{t|t+1}^*\right\}_{t=1}^T$ denotes the covariance matrix of $\left\{\alpha_{t|t+1}^*\right\}_{t=1}^T$.

3. Generate a candidate draw $z^T = \{z_t\}_{t=1}^T$, where for each $t = 1, \ldots, T$ $p_{*\alpha}\left(z_t \mid \alpha_{i-1,t}\right) = N\left(\alpha_{i-1,t}, rP_{t|t+1}^{*(i-1)}\right)$, and r is a constant. Let $p_{*\alpha}\left(z^T \mid \alpha_{i-1}^T\right) = \prod_{t=1}^T p_{*\alpha}\left(z_t \mid \alpha_{i-1,t}\right)$.

4. Compute $\theta = \frac{p\left(z^T\right) \cdot p_{*\alpha}\left(\alpha_{i-1}^T \mid z^T\right)}{p\left(\alpha_{i-1}^T\right) \cdot p_{*\alpha}\left(z^T \mid \alpha_{i-1}^T\right)}$ where $p\left(.\right)$ is the RHS of $(B.12)$ using the EKS approximation. Draw a $v \sim U\left(0, 1\right)$. Set $\alpha_i^T = z^T$ if $v < \omega$ and set $\alpha_i^T = \alpha_{i-1}^T$ otherwise, where

$$\omega \equiv \begin{cases} \min\left\{\theta, 1\right\}, & \text{if } I_\alpha\left(z^T\right) = 1 \\ 0, & \text{if } I_\alpha\left(z^T\right) = 0 \end{cases}$$

116

and $I_\alpha\left(.\right)$ is a truncation indicator.

Finally, steps 2 to 4 in this sub-loop are repeated every time step 3 of the main loop is executed.

B.3.4   The identified system

Recall the expression

$$A_t\widehat{y}_t = C_t\Sigma_t\varepsilon_t$$

The expression of the measurement equation is possible to obtain since $vec\left(A_t\widehat{y}_t\right) = vec\left(C_t\Sigma_t\varepsilon_t\right)$. Then, using the fact that $A_{t(M\times M)}$, $\widehat{y}_{t(M\times 1)}$, $\Sigma_{t(M\times M)}$, $\varepsilon_{t(M\times 1)}$, then

$$vec\left(A_t\widehat{y}_t\right) = vec\left(I_M A_t\widehat{y}_t\right) = \left(\widehat{y}_t' \otimes I_M\right)vec\left(A_t\right)$$

and also

$$vec\left(C_t\left(\Sigma_t\varepsilon_t\right)\right) = \left(\left(\Sigma_t\varepsilon_t\right)' \otimes I_M\right)vec\left(C_t\right) = \left(I_1 \otimes C_t\right)vec\left(\Sigma_t\varepsilon_t\right) = C_t\Sigma_t\varepsilon_t$$

since $A_t\widehat{y}_t$ and $C_t\Sigma_t\varepsilon_t$ are already column vectors[1]. On the other hand, following Amisano and Giannini (1997) and Hamilton (1994), we also know that the matrix of the SVAR can be decomposed as follows

$$vec\left(A_t\right) = S_A\alpha_t + s_A \tag{B.15}$$

$$vec\left(C_t^{-1}\right) = S_C F\left(c_t\right) + s_C \tag{B.16}$$

where $S_{A(M^2\times\dim\alpha)}$, $s_{A(M^2\times 1)}$, $S_{C(M^2\times\dim F(c))}$ and $s_{C(M^2\times 1)}$ are matrices filled by ones and zeros. Moreover, α_t, c_t and $F\left(c_t\right)$ are the vectors of free parameters in A_t, C_t and C_t^{-1}, respectively and $F\left(.\right) : R^{\dim(c)} \rightarrow R^{\dim F(c)}$ is in general a nonlinear invertible function. That is, we sample the vector $\left\{F\left(c_t\right)\right\}_{t=1}^T$ and if and only if $F\left(.\right)$ is invertible, then we can recover $\left\{c_t\right\}_{t=1}^T = \left\{F^{-1}\left[F\left(c_t\right)\right]\right\}_{t=1}^T$. We will denote $\widetilde{c}_t = F\left(c_t\right)$. Collecting all the results we get

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

Rewriting this equation

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

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

---

[1]In general, we have applied the property $ABd = \left(d' \otimes A\right)vec\left(B\right)$, where A is an $m \times n$ matrix, B is an $n \times q$ matrix and d is a $\left(q \times 1\right)$ vector. See details in Magnus and Neudecker (2007), chapter 2, pp. 35.

The state space form is now

$$\widetilde{y}_t = Z_{\alpha,t}\alpha_t + C_t\Sigma_t\varepsilon_t$$

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

where

$$\widetilde{y}_t \equiv \left(\widehat{y}_t' \otimes I_M\right)s_A \tag{B.17}$$

$$Z_{\alpha,t} \equiv -\left(\widehat{y}_t' \otimes I_M\right)S_A \tag{B.18}$$

On the other hand, given α_t, we proceed in the following way

$$vec\left(C_t^{-1}A_t\widehat{y}_t\right) = vec\left((\Sigma_t\varepsilon_t)\right) = \Sigma_t\varepsilon_t$$

$$\left((A_t\widehat{y}_t)' \otimes I_M\right)vec\left(C_t^{-1}\right) = \Sigma_t\varepsilon_t$$

$$\left((A_t\widehat{y}_t)' \otimes I_M\right)\left(S_C\widetilde{c}_t + s_C\right) = \Sigma_t\varepsilon_t$$

$$\left((A_t\widehat{y}_t)' \otimes I_M\right)s_C = -\left((A_t\widehat{y}_t)' \otimes I_M\right)S_C\widetilde{c}_t + \Sigma_t\varepsilon_t$$

The state space form is now

$$\overrightarrow{y}_t = Z_{C,t}\widetilde{c}_t + \Sigma_t\varepsilon_t$$

$$\widetilde{c}_t = \widetilde{c}_{t-1} + \varrho_t$$

where

$$\overrightarrow{y}_t \equiv \left((A_t\widehat{y}_t)' \otimes I_M\right)s_C \tag{B.19}$$

$$Z_{C,t} \equiv -\left((A_t\widehat{y}_t)' \otimes I_M\right)S_C \tag{B.20}$$

Moreover, for the specific case of the model presented, we have that $\widehat{y}_t$ denotes the residuals for the first stage and also the matrices

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{\pi} & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{c} & \alpha_{\pi,t}^{c} & 1 & 0 & 0 & 0 & 0 \\ \alpha_{x,t}^{SPR} & \alpha_{\pi,t}^{SPR} & \alpha_{c,t}^{SPR} & 1 & 0 & \alpha_{FFR,t}^{SPR} & 0 \\ \alpha_{x,t}^{TR} & \alpha_{\pi,t}^{TR} & \alpha_{c,t}^{TR} & 0 & 1 & \alpha_{1,t}^{d} & 0 \\ \alpha_{x,t}^{FFR} & \alpha_{\pi,t}^{FFR} & \alpha_{c,t}^{FFR} & 0 & 1 & -\alpha_{1,t}^{b} & -1 \\ \alpha_{x,t}^{NBR} & \alpha_{\pi,t}^{NBR} & \alpha_{c,t}^{NBR} & 0 & 0 & 0 & 1 \end{bmatrix} \tag{B.21}$$

and

$$C_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \phi_t^{f} & \phi_t^{d} & \phi_t^{b} & 1 \end{bmatrix} \tag{B.22}$$

118

therefore, $M = 7$, $\dim \alpha = 18$. As it is evident, C_t is a lower-triangular matrix with the main diagonal governed by ones, i.e. a unitriangular matrix. Moreover, C_t' will be unitriangular as well and in this case it can also be classified as a Frobenius matrix[2]. The inverse of a Frobenius matrix X is exactly $X^{-1} = -X$. Thus, provided by the fact that $[C_t']^{-1} = [C_t^{-1}]'$, we have that $[C_t']^{-1} = [C_t^{-1}]' = -C_t' \Rightarrow C_t^{-1} = -C_t$, hence

$$C_t^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\phi_t^f & -\phi_t^d & -\phi_t^b & 1 \end{bmatrix}$$

and $\dim F(c) = \dim \widetilde{c} = 3$. Therefore, it turns out that in this particular case the function $F(.)$ is actually a linear transformation, i.e. $\widetilde{c}_t = F(c_t) = -c_t$. That is, as long as C_t' is Frobenius matrix, $F(.)$ will be linear with a well defined $F^{-1}(.)$.

As a result, we need to define matrices S_A, s_A, S_C, s_C filled by 0s and 1s. These matrices, together with the column vectors

$$\alpha_t = \begin{bmatrix} \alpha_{x,t}^{\pi}, \alpha_{x,t}^{Pcom}, \alpha_{x,t}^{SPR}, \alpha_{x,t}^{TR}, \alpha_{x,t}^{NBR}, \alpha_{x,t}^{FFR}, \alpha_{\pi,t}^{Pc}, \alpha_{\pi,t}^{SPR}, \alpha_{\pi,t}^{TR}, \\ \alpha_{\pi,t}^{NBR}, \alpha_{\pi,t}^{FFR}, \alpha_{Pc,t}^{SPR}, \alpha_{Pc,t}^{TR}, \alpha_{Pc,t}^{NBR}, \alpha_{Pc,t}^{FFR}, \alpha_{TR,t}^{SPR}, \alpha_{FFR,t}^{SPR}, \alpha_{1,t}^{b} \end{bmatrix}'$$

and

$$\widetilde{c}_t = \begin{bmatrix} -\phi_t^f, -\phi_t^d, -\phi_t^b \end{bmatrix}'$$

are set such that equations $(B.15)$ and $(B.16)$ hold exactly.

B.4 Diagnosis of convergence of the Markov Chain to the Ergodic Distribution

In this section we closely follow the approach described in Appendix A.4. Figure B.3 depicts the IF value for each parameter of the model. Overall, serial correlation across draws does not seem to be an issue.

B.4.1 Markov Chain plots

Figures B.4 and B.5 depict the evolution of the Markov chain for the non-discarded draws for selected parameters. Recall that we have discarded the first $100,000$ draws and kept 1 for every 100 draws for the remaining $50,000$ draws.

---

[2]http://en.wikipedia.org/wiki/Frobenius_matrix
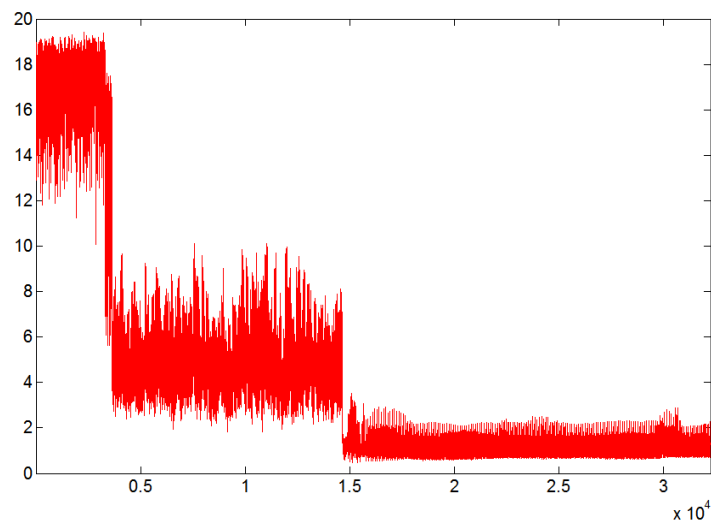
Figure B.3: Inefficiency Factor IF for each parameter in the model
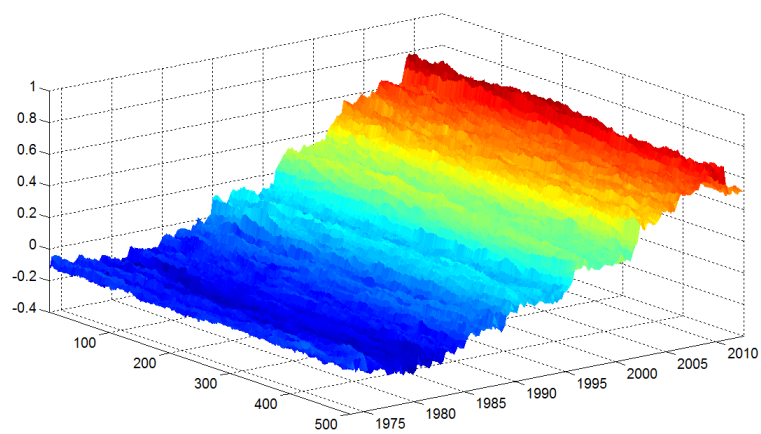
Figure B.4: MCMC draws of parameter $\alpha_{6,t}$

120
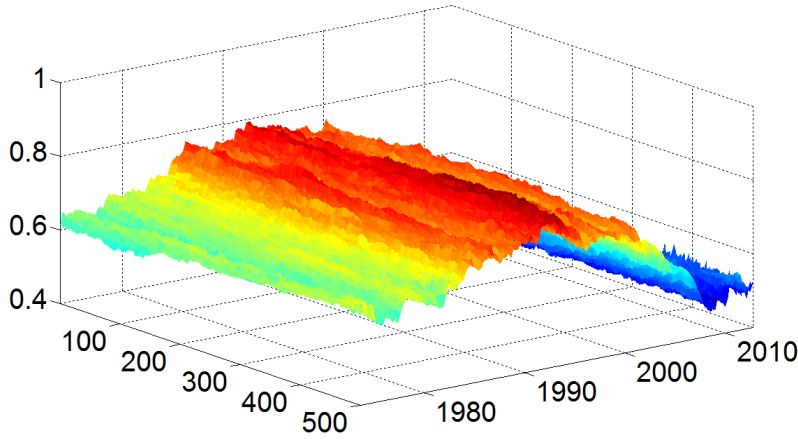
Figure B.5: MCMC draws of parameter ϕ_t^d

In all figures parameter fluctuations are small indicating that convergence to the ergodic distribution has already occurred.

Figures B.6 and B.7 present rolling estimates of the diagonal elements of the covariance matrix, where estimates at each point are obtained by adding 10 draws, and we do this for every t. we For space reasons, we only show the results for $t = 140$. In almost all the cases estimates of the variances are stable or do not fluctuate too much (values are most of them in the order of 1×10^{-3}), which suggest that the chain has reached convergence to the ergodic distributions (see Casella and Robert (2004)).

B.4.2   Histograms

Figure B.8 presents histograms of coefficient ϕ_t^d at selected dates. The empirical distribution look broadly unimodal, which is also a good indication of convergence.

121
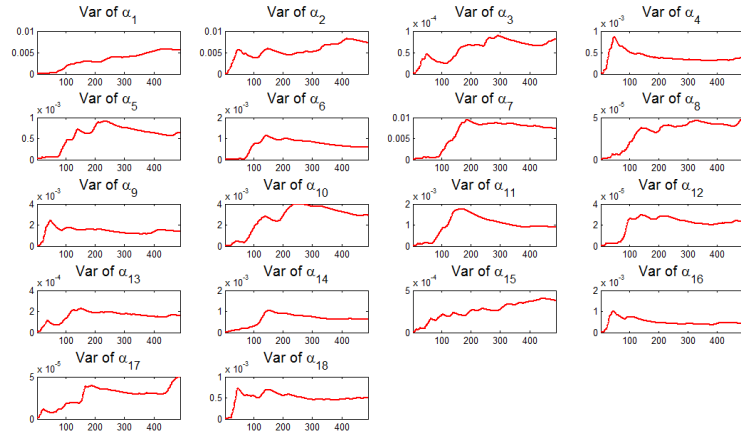
Figure B.6: Cumulative variances of vector α_t
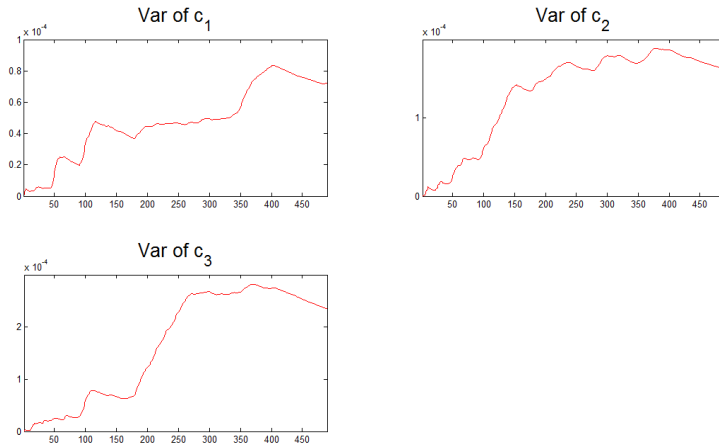
Figure B.7: Cumulative variances of vector $\widetilde{c}_t$
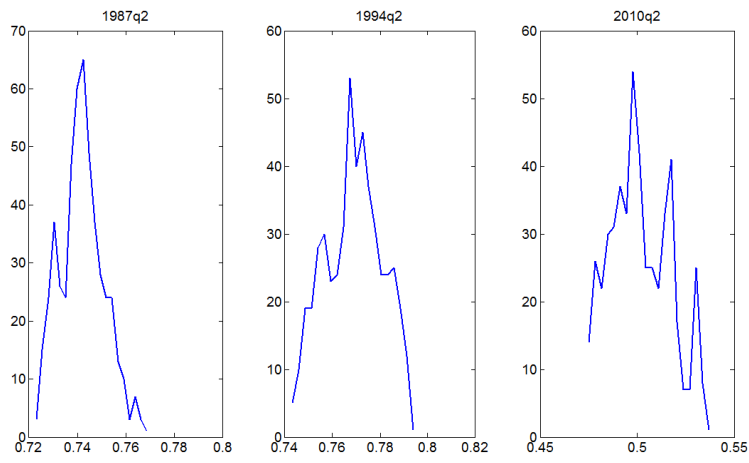
Figure B.8: Histograms of parameter ϕ_t^d

Appendix C

APPENDIX TO CHAPTER 3

C.1 Data Description

Figure C.1 depicts the Yen to Dollar Exchange Rate and the Interest rates for both Japan and United Sates.

Figure C.2 shows details regarding the Survey of Expectations. Published forecasts are for march and september. Data for december and june was linearly interpolated.

Finally, Figure C.3 provides the reference of how are expectations computed in the TANKAN Survey based on the information provided by firms in Japan.

C.2 Convergence properties of the Markov Chain

Ergodicity is a desired property for the output of a Markov Chain Monte Carlo (MCMC) experiment. This turns to be more important in the context of adaptive proposal distributions, since now draws depend on the full history of the chain, i.e. the serial correlation across draws is potentially very high. Figure C.4 shows that draws have achieved convergence to the mean of the ergodic distribution.

In addition, Figure C.5 shows rolling covariances across draws. In this case, draws have achieved convergence to the variance of the ergodic distribution.

C.3 Mixtures of Normals

C.3.1 Basic setup

Let X_1, X_2 denote random variables from the 2 component distributions

$$X_l \sim N\left(\mu_l, \sigma_l^2\right), \;\; l = 1, 2$$
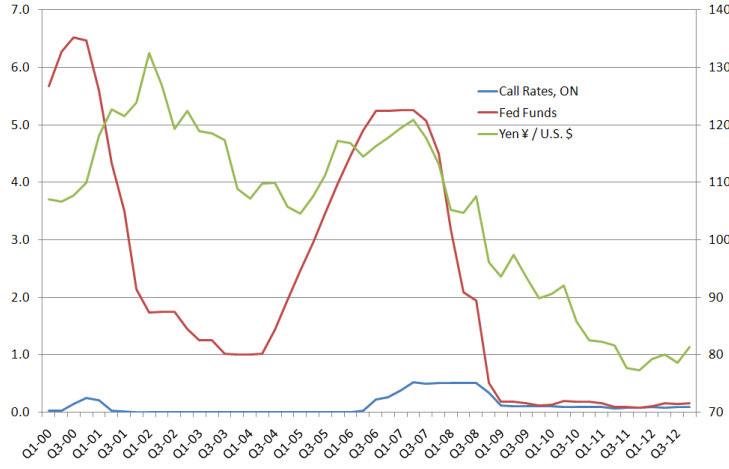
Figure C.1: ER and Interest differentials

which implies that the probability density function for each l is:

$$p_l(x) = \frac{1}{\sigma_l \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{X_l - \mu_l}{\sigma_l}\right)^2\right)$$

and also assume that the weights are ω and $1 - \omega$ and that there is a non-linear transformation for X such as $-\exp\left(-\gamma X\right)$.

The expected value of this transformation is

$$E\left[-\exp\left(-\gamma X\right)\right] = \omega E\left[-\exp\left(-\gamma X_1\right)\right] + (1 - \omega) E\left[-\exp\left(-\gamma X_2\right)\right] \quad \text{(C.1)}$$

Given the normal density, we know that for $l = 1, 2$ the following expression holds[1]:

$$
\begin{aligned}
E\left[-\exp\left(-\gamma X_l\right)\right] &= \int_{-\infty}^{\infty} -\exp\left(-\gamma X_l\right) p_l\left(X_l\right) dX_l \\
&= \frac{1}{\sigma_l \sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\gamma X_l\right) \exp\left(-\frac{1}{2}\left(\frac{X_l - \mu_l}{\sigma_l}\right)^2\right) dX_l \\
&= \frac{1}{\sigma_l \sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\gamma X_l - \frac{1}{2}\left(\frac{X_l - \mu_l}{\sigma_l}\right)^2\right) dX_l
\end{aligned}
$$

Let

$$X_l = \mu_l + \sigma_l z \Rightarrow dX_l = \sigma_l dz$$

---

[1] See also Moss (2010)

126

Details about Survey Data in Exchange Rates

1	Data is available at quarterly frequency in http://www.boj.or.jp/en/statistics/tk/index.htm/
2	Forecats are available for the end of Fiscal Year (FY) - March and middle fiscal Year - September.
3	Forecasts for December are interpolated between September of current year (half of FY) and March of next year (end of FY).
4	Forecasts for June are interpolated between March of current year (end of FY) and September of current year (half of next FY).

Figure C.2: Details about Predicted Exchange Rates by Industries data, BoJ

How predicted exchange rates are computed?

1	This a survey of predicted exchange rates by firms, and it is published by The Bank of Japan in aggregated industries level.
2	More details can be found in: http://www.boj.or.jp/en/statistics/outline/exp/tk/faqtk.htm/#p0304

Figure C.3: Reference to Predicted Exchange Rates (TANKAN), FAQ of The Bank of Japan

then

$$E\left[-\exp\left(-\gamma X_l\right)\right] = \frac{\sigma_l}{\sigma_l\sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\frac{1}{2}\left(z^2 + 2\gamma\sigma_l z + 2\gamma\mu_l\right)\right) dz$$

which means that $z \in (-\infty, \infty) \Leftrightarrow X_l \in (-\infty, \infty)$.

Complete squares:

$$E\left[-\exp\left(-\gamma X_l\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\frac{1}{2}\left(z^2 + 2\gamma\sigma_l z + \gamma^2\sigma_l^2 + 2\gamma\mu_l - \gamma^2\sigma_l^2\right)\right) dz$$

$$E\left[-\exp\left(-\gamma X_l\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\frac{1}{2}\left((z + \gamma\sigma_l)^2 + 2\gamma\mu_l - \gamma^2\sigma_l^2\right)\right) dz$$
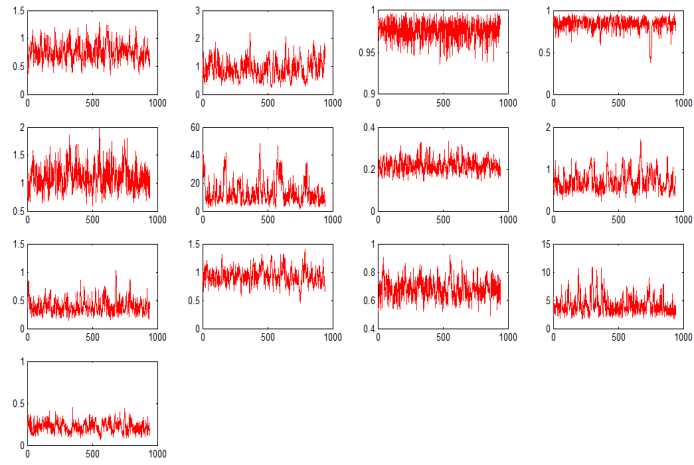
127

Figure C.4: Convergence in mean
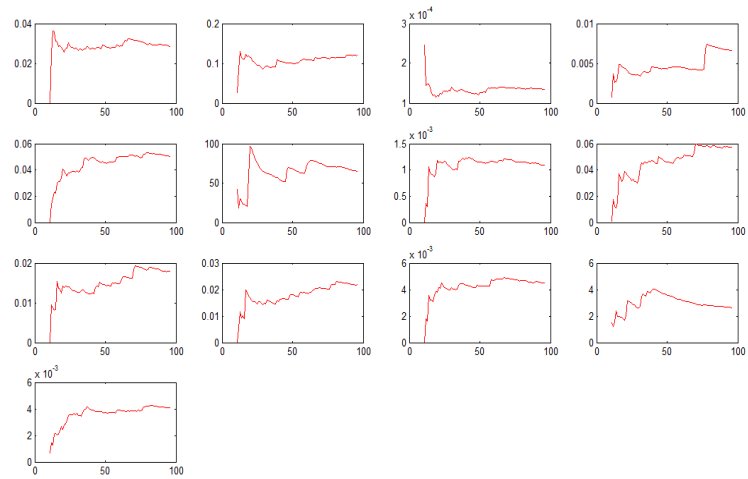
Figure C.5: Convergence in variance

$$E\left[-\exp\left(-\gamma X_l\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\frac{1}{2}\left(z+\gamma\sigma_l\right)^2\right) \exp\left(-\frac{1}{2}\left(2\gamma\mu_l - \gamma^2\sigma_l^2\right)\right) dz$$

$$E\left[-\exp\left(Z_l\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\exp\left(-\frac{1}{2}\left(z+\gamma\sigma_l\right)^2\right) \exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right) dz$$

The second exponential is independent of z, therefore:

$$\begin{aligned}
E\left[-\exp\left(-\gamma X_l\right)\right] &= -\exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right) \times \\
&\qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(z+\gamma\sigma_l\right)^2\right) dz
\end{aligned}$$

where z is a standard normal. Here we use a change of variable. Let

$$c = \frac{z+\gamma\sigma_l}{\sqrt{2}} \Rightarrow \sqrt{2}dc = dz$$

which means that $c \in (-\infty, \infty) \Leftrightarrow z \in (-\infty, \infty)$.

As a consequence:

$$\begin{aligned}
E\left[-\exp\left(-\gamma X_l\right)\right] &= -\exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right) \times \\
&\qquad \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-c^2\right) dc \\
&= -\exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right) \times \\
&\qquad \frac{1}{2} \int_{-\infty}^{\infty} \frac{2}{\sqrt{\pi}} \exp\left(-c^2\right) dc
\end{aligned}$$

Evaluating the integral:

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{2}{\sqrt{\pi}} \exp\left(-c^2\right) dc &= \lim_{c\to\infty} \text{erf}\left(c\right) - \lim_{c\to-\infty} \text{erf}\left(c\right) \\
&= 1 - (-1) \\
&= 2
\end{aligned}$$

where $\text{erf}\left(.\right)$ is the Gaussian Error Function. Using this result:

$$\begin{aligned}
E\left[-\exp\left(-\gamma X_l\right)\right] &= -\exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right) \times \frac{1}{2} \times 2 \qquad \text{(C.2)} \\
&= -\exp\left(-\gamma\left(\mu_l - \frac{\gamma}{2}\sigma_l^2\right)\right)
\end{aligned}$$

Plugging $(C.2)$ in $(C.1)$:

$$E\left[-\exp\left(-\gamma X\right)\right] = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_1^2\right)\right)\right] + (1-\omega)\left[-\exp\left(-\gamma\left(\mu_2 - \frac{\gamma}{2}\sigma_2^2\right)\right)\right]$$
$$\text{(C.3)}$$

C.3.2 Mixture of normals in the ER model

We assume that the information set $\Omega_t(j)$ can take two values according to an underlying dichotomic latent variable τ_t. More formally

$$\Omega_t(j) = \left\{ \begin{array}{ll} \Omega_t^0(j) & \text{, if } \tau_t = 0 \\ \Omega_t^1(j) & \text{, if } \tau_t = 1 \end{array} \right.$$

In addition, assume that $\omega = \Pr(\tau_t = 1)$. Before c_{t+1}^i was normally distributed with well defined mean $E_t^j c_{t+1}^j$ and variance $var_t\left(c_{t+1}^j\right)$. The latter moments were constructed conditional on the information set $\Omega_t(j)$. Thus, the resulting expression in (3.11) was straightforward. Here the result is different. Denote the conditional moments for $l = 0, 1$:

$$\begin{array}{rcl} \mu_l & = & E_{t,l}^j c_{t+1}^j = E\left(c_{t+1}^j \mid \Omega_t^l(j)\right) \qquad\qquad \text{(C.4)} \\ \sigma_{c,l}^2 & = & var_{t,l}^j c_{t+1}^j = var\left(c_{t+1}^j \mid \Omega_t^l(j)\right) \end{array}$$

Strictly speaking, c_{t+1}^j is now following a mixture of normals. Following $(C.3)$, the expected utility will be

$$E\left[-\exp\left(-\gamma c_{t+1}^j\right)\right] = \omega \left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + (1 - \omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$
(C.5)

Furthermore, we have that

$$\begin{array}{rcl} \mu_l & = & (1 + i_t)\, w_t^j + \left(E_t^i s_{t+1}^l - s_t + i_t^* - i_t\right) b_{Ft}^j - i_t \widetilde{m}_t^j \qquad\qquad \text{(C.6)} \\ & & + b_t^j E_t^j s_{t+1}^l - \dfrac{\widetilde{m}_t^j \left[\log\left(\widetilde{m}_t^j\right) - 1\right]}{\alpha} \\ & = & (1 + i_t)\, w_t^j + \left(b_{Ft}^j + b_t^j\right) E_t^j s_{t+1}^l + \left(-s_t + i_t^* - i_t\right) b_{Ft}^j \\ & & - i_t \widetilde{m}_t^j - \dfrac{\widetilde{m}_t^j \left[\log\left(\widetilde{m}_t^j\right) - 1\right]}{\alpha} \end{array}$$

and

$$\sigma_{c,l}^2 = \left(b_{Ft}^j + b_t^j\right)^2 \sigma_{t,l}^2$$

where $\sigma_{t,l}^2 = var_t\left(s_{t+1}^{j,l}\right)$, $l = 0, 1$ and because $b_t^j = \lambda_t^j$.

 FOC:

$\widetilde{m}_t^j$:

$$\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^j\right)\right]}{\partial \widetilde{m}_t^i} = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right](-\gamma) \times \quad \text{(C.7)}$$

$$\left(\frac{\partial \mu_1}{\partial \widetilde{m}_t^j} - \frac{\gamma}{2}\frac{\partial \sigma_{c,1}^2}{\partial \widetilde{m}_t^j}\right) +$$

$$(1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right](-\gamma) \times$$

$$\left(\frac{\partial \mu_0}{\partial \widetilde{m}_t^j} - \frac{\gamma}{2}\frac{\partial \sigma_{c,0}^2}{\partial \widetilde{m}_t^j}\right) = 0$$

b_{Ft}^j:

$$\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^j\right)\right]}{\partial b_{Ft}^j} = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right](-\gamma) \times \quad \text{(C.8)}$$

$$\left(\frac{\partial \mu_1}{\partial b_{Ft}^j} - \frac{\gamma}{2}\frac{\partial \sigma_{c,1}^2}{\partial b_{Ft}^j}\right) +$$

$$(1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right](-\gamma)$$

$$\times \left(\frac{\partial \mu_0}{\partial b_{Ft}^j} - \frac{\gamma}{2}\frac{\partial \sigma_{c,0}^2}{\partial b_{Ft}^j}\right)$$

$$= 0$$

For these expressions it is worth to compute the following derivatives:

$$\frac{\partial \mu_j}{\partial \widetilde{m}_t^j} = -i_t - \frac{\log\left(\widetilde{m}_t^j\right)}{\alpha} \quad \text{(C.9)}$$

$$\frac{\partial \sigma_{c,l}^2}{\partial \widetilde{m}_t^j} = 0$$

$$\Rightarrow \frac{\partial \mu_l}{\partial \widetilde{m}_t^j} - \frac{\gamma}{2}\frac{\partial \sigma_{c,l}^2}{\partial \widetilde{m}_t^j} = -i_t - \frac{\log\left(\widetilde{m}_t^j\right)}{\alpha}$$

which are, in particular, common across regimes.

131

Plugging $(C.9)$ in $(C.7)$:

$$
\begin{aligned}
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial \widetilde{m}_{t}^{j}} &= \omega\left[-\exp\left(-\gamma\left(\mu_{1}-\frac{\gamma}{2}\sigma_{c,1}^{2}\right)\right)\right](-\gamma)\times \\
&\qquad \left(-i_{t}-\frac{\log\left(\widetilde{m}_{t}^{j}\right)}{\alpha}\right)+ \\
&\qquad (1-\omega)\left[-\exp\left(-\gamma\left(\mu_{0}-\frac{\gamma}{2}\sigma_{c,0}^{2}\right)\right)\right](-\gamma)\times \\
&\qquad \left(-i_{t}-\frac{\log\left(\widetilde{m}_{t}^{j}\right)}{\alpha}\right) \\
&= 0
\end{aligned}
$$

Factorizing:

$$
\begin{aligned}
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial \widetilde{m}_{t}^{i}} &= \left\{\begin{array}{l}\omega\left[-\exp\left(-\gamma\left(\mu_{1}-\frac{\gamma}{2}\sigma_{c,1}^{2}\right)\right)\right]+ \\ (1-\omega)\left[-\exp\left(-\gamma\left(\mu_{0}-\frac{\gamma}{2}\sigma_{c,0}^{2}\right)\right)\right]\end{array}\right\}\times \\
&\qquad (-\gamma)\left(-i_{t}-\frac{\log\left(\widetilde{m}_{t}^{j}\right)}{\alpha}\right) \\
&= 0
\end{aligned}
$$

$$
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial \widetilde{m}_{t}^{j}} = E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right](-\gamma)\left(-i_{t}-\frac{\log\left(\widetilde{m}_{t}^{j}\right)}{\alpha}\right) = 0
$$

Only the last term can be equal to 0, thus,

$$
-\alpha i_{t} = \log\left(\widetilde{m}_{t}^{j}\right)
$$

which is exactly the same as (3.12). As a result, equations $(3.14)-(3.15)$ in the monetary model will remain unchanged.

On the other hand:

$$
\begin{aligned}
\frac{\partial \mu_{l}}{\partial b_{Ft}^{j}} &= E_{t}^{j}s_{t+1}^{l}-s_{t}+i_{t}^{*}-i_{t} \qquad\qquad (C.10) \\
\frac{\partial \sigma_{c,l}^{2}}{\partial b_{Ft}^{j}} &= 2\left(b_{Ft}^{j}+b_{t}^{j}\right)\sigma_{t,l}^{2} \\
&\Rightarrow \frac{\partial \mu_{l}}{\partial b_{Ft}^{j}}-\frac{\gamma}{2}\frac{\partial \sigma_{c,l}^{2}}{\partial b_{Ft}^{j}} \\
&= E_{t}^{j}s_{t+1}^{l}-s_{t}+i_{t}^{*}-i_{t}-\gamma\left(b_{Ft}^{j}+b_{t}^{j}\right)\sigma_{t,l}^{2}
\end{aligned}
$$

132

Plugging $(C.10)$ in $(C.8)$:

$$
\begin{aligned}
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial b_{Ft}^{j}} &= \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right](-\gamma) \times \\
&\quad \left(E_t^j s_{t+1}^1 - s_t + i_t^* - i_t - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,1}^2\right) + \\
&\quad (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right](-\gamma) \times \\
&\quad \left(E_t^j s_{t+1}^0 - s_t + i_t^* - i_t - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,0}^2\right) \\
&= 0
\end{aligned}
$$

Factorizing:

$$
\begin{aligned}
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial b_{Ft}^{j}} &= (-\gamma)\left\{\begin{array}{c} \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + \\ (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]\end{array}\right\} \times \\
&\quad (-s_t + i_t^* - i_t) + \\
&\quad \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right](-\gamma)\left(E_t^j s_{t+1}^1 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,1}^2\right) + \\
&\quad (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right](-\gamma)\left(E_t^j s_{t+1}^0 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,0}^2\right) \\
&= 0
\end{aligned}
$$

and simplifying:

$$
\begin{aligned}
\frac{\partial E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right]}{\partial b_{Ft}^{j}} &= (-\gamma)\,E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right](-s_t + i_t^* - i_t) \times \\
&\quad (-\gamma)\left\{\begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^1\left(j\right)\right]\left(E_t^j s_{t+1}^1 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,1}^2\right) \\ + (1-\omega)\,E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^0\left(j\right)\right]\left(E_t^j s_{t+1}^0 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,0}^2\right)\end{array}\right\} \\
&= 0
\end{aligned}
$$

Solve for b_{Ft}^j:

Re-arranging the expression:

$$
\begin{aligned}
&(\gamma)\,E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right](-s_t + i_t^* - i_t) \\
&= (-\gamma)\left\{\begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^1\left(j\right)\right]\left(E_t^j s_{t+1}^1 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,1}^2\right) \\ + (1-\omega)\,E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^0\left(j\right)\right]\left(E_t^j s_{t+1}^0 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,0}^2\right)\end{array}\right\}
\end{aligned}
$$

and simplifying:

$$
\begin{aligned}
&E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\right](-s_t + i_t^* - i_t) \\
&= -\left\{\begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^1\left(j\right)\right]\left(E_t^j s_{t+1}^1 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,1}^2\right) \\ + (1-\omega)\,E\left[-\exp\left(-\gamma c_{t+1}^{j}\right)\mid\Omega_t^0\left(j\right)\right]\left(E_t^j s_{t+1}^0 - \gamma\left(b_{Ft}^j + b_t^j\right)\sigma_{t,0}^2\right)\end{array}\right\}
\end{aligned}
$$

The expression of the RHS is equivalent to

$$-\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^1(j)\right]\left(E_t^j s_{t+1}^1\right) \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^0(j)\right]\left(E_t^j s_{t+1}^0\right) \end{array} \right\}$$
$$+\gamma\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^1(j)\right]\sigma_{t,1}^2 \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^0(j)\right]\sigma_{t,0}^2 \end{array} \right\}\left(b_{Ft}^j + b_t^j\right)$$

We conveniently re-express the LHS:

$$\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^1(j)\right]\left(-s_t + i_t^* - i_t\right) \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^0(j)\right]\left(-s_t + i_t^* - i_t\right) \end{array} \right\}$$

Combining LHS and RHS, we have as a result:

$$b_{Ft}^j = \frac{\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^1(j)\right]\left(E_t^j s_{t+1}^1 - s_t + i_t^* - i_t\right) \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^0(j)\right]\left(E_t^j s_{t+1}^0 - s_t + i_t^* - i_t\right) \end{array} \right\}}{\gamma\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^1(j)\right]\sigma_{t,1}^2 \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t^0(j)\right]\sigma_{t,0}^2 \end{array} \right\}} - b_t^j$$

(C.11)

The equation $(C.11)$ is a generalization of equation (3.16). That is, if we go back to the case where $\Omega_t^1(j) = \Omega_t^0(j) = \Omega_t(j)$, then

$$\begin{array}{rcl} E_t^j s_{t+1}^1 & = & E_t^i s_{t+1}^0 = E_t^j s_{t+1} \\ \sigma_{t,1}^2 & = & \sigma_{t,0}^2 = \sigma_t^2 \end{array}$$

so that

$$b_{Ft}^j = \frac{\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\left(E_t^j s_{t+1} - s_t + i_t^* - i_t\right) \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\left(E_t^j s_{t+1} - s_t + i_t^* - i_t\right) \end{array} \right\}}{\gamma\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\sigma_t^2 \\ +\left(1-\omega\right)E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\sigma_t^2 \end{array} \right\}} - b_t^j$$

Simplifying

$$b_{Ft}^j = \frac{\left\{E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\left(E_t^j s_{t+1} - s_t + i_t^* - i_t\right)\right\}}{\gamma\left\{E\left[-\exp\left(-\gamma c_{t+1}^{j}\right) \mid \Omega_t(j)\right]\sigma_t^2\right\}} - b_t^j$$

we get finally

$$b_{Ft}^j = \frac{\left(E_t^j s_{t+1} - s_t + i_t^* - i_t\right)}{\gamma\sigma_t^2} - b_t^j$$

which is equal to equation (3.16).

134

Furthermore, we now need to aggregate the demand of bonds. Recall that $\int_0^1 b_{Ft}^j dj = \sigma_\eta \eta_t$, where $\eta_t \sim N(0,1)$. Aggregating $(C.11)$:

$$\sigma_\eta \eta_t = \int_0^1 \left(\frac{\left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^j \right) \mid \Omega_t^1(j) \right] \left(E_t^j s_{t+1}^1 - s_t + i_t^* - i_t \right) \\ + (1-\omega) E\left[-\exp\left(-\gamma c_{t+1}^j \right) \mid \Omega_t^0(j) \right] \left(E_t^j s_{t+1}^0 - s_t + i_t^* - i_t \right) \end{array} \right\}}{\gamma \left\{ \begin{array}{c} \omega E\left[-\exp\left(-\gamma c_{t+1}^j \right) \mid \Omega_t^1(j) \right] \sigma_{t,1}^2 \\ + (1-\omega) E\left[-\exp\left(-\gamma c_{t+1}^j \right) \mid \Omega_t^0(j) \right] \sigma_{t,0}^2 \end{array} \right\}} \right) dj - b_t$$

In general, since the denominator is different across investors, aggregation cannot be performed withouth additional assumptions. In the original paper, Bacchetta and Van Wincoop (2006) assume that the next period variance of the exchange rate is common across investors. Here we assume that this variance is time varying across regimes.

There is still one additional complication. The denominator is a sort of weighted average of these variances across regimes. Weights are investor specific. Unless we can restrict the denominator to be common across investors, aggregation will be unfeasible. This issue is resolved in the next section.

C.4 Approximating the Utility Function

C.4.1 Exploiting Jensen's Inequality

Recall the objective function $(C.5)$:

$$E\left[-\exp\left(-\gamma c_{t+1}^j \right) \right] = \omega \left[-\exp\left(-\gamma \left(\mu_1 - \frac{\gamma}{2} \sigma_{c,1}^2 \right) \right) \right] + (1-\omega) \left[-\exp\left(-\gamma \left(\mu_0 - \frac{\gamma}{2} \sigma_{c,0}^2 \right) \right) \right]$$

which is a concave function. Then we know

$$c_{t+1}^j \sim \omega N\left(\mu_1, \sigma_{c,1}^2 \right) + (1-\omega) N\left(\mu_0, \sigma_{c,0}^2 \right)$$

so that:

$$\begin{array}{rcl} E_t^j \left[H\left(c_{t+1}^j \right) \right] & = & E_t^j \left[-\exp\left(-\gamma c_{t+1}^j \right) \right] \\ & = & \omega E_{t,1}^j \left[-\exp\left(-\gamma c_{t+1}^j \right) \right] \\ & & + (1-\omega) E_{t,0}^j \left[-\exp\left(-\gamma c_{t+1}^j \right) \right] \end{array}$$

Since $H(.)$ is a concave function, the Jensen's inequality says that:

$$E_t^j \left[-\exp\left(-\gamma c_{t+1}^j \right) \right] \leqslant -\exp\left(-\gamma E_t^j c_{t+1}^j \right)$$

which is actually

$$E_t^j \left[-\exp\left(-\gamma c_{t+1}^j \right) \right] \leqslant E_t^j \left\{ -\exp\left(-\gamma E_t^j c_{t+1}^j \right) \right\}$$

135

where we interpret $E_t^j c_{t+1}^j$ is a linear combination of two normal distributions. The RHS is

$$E\left\{-\exp\left(E\left[-\gamma c_{t+1}^j\right]\right)\right\} = \exp\left\{-\gamma E_t^j c_{t+1}^j + \frac{\gamma^2}{2} var_t\left(c_{t+1}^j\right)\right\}$$

where

$$
\begin{aligned}
E_t^j c_{t+1}^j &= \mu = \omega\mu_1 + (1-\omega)\mu_0 &\text{(C.12)}\\
var_t\left(c_{t+1}^j\right) &= \sigma_c^2 = \omega\left([\mu_1 - \mu]^2 + \sigma_{c,1}^2\right) + \\
&\quad\quad (1-\omega)\left([\mu_0 - \mu]^2 + \sigma_{c,0}^2\right)
\end{aligned}
$$

Using the latter notation:

$$E\left\{-\exp\left(E\left[-\gamma c_{t+1}^j\right]\right)\right\} = \exp\left\{-\gamma\left(\mu - \frac{\gamma}{2}\sigma_c^2\right)\right\} \quad\text{(C.13)}$$

Recall $(C.5)$:

$$E\left[-\exp\left(-\gamma c_{t+1}^j\right)\right] = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$

Now take the two expressions:

$$-\exp\left\{-\gamma\left(\mu - \frac{\gamma}{2}\sigma_c^2\right)\right\} \lesseqqgtr \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right]$$
$$+ (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$

The linear combination of the RHS arguments is:

$$
\begin{aligned}
&\omega\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right) + (1-\omega)\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\\
=\ &\omega\mu_1 + (1-\omega)\mu_0\\
&-\frac{\gamma}{2}\left(\omega\sigma_{c,1}^2 + (1-\omega)\sigma_{c,0}^2\right)
\end{aligned}
$$

On the other hand, the variance in the LHS argument is

$$\sigma_c^2 = \omega\sigma_{c,1}^2 + (1-\omega)\sigma_{c,0}^2 + \omega[\mu_1 - \mu]^2 + (1-\omega)[\mu_0 - \mu]^2 \quad\text{(C.14)}$$

That is, the term

$$\omega[\mu_1 - \mu]^2 + (1-\omega)[\mu_0 - \mu]^2 \geqslant 0$$

is the additional one, meaning that the variance is higher in the LHS. Since the utility function is decreasing in the variance, we have the following result:

$$-\exp\left\{-\gamma\left(\mu - \frac{\gamma}{2}\sigma_c^2\right)\right\} \leqslant \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right]$$
$$+ (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$

That is, the LHS is a lower bound for our objective function. We could maximize the LHS following the argument of the Expectation-Maximization Algorithm (see Hamilton (1989)).

Consider the new objective function $(C.13)$, which is equivalent to minimize

$$\mu - \frac{\gamma}{2}\sigma_c^2 \tag{C.15}$$

The FOCs are:

$\widetilde{m}_t^j$:

$$\frac{\partial \mu}{\partial \widetilde{m}_t^j} - \frac{\gamma}{2}\frac{\partial \sigma_c^2}{\partial \widetilde{m}_t^j} = 0$$

For the first term, recall $(C.9)$:

$$
\begin{aligned}
\frac{\partial \mu}{\partial \widetilde{m}_t^j} &= \omega \frac{\partial \mu_1}{\partial \widetilde{m}_t^j} + (1-\omega)\frac{\partial \mu_0}{\partial \widetilde{m}_t^j} \\
&= \omega\left(-i_t - \frac{\log\left(\widetilde{m}_t^j\right)}{\alpha}\right) \\
&\quad + (1-\omega)\left(-i_t - \frac{\log\left(\widetilde{m}_t^j\right)}{\alpha}\right) \\
&= -i_t - \frac{\log\left(\widetilde{m}_t^j\right)}{\alpha}
\end{aligned}
\tag{C.16}
$$

For the second term, recall also $(C.14)$

$$
\begin{aligned}
\frac{\partial \sigma_c^2}{\partial \widetilde{m}_t^j} &= \omega\frac{\partial \sigma_{c,1}^2}{\partial \widetilde{m}_t^j} + (1-\omega)\frac{\partial \sigma_{c,0}^2}{\partial \widetilde{m}_t^j} + \\
&\quad 2\omega\left[\mu_1 - \mu\right]\left(\frac{\partial \mu_1}{\partial \widetilde{m}_t^j} - \frac{\partial \mu}{\partial \widetilde{m}_t^j}\right) + \\
&\quad 2(1-\omega)\left[\mu_2 - \mu\right]\left(\frac{\partial \mu_2}{\partial \widetilde{m}_t^j} - \frac{\partial \mu}{\partial \widetilde{m}_t^j}\right)
\end{aligned}
$$

and using $(C.9)$ and $(C.16)$ we get

$$\frac{\partial \sigma_c^2}{\partial \widetilde{m}_t^j} = 0$$

thus,

$$-\alpha i_t = \log\left(\widetilde{m}_t^j\right)$$

which is exactly the same as (3.12).

137

b_{Ft}^j:

$$\frac{\partial \mu}{\partial b_{Ft}^j} - \frac{\gamma}{2} \frac{\partial \sigma_c^2}{\partial b_{Ft}^j} = 0$$

For the first term, recall $(C.10)$:

$$
\begin{aligned}
\frac{\partial \mu}{\partial b_{Ft}^j} &= \omega \frac{\partial \mu_1}{\partial b_{Ft}^j} + (1-\omega) \frac{\partial \mu_0}{\partial b_{Ft}^j} \qquad\qquad (C.17)\\
&= \omega \left(E_t^j s_{t+1}^1 - s_t + i_t^* - i_t \right) + \\
&\quad (1-\omega) \left(E_t^j s_{t+1}^0 - s_t + i_t^* - i_t \right) \\
&= \omega E_t^j s_{t+1}^1 + (1-\omega) E_t^j s_{t+1}^0 \\
&\quad -s_t + i_t^* - i_t
\end{aligned}
$$

For the second term, recall also $(C.14)$

$$
\begin{aligned}
\frac{\partial \sigma_c^2}{\partial b_{Ft}^j} &= \omega \frac{\partial \sigma_{c,1}^2}{\partial b_{Ft}^j} + (1-\omega) \frac{\partial \sigma_{c,0}^2}{\partial b_{Ft}^j} + \\
&\quad 2\omega \left[\mu_1 - \mu \right] \left(\frac{\partial \mu_1}{\partial b_{Ft}^j} - \frac{\partial \mu}{\partial b_{Ft}^j} \right) + \\
&\quad 2(1-\omega) \left[\mu_0 - \mu \right] \left(\frac{\partial \mu_0}{\partial b_{Ft}^j} - \frac{\partial \mu}{\partial b_{Ft}^j} \right)
\end{aligned}
$$

and using $(C.10)$ and $(C.17)$ we get

$$
\begin{aligned}
\frac{\partial \sigma_c^2}{\partial b_{Ft}^j} &= \omega \left(2 \left(b_{Ft}^j + b_t^j \right) \sigma_{t,1}^2 \right) + \\
&\quad (1-\omega) \left(2 \left(b_{Ft}^j + b_t^j \right) \sigma_{t,0}^2 \right) + \\
&\quad 2\omega \left[\mu_1 - \mu \right] \left(E_t^j s_{t+1}^1 - \omega E_t^j s_{t+1}^1 - (1-\omega) E_t^j s_{t+1}^0 \right) + \\
&\quad 2(1-\omega) \left[\mu_0 - \mu \right] \left(E_t^j s_{t+1}^0 - \omega E_t^j s_{t+1}^1 - (1-\omega) E_t^j s_{t+1}^0 \right)
\end{aligned}
$$

Simplifying

$$
\begin{aligned}
\frac{\partial \sigma_c^2}{\partial b_{Ft}^j} &= \omega \left(2 \left(b_{Ft}^j + b_t^j \right) \sigma_{t,1}^2 \right) + \\
&\quad (1-\omega) \left(2 \left(b_{Ft}^j + b_t^j \right) \sigma_{t,0}^2 \right) + \\
&\quad 2\omega \left[\mu_1 - \mu \right] \left(E_t^j s_{t+1}^1 - E_t^i \left(s_{t+1} \right) \right) + \\
&\quad 2(1-\omega) \left[\mu_2 - \mu \right] \left(E_t^j s_{t+1}^0 - E_t^j \left(s_{t+1} \right) \right)
\end{aligned}
$$

where

$$E_t^j \left(s_{t+1} \right) = \omega E_t^j \left(s_{t+1}^1 \right) + (1-\omega) E_t^j \left(s_{t+1}^0 \right) \qquad\qquad (C.18)$$

138

Using the definition of μ_l given by $(C.6)$: we have that

$$\mu_l - \mu = \left(b^j_{Ft} + b^j_t\right)\left(E^j_t s^1_{t+1} - E^j_t s_{t+1}\right)$$

Thus

$$
\begin{aligned}
\frac{\partial \sigma^2_c}{\partial b^j_{Ft}} &= \omega\left(2\left(b^j_{Ft} + b^j_t\right)\sigma^2_{t,1}\right) + \\
&\quad (1-\omega)\left(2\left(b^j_{Ft} + b^j_t\right)\sigma^2_{t,0}\right) + \\
&\quad 2\left(b^j_{Ft} + b^j_t\right)\omega\left(E^j_t s^1_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2 + \\
&\quad 2\left(b^j_{Ft} + b^j_t\right)(1-\omega)\left(E^j_t s^0_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2
\end{aligned}
$$

$$\frac{\partial \sigma^2_c}{\partial b^j_{Ft}} = 2\left(b^j_{Ft} + b^j_t\right)\left[\begin{array}{c}\omega\left(\sigma^2_{t,1} + \left(E^j_t s^1_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right) \\ + (1-\omega)\left(\sigma^2_{t,0} + \left(E^j_t s^0_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right)\end{array}\right] \qquad (C.19)$$

Using the results from $(C.17)$ and $(C.19)$ we get

$$
\begin{aligned}
&\omega E^j_t s^1_{t+1} + (1-\omega)E^j_t s^0_{t+1} - s_t + i^*_t - i_t \\
&\quad -\frac{\gamma}{2}2\left(b^j_{Ft} + b^j_t\right)\left[\begin{array}{c}\omega\left(\sigma^2_{t,1} + \left(E^j_t s^1_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right) \\ + (1-\omega)\left(\sigma^2_{t,0} + \left(E^j_t s^0_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right)\end{array}\right] \\
&\quad = 0
\end{aligned}
$$

Therefore:

$$b^j_{Ft} = \frac{\omega E^j_t s^1_{t+1} + (1-\omega)E^j_t s^0_{t+1} - s_t + i^*_t - i_t}{\gamma\left[\omega\left(\sigma^2_{t,1} + \left(E^j_t s^1_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right) + (1-\omega)\left(\sigma^2_{t,0} + \left(E^j_t s^0_{t+1} - E^j_t\left(s_{t+1}\right)\right)^2\right)\right]} - b^j_t$$

Notice that actually the variance of the exchange rate is:

$$
\begin{aligned}
\sigma^2_t &= var_t\left(s^j_{t+1}\right) = \omega\left(\left[E^j_t\left(s^1_{t+1}\right) - E^j_t\left(s_{t+1}\right)\right]^2 + \sigma^2_{t,1}\right) + \qquad (C.20) \\
&\quad (1-\omega)\left(\left[E^j_t\left(s^0_{t+1}\right) - E^j_t\left(s_{t+1}\right)\right]^2 + \sigma^2_{t,0}\right)
\end{aligned}
$$

Then we have that:

$$b^j_{Ft} = \frac{E^j_t\left(s_{t+1}\right) - s_t + i^*_t - i_t}{\gamma\sigma^2_t} - b^j_t$$

an expression equivalent to (3.16) with $E^i_t\left(s_{t+1}\right)$ and σ^2_t defined in $(C.18)$ and $(C.20)$.

Notice also that σ^2_t can be re-expressed more compactly as:

$$\sigma^2_t = \omega\sigma^2_{t,1} + (1-\omega)\sigma^2_{t,0} + \omega(1-\omega)\left[E^j_t\left(s^1_{t+1}\right) - E^j_t\left(s^0_{t+1}\right)\right]^2$$

139

C.4.2 Second order approximation

In this section we will not use Jensen's inequality. Instead, we will tacke the problem using Taylor approximations, which is in line with the common practice in macro literature.

Recall the objective function $(C.5)$:

$$E\left[-\exp\left(-\gamma c_{t+1}^j\right)\right] = \omega\left[-\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + (1-\omega)\left[-\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]$$

Following Yang (2011) among others, we take the Taylor expansion around $(0,0)$:

$$\log\left(a\exp\left(x\right) + b\exp\left(y\right)\right)$$
$$= \log\left(a+b\right) + \frac{ax+by}{a+b} + \frac{1}{2}\frac{ab}{\left(a+b\right)^2}\left(x-y\right)^2 + \cdots$$

it turns out that

$$\log\left\{\omega\left[\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + (1-\omega)\left[\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]\right\}$$
$$\approx \omega\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right) + (1-\omega)\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)$$
$$+ \frac{1}{2}\omega\left(1-\omega\right)\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right) + \gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)^2$$

Factorizing $-\gamma$:

$$\log\left\{\omega\left[\exp\left(-\gamma\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right)\right)\right] + (1-\omega)\left[\exp\left(-\gamma\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right)\right)\right]\right\}$$
$$\approx -\gamma\left[\begin{array}{c}\omega\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right) + (1-\omega)\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right) \\ -\frac{\gamma}{2}\omega\left(1-\omega\right)\left(\mu_1 - \mu_0 - \frac{\gamma}{2}\left(\sigma_{c,1}^2 - \sigma_{c,0}^2\right)\right)^2\end{array}\right]$$
$$\approx -\gamma\left[\omega\left(\mu_1 - \frac{\gamma}{2}\sigma_{c,1}^2\right) + (1-\omega)\left(\mu_0 - \frac{\gamma}{2}\sigma_{c,0}^2\right) - \frac{\gamma}{2}\omega\left(1-\omega\right)\left(\mu_1 - \mu_0\right)^2\right]$$

That is, since the last term contains a fourth order component, we remove it keeping our approach of the 2nd order approximation. Re-arranging the latter expression

$$\omega\mu_1 + (1-\omega)\mu_0 - \frac{\gamma}{2}\left[\omega\sigma_{c,1}^2 + (1-\omega)\sigma_{c,0}^2 + \omega\left(1-\omega\right)\left(\mu_1 - \mu_0\right)^2\right]$$

The last term is

$$\omega\left(1-\omega\right)\left(\mu_1 - \mu_0\right)^2$$

Recall $(C.12)$, then:

$$\mu_1 - \mu = -\left(1-\omega\right)\left(\mu_1 - \mu_0\right)$$

140

$$\mu_0 - \mu = (\omega) (\mu_1 - \mu_0)$$

Therefore

$$[\mu_1 - \mu]^2 = (1 - \omega)^2 (\mu_1 - \mu_0)^2$$
$$[\mu_0 - \mu]^2 = \omega^2 (\mu_1 - \mu_0)^2$$

so that:

$$
\begin{aligned}
&\omega [\mu_1 - \mu]^2 + (1 - \omega) [\mu_0 - \mu]^2 \\
=\ &\omega (1 - \omega)^2 (\mu_1 - \mu_0)^2 + (1 - \omega) \omega^2 (\mu_1 - \mu_0)^2 \\
=\ &\left[\omega (1 - \omega)^2 + (1 - \omega) \omega^2 \right] (\mu_1 - \mu_0)^2 \\
=\ &\left[\omega \left(1 - 2\omega + \omega^2 \right) + \left(\omega - \omega^2 \right) \omega \right] (\mu_1 - \mu_0)^2 \\
=\ &\omega (1 - \omega) (\mu_1 - \mu_0)^2
\end{aligned}
$$

As a result, the objective function is

$$\omega\mu_1 + (1 - \omega) \mu_0 - \frac{\gamma}{2} \left[\omega\sigma_{c,1}^2 + (1 - \omega) \sigma_{c,0}^2 + \omega [\mu_1 - \mu]^2 + (1 - \omega) [\mu_0 - \mu]^2 \right]$$

and using the definitions of $(C.12)$, then we have

$$\mu - \frac{\gamma}{2}\sigma_c^2$$

which is exactly the same as $(C.15)$. That is, using a second order approximation is equivalent to using Jensen's inequality and maximizing the lower bound.

C.5 Signal Extraction

This section closely follows the work developed by Barillas et al. (2013).

C.5.1 Filtering problem

Every period $t = 1, \ldots, T$ each agent $i \in [0, 1]$ observes the vector $Z_t(i)$:

$$Z_t(i) = D_s\left(\tau^t\right) s_t + D_{\mathbf{x}}\left(\tau^t\right) \mathbf{X}_t + \widetilde{R}\left(\tau^t\right) \mathbf{u}_t + Q\left(\tau^t\right) \boldsymbol{\eta}_t(i) \qquad \text{(C.21)}$$

where $\boldsymbol{\eta}_t(i) \sim N(0, I)$. We assume that each agent in the model is rational. Moreover, conditional on a history τ^t the model will be linear and gaussian. As a result, recalling (3.35), estimates of the state vector $\mathbf{X}_t$ using the Kalman Filter

$$\mathbf{X}_{t|t}^i = E\left[\mathbf{X}_t \mid \Omega_t(i)\right]$$

141

will be optimal, so that the Kalman Updating equation is:

$$\mathbf{X}_{t|t}^i = M(\tau^t)\mathbf{X}_{t-1|t-1}^i + K(\tau^t) \left[\underbrace{Z_t(i) - D(\tau^t) M(\tau^t)\mathbf{X}_{t-1|t-1}^i}_{\widetilde{Z}_t(i)} \right] \qquad \text{(C.22)}$$

The Kalman gain applying the standard formula is

$$K(\tau^t) = E\left[\mathbf{X}_{t|t}^j \widetilde{Z}_t(i)'\right] \times E\left[\widetilde{Z}_t(i) \widetilde{Z}_t(i)'\right]^{-1}$$

so that

$$\begin{aligned} K(\tau^t) &= \left[P(\tau^t)D(\tau^t)' + N(\tau^t)R(\tau^t)'\right] \qquad \text{(C.23)} \\ &\times \left[D(\tau^t) P(\tau^t)D(\tau^t)' + \overline{R}(\tau^t) \overline{R}(\tau^t)'\right]^{-1} \end{aligned}$$

where

$$\overline{R}(\tau^t) \equiv \left[\begin{array}{cc} R(\tau^t) & Q(\tau^t) \end{array} \right]$$

$$P(\tau^t) \equiv E\left(\mathbf{X}_t - \mathbf{X}_{t|t-1,\tau^t}^i\right)\left(\mathbf{X}_t - \mathbf{X}_{t|t-1,\tau^t}^i\right)'$$

C.5.2   Average hierarchy of expectations

Substituting $(C.21)$ in $(C.22)$ we get

$$\begin{aligned} \mathbf{X}_{t|t}^i &= \left[I - K(\tau^t)\right] D(\tau^t) M(\tau^t)\mathbf{X}_{t-1|t-1}^i \qquad \text{(C.24)} \\ &\quad + K(\tau^t)D(\tau^t)\left[M(\tau^t)\mathbf{X}_{t-1} + N(\tau^t)\mathbf{u}_t\right] \\ &\quad + K(\tau^t)\left[R(\tau^t)\mathbf{u}_t + Q(\tau^t)\boldsymbol{\eta}_t(i)\right] \end{aligned}$$

Taking averages across agents we get $\mathbf{X}_{t|t} \equiv \int \mathbf{X}_{t|t}^i di$, so that

$$\begin{aligned} \mathbf{X}_{t|t} &= \left[I - K(\tau^t)\right] D(\tau^t) M(\tau^t)\mathbf{X}_{t-1|t-1} \\ &\quad + K(\tau^t)D(\tau^t)\left[M(\tau^t)\mathbf{X}_{t-1} + N(\tau^t)\mathbf{u}_t\right] \\ &\quad + K(\tau^t)R(\tau^t)\mathbf{u}_t \end{aligned}$$

since $\int \boldsymbol{\eta}_t(i) di = 0$.

142

Then we stack this posterior estimates with the law of motion of fundamentals:

$$
\mathbf{X}_t \;=\; \left[\begin{array}{c} \mathbf{x}_t \\ \mathbf{X}_{t|t} \end{array}\right] = \left[\begin{array}{cc} \boldsymbol{\rho} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right]\left[\begin{array}{c} \mathbf{x}_{t-1} \\ \mathbf{X}_{t-1|t-1} \end{array}\right] + \tag{C.25}
$$

$$
\left[\begin{array}{c} \mathbf{0} \\ K(\tau^t)D\left(\tau^t\right)M(\tau^t) \end{array}\right]\left[\begin{array}{c} \mathbf{x}_{t-1} \\ \mathbf{X}_{t-1|t-1} \end{array}\right] +
$$

$$
\left[\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [I - K(\tau^t)]\,D\left(\tau^t\right)M(\tau^t) \end{array}\right]\left[\begin{array}{c} \mathbf{x}_{t-1} \\ \mathbf{X}_{t-1|t-1} \end{array}\right] +
$$

$$
\left[\begin{array}{c} \nu(\tau_t) \\ K(\tau^t)D\left(\tau^t\right)N(\tau^t) \end{array}\right]\mathbf{u}_t + K(\tau^t)R\left(\tau^t\right)\mathbf{u}_t
$$

Finally we equate undetermined coefficients using equation (**??**)

$$
M(\tau^t) \;=\; \left[\begin{array}{cc} \boldsymbol{\rho} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right] + \left[\begin{array}{c} \mathbf{0} \\ K(\tau^t)D\left(\tau^t\right)M(\tau^t) \end{array}\right] \tag{C.26}
$$

$$
+ \left[\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [I - K(\tau^t)]\,D\left(\tau^t\right)M(\tau^t) \end{array}\right]
$$

$$
N(\tau^t) = \left[\begin{array}{c} \nu(\tau_t) \\ K(\tau^t)D\left(\tau^t\right)N(\tau^t) \end{array}\right] + K(\tau^t)R\left(\tau^t\right) \tag{C.27}
$$

143

Bibliography

Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*(3), 629–57.

Allen, F., S. Morris, and H. S. Shin (2006). Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies 19*(3), 719–752.

Amisano, G. and C. Giannini (1997). *Topics in Structural VAR Econometrics*. Springer 2nd Rev. and Enlarged Edition.

An, S. and F. Schorfheide (2007). Bayesian analysis of DSGE models. *Econometric Reviews, Taylor and Francis Journals 26*(2-4), 113–172.

Bacchetta, P., E. Mertens, and E. Van Wincoop (2009). Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance 28*, 406–426.

Bacchetta, P. and E. Van Wincoop (2006). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review 96*(3), 552–576.

Bacchetta, P. and E. Van Wincoop (2009). On the unstable relationship between exchange rates and macroeconomic fundamentals. NBER Working Paper No. 15008.

Bacchetta, P., E. Van Wincoop, and T. Beutler (2010). *Can Parameter Instability Explain the Meese-Rogoff Puzzle?* L. Reichlin and K.West (eds.), NBER International Seminar on Macroeconomics, University of Chicago Press.

Barillas, F. and K. Nimark (2012). Speculation, risk premia and expectations in the yield curve. Working Paper, Universitat Pompeu Fabra.

Barillas, F., K. Nimark, and F. Pérez Forero (2013). Heterogeneous information and regime switches: A structural empirical framework. Manuscript.

Barsky, R. and E. Sims (2012). News shocks and the business cycle. *Journal of Monetary Economics 58*, 273–289.

Baumeister, C. and L. Benati (2012). Unconventional monetary policy and the great recession - estimating the impact of a compression in the yield spread at the zero lower bound. Forthcoming in the International Journal of Central Banking.

Beckworth, D. (2011). The stance of US monetary policy: An alternative measure. *Applied Economics Letters 18*(2), 121–125.

Bernanke, B. S. and A. S. Blinder (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review 82*(4), 901–921.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 120*(1), 387–422.

Bernanke, B. S. and I. Mihov (1998a). The liquidity effect and long-run neutrality. *Carnegie-Rochester Conference Series on Public Policy 49*, 149–194.

Bernanke, B. S. and I. Mihov (1998b). Measuring monetary policy. *Quarterly Journal of Economics 113*(3), 869–902.

Blinder, A. S. (2010). Quantitative easing: Entrance and exit strategies. CEPS Working Paper 204.

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

Borio, C. and P. Disyatat (2010). Unconventional monetary policies: An appraisal. *The Manchester School 78*(1), 53–89.

Canova, F. (1993). Modelling and forecasting exchange rates with a bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control 17*, 233–261.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press.

Canova, F. and M. Ciccarelli (2009). Estimating multicountry VAR models. *International Economic Review 50*(3), 929–959.

Canova, F., M. Ciccarelli, and E. Ortega (2012). Do institutional changes affect business cycles? evidence from europe. Universitat Pompeu Fabra Working Paper 1158.

146

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

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

Canova, F. and L. Gambetti (2009). Structural changes in the US economy: Is there a role for monetary policy? *Journal of Economic Dynamics and Control 33*, 477–490.

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*(2-3), 369–388.

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

Canova, F. and F. J. Pérez Forero (2013). A general algorithm for estimating structural vars VARs. Manuscript.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Casella, R. and C. S. Robert (2004). *Monte Carlo Statistical Methods* (Second ed.). Springer.

Cecioni, M., G. Ferrero, and A. Secchi (2011). Unconventional monetary policy in theory and in practice. Banca D'Italia: Questioni di Economia e Finanza (Occasional papers) 102.

Chang, Y., S.-B. Kim, and F. Schorfheide (2010). Labor-market heterogeneity, aggregation, and the lucas critique. NBER Working Paper No. 16401.

Chen, H., V. Cúrdia, and A. Ferrero (2012). The macroeconomic effects of large-scale asset purchase programmes. *The Economic Journal 122*(564), F289–F315.

Chib, S. (2001). *Markov Chain Monte Carlo Methods: Computation and Inference*, Chapter 57, pp. 3569–3649. J.J. Heckman and E.E. Leamer (ed.) Handbook of Econometrics.

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*(2), 339–360.

Christensen, J. H. E. and G. D. Rudebusch (2012). The response of interest rates to US and UK quantitative easing. *The Economic Journal 122*(564), F385–F414.

Christiano, L. and M. Eichenbaum (1992). *Identification and the Liquidity Effect of a Monetary Policy Shock*. in A. Cukierman, Z. Hercowitz and L. Leiderman, eds., Political Economy, Growth, and Business Cycles (Cambridge MA: MIT Press).

Christiano, L., M. Eichenbaum, and C. Evans (1999). *Monetary Policy Shocks: What Have We Learned and to What End?* Handbook of Macroeconomics, Ed.1,Vol. 1,Ch.II. Edited by J.B. Taylor and M. Woodford. Elsevier Science B.V. All rights reserved.

Christiano, L. J. and M. Eichenbaum (1991). Identification and the liquidity effect of a monetary policy shock. NBER Working Paper Nï£¡ 3920.

Chung, H., J.-P. Laforte, D. Reifschneider, and J. C. Williams (2012). Have we underestimated the likelihood and severity of zero lower bound events? *Journal of Money, Credit and Banking 47*(1), 47–82.

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*(2), 262–302.

Cogley, T. and T. Yagihashi (2010). Are DSGE approximating models invariant to shifts in policy? *The B.E. Journal of Macroeconomics 10*(1), 27.

Cosimano, T. F. and D. W. Jansen (1988). Federal reserve policy, 1975-1985: An empirical analysis. *Journal of Macroeconomics 10*(1), 27–47.

Cosimano, T. F. and R. G. Sheehan (1994). The federal reserve operating procedure, 1984-1990: An empirical analysis. *Journal of Macroeconomics 16*(4), 573–588.

Creel, D. (2012). Survey of sequential monte carlo methods for economics and finance. *Econometric Reviews 31*, 245–296.

Cúrdia, V. and M. Woodford (2011). The central-bank balance sheet as an instrument of monetary policy. *Journal of Monetary Economics 58*, 54–79.

Del Negro, M. (2003). Discussion of cogley and sargent's "drifts and volatilities: Monetary policy and outcomes in the post WWII u.s.". Federal Reserve Bank of Atlanta. Working Paper 2003-26.

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *The Journal of Political Economy 84*(6), 1161–1176.

Eichenbaum, M. and C. L. Evans (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates. *The Quarterly Journal of Economics 110*(4), 975–1009.

Engel, C. (1994). Can the markov switching model forecast exchange rates? *Journal of International Economics 36*, 151–165.

Engel, C. and J. D. Hamilton (1990). Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review 80*(4), 689–713.

Engel, C., J. Wang, and J. Wu (2010). Long-horizon forecasts of asset prices when the discount factor is close to unity. Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute. Working Paper No. 36.

Engel, C. and K. D. West (2005). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517.

Evans, M. D. D. and K. K. Lewis (1995). Do long-term swings in the dollar affect estimates of the risk premia? *The Review of Financial Studies 8*(3), 709–742.

Favero, C. A. (2001). *Applied Macroeconometrics*. Oxford University Press.

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 DP 7813.

Fiorentini, G., C. Planas, and A. Rossi (2012). Efficient MCMC sampling in dynamic mixture models. *Statistics and Computing September*, 1–13.

Gagnon, J., M. Raskin, J. Remache, and B. Sack (2010). Large-scale asset purchases by the federal reserve: Did they work? Federal Reserve Bank of New York Staff Reports 441.

Gagnon, J., M. Raskin, J. Remache, and B. Sack (2011). The financial market effects of the federal reserveï£¡s large-scale asset purchases. *International Journal of Central Banking 7*(1), 3–43.

Gali, J. (1991). How well does the islm model fits the postwar us data? *Quarterly Journal of Economics 106*, 709–738.

Galí, J. and L. Gambetti (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics 1*(1), 26–57.

Gertler, M. and P. Karadi (2011). A model of unconventional monetary policy. *Journal of Monetary Economics 58*, 17–34.

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments*. In Bayesian Statistics 4 (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Geweke, J. (1999). Using simulation methods for bayesian econometric models: Inference, development,and communication. *Econometric Reviews 18*(1), 1–73.

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.

Giordani, P. and R. Kohn (2010). Adaptive independent metropolis-hastings by fast estimation of mixtures of normals. *Journal of Computational and Graphical Statistics 19*(2), 243–259.

Goodfriend, M. and R. G. King (2005). The incredible volcker disinflation. *Journal of Monetary Economics 52*, 981–1015.

Gordon, D. B. and E. M. Leeper (1994). The dynamic impacts of monetary policy: An exercise in tentative identification. *Journal of Political Economy 102*(6), 1228–1247.

Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive metropolis algorithm. *Bernoulli 7*(2), 223–242.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

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

Hamilton, J. D. and J. Wu (2012). The effectiveness of alternative monetary policy tools in a zero lower bound environment. *Journal of Money, Credit and Banking 44*(1), 3–46.

Herbst, E. and F. Schorfheide (2013). Sequential monte carlo sampling for dsge models. University of Pennsylvania, manuscript.

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

Kaminsky, G. (1993). Is there a peso problem? evidence from the Dollar/Pound exchange rate, 1976-1987. *American Economic Review 83*(3), 450–472.

Kasa, K. (2000). Forecasting the forecasts of others in the frequency domain. *Review of Economic Dynamics 3*, 726–756.

Kasa, K., T. B. Walker, and C. H. Whiteman (2007). Asset prices in a time series model with perpetually disparately informed, competitive traders. Unpuplished manuscript.

Kilian, L. (2012). Structural vector autoregressions. Unpublished manuscript.

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

Kim, S. (2005). Monetary policy, foreign exchange policy, and delayed overshooting. *Journal of Money, Credit, and Banking 37*(4), 775–782.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Kociecki, A. Rubaszek, M. and M. Ca' Zorzi (2013). Bayesian analysis of recursive svar models with overidentifying restrictions. Bank of Poland, manuscript.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Koop, G. and S. M. Potter (2011). Time varying vars with inequality restrictions. *Journal of Economic Dynamics and Control 35*(7), 1126–1138.

Lakdawala, A. (2011). Changes in federal reserve preferences. Job Market Paper UCSD.

Lanne, M. and H. Lütkepohl (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking 40*(6), 1131–1149.

Leeper, E., C. Sims, and T. Zha (1996). What does monetary policy do? *Brookings Papers on Economic Activity 2*, 1–78.

Lenza, M., H. Pill, and L. Reichlin (2010). Monetary policy in exceptional times. European Central Bank, Working Paper Series 1253.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Magnus, J. R. and H. Neudecker (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Third Edition ed.). Wiley Series in Probability and Statistics. John Wiley and Sons.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics 14*(1-2), 3–24.

Melosi, L. (2012). Estimating models with dispersed information. Unpublished manuscript.

Moss, C. B. (2010). Closed form solutions to expected utility: Lecture XIV. AEB 6182 - Expected Utility - University of Florida.

Mumtaz, H. and P. Surico (2009). Time-varying yield curve dynamics and monetary policy. *Journal of Applied Econometrics 24*(6), 895–913.

Nimark, K. (2011). Dynamic higher order expectations. Working Paper, Universitat Pompeu Fabra.

Nimark, K. (2012a). Man-bites-dog driven business cycles. Working Paper, Universitat Pompeu Fabra.

Nimark, K. (2012b). Speculative dynamics in the term structure of interest rates. Working Paper, Universitat Pompeu Fabra.

Patton, A. J. and A. Timmermann (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics 57*(7), 803–820.

Peersman, G. (2011). Macroeconomic effects of unconventional monetary policy in the euro area. ECB Working Paper 1397.

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

Reis, R. (2009). Interpreting the unconventional u.s. monetary policy of 2007-09: Comments and discussion. *Brookings Papers on Economic Activity 40*(2), 119–165.

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*(3), 324–330.

Rossi, B. (2005). Testing long-horizon predictive ability with high persistence, and the meese-rogoff puzzle. *International Economic Review 46*(1), 61–92.

Rossi, B. (2006). Are exchange rates really random walks? some evidence robust to parameter instability. *Macroeconomic Dynamics 10*, 20–38.

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.

Sargent, T. J. (1991). Equilibrium with signal extraction from endogenous variables. *Journal of Economic Dynamics and Control 15*, 245–273.

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

Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*(4), 949–968. Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.

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

Singleton, K. J. (1987). Asset prices in a time series model with disparately informed, competitive traders. In W. A. Barnett and K. J. Singleton (Eds.), *New Approaches to Monetary Economics*, pp. 249–272. Cambridge, UK: Cambridge University Press.

Strid, I., P. Giordani, and R. Kohn (2010). Adaptive hybrid metropolis-hastings samplers for DSGE models. SSE/EFI Working Paper Series in Economics and Finance No. 724.

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

Swanson, E. T. and J. C. Williams (2012). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. FRB San Francisco Working Paper 2012-02.

Townsend, R. M. (1983). Forecasting the forecasts of others. *The Journal of Political Economy 91*(4), 546–588.

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.

Vives, X. (2010). *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press.

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

Waggoner, D. F. and T. Zha (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics 81*(4), 639–651.

Williams, J. C. (2011). Unconventional monetary policy: Lessons from the past three years. FRBSF Economic Letter 2011-31 (October 3).

Williams, J. C. (2012a). The federal reserveï£¡s unconventional policies. FRBSF Economic Letter 2012-34 (November 13).

Williams, J. C. (2012b). Monetary policy, money, and inflation. FRBSF Economic Letter 2012-21 (July 9).

Wolff, C. C. (1987). Time varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business and Economic Statistics 5*(1), 87–97.

Yang, W. (2011). Business cycles and regime-shift risk. Available at SSRN: http://ssrn.com/abstract=1991694 or http://dx.doi.org/10.2139/ssrn.1991694.

154