

*Not for Publication*

**Technical Appendix of: “Policy  
Rules, Regime Switches, and Trend  
Inflation: An Empirical Investigation  
for the U.S.”**

## 1 MCMC methodology

In the following, we describe our inferential solution for the model

$$i_t = (1 - \rho_{S_{1,t}})[\bar{r} + \pi_t + \alpha_{S_{1,t}}z_t + \beta_{S_{1,t}}y_t] + \rho_{S_{1,t}}i_{t-1} + \epsilon_{S_{2,t}} \quad (1)$$

$$\pi_t^* = (1 - \rho_\pi)\pi_{LR} + \rho_\pi\pi_{t-1}^* + \xi_{S_{2,t}} \quad (2)$$

$$\pi_t = \pi_t^* + z_t \quad (3)$$

$$z_t = \phi_{S_{1,t}}z_{t-1} + \eta_{S_{2,t}} \quad (4)$$

$$\epsilon_{S_{2,t}} \sim \mathcal{N}(0, \sigma_{\epsilon_{S_{2,t}}}^2), \xi_{S_{2,t}} \sim \mathcal{N}(0, \sigma_{\xi_{S_{2,t}}}^2), \eta_{S_{2,t}} \sim \mathcal{N}(0, \sigma_{\eta_{S_{2,t}}}^2) \quad (5)$$

In particular  $S_{1,t}$ ,  $S_{2,t}$  are the unobservable two-state first order Markov chains with transition probability matrix  $P_k = \{p_{k,ij}\}$ ,  $k \in \{1, 2\}$ ,  $i, j \in$

$\{0, 1\}$ .

The goal of the inferential procedure is to estimate the latent switching regime processes  $\mathbf{S}_k = \{S_{k,t}, t = 1, \dots, T\}$ , the unobservable inflation target  $\boldsymbol{\pi}^* = \{\pi_t^*, t = 1, \dots, T\}$  and the parameters  $\boldsymbol{\theta}$ . The observed interest rate is denoted by  $\mathbf{I} = \{i_t, t = 1, \dots, T\}$ , while  $\mathbf{X} = (\boldsymbol{\pi}, \mathbf{y})$ , where  $\boldsymbol{\pi} = \{\pi_t, t = 1, \dots, T\}$  is the inflation rate and  $\mathbf{y} = \{y_t, t = 1, \dots, T\}$  is the output gap.

Since for non-linear latent factor models the likelihood function is not available in closed form, inference has to be based on approximations or numerical evaluations. Markov Chain Monte Carlo (MCMC) basically avoids the need to directly computing the likelihood function that is expressed in terms of a highly multivariate integral. The basic idea behind MCMC is to build a Markov chain transition kernel starting from a given initial point and with limiting invariant distribution equal to the posterior distribution of the quantities of interest. Under suitable conditions [see Robert and Casella, 1999, ch. 6-7], such a transition kernel converges in distribution to the target posterior density. In practice, the chain is updated to a new position by first simulating from a given proposal distribution, then by eventually accepting the move. MCMC for switching regime ARMA models have been introduced in Albert and Chib [1993] and in McCulloch and Tsay [1993] whereas their inferential approach have been successively generalized in Billio et al. [1999] and in Kim and Nelson [1999]. In particular Frühwirth-Schnatter [2001] proposes an MCMC-based solution to address the parameters identification problems related to switching regime models [see also Frühwirth-Schnatter, 2006 for a comprehensive treatment of this topic].

In our framework, given the starting vector  $(\boldsymbol{\theta}^{(0)}, \mathbf{S}_1^{(0)}, \mathbf{S}_2^{(0)}, \boldsymbol{\pi}^{*(0)})$  we simulate through MCMC the trajectory of the Markov chain  $\{\boldsymbol{\theta}^{(j)}, \mathbf{S}^{(j)}, \mathbf{S}_2^{(j)}, \boldsymbol{\pi}^{*(j)}, j = 1, \dots, n\}$  whose draws converge to the posterior distribution  $p(\boldsymbol{\theta}, \mathbf{S}_1, \mathbf{S}_2, \boldsymbol{\pi}^* | \mathbf{I}, \mathbf{X})$ . Once convergence is achieved, inference can be based on the generated serially dependent sample. More precisely, estimates of the latent factors and of the posterior mean of  $\boldsymbol{\theta}$  are obtained by averaging over the realization of the chains, i.e.,  $\hat{Pr}[S_{k,t} = 1 | \mathbf{I}, \mathbf{X}] = n^{-1} \sum_{j=1}^n S_{k,t}^{(j)}$ ,  $\hat{\pi}_t^* = n^{-1} \sum_{j=1}^n \pi_t^{*(j)}$  and  $\hat{\boldsymbol{\theta}} = n^{-1} \sum_{j=1}^n \boldsymbol{\theta}^{(j)}$ . To account for serial correlation in the draws, we estimate the numerical standard error of the sample posterior mean using the approach implemented in Kim et al. [1998]. The analysis is implemented through MCMC, running the algorithm for 50,000 iterations and discarding the first 25,000 as burn-in. In our experience, this choice for the burn-in to remove the effects of initial conditions is more than adequate.

Moving the whole vector  $(\boldsymbol{\theta}, \mathbf{S}_1, \mathbf{S}_2, \boldsymbol{\pi}^*)$  in block can be inefficient, since it is highly multivariate. We implement our MCMC strategy through the Gibbs sampler, an algorithm that efficiently simulates each element or block<sup>1</sup> of  $(\boldsymbol{\theta}, \mathbf{S}_1, \mathbf{S}_2, \boldsymbol{\pi}^*)$  in turn from its full conditional distribution, i.e., the distribution conditioned on the data and the remaining components of the vector. In this case the acceptance probability of moving each sub-block of the vector is 1. As suggested in Shephard [1994] and Carter and Kohn [1994] amongst others, updating the whole latent process  $\mathbf{S}_k$  in block from its joint distribution given the data and the other parameters should reduce the autocorrelation between states and then speed up the convergence of the chain to its invariant distribution.

---

<sup>1</sup>It is worth noting that a block of parameters can be a sub-vector of parameters or a singleton.

To take care of the regime switches, we provide an efficient algorithm based on the multi-move Gibbs sampler proposed in Chib [1996] to update the states  $S_{1,t}$  and  $S_{2,t}$ ,  $t = 1, \dots, T$ , whereas we update the parameters and the latent process  $\pi_t^*$  one component at a time.

To simplify our notation, in the following we call  $\theta_j$  the generic  $j$ -th block of the vector  $\boldsymbol{\theta} = \{\theta_j, j = 1, \dots, J\}$ ,  $\boldsymbol{\theta}_{-j^-} = (\theta_1, \dots, \theta_{j-1})$  and  $\boldsymbol{\theta}_{-j^+} = (\theta_{j+1}, \dots, \theta_J)$ . We summarize the algorithm as follows:

#### MCMC algorithm

- Initialize the chain at  $(\boldsymbol{\theta}^{(0)}, \mathbf{S}_1^{(0)}, \mathbf{S}_2^{(0)}, \boldsymbol{\pi}^{*(0)})$
- At step  $j = 1, \dots, n$ 
  - Update  $\boldsymbol{\theta}$  one component at a time from the full conditional  $p(\boldsymbol{\theta}_i | \mathbf{S}_1^{(j-1)}, \mathbf{S}_2^{(j-1)}, \boldsymbol{\pi}^{(j-1)}, \boldsymbol{\theta}_{-i^-}^{(j)}, \boldsymbol{\theta}_{-i^+}^{(j-1)}, \mathbf{I}, \mathbf{X})$  through the Gibbs sampler algorithm;
  - Draw  $(S_{1,1}^{(j)}, \dots, S_{1,T}^{(j)})$  in block from  $p(\mathbf{S}_1 | \mathbf{S}_2^{(j-1)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\pi}^{(j-1)}, \mathbf{I}, \mathbf{X})$ ;
  - Draw  $(S_{2,1}^{(j)}, \dots, S_{2,T}^{(j)})$  in block from  $p(\mathbf{S}_2 | \boldsymbol{\pi}^{(j-1)}, \boldsymbol{\theta}^{(j)}, \mathbf{S}_1^{(j)}, \mathbf{I}, \mathbf{X})$ ;
  - Draw  $(\pi_1^{*(j)}, \dots, \pi_T^{*(j)})$  one-at-a-time from the conditional  $p(\pi_t^* | \mathbf{S}_1^{(j)}, \mathbf{S}_2^{(j)}, \pi_{t-1}^{*(j)}, \pi_{t+1}^{*(j-1)}, \boldsymbol{\theta}^{(j)}, \mathbf{I}, \mathbf{X})$ , through Gibbs sampler.

In the next subsections we describe the details of the algorithm.

## 1.1 Updating the parameters

The use of conjugate priors makes it straightforward to update  $\boldsymbol{\theta}$ . It is easy to show that the conditional posteriors of  $\sigma_{\epsilon,i}^2$ ,  $\sigma_{\xi,i}^2$  and  $\sigma_{\eta,i}^2$  are Inverse Gamma,

$p_{k,ij}$  are Beta, whereas all the remaining parameters are Gaussian. Since the full conditional distribution can be simulated directly, each sub-movement of the chain is accepted with probability 1.

The main issue when estimating regime switching models is identification, since the labeling of the states can be permuted without modifying the marginal likelihood. It is thus common in the literature to impose some constraints on the parameters' space. In particular, we find it useful to identify  $S_{2,t}$  by imposing  $\sigma_{\epsilon S_2=0}^2 > \sigma_{\epsilon S_2=1}^2$  and we implement it in our algorithm through the permutation sampler of Frühwirth-Schnatter [2001]. Furthermore, to identify the first regime  $S_{1,t}$  we hypothesize:

$$(i) \quad 0 < \alpha_{S_1=0} < +\infty \text{ and } -1 < \alpha_{S_1=1} < 0;$$

$$(ii) \quad \phi_{S_1=0} \sim \mathcal{N}(0.4, 0.1) \text{ and } \phi_{S_1=1} \sim \mathcal{N}(0.6, 0.1)$$

We notice that (i)-(ii) are informative to identify the two states. If for instance we reverse the constraint, i.e.  $0 < \alpha_{S_1=1} < +\infty$  and  $-1 < \alpha_{S_1=0} < 0$  we obtain the exact reverse labeling of the probability of being in  $S_1$ .

## 1.2 Updating $\mathbf{S}_k$

We update  $\mathbf{S}_k = (S_{k,1}, \dots, S_{k,T})$  in block, moving the two vectors according to the algorithm proposed in Chib [1996]. In the following we focus on  $\mathbf{S}_1$  since the procedure for  $\mathbf{S}_2$  is in principle the same. To keep the notation concise, we define a generic vector  $\mathbf{S}_{k,t_1:t_2} = (S_{t_1}, \dots, S_{t_2})$ . The method exploits the following decomposition

$$p(\mathbf{S}_1 | \boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_2, \boldsymbol{\theta}) = p(S_{1,T} | \boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_2, \boldsymbol{\theta}) \prod_{t=1}^{T-1} p(S_{1,t} | \boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_2, \mathbf{S}_{1,t+1:T}, \boldsymbol{\theta}), \quad (6)$$

in which the generic term of the product is

$$p(S_{1,t}|\boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_{1,t+1:T}, \mathbf{S}_2, \boldsymbol{\theta}) \propto p(S_{1,t}|\mathbf{I}_{1:t}, \boldsymbol{\pi}_{1:t}^*, \mathbf{X}_{1:t}, \mathbf{S}_{2,1:t}, \boldsymbol{\theta})p(S_{1,t+1}|S_{1,t}, \boldsymbol{\theta}). \quad (7)$$

In particular  $p(S_{1,t}|\mathbf{I}_{1:t}, \boldsymbol{\pi}_{1:t}^*, \mathbf{X}_{1:t}, \mathbf{S}_{2,1:t}, \boldsymbol{\theta})$  can be written as

$$p(S_{1,t}|\mathbf{I}_{1:t}, \boldsymbol{\pi}_{1:t}^*, \mathbf{X}_{1:t}, \mathbf{S}_{2,1:t}, \boldsymbol{\theta}) \propto p(S_{1,t}|\mathbf{I}_{1:t-1}, \boldsymbol{\pi}_{1:t-1}^*, \mathbf{X}_{1:t-1}, \mathbf{S}_{2,1:t-1}, \boldsymbol{\theta})p(i_t, \pi_t, \pi_t^*, S_{2,t}|\boldsymbol{\pi}_{1:t-1}^*, \mathbf{X}_{1:t-1}, \mathbf{S}_{2,1:t-1}, \boldsymbol{\theta}) \quad (8)$$

where

$$\begin{aligned} p(S_{1,t}|\mathbf{I}_{1:t-1}, \boldsymbol{\pi}_{1:t-1}^*, \mathbf{X}_{1:t-1}, \mathbf{S}_{2,1:t-1}, \boldsymbol{\theta}) &= \\ &= \sum_{j=0}^1 p(S_{1,t}|S_{1,t-1} = j, \boldsymbol{\theta})p(S_{1,t-1} = j|\mathbf{I}_{1:t-1}, \boldsymbol{\pi}_{1:t-1}^*, \mathbf{X}_{1:t-1}, \mathbf{S}_{2,1:t-1}, \boldsymbol{\theta}) \end{aligned} \quad (9)$$

These two latter distributions can be numerically evaluated in a recursive fashion by setting the distribution of the initial state  $S_{1,1}$  as the stationary distribution of the Markov chain,  $p(S_{1,1}|\boldsymbol{\theta})$ . Once all these quantities are computed,  $S_{1,T}$  is sampled from  $p(S_{1,T}|\boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_2, \boldsymbol{\theta})$ , that is a Binomial random variable, while the remaining states can be directly simulated from  $p(S_{1,t}|\boldsymbol{\pi}^*, \mathbf{I}, \mathbf{X}, \mathbf{S}_2, S_{1,t+1:T}, \boldsymbol{\theta})$ , starting from  $S_{1,T-1}$  until  $S_{1,1}$ .

## 2 Marginal Likelihood computation

In Bayesian statistics it is common practice to use the marginal likelihood to measure the goodness-of-fit. This quantity is defined as

$$\begin{aligned} m(\mathbf{I}, \mathbf{X}|\mathcal{M}) &= \\ &= \int p(\mathbf{I}, \mathbf{X}|\boldsymbol{\pi}^*, \mathbf{S}_1, \mathbf{S}_2, \boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\pi}^*, \mathbf{S}_1, \mathbf{S}_2|\boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}d\boldsymbol{\pi}^*d\mathbf{S}_1d\mathbf{S}_2, \end{aligned} \quad (10)$$

where  $(\mathcal{M}, \boldsymbol{\theta})$  indicates a given model  $\mathcal{M}$  and its parameters and  $p(\boldsymbol{\theta}|\mathcal{M})$  is the set of prior distributions. In the following we suppress the model index  $\mathcal{M}$  for conciseness. Many techniques have been proposed in the literature to evaluate the marginal likelihood. A review on some Monte Carlo alternatives is given in Han and Carlin [2001].

We compute the marginal likelihood through the Modified Harmonic Mean (MHM) of Gelfand and Dey [1994] based on the following result

$$p(\mathbf{I}, \mathbf{X})^{-1} = \int \frac{h(\boldsymbol{\theta})}{p(\mathbf{I}, \mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}p(\boldsymbol{\theta}|\mathbf{I}, \mathbf{X})d\boldsymbol{\theta} \quad (11)$$

in which  $h(\boldsymbol{\theta})$  is a weighting function. A Monte Carlo estimate of (11) is

$$\hat{p}(\mathbf{I}, \mathbf{X})^{-1} = \frac{1}{N} \sum_{i=1}^N \frac{h(\boldsymbol{\theta}^{(i)})}{p(\mathbf{I}, \mathbf{X}|\boldsymbol{\theta}^{(i)})p(\boldsymbol{\theta}^{(i)})} \quad (12)$$

in which  $\boldsymbol{\theta}^{(i)}$  are draws from the posterior distribution.

The computation of the weighting function  $h(\boldsymbol{\theta})$  plays a key-role in this context. Geweke [1999] proposes a weighting function based on a truncated multivariate normal random variable centered at the posterior mean and with dispersion the variance covariance matrix. Differently, here we consider the

generalization of Sims et al. [2008]. In this framework, the  $h(\boldsymbol{\theta})$  is a truncated elliptical distribution centered at the posterior mode of  $\boldsymbol{\theta}$ , namely  $\hat{\boldsymbol{\theta}}$ , and with dispersion matrix  $\hat{\Omega} = \frac{1}{N} \sum_{i=1}^N (\boldsymbol{\theta}^{(i)} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(i)} - \hat{\boldsymbol{\theta}})'$ . As pointed out in Sims et al. [2008] this choice results in a considerable improvement for switching regime models. A crucial point for the estimation of the marginal likelihood is the evaluation of  $p(\mathbf{I}, \mathbf{X} | \boldsymbol{\theta}^{(i)})$ . We recur to the approximated solution proposed in Kim and Nelson [1999] (pp. 97-106) in which the Kalman and the Hamilton filters are merged together, and in which the mixture component is collapsed at the end of each iteration to keep the algorithm feasible. To apply this method, it is convenient to rewrite the model through the following state space representation

$$\mathbf{y}_t = \Gamma_t^{-1} \mu_{y,t} + \Gamma_t^{-1} H_t \Pi_t^* + \Gamma_t^{-1} \tilde{\epsilon}_t \quad (13)$$

$$\Pi_t^* = \begin{bmatrix} \pi_t^* \\ \pi_{t-1}^* \end{bmatrix} = \begin{bmatrix} (1 - \rho_\pi) \pi^{LR} \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_\pi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \pi_{t-1}^* \\ \pi_{t-2}^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_{S_2,t} \\ 0 \end{bmatrix} \quad (14)$$

$$\text{in which } \mathbf{y}_t = \begin{bmatrix} i_t \\ \pi_t \end{bmatrix}, \mu_{y,t} = \begin{bmatrix} (1 - \rho_{S_1,t})[\bar{r} + \beta_{S_1,t}] + \rho_{S_1,t} i_{t-1} \\ \phi_{S_1,t} \pi_{t-1} \end{bmatrix},$$

$$\Gamma_t = \begin{bmatrix} 1 & -(1 - \rho_{S_1,t})(1 + \alpha_{S_1,t}) \\ 0 & 1 \end{bmatrix}, H_t = \begin{bmatrix} -(1 - \rho_{S_1,t})\alpha_{S_1,t} & 0 \\ 1 & \phi_{S_1,t} \end{bmatrix} \text{ and}$$

$$\tilde{\epsilon}_t = \begin{bmatrix} \epsilon_{S_2,t} \\ \eta_{S_2,t} \end{bmatrix}.$$

Furthermore, it is also useful to define a new switching regime process  $\mathbf{S}$

as follows:

$$S_t = \begin{cases} 0 & \text{if } S_{1,t} = 0 \text{ and } S_{2,t} = 0 \\ 1 & \text{if } S_{1,t} = 0 \text{ and } S_{2,t} = 1 \\ 2 & \text{if } S_{1,t} = 1 \text{ and } S_{2,t} = 0 \\ 3 & \text{if } S_{1,t} = 1 \text{ and } S_{2,t} = 1 \end{cases}$$

with transition probabilities  $P = P_1 \otimes P_2$ .

## References

- J. H. Albert and S. Chib. Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics*, 11:1–15, 1993.
- M. Billio, A. Monfort, and C.P. Robert. Bayesian estimation of Switching ARMA models. *Journal of Econometrics*, 93:229–255, 1999.
- C. K. Carter and R. Kohn. On Gibbs Sampling for State Space Models. *Biometrika*, 81:541–553, 1994.
- S. Chib. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75:79–97, 1996.
- S. Frühwirth-Schnatter. Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association*, 96:194–209, 2001.
- S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, Berlin, 2006.

- A. Gelfand and D. Dey. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society, Ser. B*, 56:501–514, 1994.
- J. Geweke. Using Simulation Methods for Bayesian Econometric Models: inference, development, and communication. *Econometric Reviews*, 18:1–73, 1999.
- C. Han and B.P. Carlin. MCMC Methods for Computing Bayes Factors: A Comparative Review. *Journal of the American Statistical Association*, 96:1122–1132, 2001.
- C. J. Kim and C.R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, 1999.
- S.N. Kim, N. Shephard, and S. Chib. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65:361–393, 1998.
- R.E. McCulloch and R. S. Tsay. Statistical Inference of Macroeconomic time series via Markov Switching models. *Journal of Time Series Analysis*, 15:523–539, 1993.
- C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, Berlin, 1999.
- N. Shephard. Partial Non-Gaussian State Space. *Biometrika*, 81:115–131, 1994.

C. A. Sims, D.F. Waggoner, and T. Zha. Methods for Inference in Large Multiple-Equation Markov-Switching Models. *Journal of Econometrics*, 146:255–274, 2008.