Essays on Empirical Macroeconomics

Yu-Fan Huang

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Richard Startz, Chair

Chang-Jin Kim

Yu-Chin Chen

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on Empirical Macroeconomics

Yu-Fan Huang

Chair of the Supervisory Committee:
Professor Richard Startz
Economics

The nonlinear time series analysis has been applied widely in many empirical macroeconomic studies and deepened our understanding of the economy. One of the traditional issues of most importance is how to describe the monetary policy conducted by the central bank. A *monetary policy reaction function* is a hypothetical function connecting the central bank's monetary policy instrument with the central bank's objectives, e.g., the stability of the output and inflation. The first two chapters in this dissertation add to the literature of the nonlinear monetary policy reaction function through the lens of the Taylor rule using U.S. data. On the other hand, understanding the macroeconomic dynamics has long been the main motivation of using nonlinear time series analysis. The state space model is one of the most useful techniques which can be applied to a wide range of empirical problems. The third chapter in this dissertation is concerned with a class of the state space models whose disturbances follow stochastic volatility processes.

For reasons which will become clear as this dissertation proceeds, I adopt Bayesian approaches, which are especially useful for comparing several competing empirical models and for estimating models that are nonlinear in latent variables, to deal with all issues under consideration.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Dick Startz, who provided valuable guidance and comments during my Ph.D. studies. I would like to offer my appreciation to Chang-Jin Kim and Yu-Chin Chen for their kind help and intriguing discussions. Financial support from the Buechel Fellowship is gratefully acknowledged. My special thanks are extended to Sui Luo, with whom I have been exchanging ideas and solving problems for three years. Finally, relentless encouragement and supports from my family are particularly appreciated.

DEDICATION

to my parents

Chapter 1

AIMING LOW OR REACTING STRONGLY? TIME-VARYING POLICY TARGETS AND REGIME SWITCHES IN U.S. MONETARY POLICY

This paper is concerned with the sources of time variation in U.S. monetary policy through the lens of the Taylor rule. Two important aspects of the Taylor rule are examined: the inflation target and the response coefficients. Allowing either or both of them to vary over time leads to four versions of the Taylor rule, which are empirically compared within a Bayesian framework. The results indicate that the most distinctive feature of U.S. monetary policy is the time-varying inflation targets, and that response coefficients have not been switching over time. The implications are twofold. First, aiming for a low inflation target is the most plausible channel through which U.S. monetary policy contributed to the Great Moderation. Second, assuming a constant inflation target might lead to a biased conclusion that the response coefficients have been switching over time.

Keywords. Taylor rule, Time-varying inflation target, Regime-switching response coefficients, Bayes factor.

1.1 Introduction

It has been recognized that the stability of the macroeconomy can be enhanced through proper economic policy actions, among which the monetary policy plays an indispensable role. In order to understand the main driving forces of the Great Moderation, a substantial reduction in the volatility of business cycle fluctuations starting in the mid-1980s, there has been burgeoning theoretical research on how the changes in the monetary policy *targets* and *reactions* affect the macroeconomic stability.[1] However, empirical studies have not yet

[1]For example, according to Coibion and Gorodnichenko [2011], changes in the monetary policy targets, especially the inflation target, could dramatically alter the determinacy region of a macroeconomic model and, therefore, the inference about the driving force of the Great Moderation.

provided an unambiguous answer to the following question: Is it the monetary policy targets or reactions that have been varying over the past five decades?

The broad and general notion of monetary policy can be formalized by the Taylor rule (Taylor [1993]), a hypothetical but representative policy rule. Under the simplest Taylor rule, the central bank adjusts the short-term nominal interest rate i_t around its equilibrium level $\bar{i}_t$ in response to the output gap x_t and to the deviation of inflation π_t from its target levels π^* according to:

$$i_t = \bar{i}_t + \alpha_\pi(\pi_t - \pi^*) + \alpha_x x_t. \tag{1.1}$$

Note that the inflation target *define* the gaps, $(\pi_t - \pi^*)$, meaning that the estimates of the response coefficient, α_π, depend to a great extent on how the unobserved inflation target are specified. Through the lens of the Taylor rule, the research question translates to investigating whether π^* or α_π is actually time-varying.

We propose an empirical framework based on a Markov-switching Vector Autoregression model, in which we impose restrictions so that the equation for the short-run nominal interest rate corresponds to the Taylor rule similar to equation (1.1). Most importantly, we do not assume that the monetary policy targets and reactions switch simultaneously, so that our empirical framework allows us to capture infrequent changes in monetary policy reactions under the different behaviors of the monetary policy targets. We tackle our question using a simple strategy based on the model comparison: we construct two classes of the Taylor rule: one assuming constant policy targets and one assuming time-varying policy targets. Within each class, two versions of the Taylor rule are estimated and compared: one with constant coefficients and one with Markov-switching coefficients. Examining these versions of the Taylor rule allows us to learn whether the Markov-switching response coefficients or time-varying policy targets are more preferred by the data.

Despite our simple empirical strategy, we are confronted with three technical issues along the way. First, to compare models within each class is equivalent to test the Markov-switching coefficients of the monetary policy rule. It is a challenging task within a frequentist framework in that some nuisance parameters are not identified under the null hypothesis of no regime switch, so the asymptotic distributions of the usual test statistics are nonstandard. Second, introducing the time-varying policy targets into the Taylor rule might cause an identification problem, implying that estimating time-varying policy targets and Markov-switching response coefficients jointly needs extremely strong prior beliefs on the values of parameters. Third, according to Sims and Zha [2006], failure to allow for heteroskedasticity properly can strongly bias statistical evidence in favor of significant shifts in response coefficients describing monetary policy. Furthermore, even when the heteroskedasticity is taken into account, it may be problematic if the coefficients in the MS-VAR are assumed to switch simultaneously with volatilities in shocks.

To deal with these difficulties, we cast the estimation within a Bayesian framework. One of the advantages of the Bayesian approach is that the model comparison or selection is simply conducted via the Bayes factor, which can be approximated accurately through the numerical simulation. Instead of estimating the monetary policy targets and reactions jointly, we adopt a two-step approach for the class of models assuming time-varying policy targets: in the first step, we obtain a set of estimates of the monetary policy targets using methods proposed in the literature; then we estimate the reaction parameters based on the estimated policy targets. Finally, we follow Bianchi [2013] and assume that coefficients of the empirical model and volatilities of shocks evolve as two independent Markov processes.

Our first empirical finding indicates that the most distinctive feature of U.S. monetary policy is the time-varying policy targets, especially the inflation targets. Among all models under our consideration, the "best-fit" model selected by the Bayes factor is the Taylor rule with constant response coefficients and time-varying policy targets. The implication of this result is two-fold: firstly, consistent with Sims and Zha [2006], we find no evidence of regime switches in the response coefficients. Therefore, we cannot corroborate the "good-policy" explanation of the Great Moderation proposed byClarida et al. [2000], which argued that U.S. monetary policy reactions changed a great deal and that the policy rule apparently followed in the 1970s implied "sunspot" fluctuations of an arbitrary large size. Secondly, the evidence for the time-varying inflation targets highlights the alternative good-policy explanation proposed by Coibion and Gorodnichenko [2011]: by lowering the inflation target, the stability of the economy could be enhanced even with no change in the response of the central bank to macroeconomic variables.

Our empirical results also show that assuming constant monetary policy targets might lead to the spurious evidence for the Markov-switching response coefficients. Within the class of models assuming constant policy targets, the model comparison indicates that the Taylor rule with Markov-switching response coefficients is "very strongly" preferred by the data according to the Bayes factor. We find, however, the opposite result within the class of models assuming time-varying policy targets. Probing deeper into the reason for these opposite findings, we find that the response coefficients might be forced to switch in the model in which the inflation targets are misspecified. It has been found, e.g., Ireland [2007], that the inflation target was high in the 1970s and has been low since early 1980s. Therefore, if the inflation targets are assumed to be constant over time, the inflation gap in the 1970s will be over-estimated, and the response coefficient will be underestimated. As a result, the estimated response coefficient of the inflation gap can be lower than it is in other periods.[2] We thus argue that special attention should be paid to the time-varying monetary policy targets embedded in an empirical monetary policy reaction function, e.g., the Taylor rule, especially when some coefficients are switching over time.

The outline of the remainder of the paper is as follows. Section 2 introduces the empirical frameworks and defines the classes of models under consideration. In section 3, we present the estimated time-varying policy targets and the procedure for estimating and comparing models under consideration. We discuss the empirical results in section 4. Section 5 concludes.

1.2 Model Specification and Class of Models

We use a Markov-Switching Vector Autoregression (MS-VAR) model of the form:

$$A(s_{c,t})\mathbf{Y}_t = \Phi_0(s_{c,t}) + \Phi_1(s_{c,t})\mathbf{Y}_{t-1} + \ldots + \Phi_p(s_{c,t})\mathbf{Y}_{t-p} + \boldsymbol{\epsilon}_t, \tag{1.2}$$

$$\epsilon_{k,t} \overset{i.i.d}{\sim} N(0, \sigma_k^2(s_{v,t})), \quad k = y, \pi, i. \tag{1.3}$$

[2] Cogley et al. [2010] also found similar results. By allowing inflation targets drifting overtime, they estimated a DSGE model for two subsamples: 1960:Q1 to 1979:Q3 and 1982:Q4to 2006:Q4, excluding the years of monetary targeting, and found that the response coefficients of inflation gaps in these two samples are very closed. They argued that the presence of a time-varying inflation target may reduces the difference between reactions to inflation in the two subsamples.

where $\mathbf{Y}_t$ is a 3×1 vector of endogenous variables, consisting of a measure of overall slack in the real economy (x), the inflation rate (π), and the short-run nominal interest rate (i).[3] A is a lower diagonal matrix with ones on the diagonal. The recursive identification scheme is also used by many empirical studies, e.g., Christiano et al. [2005]. To avoid potential bias caused by the heteroskedasticity of the macroeconomic time series, we follow Sims et al. [2008] and Bianchi [2013] and assume that the coefficients and the volatilities of the shocks in the MS-VAR model evolve independently as Markov processes. That is, $s_{c,t}$ and $s_{v,t}$ are two independent unobserved two-state Markov-switching variables that evolve according to transition probabilities given below:

$$\begin{aligned}
&Pr[s_{j,t} = 1 | s_{j,t-1} = 1] = p_{11,j} \\
&Pr[s_{j,t} = 0 | s_{j,t-1} = 1] = 1 - p_{11,j} \\
&Pr[s_{j,t} = 0 | s_{j,t-1} = 0] = p_{00,j} \\
&Pr[s_{j,t} = 1 | s_{j,t-1} = 0] = 1 - p_{00,j},
\end{aligned}$$

and a generic Markov-switching coefficient, θ, evolves as follow:

$$\theta(s_{j,t}) = \theta_0(1 - s_{j,t}) + \theta_1 s_{j,t}, \tag{1.4}$$

for $j = c, v$.

The MS-VAR model, equations (1.2), (1.3), and the evolving processes of $s_{c,t}$ and $s_{v,t}$, forms the basis of our empirical framework. Note that the lower triangular contemporaneous matrix, A, implies that the output growth and inflation do not respond instantaneously to a monetary policy shock, while the short-run nominal interest rate responds contemporaneously to the output growth and inflation. In line with the Taylor rule, the third equation in system (1.2) thus can be interpreted as the monetary policy reaction function. However, all the policy targets embedded in the Taylor rule are lumped into a intercept term of the system (1.2) , leading to the difficulty of clearly distinguishing changes in the monetary policy targets and reactions. To deal with this difficulty, we impose restrictions

[3] Sims et al. [2008] adopted a similar three-variables recursive MS-VAR to investigate whether the coefficients in the policy equation have changed or the variances for structural shocks have changed.

on coefficients of the third equation in system (1.2) based on the following four versions of the Taylor rule.

1.2.1 A constant coefficient Taylor rule with constant policy targets

Let i_t^* denote the Fed's target rate for the Federal Funds rate in period t:

$$i_t^* = \tilde{i} + \alpha_\pi[\pi_t - \pi^*] + \alpha_x x_t, \tag{1.5}$$

where π^* is the long-run inflation target, so $\tilde{i}$, by construction, is the desired nominal interest rate when both gaps are closed. We follow Clarida et al. [2000] and assume $\tilde{i} = rr + \pi^*$, and rr is the neutral, or equilibrium, real interest rate, which is determined by non-monetary factors in the long run. However, this specification assumes an immediate adjustment of the actual Funds rate to its target level, and thus ignores the Federal Reserve's tendency to smooth changes in the Funds rate. We follow Clarida et al. [2000] and assume the actual Federal Funds rate, i_t, is set by

$$i_t = (1 - \rho(1))i_t^* + \rho(L)i_{t-1} + \epsilon_{i,t}. \tag{1.6}$$

$\epsilon_{i,t}$ is the shock to the Federal Funds rate, reflecting possible randomness in policy actions or the Fed's imperfect control over interest rates. $\rho(L) = \rho_1 + \rho_2 L + \ldots + \rho_p L^{p-1}$ is a lag polynomial.

Combine (1.6) and (1.5), we obtain the following monetary policy reaction function:

$$i_t = \beta_0 + \beta_\pi \pi_t + \beta_y y_t + \rho(L)i_{t-1} + \epsilon_{i,t}, \tag{1.7}$$

where

$$\beta_0 = (1 - \rho(1))[rr + (1 - \alpha_\pi)\pi^*],$$
$$\beta_\pi = (1 - \rho(1))\alpha_\pi, \quad \text{and} \quad \beta_x = (1 - \rho(1))\alpha_x.$$

Note that equation (1.7) corresponds to the (restricted) interest rate equation in MS-VAR model[4]. The restrictions implied by the Taylor rule are that coefficients of lagged π_t and x_t in the interest rate equation are zero.

[4]It can be seen by letting $\beta_\pi = -a_{3,1}$ and $\beta_x = -a_{3,2}$, where $a_{i,j}$ is the (i,j) element of matrix A.

In what follows we introduce the other three versions of the Taylor rule, each of which is differentiated by its specification of the target rate, i.e., equation (1.5). Other model specifications, i.e., the desired nominal interest rate $\tilde{i} = rr + \pi^*$ and interest rate smoothing behaviors, are retained for all cases under consideration.

1.2.2 A regime-switching coefficient Taylor rule with constant policy targets

In this case, we allow the long-run responses to the inflation gap and the output growth gap in equation (1.5) to be Markov-switching as follows

$$\begin{aligned} i_t^* &= \tilde{i} + \alpha_\pi(s_{c,t})[\pi_t - \pi^*] + \alpha_x(s_{c,t})x_t, \\ &= \alpha_0(s_{c,t}) + \alpha_\pi(s_{c,t})\pi_t + \alpha_x(s_{c,t})x_t \end{aligned} \tag{1.8}$$

where $\alpha_0(s_{c,t}) = rr + (1 - \alpha_\pi(s_{c,t}))\pi^*$. $s_{c,t}$ is a unobserved two-state Markov-switching variable, and $\alpha_\pi(s_{c,t})$ and $\alpha_y(s_{c,t})$ evolve according to equation (1.4). Models similar to (1.8) have been used in the literature to investigate the regime-switching Monetary policy rule. For example, Davig and Leeper [2011] assume away the interest rate smoothing behavior and augment equation (1.8) with a error term. Bunzel and Enders [2010] consider the interest rate smoothing behavior and investigate the possibility that the Taylor rule should be formulated as a threshold process.

1.2.3 A constant coefficient Taylor rule with time-varying policy targets

Based on the linear Taylor rule (1.5), we allow the policy targets to be time-varying and obtain the third version of the Taylor rule as follows:

$$i_t^* = \tilde{i}_t + \alpha_\pi[\pi_t - \pi_t^*] + \alpha_x x_t, \tag{1.9}$$

where $\tilde{i}_t = r_t + \pi_t^*$.

Incorporating the time-varying policy targets involves in a technical difficulty. Under the same assumptions of the desired nominal interest rate and the interest rate smoothing behavior, the intercept term is time-varying because of time-varying policy targets. Simple regression technique cannot be used to estimate coefficients α_π and α_y, Instead, a filtering technique, such as Kalman filter, should be used to cope with the time-varying policy targets.

For example, Leigh [2008] use a state space model and the Kalman filter to estimate a model similar to (1.9) for the post-Volcker period. Even though under our model specification the Kalman filter cannot be directly applied because of the Markov-switching volatilities, the Kim [1994]'s filter can be used to deal with a general linear state space model with Markov-switching parameters.

We adopt, however, an alternative two-step approach for these models with time-varying policy targets because of a concern about the identification. For illustration, we consider the simplest Taylor rule as follows:

$$i_t = \alpha_\pi g_t + \epsilon_t, \tag{1.10}$$

where $g_t \equiv [\pi_t - \pi_t^*]$ is the inflation gap. Under this example, by letting $\pi_t^* = \pi_t - c g_t$, we could always multiply α_π by a factor c without changing the value of the likelihood function. For example, assume we have three observations of the inflation rate: {2%, 2.5%, 3%}, the following two outcomes are observational equivalent:

1. $\alpha = 2$, and $\pi_t^* = \{2\%, 2\%, 2\%\}$, and

2. $\alpha = 1$, and $\pi_t^* = \{2\%, 1.5\%, 1\%\}$.

The value of the likelihood function evaluated at these two sets of estimates are the same, so the likelihood-based inference cannot distinguish these two cases from each other. However, under the assumption of constant π^*, the other cases are ruled out except the first one. From this point of view, the assumption of constant policy targets can be taken as an important identification assumption.

The two-step approach works as follows: First obtain estimates of time-varying policy targets using methods that have been adopted successfully in the literature; Second, estimate coefficients based on estimated time-varying policy targets. The two-step approach has several appealing features: it reduces the computational cost substantially, given that we have paid the "fixed cost" in the first step; it allows us to check the robustness of our results with respect to various estimates of policy targets found in the literature; and it achieves better identification.

1.2.4 A regime-switching coefficient Taylor rule with time-varying policy targets

In our last case we allows the coefficients in (1.9) to be regime-switching as follows:

$$i_t^* = \tilde{i}_t + \alpha_\pi(s_{c,t})[\pi_t - \pi_t^*] + \alpha_x(s_{c,t})x_t, \tag{1.11}$$

where $\tilde{i}_t = r_t + \pi_t^*$. $s_{c,t}$ is unobserved two-state Markov-switching variable, and $\alpha_\pi(s_{c,t})$ and $\alpha_y(s_{c,t})$ evolve according to equation (1.4).

Estimating MS-VAR model embedding equation (1.11) and heteroskedasticity is a challenging task because it involves in computing and maximizing complicated likelihood function, which possibly has many local maximum. Furthermore, the identification issue becomes more severe because there are more possible cases that deliver the same value of the likelihood function. For example, given the simplest Taylor rule (1.10), in additional to the two possibilities provided above, we could have

3. $\alpha_t = \{2, 1, 1\}$, and $\pi_t^* = \{2\%, 1.5\%, 1\%\}$, and

4. $\alpha_t = \{1, 2, 2\}$, and $\pi_t^* = \{2\%, 2\%, 2\%\}$.

All these four outcomes are observationally equivalent. In fact, the number of possibilities can be expanded without limit.

Castelnuovo et al. [2013] made an interesting attempt to estimate regime-switching coefficients and time-varying policy targets jointly within a Bayesian framework. They incorporate a trend-cycle decomposition of the inflation to help identify the trend inflation, which is interpreted as the inflation target. Furthermore, the Bayesian inference combines researchers' prior beliefs and data information to form posterior distributions of parameters, so it provides an opportunity to achieve identification through the use of strong prior beliefs on the parameter space. Because our main interest is in the evidence of regime switches in coefficients, we do not employ strong prior beliefs on the coefficients and, instead, use the two step approach to cope with the identification issue.

1.3 Model Estimation and Comparison

To compare models with constant and regime-switching coefficients is essentially to test the hypotheses as follows:

$$H_0: \quad \alpha_{0,\pi} = \alpha_{1,\pi}, \quad \alpha_{0,x} = \alpha_{1,x}$$
$$H_1: \quad \alpha_{0,\pi} \neq \alpha_{1,\pi}, \quad \alpha_{0,x} \neq \alpha_{1,x}$$

However, in our empirical framework, the transition probabilities of Markov processes are nuisance parameters that do not exist under the null hypothesis of no regime switch. Thus, the problem is nonstandard, and the classical asymptotic results do not hold for the likelihood ratio test statistic. As such, we could therefore apply the testing procedure proposed by Hansen [1992] and Garcia [1998]. Alternatively, we cast the problem into a Bayesian framework because of several important features. First, the nuisance parameters that exist under the alternative but not under the null do not pose any special problem. The main issue in Bayesian model selection comes down to calculating the marginal likelihood for each model, obtained by numerically integrating the nuisance parameters out of the joint density. This procedure works whether the nuisance parameters exist under the null hypothesis or not. Second, as pointed out by Koop and Potter [1998], classical solutions also must integrate out the nuisance parameters, but the integration is with respect to an arbitrary distribution. In the Bayesian approach, there is a sense in which information in the data is used to integrate out nuisance parameters.

In this section, we first discuss estimation of the time-varying policy targets needed for the two-step approach. Then we introduce the Bayesian method for estimating model parameters and the computation method for the Bayes factor.

1.3.1 Estimation of Time-varying Policy Targets

In additional to the inflation target π_t^*, we also need a set of estimates of the neutral real interest rate because of our assumption of the desired nominal interest rate. The estimates of the neutral real interest rates are obtained according to the results from Laubach and Williams [2003], who find that the two-sided Kalman filter estimates and the two-sided estimates from the Hodrick-Prescott filters with the smoothing parameter of 6400 are very similar. Therefore, we let rr_t be equal to the estimates form the HP filters of ex post real rate, defined as the difference between the funds rate and the realized inflation rate. Specifically, we approximate the target growth rate of real GDP by the Hodrick-Prescott trend of the growth rate of real GDP. Finally, following Cogley and Sbordone [2008], we approximate long-run inflation target as the Beveridge-Nelson trend component of inflation from a reduced-form time-varying parameter VAR (TVPVAR) described briefly as follows:[5] Assume y_t, a $m \times 1$ vector, follows a TVPVAR process:

$$Z_t = \mu_t + F_t Z_{t-1} + E_t, \quad \epsilon_t \sim N(0, \Omega_t) \tag{1.12}$$

where $Z_t = [y_t', \ldots, y_{t-p+1}']'$. We follow Cogley and Sbordone [2008]'s method, to define trend inflation as the level to which inflation is expected to settle after short-run fluctuations die out, $\pi_t^* = \lim_{j\to\infty} E_t \pi_{t+j}$. They approximate this by calculating a local-to date t estimate of mean inflation from the VAR,

$$\pi_t^* = e_\pi'(I - F_t)^{-1}\mu_t.$$

where e_π is a selector vector. [6]

Figure (1.1) shows estimates of the desired interest rate, $rr_t + \pi_t^*$, the inflation target, π_t^*, and the output growth target, y_t^*, respectively. The estimated output growth target is stable around the 3.25%, the sample average of the annualized growth rate of real GDP over our sample period. The estimated inflation targets raised substantially in the period of Great Inflation, 1970s, and lowered down at the beginning of Paul Volcker's tenure as its Chairman. This finding is consistent with Paul Volcker's well-known disinflation policy in early 1980s. After Alan Greenespan took office on August, 1987, the inflation target

[5]They interpret movements in trend inflation as changes in the Federal Reserve's inflation target.

[6]The data we used for estimating TVPVAR is the same as those used by Cogley and Sargent [2005], consisting of unemployment rates, inflation rates, and interest rates. Unemployment rates are measured by all civilian unemployment rate, inflation rates by annualized quarterly rate of change of Chain-type GDP deflater. and the interest rates by the yield on 3-Month Treasury bill rates. The sample spans from 1948:Q1 to 2010:Q4, and we work with VAR(2) representation for nominal interest, quarterly inflation, and logit of unemployment. The MCMC method used to estimate TVPVAR can be found in Cogley and Sargent [2005] and Koop and Korobilis [2010], and we thank authors for generously providing Matlab codes for replication.

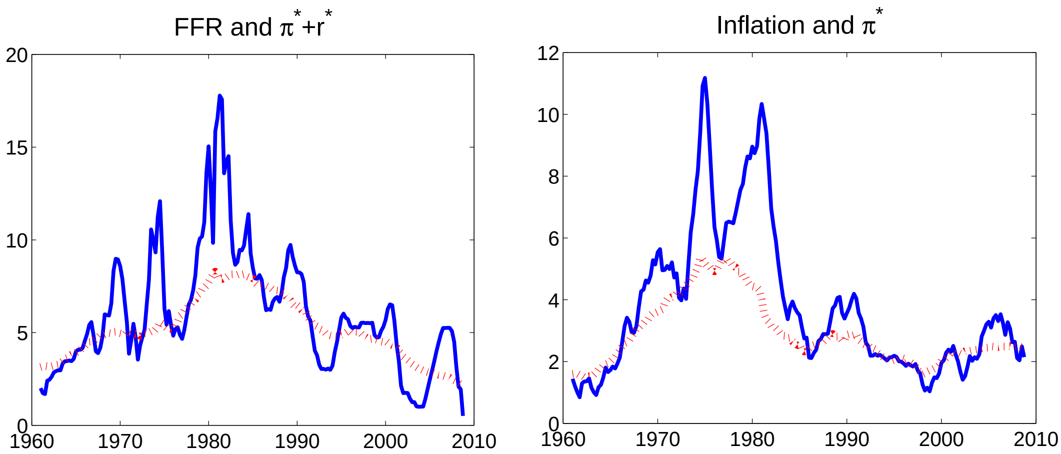


Figure 1.1: Data and Time-varying Policy Targets

had been around 2%, a value known to be the target for most advanced country central banks (Romer and Romer [2002]) and a sufficient cushion to make the zero lower bound unimportant in a world of small shocks (Blanchard et al. [2010]).

1.3.2 Estimation of Regime-switching Coefficients

For all models we estimate and compare, we draw samples from the the marginal posterior distributions for the following parameters:

$$\begin{aligned} \mathbf{\Phi} &= [A(j), \Phi_0(j), \ldots, \Phi_p(j)], \quad \boldsymbol{\sigma} = [\sigma_1^2(j), \sigma_2^2(j), \sigma_3^2(j)], \\ \mathbf{S_c} &= [s_{c,1}, \ldots, s_{c,T}], \quad \mathbf{S_v} = [s_{v,1}, \ldots, s_{v,T}], \\ \mathbf{p} &= [p_{00,c}, p_{11,c}, p_{00,v}, p_{11,v}], \end{aligned}$$

for $j = 0, 1$, given appropriate prior distributions of them. These marginal posterior distributions may be obtained from the joint posterior distribution

$$f(\mathbf{\Phi}, \boldsymbol{\sigma}, \mathbf{S_c}, \mathbf{S_v}, \mathbf{p}|Data).$$

However, the hierarchical nature of the model allows us to easily employ the Gibbs sampling to obtain the marginal posterior distributions of interest. This is done by successively sampling from the full conditional densities. The following describes the Gibbs sampling procedure:

Choose a starting values, for $g = 1, \ldots, G$:

(1) Generate $\mathbf{\Phi}^{(g)}$ from $f(\mathbf{\Phi}|Data, \boldsymbol{\sigma}^{(g-1)}, \mathbf{S_c}^{(g-1)}, \mathbf{S_v}^{(g-1)})$.

(2) Generate $\boldsymbol{\sigma}^{(g)}$ from $f(\boldsymbol{\sigma}|Data, \mathbf{\Phi}^{(g)}, \mathbf{S_c}^{(g-1)}, \mathbf{S_v}^{(g-1)})$.

(3) Generate $\mathbf{S_c}^{(g)}$ from $f(\mathbf{S_c}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_v}^{(g-1)}, \mathbf{p}^{(g-1)})$.

(4) Generate $\mathbf{S_v}^{(g)}$ from $f(\mathbf{S_v}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_c}^{(g)}, \mathbf{p}^{(g-1)})$.

(5) Generate $\mathbf{p}^{(g)}$ from $f(\mathbf{p}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_c}^{(g)}, \mathbf{S_v}^{(g)})$.

Because of the recursive structure, the MS-VAR (1.2) can be estimated equation-by-equation. Therefore, conditional on unobserved states and variances of structural shocks, the conditional density of $\mathbf{\Phi}$ can be further broken down into three blocks, one for each equation in (1.2). For the equation of the nominal interest rate, on which we impose the restrictions implied by the Taylor rule, the conditional density is slightly different but is still known up to a normalizing constant, so the Gibbs sampling can be used. Other steps are relatively standard,[7] so we leave the details of the posterior simulation in the appendix.

1.3.3 Model Comparison

Let ω be the model indicator parameter. Thus, when $\omega = k$, we assume that the data have arisen from case k, according to a probability function (marginal likelihood) $m(Y_T|\omega = k)$. The Bayes factor is defined as the ratio of marginal likelihoods (or marginal data density) for models under consideration:

$$B_{k,h} = \frac{m(Y_T|\omega = k)}{m(Y_T|\omega = h)}, \quad k \neq h,$$

[7] See Kim and Nelson [1999] for extensive introduction to the Bayesian approach of Markov Switching models.

where $B_{k,h}$ refers to the Bayes factor in favor of case k. The Bayes factor, according to Kass and Raftery [1995], 'is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another'. Given the marginal likelihoods and Bayes factors, the comparison between models can be made according to the criteria provided by Kass and Raftery [1995]:

$2 \ln B_{i,j}$	Evidence against M_j
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
>10	Very strong.

Since in most of the case we do not know the exact form for $m(Y_T|\omega = i)$, recently literature of Bayesian model comparison relies heavily on simulation approximation. The modified harmonic mean (MHM) method of Dey [1994] has been a widely used method for computing the marginal data density. Let θ as a collection of all the free parameters in the model. Denote the likelihood function by $p(Y_T|\theta)$ and the prior density by $p(\theta)$. Given these two objects, the marginal likelihood is defined as

$$m(Y_T|\omega = i) = \int p(Y_T|\theta)p(\theta)d\theta.$$

The MHM method used to approximate the above integral numerically is based on the following theorem

$$m(Y_T|\omega = i)^{-1} = \int_\Theta \frac{h(\theta)}{p(Y_T|\theta)p(\theta)} p(\theta|Y_T)d\theta, \tag{1.13}$$

where Θ is the support of the posterior probability density and $h(\theta)$, often called a weighting function, is any probability density whose support is contained in Θ. Denote

$$g(\theta) = \frac{h(\theta)}{p(Y_T|\theta)p(\theta)}.$$

A numerical evaluation of the integral on the right hand side of (1.13) can be accomplished in principle through the Monte Carlo (MC) integration

$$m(Y_T|\omega = i)^{-1} = \frac{1}{N}\sum_{n=1}^{N} g(\theta^{(n)}), \tag{1.14}$$

where $\theta^{(i)}$ is the ith draw of θ from the posterior distribution $p(\theta|Y_T)$. If $g(\theta)$ is bounded above, the rate of convergence from this MC approximation is likely to be practical.

Geweke [1999] proposes an implementation with $h(.)$ constructed from the posterior simulator. The sample mean $\bar{\theta}$ and sample covariance matrix $\bar{\Omega}$ can be calculated from draws of θ from the posterior simulator. The weighting function is chosen to be a truncated multivariate Gaussian density with mean $\bar{\theta}$ and covariance $\bar{\Omega}$. The tail of the Gaussian distribution is truncated to ensure that the support of the weighting function is contained in the support of posterior. However, Sims et al. [2008] argue that if the parameters are allowed to vary over time, the posterior density tends to be non-Gaussian can be very low at the sample mean, especially when the posterior density has multiple peaks. Therefore, a truncated Gaussian density function tends to be a poor local approximation to the non-Gaussian posterior density. Furthermore, the likelihood can get close to zero in the interior points of the parameter space. To deal with these problems, Sims et al. [2008] propose a new implementation of the MHM method based on a family of elliptical distribution centered at posterior mode, $\hat{\theta}$, and scaled by

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\left(\theta^{(i)} - \hat{\theta}\right)\left(\theta^{(i)} - \hat{\theta}\right)'$$

Interested readers are referred to Sims et al. [2008] for thorough discussion and computational details.

1.4 Empirical Results

1.4.1 Data and Priors

We use U.S. quarterly data spanning from 1961:1 to 2008:4 in this section. We measure the overall slack in the real economy as $x_t = y_t - \tau_t$, where y_t represents the 4-quarter growth rate of the real GDP and τ_t is the Hodrick-Prescott trend of y_t. The inflation rate is measured by the 4-quarter growth rate of the GDP deflator, and we use the effective federal funds rates as the short-run nominal interest rate.

We assume that parameters are a-prior independently distributed across blocks, so that the joint prior can be expressed as the product of marginal priors,

$$f(\mathbf{\Phi}, \boldsymbol{\sigma}, \mathbf{p}) = f(\mathbf{\Phi}) f(\boldsymbol{\sigma}) f(\mathbf{p})$$

We employ Normal priors for $\mathbf{\Phi}$, inverted Gamma distributions for $\boldsymbol{\sigma}$, and, finally, Beta distributions for $\mathbf{p}$. The prior mean of the coefficients of the growth rate and the inflation equation are all zero except for the elements corresponding to the first own lag of the dependent variable in each equation, and prior variance are set to be 1. These non-zero elements are set to be 0.9, expressing a belief that the individual variable follows AR(1) process. For the interest rate equation, on which the Taylor rule is imposed, the prior mean and variance are set as follows: $\alpha_{i,x} \sim N(0.5, 0.5^2)$ for $i = 1, 2$, $\alpha_{1,\pi} \sim N(1, 0.5^2)$ and $\alpha_{2,\pi} \sim N(1.5, 0.5^2)$. The priors for the response to the output growth are symmetric across regimes, but we set asymmetric priors for the responses to inflation. This is meant to reflect the a-priori belief that the inflationary stance of the Federal Reserve has changed over time. However, the priors are very loose, implying a substantial degree of overlapping. The smoothness parameters, ρ is assumed to follow a normal distribution: $\rho \sim N(0.6, 0.2^2)$. The priors for all elements in $\boldsymbol{\sigma}$ are inverted Gamma $IG(1, 1)$. Finally, the priors for $\mathbf{p}$ are $Beta(50, 2)$, implying the a-prior expected duration of each regime to be 26 quarters, or 6.5 years.

Throughout this section, all inferences are based on 50,000 Gibbs simulations, after discarding the initial 10,000 Gibbs simulations in order to mitigate the effects of initial conditions. The number of lags in the MS-VAR is chosen to be 2 according to Bayes factors.[8]

While estimating models we allow only the coefficients in the interest rate equation are switching over time. This assumption is motivated by one of Sims and Zha [2006]'s findings that the best-fit model among those that do allow coefficients to change is one that constrains the changes to occur only in the monetary policy equation, while coefficients in the other equations remain constant.

Table 1.1: Log Marginal Likelihood of Different Versions of the Taylor Rule

	Cont. Coef.	MS Coef.
Cont. Targets	-521.4319	-516.3000
TV Targets	-505.8665	-513.5657

* MS Coef. refers to Markov-switching response coefficient. TV Targets refers to time-varying policy targets.

1.4.2 The "Best-fit" Monetary Policy Reaction Function

Table (1.1) shows log marginal likelihoods of all four models under consideration. The logarithm of the Bayes factor comparing two versions of the Taylor rule is the difference between the log marginal likelihoods; therefore, table (1.1) clearly indicates that the "best-fit" version of the Taylor rule is one with *constant* response coefficients and *time-varying* policy targets, and that the evidence is very strong, according to Kass and Raftery [1995]'s criteria.

Knowing the best-fit version of the Taylor rule is not sufficient to answer our main question: Is it the monetary policy targets or reactions which have been varying over the past five decades? By comparing the two versions of the Taylor rule assuming constant policy targets, whose log marginal likelihoods are listed in the first row of table (1.1), we find that the Markov-switching response coefficients is "very strongly" preferred by the U.S. data. Therefore, even though the best-fit model is not characterized by Markov-switching response coefficients, we cannot eliminate the possibility that the monetary policy reactions have been changing as well.

The fact that the version of the Taylor rule with both time-varying policy targets and reactions is not selected by the Bayes factor provides a clue to solve this problem. If the Taylor rules incorporating time variation in policy targets or reactions *along* could have actually improved upon the benchmark Taylor rule, how could the Taylor rule characterized by *both* features not be selected as the best-fit model? One possible reason is that one of

[8] Marginal likelihoods of models with different lags are not reported but available upon request.

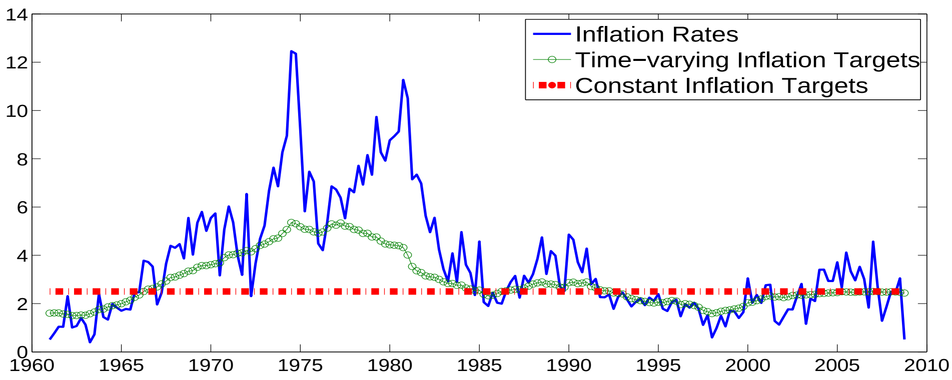


Figure 1.2: Inflation Rates and Inflation Targets

the two features does not in fact help describing the U.S. monetary policy behaviors, and it seems so just because the other feature is mis-specified.

In Figure (1.2), we plot inflation rates, measured by annualized quarterly change of GDP deflators, and two sets of estimates of inflation targets to contrast the behaviors of the inflation gap.[9] Notice that in 1970s, the inflation gap implied by the constant inflation target was much larger than that implied by the time-varying inflation target, revealing a possibility that the evidence for Markov-switching response coefficients is just a consequence of mis-specification. That is, in the case where the time variation in the inflation target is not properly taken account for, the Fed's responses to inflation gaps can exhibit significant time variations even if in fact they are better described by constant coefficients.

1.4.3 Are response coefficients Markov-switching?

To verify our explanation of the puzzling results from model comparisons, we further investigate estimates of parameters and smoothed probabilities of $s_{c,t}$ in all four cases. Table (1.2) shows estimation results from models assuming constant policy targets, that is, equations (1.5) and (1.8). Note that estimates of αs are the Fed's "long-run" responses to respective

[9]The constant inflation target is set to be 2.5 %, which is close to the value found in Bianchi [2013], and the time-varying inflation targets are based on our estimates using a time-varying parameter vector autoregression model as in Cogley and Sbordone [2008].

Table 1.2: Estimation Results from Models Assuming Constant Policy Targets

	Constant Coefficients			Regime-Switching Coefficients		
	Mean	SD	90% Bands	Mean	SD	90% Bands
$\alpha_y(s_c = 0)$	1.58	0.38	(0.97, 2.21)	0.82	0.31	(0.36, 1.36)
$\alpha_y(s_c = 1)$	-	-	-	0.94	0.36	(0.40, 1.58)
$\alpha_\pi(s_c = 0)$	1.56	0.23	(1.19, 1.94)	2.15	0.33	(1.62, 2.72)
$\alpha_\pi(s_c = 1)$	-	-	-	1.12	0.23	(0.75, 1.48)
$\alpha_0(s_c = 0)$	0.46	0.54	(-0.43, 1.34)	1.04	0.69	(-0.11, 2.14)
$\alpha_0(s_c = 1)$	-	-	-	-0.04	0.84	(-1.42, 1.34)
ρ_1	1.31	0.06	(1.21, 1.42)	1.25	0.07	(1.14, 1.36)
ρ_2	-0.36	0.06	(-0.48, -0.29)	-0.35	0.06	(-0.45, -0.24)

* SD refer to standard deviation. 90% Bands refers to 90% posterior probability bands.

targeting variable, and the contemporaneous, or short-run, responses can be obtained by $(1-\rho)\alpha$. Concerning the parameters of the Taylor rule, the Federal Funds rate reacts to deviations of inflation from its target more strongly under Regime 1 ($s_c = 0$) than it does under Regime 2 ($s_c = 1$). Thus I shall refer to Regime 1 as the Hawk regime, while Regime 2 will be the Dove regime. The output growth seems to play a similar role in both cases: the long run response to output growth gap is not significantly different between two regimes since 90 % bands of them overlap substantially.

Figure (1.3) shows the smoothed probabilities assigned to the Dove regime. The monetary policy had been under Dove regime for most of the pre-1980 period. After late 1970s, when long period of passive monetary policy ended, the stance of the monetary policy had been more anti-inflationary for most the 1980s and the 1990s, except that Dove regime recurred in the first half of 1990s. These regime-switching behaviors of the monetary policy before 2000 are, by and large, consistent with Clarida et al. [2000]'s empirical results: interest rate policy in the Volcker-Greenspan period appears to have been much more sensitive

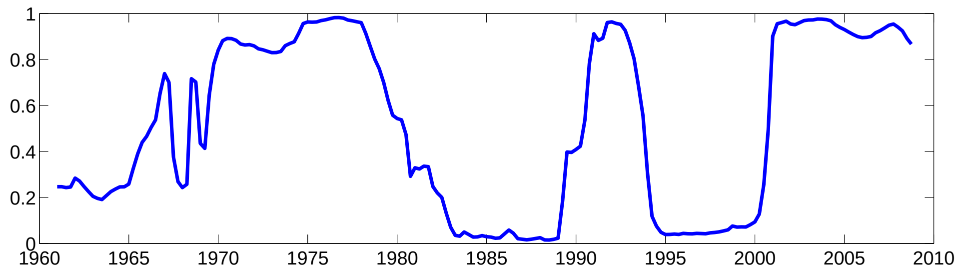


Figure 1.3: State Probabilities

to changes in inflation than in the pre-Volcker period. However, the Dove regime has returned to dominate since the beginning of 2000s. This result corroborates the view of Sims and Zha [2006] that monetary policy regime changes are better modeled as stochastic and reversible.

The results from models assuming constant policy targets point to a clear conclusion that the monetary policy was subject to significant regime switches. However, we notice that the time when the regime switches remarkably coincides with the time when the inflation target is subject to change. For example, it is well-known that the inflation target dropped substantially after Paul Volcker took office in October 1979, e.g., Kozicki and Tinsley [2009] and Cogley et al. [2010]. Furthermore, as shown in figure (1.2), in the mid of 1990s, the inflation target declined significantly after the recession of the early 1990s.[10] Even though the estimates of the inflation targets suffer from a great amount of uncertainty, it cannot be ruled out that the regimes detected in figure (1.3) actually capture time variations in the policy targets instead of that in the responses coefficients.

Table (1.3) shows estimation results obtained from models assuming time-varying policy targets. These estimates are, by and large, similar to those from models assuming constant policy targets, but with a greater degree of uncertainty. Notice that the long-run response to the inflation gap enlarges under Hawk regime, but the 90% bands of α_π under two regimes

[10]The decline of the inflation target in the mid of 1990s is consistent with the idea of the opportunistic approach to disinflation, by Orphanides and Wilcox [2002], which states that the central bank should wait for exogenous circumstances - e.g. favorable supply shocks and unforeseen recessions - to deliver the desired additional reduction in inflation.

Table 1.3: Estimation Results from Models Assuming Time-varying Policy Target

	Constant Coefficients			Regime-Switching Coefficients		
	Mean	SD	90% Bands	Mean	SD	90% Bands
$\alpha_y(s_c = 0)$	0.72	0.21	(0.42, 1.09)	0.13	0.48	(-0.62, 0.95)
$\alpha_y(s_c = 1)$	-	-	-	0.75	0.27	(0.38, 1.23)
$\alpha_\pi(s_c = 0)$	1.33	0.26	(0.92, 1.76)	2.67	0.69	(1.42, 3.68)
$\alpha_\pi(s_c = 1)$	-	-	-	0.77	0.30	(0.25, 1.24)
ρ_1	1.28	0.06	(1.18, 1.37)	1.19	0.06	(1.09, 1.29)
ρ_2	-0.41	0.05	(-0.50, -0.33)	-0.36	0.06	(-0.45, -0.27)

* SD refer to standard deviation. 90% Bands refers to 90% posterior probability bands.

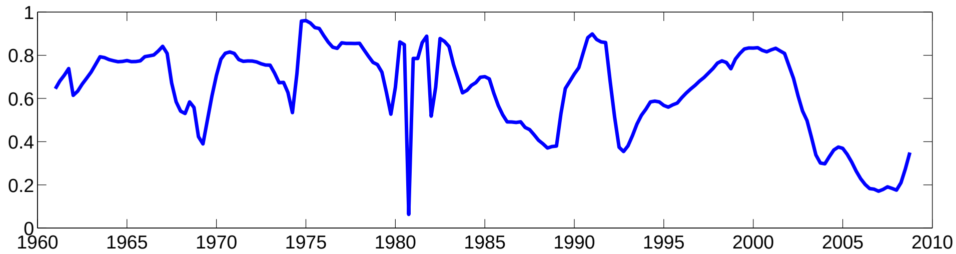


Figure 1.4: State Probabilities

overlaps substantially. Furthermore, figure (1.4) shows the smoothed probabilities assigned to the Dove regime over our sample period. We could see that these smoothed probabilities do not exhibit obvious regime-switching patterns as shown clearly in figure (1.3).

To sum up, the empirical results shown above point to following two important messages: 1) When the monetary policy targets are assume to be constant over time, the evidence for Markov-switching response coefficients is very strong; 2) Once the monetary policy targets are considered to be time-varying, there is no evidence for regime switch in response coefficients. They imply that improvements in the marginal likelihood we observed in the first row of table (1.1) is possibly a consequence of the mis-specified monetary policy targets.

Table 1.4: Log Marginal Likelihood under Different Priors

	CC-CT	RS-CT	CC-TT	RS-TT
Prior A.1	-521.6667	-516.9395	-506.0170	-513.2383
Prior A.2	-520.0728	-517.1238	-506.7362	-513.6370
Prior A.3	-529.9243	-517.0785	-505.1758	-506.4708

* CC refers to constant coefficients. RS refers to regime-switching coefficient. CT refers to constant policy targets. TT refers to time-varying policy targets.

Therefore, a better strategy to understand the role played by the monetary policy reactions might be to compare the second row of table (1.1), which leads to the answer to our main question: it is the monetary policy targets, *not reactions*, which has been varying over the past five decades.

1.4.4 Robustness Check

Within the Bayesian framework, inferences are sometimes dependent upon the priors employed, so it is important to evaluate the Bayes factor over a range of possibilities. We use informative and proper priors to ensure that the Bayes factor is well defined, but hyperparameters we specified imply disperse prior distributions except for two sets of parameters: long-run response coefficients $\boldsymbol{\alpha}$ and transition probabilities $\mathbf{p}$. They are important parameters which determine how the monetary policy rule evolves over time, thus it is crucial to check whether our results hinge on the prior distributions we specify.

We specify the following alternative priors to reflect different prior beliefs about how the monetary policy might evolve overtime:

- Prior A.1: $\mathbf{p} \sim$ Beta(10,1). Others remain the same.

- Prior A.2: $\alpha_{i,y} \sim N(0.5, 1)$ for $i = 1, 2$, $\alpha_{1,\pi} \sim N(1, 1)$ and $\alpha_{2,\pi} \sim N(1.5, 1)$. Other remain the same.

- Prior A.3: $\alpha_{i,y} \sim N(0.5, 0.1^2)$ for $i = 1, 2$, $\alpha_{1,\pi} \sim N(1, 0.1^2)$ and $\alpha_{2,\pi} \sim N(1.5, 0.1^2)$. Other remain the same.

Prior A.1 implies a wide range of the expected duration of each regime: 90% interval is $[3.86, 195.46]$, or 0.97 to 48.87 years. Prior A.2 increases the prior variance of each response coefficient, reflecting the prior beliefs that response coefficients are *not* different across regimes. On the other hand, prior A.3 decreases the prior variance of each response coefficient, reflecting the prior beliefs that response coefficients are *very* different across regimes. Table (1.4) shows the log marginal likelihood computed using three alternative priors: The constant coefficient Taylor rule with time-varying policy targets is always preferred by the data under all cases. Even under the prior A.3 which imply very strong prior beliefs, the regime-switching coefficient Taylor rule still cannot improve upon the constant coefficient Taylor rule.

1.5 Conclusion

Searching for regime switches in the monetary policy is complicated by the time-varying monetary policy targets. Despite its appealing ability to extremely simplify empirical model, assuming constant policy targets might result in misleading evidence for regime switches in coefficients of the Taylor rule. In this paper, we show that time-varying policy targets play a crucial role in U.S. monetary policy in the past five decades. The Bayesian model comparison indicates that constant coefficient Taylor rule with time-varying policy targets is very strongly preferred by the data. Furthermore, once we incorporate the time-varying policy targets in the Taylor rule, the response coefficients of the monetary policy rule were not subject to regime switches, even under the very strong prior beliefs about the regime-switching response coefficients.

Our empirical results highlight the role of U.S. monetary policy in the Great Moderation. Sims and Zha [2006] argue against the original "good policy" explanation proposed by Clarida et al. [2000] and shows that there were no regime switch in the response coefficients describing U.S. monetary policy. From this perspective, we agree with Sims and Zha [2006] in that our results also show no evidence for the regime switch in the response coefficients.

However, we find empirical supports for the alternative “good policy” explanation for the Great Moderation proposed by the Coibion and Gorodnichenko [2011], who argued that by lowering the inflation target, the monetary authority could enhance the stability of the U.S. economy even with no change in the response of the central bank to macroeconomic variables.

Our analysis also sheds some light on modeling nonlinear economic time series, e.g. smooth transition autoregression and Markov-switching autoregression, which have been used to understand the nature of the economy over the course of a business cycle. Because many macroeconomic models assume that some macroeconomic variables only deviate from their “neutral levels” temporarily and refer to the deviation as “gap”, statistical properties of the gap come to the center of economists’ attentions. The lesson we learned from this paper is that special attention should be given to the neutral level of the macroeconomic variable under consideration, especially when the neutral level is arguably changing over time.

Chapter 2

WAS THERE NONLINEARITY IN U.S. MONETARY POLICY? A BAYESIAN APPROACH

In this paper, several versions of nonlinear forward-looking Taylor rule are estimated and compared within a Bayesian framework. Endogeneity problem caused by forward-looking variables are dealt with using a Bayesian Instrumental Variable estimator. Empirical evidence indicates that regime-switching behaviors of the Federal Reserve emerged after 1990. The regime is characterized by the sign of the output gap, meaning that the Fed reacts to inflation and output gap aggressively during a recession. Without a concern of recession, adjustments of the federal funds rate become very persistent and the volatility of monetary policy shocks shrinks substantially. Furthermore, the best-fit nonlinear Taylor rule can describe U.S. monetary policy better than a linear Taylor rule.

Keywords. Nonlinear Taylor rule, Forward-looking, Regime-switching, Bayesian Instrumental Variable.

2.1 Introduction

Linking movements in the federal funds rate to macroeconomic variables by a linear equation, or so-called policy rule, help economists and the general public understand how U.S. monetary policy is conducted. However, without committing to a specific rule, the Federal Reserve's (Fed) decisions on adjusting the federal funds rate might not be invariant to the ongoing economic situation. It is not hard to imagine a scenario that the Fed's reactions are more accommodative when key macroeconomic variables are wandering inside a certain "comfort zone", and aggressive when outside. In this paper, we empirically investigate the possibility that U.S. monetary policy could be described better by a nonlinear Taylor rule.

As put by Taylor [1993], Taylor rule is a "hypothetical but representative policy rule" which links the federal funds rate to the inflation gap and the output gap. It has been

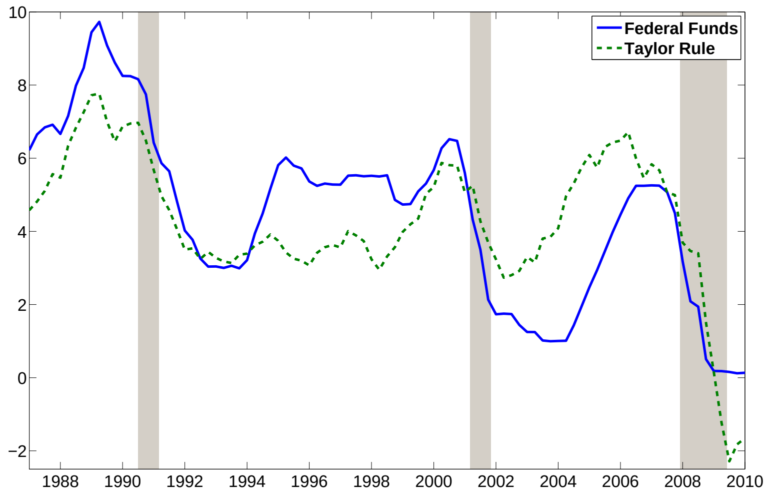


Figure 2.1: The Classic Taylor Rule

used widely in many theoretical macro models as a equation describing how the short-term nominal interest rate is adjusted. In figure 2.1, we plot the quarterly federal funds rate and the policy rate constructed using the classic Taylor rule, $r_t = 2 + \pi_t + 0.5(\pi_t - 2) + 0.5y_t$, where π_t is the inflation rate and y_t is the output gap[1]. Gray bars represent the NBER recession dates. Despite that the policy rate moved closely with the federal funds rate around these recession dates, there are periods where the Taylor rule does not capture movements of the federal funds rate well. Notably, the period from 2002 to 2006, according to Taylor [2011], is "a period during which policy deviated from the practice of at least the previous two decades." Furthermore, Bernanke [2006] also notes that the tightening cycle began at the end of 2004 is notably different from previous experiences.

Motivated by the periodic failure to capture movements in the federal funds rate, we are interested in whether the nonlinear Taylor rule can serve an alternative to the linear Taylor. We use the forward-looking Taylor rule proposed by Clarida et al. [2000] as the benchmark linear model of U.S. monetary policy. The nonlinearity is introduced by augmenting the

[1] As Taylor [1993], we use yearly rate of change of the GDP deflator as a measure of inflation and CBO's estimates of output gap

linear model with additional terms capturing changes in behaviors between "regimes", which are pre-determined and implied by theoretical models of optimal monetary policy. For example, Bec et al. [2002] assume Fed's preference on output gap is asymmetric and find that the optimal reaction to positive and negative output gap are different. In this case, we can define the regime according to the sign of the output gap. We also examine the possibility that U.S. monetary policy underwent a structural break in the mid of our sample. In such cases, regime-switching behaviors would only emerge before or after a pre-specified break date.

We cast the estimation into a Bayesian framework, within which marginal likelihoods can be used to compare models directly and, in the case where several competing models are similar, serve as the weights for averaging models. The endogeneity problem caused by replacing forward-looking variables in the Taylor rule by ex-post realized data is dealt with using the standard Instrumental Variable (IV) estimation to. Following Cogley and Startz [2012], we cast the IV equations into a seemingly unrelated regressions (SUR) system with nonlinear parameters, so Gibbs sampling can be used to draw samples from posterior distributions of parameters.

Our empirical results indicate that the Fed's reactions might be different when the economy is producing above potential output, and this regime-switching behavior emerged after 1990. Estimates obtained from the best-fit model imply that changes in reactions are two-fold: When the Fed perceives a concern of recession, it (1) adjusts the funds rate relatively quicker and (2) reacts to macroeconomic variables more aggressively than it does otherwise. In addition, in the regime where the output gap is positive, the Fed's reactions are characterized by greater persistence, defined as the sum of AR coefficients, and smaller volatility of monetary policy shocks. This finding is in line with Basistha and Startz [2004], who find that the 1990s saw a longer duration in the regime where the desired federal funds rate is static, and Bernanke [2006]'s observation that policy moved more gradually than it did before 2004. We also compare the performance of linear and nonlinear Taylor rule on capturing movements of the Federal Funds rate. The best-fit nonlinear Taylor rule does better describe the U.S. monetary policy reactions than the linear Taylor rule.

The outline of the remainder of the paper is as follows. Section 2 introduces the empirical frameworks and specify nonlinear Taylor rules under consideration. In section 3, we present the Bayesian IV estimation using Gibbs sampler. We discuss the empirical results and implications in section 4. Section 5 concludes.

2.2 Modeling Monetary Policy Reactions

2.2.1 A forward-looking Taylor rule

Our empirical framework is based on a linear forward-looking Taylor rule with interest rate smoothing:

$$\begin{aligned} i_t =&(1-\rho(1))[\tilde{i} + \alpha_\pi E_t(\pi_{t+1} - \pi^*) + \alpha_y E_t(y_{t+1})] \\ &+ \rho(L) i_{t-1} + m_t, \end{aligned} \tag{2.1}$$

where $\rho(L) = \rho_1 L + \ldots + \rho_p L^p$ is the lag polynomial, and m_t is a random disturbance term, or the monetary policy shock. π_t and y_t are the inflation and the output gap, so $\tilde{i}$ by construction is the desired nominal interest rate when both inflation gap and output gap are closed. We follow Clarida et al. [2000] and assume that $\tilde{i} = \bar{r} + \pi^*$, where $\bar{r}$ is the neutral, or equilibrium, real interest rate, and π^* is the long-run inflation target. $E_t(x_{t+1})$ refer to the 1-step-ahead predictions of variable x_t formed by the Fed conditional on the information available at the beginning of time t. Replacing expected values by realized future values, we have

$$i_t =(1-\rho(1))[\tilde{i} + \alpha_\pi(\pi_{t+1} - \pi^*) + \alpha_y y_{t+1}] + \rho(L) i_{t-1} + e_t, \tag{2.2}$$

where

$$e_t = m_t - (1-\rho(1))[\alpha_\pi(\pi_{t+1} - E_t(\pi_{t+1})) + \alpha_y(y_{t+1} - E_t(y_{t+1}))].$$

Note that regressors, π_{t+1} and y_{t+1}, in equation (2.2) are correlated with the disturbance term e_t; therefore, OLS estimator is not consistent. We follow Kim and Nelson [2006] and deal with the endogeneity problem using Instrumental Variable (IV) estimation.

The Taylor rule contains two unobserved policy variables: $\bar{r}$ and π^*. In the original paper, Taylor [1993] assumes $\bar{r}$ and π^* are both constant and equal to 2 %, which are reasonable measures for short sample period. We do not follow this specification because assuming constant π^* may introduce bias to estimates of response coefficients in the period when the inflation was higher than the sample average. It has been found that inflation targets were higher in 1970s other periods, see, for example, Ireland [2007], thus assuming constant inflation targets would overstate inflation gaps in 1970s and understate the Fed's reactions to inflation gaps. To avoid these problems, we obtain a set of measures of time-varying $\bar{r}_t$ and π_t^* prior to all estimation, and rewrite equation (2.2) as follows

$$\begin{aligned}(i_t - \tilde{i}_t) =& \beta_\pi g_{t+1} + \beta_y y_{t+1} + \rho(1)(i_{t-1} - \tilde{i}_t) \\ &+ \rho_2 \Delta i_{t-1} + \ldots + \rho_p \Delta i_{t-p} + e_t,\end{aligned} \tag{2.3}$$

where $g_{t+1} = \pi_{t+1} - \pi_t^*$ is the inflation gap. Note that rewriting the lagged terms of i_t into the form of the Dickey-Fuller regression does not change the equation fundamentally. It only changes the interpretation of autoregressive coefficients – $\rho(1)$ can be interpreted as the persistence, or smoothness, of monetary policy reactions. We will discuss the selection of $\bar{r}_t$ and π_t^* latter.

By defining z_t as a $q \times 1$ vector of instrumental variables, we have the following instrumenting equations:

$$\begin{bmatrix} g_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} z_t' & 0 \\ 0 & z_t' \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \nu_{1,t} \\ \nu_{2,t} \end{bmatrix}, \tag{2.4}$$

and $[e_t, \nu_{1,t}, \nu_{2,t}]' \sim N(0, \Sigma)$. For clarity of discussion, we rewrite equations (2.3) and (2.4) in the following compact form:

$$i_t^* = X_t' \beta + e_t$$

$$X_{nt} = Z_t' \Gamma + \nu_t$$

and $[e_t, \nu_t]' \sim N(\Sigma)$ and $i_t^* = i_t - \tilde{i}_t$. $X_t = [X_{nt}, X_{xt}]$ is a $(k_1 + k_2) \times 1$ vector of the all regressors of equation (2.3), and X_{nt} and X_{xt} are endogenous and exogenous regressors, respectively.

2.2.2 Forms of nonlinearity

In this subsection, we introduce several versions of the nonlinear Taylor rule and cast them into our framework. It has been shown that under assumptions of quadratic loss function and linear Phillips Curve, the optimal monetary policy rule is linear. However, various versions of the Taylor rule can be derived even with a slight adjustment of one of these two assumptions. Many empirical studies have found evidence supporting one specific version of the Taylor rule[2], but summarizing these results does not help us understand which one can best describe realized federal funds rates over past decades. One of our contribution is that we estimate several versions of the nonlinear Taylor rule within a unified framework which allows us to compare these rules directly.

A nonlinear Taylor rule assumes that response coefficients are switching between two regimes, governed by a state variable S_t, and equations (2.3) and (2.4) becomes:

$$i_t^* = X_t'\beta(S_t) + e_t \tag{2.5}$$

$$X_{nt} = Z_t'\Gamma(S_t) + \nu_t \tag{2.6}$$

and $[e_t, \nu_t]' \sim N(0, \Sigma(S_t))$. $S_t = 0, 1$ and for any generic parameter, $\theta(S_t) = (1 - S_t)\theta_0 + S_t\theta_1$. If we assume that S_t is depending on a set of macroeconomic variables, equations (2.5) and (2.6) become threshold-switching processes. The forms of nonlinearity we considered are listed as follows:

(a). $S_t = 0$ if $y_t \geq 0$.

(b). $S_t = 0$ if $\pi_t \leq \pi_t^h$.

(c). $S_t = 0$ if $y_t \geq 0$ or $\pi_t \leq \pi_t^h$.

(d). $S_t = 0$ if $|\pi_t - \pi_t^I| \leq \delta$

The theoretical foundation of the threshold-type Taylor rule is based on the assumption of the policy maker's loss function in optimal monetary policy analysis. Gertler et al. [1999] shows that under the quadratic loss function, $L_{q,j} = (\pi - \pi^*)^2 + \kappa_j y^2$, the optimal monetary policy rule is linear. Given assumptions of the rest of the economy unchanged, a

[2]Osborn et al. [2005] make a attempt to test nonlinearity within several versions of the Taylor rule and find that the inclusion of the interaction between inflation deviations and the output gap appears to characterize the nonlinear policy rule adequately.

nonlinear monetary policy rule can be derived by assuming alternative loss function. Model (a) corresponds to Bec et al. [2002]'s business-cycle dependent monetary policy rule, which is derived based on a asymmetric loss function as $L_B = I(y_t > 0)L_{q,1} + I(y_t \leq 0)L_{q,2}$. This loss function implies that the monetary authority may adopt a more (less) aggressive reaction to a deviation of inflation and output from their targeted levels, depending on the state of the economy.[3] Similar idea applies to model (b) and (c), where π_t^h is inherited inflation and depends on the history of past inflation. Model (d) represents Orphanides and Wilcox [2002]'s opportunistic approach to disinflation, under which a central bank controls inflation aggressively when inflation is far from its target, but concentrates more on output stabilization when inflation is close to its targets. The underlying loss function is assumed as $L_O = (\pi - \pi^*)^2 + \kappa|y|$. According to Aksoy et al. [2006], π^I is an intermediate target which corresponds to a weighted average of the long-run target, π^*, and inherited inflation, π^h. That is, $\pi^I = (1 - \mu)\pi^* + \mu\pi^h$.

2.2.3 Extensions

Models represented by equations (2.5), (2.6) and S_t introduced above might be too restrictive because they assume that 1) the Fed has been reacting nonlinearly throughout the entire sample periods and 2) the covariance matrix are also switching with monetary policy regime. In this subsection, we introduce some extensions of our basic nonlinear Taylor rules along these two dimensions.

Subsample-specific nonlinearity

To incorporate the possibility of period-specific nonlinearity, we consider the cases where nonlinearity emerges only in a subsample. Taking form (a) as an example, we consider the following alternatives:

- $S_t = 0$ if $y_t \geq 0$.

---

[3] They empirically find that the Fed seems to be essentially concerned with inflationary pressures during economic booms – trying to avoid any economic overheating – whereas contractions are essentially devoted to economic activity stabilization for the post-Volcker periods.

- $S_t = 0$ if $y_t \geq 0$ and $I(t \leq Tb)$.

- $S_t = 0$ if $y_t \geq 0$ and $I(t > Tb)$.

where $I(\bullet)$ is a indicator function that is equal to one if the enclosed statement is true, and Tb is a break date.

The dating of the breakpoint is treated as known and certain, and we specify the known breakpoint at 1982:4 or 1990:1 in this paper. The first break date 1982:4 is a "popular" option in the literature since it has been believed that U.S. monetary underwent a structural change after Paul Volcker's disinflation policy. For example, Osborn et al. [2005], using sample excluding the years of monetary targeting, find that while there is significant evidence of nonlinearity for the period to 1979, there is little such evidence for the subsequent period. Although it is a sensible assumption, by splitting sample into pre- and post- Volcker periods, one might miss another form of nonlinearity studied by Basistha and Startz [2004]. They model the desired Federal Funds rate in a two-regime setting, one when the Fed makes no change and the other when the Fed is moving the desired rate to a new level, and show that the probability of being in the static-desired-rate regime increased significantly in the 1990s. In fact, the second breakpoint, 1990:1, is the break date used by Basistha and Startz [2004].

The Great Moderation

One of the most prominent features of U.S. economy for the past four decades is the Great Moderation, a substantial reduction in the volatility of business cycle fluctuations starting in the mid-1980s. Even though the Great Moderation does not affect consistency of OLS-based estimators in the context of linear regression, it might have enormous effects on the model selection and comparison.[4] Since all models introduced so far assume that the covariance matrix of error terms is regime-switching, variations in volatilities caused by the Great Moderation are not incorporated. To avoid misleading model comparison, we

[4] According to Sims and Zha [2006], failure to take account for the heteroscedasticity can strongly bias statistical evidence in favor of significant shifts in response coefficients describing monetary policy.

also estimate another version which takes into account the Great Moderation for all models under consideration.

The models with the Great Moderation can be written as follows:

$$i_t^* = X_t'\beta + e_t, \tag{2.7}$$

$$X_{nt} = Z_t'\Gamma + \nu_t, \tag{2.8}$$

$$[e_t, \nu_t]' \sim N(0, (1 - d_t)\lambda\Sigma + d_t\Sigma), \tag{2.9}$$

where $d_t = 1$ if $t > 1984Q4$ and $\lambda > 1$. Note that λ scales up or down the entire covariance matrix Σ, while maintaining the correlation between errors unchanged. The greatest advantage of this specification is that it substantially reduces the number of parameters. However, this benefit comes at a cost of reduced flexibility – we cannot capture idiosyncratic changes of each element in the covariance matrix.

2.3 Bayesian Inference and Model Selection Procedure

Our primary goal is to compare some non-nested models represented by different specifications of S_t. Within a frequentist framework, it can be done by establishing several hypotheses in which one of the models are assumed to be true, and one can test those hypotheses (see Pesaran and Weeks [2007] for extensive review). Under the context of linear regression models, test statistics based on i.i.d. errors has been developed by, for instance, Davidson and MacKinnon [1981] and Ericsson [1983]. Recently, test non-nested regression models with exogenous regressors using heteroscedasticity/autocorrelation-consistent (HAC) estimator has been proposed (see Choi and Kiefer [2008] and Godfrey [2011]). Even though their results are not ready to be extended to regression models with endogenous regressors, testing non-nested IV regressions using Generalized Method of Moment (GMM) is possible (Hall and Pelletier [2011]).

We instead cast the estimation within a Bayesian framework in order to take advantage of important features this framework provides. First, unlike the classical likelihood-ratio test, non-nested models do not pose any special problem within Bayesian model comparison. The main issue is to compute the marginal likelihood of each model, which does not have closed form in most of cases. Fortunately, simulation-based approaches can be used to accurately approximate the marginal likelihood. Second, Bayesian approach deal with model uncertainty in a coherent way. According to Davidson and MacKinnon [2004], testing each of two non-nested models against the other may or may not allow us to choose one model over the other. When a pairwise classical test does not reject either model, data appear to be compatible with both models. Within a Bayesian framework, we can examine the extent to which a model is preferred by the data for all models, and marginal likelihoods serve as the weights for model averaging if multiple models are equally preferred.

Third, classical test for nonlinearity within each model is an nonstandard problem. Some nuisance parameters, for example, the variance-covariance matrix in one of the regime, are not identified under the null hypothesis of no nonlinearity, so the asymptotic distributions of the usual test statistics are nonstandard. However, the nuisance parameters that exist under the alternative but not under the null do not pose any special problem within a Bayesian framework. As pointed out by Koop and Potter [1998], classical solutions also must integrate out the nuisance parameters, but the integration is with respect to an arbitrary distribution. In the Bayesian approach, there is a sense in which information in the data is used to integrate out nuisance parameters.

In this section, we introduce a Gibbs sampler for a Bayesian Instrumental Variable (BIV) estimator with normal errors and priors. The algorithm is developed by Cogley and Startz [2012][5], who observe while the seemingly unrelated regression formulation of instrumental variables is nonlinear in its parameters, Gibbs sampling is possible if parameters are blocked in the right way. Here we describe a Gibbs sampler for a model with multiple instruments and endogenous right-hand variables. Note that the state variable, S_t, is not endogenously determined by our model, so it does not introduce any difficulty other than doubling up the number of parameters. For the clarity of discussion, we suppress the notation S_t in the following derivations without loss of generality.

[5]The Matlab code can be obtained at http://startz.weebly.com/.

2.3.1 A Gibbs Sampler

Define z_t is a vector of valid instrumental variables, then the IV regression are

$$i_t^* = X_t'\beta + e_t \tag{2.10}$$

$$X_t = (I_k \otimes z_t)\gamma + \nu_t, \tag{2.11}$$

$$\begin{bmatrix} e_t \\ \nu_t \end{bmatrix} \sim N(0, \Sigma).$$

X_t is a k vector of endogenous regressors, γ is a $kq \times 1$ vector and q is the number of instrumental variables[6].

We estimate the model using the control function approach, see Kim [2009] for extensive review. The main idea behind the control function approach is to model the dependence of disturbance terms in a way that allows us to construct a function such that, conditional on the function, the endogeneity problem in the regression equation of interest disappears. Under the normality assumption, it is easy to construct the control function to eliminate endogeneity by the fact that $e_t|\nu_t \sim N(c_t, H)$, where c_t and H are conditional mean and conditional variance computed as follows:

$$c_t = \Sigma_{e\nu}\Sigma_{\nu\nu}^{-1}\nu_t$$

$$H = \sigma_e^2 - \Sigma_{e\nu}\Sigma_{\nu\nu}^{-1}\Sigma_{\nu e},$$

$$\Sigma = \begin{bmatrix} \sigma_e^2 & \Sigma_{e\nu} \\ \Sigma_{\nu e} & \Sigma_{\nu\nu} \end{bmatrix},$$

Note that c_t is a function of disturbances of IV equations so it is not a fixed number. The control function is $e_t = c_t + chol(H)\epsilon_t$, where ϵ_t is an i.i.d. disturbance following N(0,1). Substitute the control function into equation (2.10), then we obtain a regression equation with i.i.d. disturbance.

Gibbs sampling is accomplished in three blocks: $\Sigma|\gamma, \beta$, $\beta|\gamma, \Sigma$, and $\gamma|\beta, \Sigma$. We assume that parameters are independent a priori across blocks,

$$p(\beta, \Gamma, \Sigma) = p(\beta)p(\gamma)p(\Sigma). \tag{2.12}$$

[6] Without loss of generality, in this subsection we assume all regressors are endogenous. Incorporating exogenous regressors does not cause any problem in what follows.

Marginal priors are specified so that the conditional posterior for each block has a convenient form. The data are denoted $D = (y, X, Z)$. The Gibbs sampler involves the following steps:

Choose a starting values, for $s = 1, \ldots, S$:

(1) Take a random draw, $\beta^{(s)}$ from $\beta | D, \gamma^{(s-1)}, \Sigma^{(s-1)}$.

(2) Take a random draw, $\gamma^{(s)}$ from $\gamma | D, \beta^{(s)}, \Sigma^{(s-1)}$.

(3) Take a random draw, $\Sigma^{(s)}$ from $\Sigma | D, \beta^{(s)}, \gamma^{(s)}$.

We leave details of the posterior simulation in the appendix.

2.3.2 Models with the Great Moderation

Incorporating the Great Moderation can be dealt with by adding one more block to the Gibbs sampler. Conditional on λ, we can divide both sides of equations (2.7) and (2.8) by $(1 - d_t)\sqrt{\lambda}$, and draw posterior samples of β, Γ, and Σ using the Gibbs sampler introduced in previous subsection. Therefore, we only need the conditional distribution of $\lambda | \beta, \Gamma, \Sigma, D$ to complete the Gibbs sampler.

We assume independent inverse gamma priors for λ. Let $u_t = [e_t, \nu_t']'$. Conditional on β and Γ, u_t are observable, and we can transform these residuals according to

$$chol(\Sigma)^{-1} u_t \equiv \tilde{u}_t \sim N(0, (1 - d)\lambda I_{k+1} + dI_{k+1}), \tag{2.13}$$

where *chol* is the Cholesky decomposition of a positive definite matrix. Thus the likelihood function of λ depends only on the values of u_t for which $d_t = 0$. Let $\tilde{U}$ denote a $Td \times 1$ vector containing all u_t for which $d_t = 0$, and the likelihood of $\tilde{U}$ is

$$f(\tilde{U}|\lambda) = C \left(\frac{1}{\lambda}\right)^{Td/2} (\tilde{U}'\tilde{U})^{-0.5} exp\left(\frac{-\tilde{U}'\tilde{U}}{2\lambda}\right),$$

where C is a constant. It follows immediately that the conditional posterior of λ_t is characterized by

$$\lambda | \beta, \Gamma, \Sigma, D \sim IG\left(\frac{Td}{2} + a0, \quad \frac{\tilde{U}'\tilde{U}}{2} + b0\right) \tag{2.14}$$

where a_0 and b_0 are hyperparameters of the prior distribution.

2.3.3 Bayesian Model Comparison

Bayesian model comparison is based on the Bayes Factor defined as follows

$$B_{i,j} = \frac{Pr(D|M_i)}{Pr(D|M_j)}.$$

where D represents data, and $Pr(D|M_i)$ is the marginal likelihood for Model i. According to Kass and Raftery [1995], Bayes factor "is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another". In most of cases, however, the exact form for $Pr(D|M_i)$ is not feasible, and recently literature of Bayesian model comparison relies heavily on simulation approximation. In this paper, we adopt Chibs [1995]'s method to compute the marginal likelihood and give the computational details in appendix.

Given Bayes factors comparing two models, decisions can be made according to a convenient criteria, provided by Kass and Raftery [1995], as follows:

$2 \ln B_{i,j}$	Evidence against M_j
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
>10	Very strong.

It is more intuitive to interpret these cutoff points by how much more likely the data support model A versus model B. If two times the Bayes Factor comparing model A and B equals 2, 6, or 10, it implies that model A is about 2.7, 21, or 22000 times more likely than model B to be the correct model.

2.4 Empirical Results

2.4.1 Data

The data set we used in the empirical study of the nonlinear Taylor rule consists of inflation, the output gap, and the effective federal funds rates form 1967Q1 to 2010Q1. Inflation is measured by annualized quarterly percentage change of GDP deflator. In order to account for the fact that real GDP is often subject to substantial revisions, we follow Croushore and Stark [2001] and use real time value of real GDP available at the Federal Reserve Bank of Philadelphia. The output gap is measure as the percentage deviation of the real time real GDP form its Hodrick-Prescott (HP) trend.[7] Instrumental variables include 4 lags of the following variables: inflation, the real time output gaps, and the effective federal funds rates.

It has been recognized that there were periods when the Fed focused more on monetary-aggregate-targeting rather than interest-rate-targeting policy.[8]. Using interest-rate-targeting reaction function, e.g., Taylor rule, to describe U.S. monetary reaction during these periods might not be appropriate for two reasons: 1) The comovements between inflation and the nominal interest rate do not reflect intentional monetary policy reactions. 2) Extremely volatile nominal interest rates might distort estimates of coefficients of a reactions function. Therefore, we exclude the sample from 1979Q4 to 1982Q4 in this section.

We let $\bar{r}_t$ be equal to the estimates form the HP filters of ex post real rate, defined as the difference between the funds rate and the realized inflation rate[9]. We approximate long-run inflation target as the Beveridge-Nelson trend component of inflation from a reduced-form time-varying parameter VAR (TVPVAR) described briefly as follows

$$Z_t = \mu_t + F_t Z_{t-1} + E_t, \quad \epsilon_t \sim N(0, \Sigma_t) \tag{2.15}$$

where $Z_t = [y'_t, \dots, y'_{t-p+1}]'$, and y_t is a $m \times 1$ vector containing inflation. Following Cogley and Sbordone [2008]'s method, we define trend inflation as the level to which inflation is expected to settle after short-run fluctuations die out[10].

---

[7]Bunzel and Enders [2010] and Martin and Milas [2010] both use the same method. Bunzel and Enders [2010] provide step-by-step details of constructing the real time output gap.

[8]For example, Sims and Zha [2006] find a monetary regime corresponds to the Volcker reserve targeting period, roughly 1979Q4 to 1982Q4, and shows clearly the targeting of monetary aggregates, rather than interest rates.

[9]The estimates of $\bar{r}_t$ are obtained according to the results from Laubach and Williams [2003], who find that the two-sided Kalman filter estimates and the two-sided estimates from the Hodrick-Prescott filters with the smoothing parameter of 6400 are very similar.

[10]It can be approximated this by calculating a local-to date t estimate of mean inflation from the VAR,

$$\pi^*_t = e'_\pi (I - F_t)^{-1} \mu_t.$$

2.4.2 Priors

The parameters are assumed to be independent across blocks, so that the joint prior can be expressed as the product of marginal priors,

$$f(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \Sigma) = f(\boldsymbol{\beta}) f(\boldsymbol{\Gamma}) f(\Sigma)$$

The prior for $\boldsymbol{\beta}$ is independent normal distribution. We specify a 'no-difference' prior in that prior means of coefficients of the monetary policy rules are identical in two regimes, i.e., $\beta_{S_t=i,j} \sim N(0.2, 1)$ for $i = 0, 1$ and $j = \pi, y$. The smoothness parameters, $\rho_{S_t=i,j}$, is assumed to follow a normal distribution: $\rho_{S_t=i,1} + \ldots + \rho_{S_t=i,p} \sim N(0.7, 1)$ and $\rho_{S_t=i,p} \sim N(0, 1)$ for $p > 1$. The prior distribution of $\boldsymbol{\Gamma}$, coefficients of the IV regressions, are normal, and the prior mean of $\boldsymbol{\Gamma}$ are all zero except for the elements corresponding to the first own lag of the dependent variable in each equation. These elements are set to be 0.7. Prior variance are set to be 1. Our prior for Σ is Inverse-Wishart,

$$f(\Sigma) = IW(0.1 I_{k+1}, 5).$$

Finally, we assume[11]

$$f(\lambda) = IG(2, 1).$$

The Gibbs sampler runs 15,000 times for all models and the first 5,000 draws are discarded to eliminate the effects of initial values. This procedure gives us 10,000 posterior samples, upon which the following analyses are based.

2.4.3 Were there nonlinearity in U.S. monetary policy?

Table (2.1) shows the marginal likelihoods of all models under consideration. In additional to nonlinear Taylor rules, we also report the marginal likelihood of the linear Taylor rule.

The first finding is that nonlinearity in the Taylor rule is supported by the U.S. data. In the baseline cases, Many alternative nonlinear Taylor rules have greater marginal likelihoods than the linear Taylor rule does. The evidence for nonlinearity weakens in the cases

---

where e_π is a selector vector. Following Cogley and Sargent [2005], we work with VAR(2) representation for nominal interest, quarterly inflation, and logit of unemployment. We thank Tim Cogley for generously providing Matlab codes for replication.

[11] The inverse gamma density $IG(a, b)$ is $f_{IG}(z; a, b) = \frac{b^a}{\Gamma(a)} z^{-(a+1)} exp(-b/z)$.

Table 2.1: Marginal Likelihoods

• **Baseline.**

		Tb=1983		Tb=1990	
	All	**Post-Tb**	**Pre-Tb**	**Post-Tb**	**Pre-Tb**
(a)	-649.3519	-635.4586	-662.0831	**-616.8803**	-667.8296
(b)	-664.6001	-657.3281	-675.2080	-653.9156	-671.0012
(c)	-651.0741	-628.9552	-666.7968	-627.2781	-666.8096
(d)	-633.4804	-622.9028	-644.5577	-638.3988	-674.4710

No Switching: -638.3142

• **Incorporating the Great Moderation.**

		Tb=1983		Tb=1990	
	All	**Post-Tb**	**Pre-Tb**	**Post-Tb**	**Pre-Tb**
(a)	-654.8705	-650.4744	-649.8828	-625.1804	-665.1430
(b)	-666.7784	-663.3320	-648.4467	-660.5057	-648.9072
(c)	-649.4301	-650.6145	-652.4276	-642.9967	-657.7126
(d)	-654.6445	-645.4849	-641.6545	-651.2638	-649.4522

No Switching: -621.2624

* (a) $S_t = 1$ if $y_t < 0$. (b) $S_t = 1$ if $\pi_t > \pi_t^h$. (c)$S_t = 0$ if $y_t \geq 0$ or $\pi_t \leq \pi_t^h$. (d) $S_t = 1$ if $|\pi_t - \pi_t^I| > \delta$. We use $\pi_t^h = \sum_{j=1}^4 \pi_{t-j}$ as in Martin and Milas [2010]. Log marginal likelihoods are approximated using Chibs [1995]'s method and evaluated at posterior medians of the parameters. "Post-Tb" refers to the model in which nonlinearity only emerges in periods after Tb. μ and δ in (d) are treated as hyperparameters and set to be equal to 0.6 and 0.5 respectively, based on the estimates of Martin and Milas [2010].

incorporating the Great Moderation. However, nothing prevents us from comparing models across cases, so all models listed in table (2.1) can be directly compared. We thus conclude that U.S. data strongly support the nonlinear Taylor rule.

The second noteworthy pattern found in table (2.1) is that nonlinearity might not be an all-time phenomenon. For all forms of nonlinearity, there is at least one model with subsample-specific nonlinearity can beat strongly the model assuming nonlinearity in full sample. Furthermore, how the sample is divided could affect the model selection. By dividing sample into post- and pre-Volcker periods, the model which implied by the opportunistic approach of disinflation would be selected. However, by dividing sample into pre- and post-Greenspan sample[12], we will select the model implied by the asymmetric loss function.

The best-fit model selected by marginal likelihoods suggests that the Fed's behaviors change when the output gap turns to positive, and this regime-switching behavior emerged in the beginning of 1990. Table (2.2) reports estimates of parameters of the best-fit model. Several differences between the two regimes stand out. First, the contemporary responses coefficients of the inflation gap and output gap are different across regimes. Generally speaking, the Fed's reactions are more accommodative in the comfort zone. Second, the federal funds rate is much more persistent in the comfort zone than it is otherwise, meaning that the Fed will adjust the federal funds rate in a more rapid way when it perceives a concern of recession. Third, volatilities of the error terms are much smaller in the comfort zone. Note that we cannot interpret Ω_i as volatility of the monetary policy shocks because e_t is equation (2.3) is a composition of the monetary policy shock and expectation errors. Therefore, the volatility of the monetary policy shocks should be even smaller than 0.06.

To sum up, estimates reported in table (2.2) imply that the monetary policy reactions are regime-switching and depending on whether the economy is producing below potential output. After 1990, the Fed seems to be more accommodative to inflation gap and output gap and adjusts the federal fund rate in a smooth and relatively predictable way when economy is booming.

[12]Allen Greenspan took office in August 1987, which is close to our break date 1990Q1.

Table 2.2: "Best-fit" nonlinear Taylor rule

	$y_{t-1} \geq 0$ and $t \geq 1990Q1$			Otherwise		
	Mean	SD	90% Bands	Mean	SD	90% Bands
β_π	0.15	0.24	(-0.26, 0.50)	0.38	0.12	(0.19, 0.59)
β_y	0.09	0.05	(0.01, 0.19)	0.11	0.05	(0.03, 0.19)
$\rho(1)$	0.98	0.03	(0.94, 1.04)	0.75	0.07	(0.62, 0.86)
ρ_2	-0.49	0.19	(-0.80, -0.18)	-0.26	0.09	(-0.41, -0.11)
Ω_i	0.07	0.04	(0.04, 0.15)	0.54	0.11	(0.40, 0.74)
Ω_π	0.38	0.09	(0.25, 0.54)	1.37	0.20	(1.08, 1.72)
Ω_y	0.12	0.03	(0.08, 0.18)	0.67	0.10	(0.53, 0.84)
$\Omega_{i,\pi}$	-0.05	0.11	(-0.20, 0.14)	-0.30	0.19	(-0.63, -4e^{-3})
$\Omega_{i,y}$	-0.01	0.02	(-0.05, 0.02)	0.03	0.07	(-0.09, 0.15)
$\Omega_{\pi,y}$	-0.03	0.04	(-0.09, 0.03)	0.04	0.09	(-0.11, 0.20)

$\rho(1)$ refers to the sum of all AR coefficients, SD to standard deviation, and 90% Bands to 90% posterior probability bands.

2.4.4 Describing U.S. Monetary Policy Using Nonlinear Taylor

The marginal likelihood reflects model "fitness" of the entire SUR system, not just of the monetary policy rule. Therefore, it is possible that the best-fit model chosen by the marginal likelihood might not be able to describe U.S. monetary policy reactions better than linear Taylor rule is. In this subsection, we examine the ability of the best-fit nonlinear Taylor rule to describe the monetary policy reactions over the past two decades[13].

We set up a small macro model based on a reduced form VAR, in conjunction with an explicit description of the process, the Taylor rule, generating short-term nominal rates. The model consists of two parts: 1) monetary policy and 2) inflation and real activity. We assume the monetary policy follows the forward-looking Taylor rules

$$i_t = (1 - \rho(1))[\tilde{i} + \alpha_\pi E_t(\pi_{t+1} - \pi^*) + \alpha_y E_t(y_{t+1})] + \rho(L) i_{t-1} + m_t.$$

Coefficients of the monetary policy rule are set to be the posterior mean of the corresponding model.. The inflation gap and the output gap are determined by simple reduced form equations from a VAR with four lags each of the inflation gap, the output gap, and the federal funds rate. That is,

$$\begin{bmatrix} \pi_t - \pi^* \\ y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} \pi_t - \pi^* \\ y_t \\ i_t - \tilde{i} \end{bmatrix} + e_t. \tag{2.16}$$

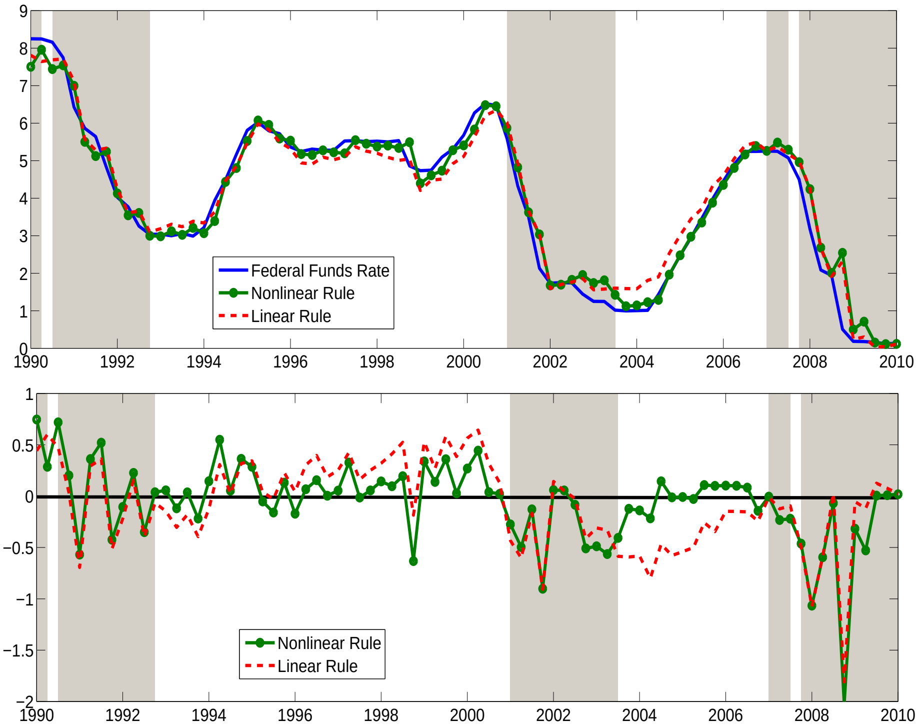
Throughout the rest of this subsection, we set $\pi^* = 2.5\%$ and $\bar{r} = 2\%$, so that $\tilde{i} = 4.5\%$. And we obtain VAR coefficients using OLS equation-by-equation. For nonlinear Taylor rule, we estimate the VAR for each regime because coefficients of the reduced form VAR are functions of coefficients describing the monetary policy.

Given parameter values, this simple macro model can be easily solved for the endogenous

---

[13]We focus on the past two decades since our best-fit model indicates that regime-switching behaviors emerged after 1990

Figure 2.2: Linear and Nonlinear Taylor rule



variables[14]. The solution takes the following form:

$$\begin{bmatrix} i_t - \tilde{i} \\ \pi_t - \pi^* \\ y_t \end{bmatrix} = \Psi(L) \begin{bmatrix} i_{t-1} - \tilde{i} \\ \pi_{t-1} - \pi^* \\ y_{t-1} \end{bmatrix} + e_t. \tag{2.17}$$

As pointed out by Favero [2006], the interpretation of the reduced form could be different but different interpretation does not affect the resulting forecast short-term rates. For the purpose of forecasting, we rewrite the solution in state space form as $X_t = FX_{t-1} + E_t$, where $X_t = [i_t - \tilde{i}, \pi_t - \pi^*, y_t, \ldots, i_{t-3} - \tilde{i}, \pi_{t-3} - \pi^*, y_{t-3}]'$. Therefore, the 1-step ahead forecast of policy rate at time t is the first element of FX_{t-1} plus $\tilde{i}$.

We report actual and fitted federal funds rates using the linear and the nonlinear Taylor rules in the upper panel of figure 2.2. The lower panel show the difference between actual

[14]We use Uhlig's toolkit which are available at http://www2.wiwi.hu-berlin.de/institute/wpol/html/toolkit.htm

and fitted federal funds rates. Gray bars indicates times when the output gap was negative. The nonlinear Taylor rule captures adjustments of the federal funds rate relatively closer than the linear Taylor rule does, especially in periods where the economy was producing above the potential. The most significant difference was in the period between 2003 and 2007. The linear Taylor rule consistently overstates adjustments of the federal funds rate, while the nonlinear rule fit rather well. In fact, as put by Bernanke [2006], the tightening cycle began at the end of June 2004 is notable in that its onset was delayed for longer than many observer expected and that policy moved gradually. The fitted federal funds rates using nonlinear Taylor rules are indeed consistent with Bernanke's observations.

2.5 Conclusion

Monetary policy reactions may not be symmetric as described by a simple Taylor. We confirm this possibility by means of the Bayesian model comparison. Several theoretical founded nonlinear Taylor rule are estimated and compared within a Bayesian framework, and endogeneity problem is dealt with using a Bayesian Instrumental Variable estimator. Empirical results indicate that monetary policy reactions are regime-switching after 1990 and depending on whether the economy is producing below potential output. On the other hand, the Fed is more accommodative to inflation gap and output gap and adjusts the federal fund rate in a smooth and relatively predictable way when economy is booming. We show that the nonlinear Taylor can describe the behaviors of the federal funds rate better than the linear Taylor rule.

Chapter 3

AN EFFICIENT BAYESIAN INFERENCE FOR THE STATE SPACE MODELS WITH THE STOCHASTIC VOLATILITY

This paper is concerned with the Bayesian inference of a class of linear state space models whose disturbances follow stochastic volatility processes. The non-Gaussian disturbances in the model are approximated using a mixture of normals, as in Kim et al. [1998]. The conventional auxiliary mixture sampler treats the mixture indicator variable as a latent variable and simulates the model using the data augmentation method. Alternatively, we suggest a Metropolis-Hasting-within-Gibbs (MHGibbs) sampler, which is implemented without conditioning on the mixture indicator variable. Using simulated data, we show that the MHGibbs sampler significantly outperforms the conventional auxiliary mixture sampler by a factor from 4 to 11. We revisit the unobserved component model with the stochastic volatility used by Stock and Watson [2007] as an empirical application.

Keywords. Bayesian Inference, Metropolis-Hasting, Gibbs sampler, State Space Model, Stochastic Volatility, Auxiliary Mixture Sampler.

3.1 Introduction

The state space model has become a workhorse for statistical analysis of time series in macroeconomics and finance. The traditional estimation method of the linear Gaussian state space model relies on the prediction error decomposition of the likelihood function, which can be effectively evaluated using the Kalman filter. Maximizing the likelihood function with respect to the model parameters leads to the the maximum likelihood estimator (MLE). However, implementing the MLE for the linear state space models whose disturbances follow stochastic volatility processes (SS-SV model hereinafter) is difficult in that some of the latent variables enter nonlinearly and that there is no analytical form for the likelihood function.[1]

[1]Alternative frequentist approaches, e.g., Koopman and Bos [2004], and simulation-based filtering techniques, e.g., Creal [2012], have been proposed to deal with this difficulty.

In this paper, we focus on the Bayesian Markov-Chain Monte Marlo (MCMC) approaches and suggest a simple algorithm which requires neither a new filtering technique nor a novel sampler, while improving the efficiency to a great extent.

The Bayesian MCMC methods allow computational problems to be divided into a number of tractable subproblems which are solved iteratively. The SS-SV model thus can be dealt with by dividing it into the SS-block and the SV-block. Conditional on the knowledge of the SV-block, the SS-SV model becomes a linear Gaussian state space model, and posterior samples can be drawn using the simulation smoothers as surveyed in Giordani et al. [2011]. As for the SV-block, Kim et al. [1998] show that the similar simulation smoother can also be applied to the stochastic volatility process by approximating non-Gaussian errors using a mixture of normals. This approach requires to update an auxiliary indicator variable, indicating which component of the mixture of normals the error belongs to, in each iteration of the MCMC algorithm, and thus is called the auxiliary mixture sampler (AMS hereinafter). The AMS has been applied successfully in many empirical studies[2].

Caution should be exercised, however, when using the AMS to estimate a SS-SV model. Our simulation study indicate that the generic AMS can fail to capture the true values of the model parameters and the latent variables in a simple local level model with a stochastic volatility process. We propose two alternative AMS: (1) the iterated AMS which runs the SV-block for multiple times in each iteration of the AMS, and (2) the disturbance-based AMS which reformulates the sampling scheme in terms of the disturbances of the stochastic volatility process as in Bos and Shephard [2006]. Simulation study shows that both methods generate MCMC draws which converge to the same posterior distributions covering the true values of the model parameters.

The main contribution of this paper to is to add to the literature by suggesting a Metropolis-Hasting-within-Gibbs sampler (MHGibbs hereinafter) which does not need to draw the auxiliary indicator variable. As pointed out by Carter and Kohn [1996], there might be a strong dependence between the indicator variable and the latent variable in some models which may cause inefficient MCMC draws. Therefore, the efficiency of the

[2] See Koop and Korobilis [2010] for an excellent introduction to the AMS for the SS-SV model and several applications in empirical macroeconomics.

AMS could be enhanced by marginalizing the mixture indicator variable. We note that conditioning on the other part of the model, the SS-block forms a linear Gaussian state space model, while the SV-block forms a *switching* linear Gaussian state space model. The likelihood functions in both blocks can be thus evaluated using the prediction error decomposition and conventional filtering techniques, i.e., the Kalman filter and the Kim's filter (Kim [1994])[3]. Accordingly, a Metropolis-Hasting algorithm can be used to draw the parameters for either SS-block or SV-block, and we can obtain a posterior draw of the latent variables using the simulation smoothers.

The MHGibbs sampler is implemented without drawing the auxiliary indicator variable and approximates the non-Gaussian errors *within* the filtering procedure. This feature has the potential to lead to fast exploration of the parameter space and to improve the efficiency of the MCMC algorithm. The simulation study shows that MCMC draws generated by the MHGibbs sampler also cover to the true value without the need of multiple runs in each iteration. Furthermore, the inefficiency analysis indicates that the MHGibbs sampler outperforms the two versions of the AMS by a factor from 4 to 11. We apply the MHGibbs sampler to fully estimate the unobserved component model with stochastic volatility and find similar results as in Stock and Watson [2007] except for the greater variability in the transitory component of U.S. inflation.

The outline of the remainder of the paper is as follows. Section 2 review the conventional Gibbs sampler and introduce the novel MHGibbs sampler for a generic SS-SV model. In section 3, we compare the performance of the Gibbs sampler and the MHGibbs sampler using simulated data. We apply the MHGibbs sampler to a unobserved component model of U.S. inflation in section 4. Section 5 concludes.

---

[3] Kim's filter is an extension of the Kalman filter dealing with a class of switching linear state space models. See Kim and Nelson [1999] for the introduction and examples.

3.2 The State Space Model with Stochastic Volatility

3.2.1 General Specification

The Gaussian SS-SV model can be formulated as follows:

$$y_t = c_t + H\beta_t + Re_t \quad \text{(Measurement equation)} \tag{3.1}$$

$$\beta_t = d_t + F\beta_{t-1} + Wu_t \quad \text{(Transition equation)}, \tag{3.2}$$

where y_t is a $p \times 1$ vector of observations, β_t is a $q \times 1$ vector of latent variables. In what follows we assume that $q = p = 1$ without loss of generality. Assume $e_t \sim N(0, \sigma^2_{e,t})$ and $u_t \sim N(0, \sigma^2_{u,t})$, c_t, H, R, d_t, F, W are possibly time-varying conformable matrices containing unknown parameters. Note that the volatilities of the errors are time-varying in equations (3.1) and (3.2). Following the stochastic volatility literature, we assume that the variances of the errors, $\sigma^2_{e,t}$ and $\sigma^2_{u,t}$, follow two latent stochastic processes. To ensure that these processes always generate non-negative values for variances, we assume

$$e_t = \exp(h_{e,t}/2)e^*_t, \tag{3.3}$$

$$u_t = \exp(h_{u,t}/2)u^*_t, \quad \text{and} \tag{3.4}$$

$$h_{j,t} = h_{j,t-1} + \eta_{j,t} \quad \text{for} \quad j = e, u, \tag{3.5}$$

where e^*_t and u^*_t are both i.i.d. N(0,1) and $\eta_{j,t} \sim N(0, \sigma^2_{\eta,j})$. Note that $h_{j,t}$ can also be assumed to follow a stationary autoregressive process at the cost of extra parameters. In this paper, we focus on the random walk process used extensively in macroeconomic studies, e.g., Stock and Watson [2007]. As pointed out by Bos and Shephard [2006], when the aim is solely to smooth the data, rather than predict future values, it often makes sense to use a random walk process.

Many empirical macroeconomic models are nested in the SS-SV model represented by equations (3.1) to (3.5). For example, Cogley and Sargent [2005] first incorporate the stochastic volatility into a Time-varying vector autoregressive model. Stock and Watson [2007] introduce the stochastic volatility into a trend-cycle decomposition of U.S. inflation. Del Negro and Schorfheide [2011] summarize some applications of the stochastic volatility in DSGE models. However, estimating SS-SV model is a challenging task because the likelihood function cannot be directly evaluated through the Kalman filter. Therefore, the Maximum Likelihood Estimation (MLE) will not be feasible unless some simulation-based techniques, e.g., particle filter, are used to approximate the likelihood. Among many estimation methods proposed in the literature, Bayesian MCMC methods have gained increasing popularity, probably because of the ease of implementation. In what follows, we first introduce the widely-used Gibbs sampler, namely, the auxiliary mixture sampler (AMS hereinafter), and then introduce the MHGibbs sampler.

3.2.2 The Auxiliary Mixture Sampler

The AMS uses the fact that the SS-SV model is a *conditionally* Gaussian state space model, introduced by Carter and Kohn [1994], and this class has a convenient blocking structure which allows us to implement MCMC techniques. Let parameters contained in c, F, H, d, R, and W be denoted by θ_{ss} and those in equations (3.3) to (3.5) by θ_{sv}. $\beta^T = [\beta_0, \beta_1, \dots, \beta_T]'$ and $h^T = [h_0, h_1, \dots, h_T]'$ are vectors of latent stochastic processes. Let D denote available data used for empirical studies. In principle, the AMS consists of the following three steps:

1. Draw random sample from $\theta_{ss}, \theta_{sv} | D, \beta^T, h^T$.

2. Draw random sample from $\beta^T | D, \theta_{ss}, \theta_{sv}, h^T$.

3. Draw random sample from $h^T | D, \theta_{ss}, \theta_{sv}, \beta^T$.

The first step is usually standard. We can rewrite the conditional density of θ_{ss}, θ_{sv} as follows:

$$
\begin{aligned}
f(\theta_{ss}, \theta_{sv} | D, \beta^T, h^T) &= f(\theta_{ss} | D, \theta_{sv}, \beta^T, h^T) f(\theta_{sv} | D, \beta^T, h^T), \\
&= f(\theta_{ss} | D, \beta^T, h^T) f(\theta_{sv} | D, h^T).
\end{aligned}
$$

Note that h^T contains all information on θ_{sv}, and that conditioning on h^T and β^T, θ_{sv} is irrelevant to θ_{ss}. Thus the second equality holds. With Normal and inverse-Gamma priors, $f(\theta_{ss} | D, \beta^T, h^T)$ and $f(\theta_{sv} | D, h^T)$ usually are well-known distributions, such as Normal and inverse-Gamma, so drawing samples from these conditional distributions is straightforward.

The second step can be done easily by the fact that conditional on h^T, $\sigma^2_{e,t}$ and $\sigma^2_{u,t}$ are known, so equations (3.1) and (3.2) form a Gaussian linear state space model. An efficient way to sample β^T is multi-move sampling, which starts by representing the joint density as the product of $T+1$ conditional densities:

$$f(\beta^T|D) = f(\beta_T|D)\prod_{t=0}^{T-1} f(\beta_t|\beta_{t+1},\ldots,\beta_T,D).$$

Note we suppress conditioning on $\theta_{ss}, \theta_{sv}, h^T$ for notional clarity. Because of the Markov property of β_t, it can be shown that

$$f(\beta_t|\beta_{t+1},\ldots,\beta_T,D) \propto f(\beta_{t+1}|\beta_t)f(\beta_t|y_t).$$

The above two equations motivate the Forward-Filtering Backward-Sampling (FFBS henceforth) sampling scheme, e.g., Carter and Kohn [1994]. See appendix or Kim and Nelson [1999] for details of the implementation of the FFBS.

The third step is challenging in that equations (3.3) to (3.5) do not form a linear state space model. Following Harvey et al. [1994], we transform these equations by:

$$\begin{aligned}(y_t - c_t - H\beta_t)^2 &= e_t^2 = \exp(h_{e,t})e_t^{*2}\\ (\beta_t - d_t - F\beta_{t-1})^2 &= u_t^2 = \exp(h_{u,t})u_t^{*2}.\end{aligned}$$

Notice that conditional on θ_{ss} and β^T, errors e_t and u_t are conditionally observable. Take log to above two equations, we obtain

$$\ln e_t^2 = h_{et} + \ln e_t^{*2} \equiv h_{et} + z_{et} \tag{3.6}$$

$$\ln u_t^2 = h_{ut} + \ln u_t^{*2} \equiv h_{ut} + z_{ut}. \tag{3.7}$$

Even though equations (3.6), (3.7), and (3.5) form a linear state space model, the errors in these two equations follow log-χ^2 distribution, meaning that the FFBS cannot be applied to draw sample from $h^T, |D, \theta_{ss}, \theta_{sv}, \beta^T$.

Kim et al. [1998] deal with this difficulty by approximating the log-χ^2 distribution by a 7-components mixture normal distribution with the representation

$$\begin{aligned} z_{j,t}|s_{j,t} = i &\sim N(m_i, \nu_i^2),\\ Pr(s_{j,t} = i) &= q_i. \end{aligned} \tag{3.8}$$

Table 3.1: Selection of the mixing distribution to be logχ^2

i	q_i	m_i	ν_i^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	0.85173	0.62699
6	0.18842	1.97278	0.98583
7	0.12047	3.46788	1.57469
8	0.05591	5.55246	2.54498
9	0.01575	8.68384	4.16591
10	0.00115	14.65000	7.33342

for $j = e, u$, where s_t is an unobserved mixture component indicator with probability q_i for $i = 1, 2, \ldots, 7$. q_i, m_i, ν_i^2 are known constants that are chosen to approximate closely the log-χ^2 distribution. Omori et al. [2007] improve the already accurate approximation using a 10-components mixture normal distribution, which is used in this paper. The weights, means and variances are given in Table 3.1.

The mixture-normal approximation allows us to rewrite equations (3.6) and (3.7) as follows:

$$\ln e_t^2 = m_{s_{e,t}} + h_{et} + \nu_{s_{e,t}} z_{e,t}^* \tag{3.9}$$

$$\ln u_t^2 = m_{s_{u,t}} + h_{ut} + \nu_{s_{u,t}} z_{u,t}^*, \tag{3.10}$$

where $z_{e,t}^*$ and $z_{e,t}^*$ are both standard normal. Equations (3.9), (3.10) and (3.5) fall into the category of the switching linear Gaussian state space model[4] in that the parameter matrices are switching among 10 states. The basic MCMC method for the switching linear Gaussian state space model is the data augmentation scheme, which updates the indicator

[4]For textbook introduction and discussion, see Kim and Nelson [1999] and Fruhwirth-Schnatter [2006].

variables, $s_{e,t}$ and $s_{u,t}$, in each iteration. Note that conditional on $s^T = [s_1, \ldots, s_T]'$, h^T can be simulated using the FFBS discussed above. Finally, conditional on all the model parameters and the latent variables, s^T also can be updated using the multi-move sampling similar to that used for drawing β^T. Specifically, draw s^T period-by-period according to the following probabilities:

$$pr(s_t = i|y_t) = \frac{f_N(\ln e_t^2 - h_{et}|s_t = i)q_i}{\sum_{k=1}^{10} f_N(\ln e_t^2 - h_{et}|s_t = k)q_k}.$$

We summarize the Gibbs sampler in algorithm 1. See appendix for details of the full conditional distributions.

Algorithm 1 The AMS for the SS-SV model

1. Draw random sample from $\theta_{ss}, \theta_{sv}|D, \beta^T, h^T$ by drawing
   (a) θ_{ss} from $\theta_{ss}|D, \beta^T, h^T$ and
   (b) θ_{sv} from $\theta_{sv}|D, h^T$.
2. Draw random sample from $\beta^T|D, \theta_{ss}, \theta_{sv}, h^T$.
3. Draw random sample from $h^T, s^T|D, \theta_{ss}, \theta_{sv}, \beta^T$ by drawing
   (a) h^T from $h^T|D, \theta_{ss}, \theta_{sv}, \beta^T, s^T$.
   (b) s^T from $s^T|D, \theta_{ss}, \theta_{sv}, \beta^T, h^T$.

3.2.3 The Disturbance-based Auxiliary Mixture Sampler

The AMS outlined in Algorithm 1 implies, by construction, high dependence between parameters and the latent variables, leading to loss of the efficiency. Bos and Shephard [2006] note that the slow mixing of the MCMC samplers is caused in part by the choice of conditioning variables in the Gibbs chain. That is, the latent variables, h^T and β^T, are very informative on the parameters governing the evolving processes, θ_{ss} and θ_{sv}, leading to slow exploration of the parameter space. To "break" the high dependence caused by inefficient blocking of the Gibbs sampler, they propose a simple reformulation of the sampling scheme in terms of the errors of the stochastic volatility process.

The stochastic volatility process was defined in terms of

$$h_t = h_{t-1} + \eta_t, \quad \eta_t = \sigma_\eta \omega_t, \tag{3.11}$$

where ω_t is a white noise. Note that there is a one-to-one relation between the h_t and the white noise disturbance ω_t. Therefore, the conditioning in the block sampler can also be done on ω_t, which by construction contains less information on the value of the parameters.

Inspired by Bos and Shephard [2006], we modify the AMS as outlined in Algorithm 2. Note that $\omega^T = [\omega_1, \ldots, \omega_T]'$ is the vectors of the white noise disturbances in the stochastic volatility processes. Because the modified AMS is based on the white noise disturbance, we called Algorithm 2 as the disturbance-based auxiliary mixture sampler (DAMS hereinafter).

Algorithm 2 Disturbance-based AMS for the SS-SV model

1. Draw random sample from $\theta_{ss}, \theta_{sv} | D, \omega^T$
2. Draw random sample from $\beta^T, | D, \theta_{ss}, \theta_{sv}, \omega^T$.
3. Draw random sample from $h^T, s^T | D, \theta_{ss}, \theta_{sv}, \beta^T$ by drawing
   (a) h^T from $h^T | D, \theta_{ss}, \theta_{sv}, \beta^T, s^T$. Reconstruct ω^T.
   (b) s^T from $s^T | D, \theta_{ss}, \theta_{sv}, \beta^T, \omega^T$.

Note that the step 1 is different from step 1 of AMS for now we are no longer conditioning on h^T. Instead, we are conditioning on the standardized disturbances for the h_t process. The consequence is that as the parameters change so do the conditional variances, $exp(h_t/2)$, and that the full conditional densities needed for the Gibbs sampler are not known in closed form. Therefore, we use the Metropolis-Hasting algorithm for step 1 of the DAMS.[5] Another

[5] Given any value for θ_{ss} and θ_{sv}, we can construct the the conditional variances, $exp(h_t/2)$, according to ω^T, and obtain the likelihood by the Kalman filter and the prediction error decomposition.

trivial difference is that once given a draw of h^T in step 3.a, we need to reconstruct ω^T by standardizing $h_t - h_{t-1}$.

3.2.4 The Metropolis-Hasting-within-Gibbs Sampler

While the AMS works well for many cases, the indicator variable, s_t, is not itself a quantity of economic interest. It just serves as a device to approximate the non-Gaussian errors as close as possible. Although the data augmentation scheme which draws samples for s_t enables the use of the Kalman filter and the FFBS, as pointed out by Carter and Kohn [1996], there might be a strong dependence between the indicator variable and the latent variable in some models which may cause inefficient MCMC draws. In this subsection, we suggest a simple MCMC method which breaks this dependence and is implemented without conditioning on the indicator variable.

The MHGibbs is motivated by the fact that drawing the auxiliary indicator variable is not essential for implementing the MCMC method. Take equation (3.9) as an example to illustrate the MHGibbs algorithm. Let $y_t^* = \ln e_t^2$, we have:

$$y_t^* = m_{s_{e,t}} + h_{et} + \nu_{s_{e,t}} z_{e,t}^* \tag{3.12}$$

$$h_{e,t} = h_{e,t-1} + \eta_{e,t}, \quad \eta_{e,t} \sim N(0, \sigma_{\eta,e}^2). \tag{3.13}$$

Conditioning on s^T, drawing samples for $h_{e,t}$ using the efficient FFBS requires a set of filtered means and variances, denoted by $E(h_{e,t}|\Psi_t, s_t)$ and $V(h_{e,t}|\Psi_t, s_t)$, where $\Psi_t = (y_1^*, \ldots, y_t^*)$ representing the information available up to time t. Because the indicator variable is a discrete-valued random variable, we can integrate s_t out and obtain the following filtered means and variances:

$$E(h_{e,t}|\Psi_t) = \sum_{i=1}^{K} E(h_{e,t}|s_t = i, \Psi_t) pr(s_t = i|\Psi_t) \equiv h_{e,t|t} \tag{3.14}$$

$$V(h_{e,t}|\Psi_t) = \sum_{i=1}^{K} \left\{ V(h_{e,t}|\Psi_t, s_t = i) + (E(h_{e,t}|\Psi_t, s_t = i) - h_{e,t|t})^2 \right\} pr(s_t = j|\Psi_t) \tag{3.15}$$

Note that we suppress the conditioning of β^T and θ_{sv} for the notational clarity. Thus drawing samples for $h_{e,t}$ using the efficient FFBS *without* conditioning on s^T is practicable if we can obtain the filter means and variances for $t = 1, \ldots, T$.

Without conditioning on s^T, however, Kalman filter is infeasible in practice because the number of components in the filtering density is increasing exponentially fast. Running the exact recursive filter requires combining all K^{t-1} (K=10 in our approximation) normal posterior density with each of the K states of s_t, running in total K^t parallel Kalman filters. This problem has been long studied in the literature of switching state space model, see Kim and Nelson [1999] and Fruhwirth-Schnatter [2006]. Notably, Kim [1994] develops a filtering scheme for a class of the state space models with Markov-switching parameters. Because the mixture process is a special case of the Markov-switching process, a simplified version of the Kim's filter can be used to compute the filtered means and variance, i.e., equations (3.14) and (3.15). See appendix for the details of the simplified Kim's filter.

An important "by-product" obtained by a practice of the Kim's filter is the likelihood of $y_1^*, \ldots, y_T^*$ using the prediction error decomposition as follows:

$$f(y_1^*, \ldots, y_T^* | \sigma_{\eta,e}^2) = \prod_{t=1}^{T} f(y_t^* | \Psi_{t-1}, \sigma_{\eta,e}^2).$$

where the one-step ahead density, $f(y_t^* | \Psi_{t-1}, \sigma_{\eta,e}^2)$, is normal with the filtered mean and variance computed as in (3.14) and (3.15). According to the basic Bayes' rule, the posterior distribution of $\sigma_{\eta,e}^2$ is proportional to the product of the likelihood and the prior density, that is,

$$f(\sigma_{\eta,e}^2 | y_1^*, \ldots, y_T^*) \propto f(y_1^*, \ldots, y_T^* | \sigma_{\eta,e}^2) f(\sigma_{\eta,e}^2),$$

where $f(\sigma_{\eta,e}^2)$ is the prior density of $\sigma_{\eta,e}^2$ which is usually assumed to be a well-known distribution. Therefore, it is possible to draw sample of $\sigma_{\eta,e}^2$ without conditioning on h^T and s^T according to the above posterior kernel . Let $\tilde{\sigma}_{\eta,e}^2$ be a draw from some proposal density $q(\sigma_{\eta,e}^2; \sigma_{\eta,e}^{2(s-1)})$, and $\sigma_{\eta,e}^{2(s-1)}$ be the previous draw of the MCMC chain. Accept the proposed draw with the following probability

$$\alpha = min \left\{ 1, \frac{f(y_1^*, \ldots, y_T^* | \tilde{\sigma}_{\eta,e}^2) f(\tilde{\sigma}_{\eta,e}^2)}{f(y_1^*, \ldots, y_T^* | \sigma_{\eta,e}^{2(s-1)}) f(\sigma_{\eta,e}^{2(s-1)})} \times \frac{q(\sigma_{\eta,e}^{2(s-1)}; \tilde{\sigma}_{\eta,e}^2)}{q(\tilde{\sigma}_{\eta,e}^2; \sigma_{\eta,e}^{2(s-1)})} \right\}$$

If the proposed draw is accepted, the filtered means and variances associated with $\tilde{\sigma}_{\eta,e}^2$ are used to as the input of the FFBS to generate $h_{e,t}$. Otherwise, generate $h_{e,t}$ according to the filtered means and variances associated with $\sigma_{\eta,e}^{2(s-1)}$.

We summarize the MHGibbs sampler in Algorithm 3. Comparing to the AMS, there are two main differences which may improve the efficiency of the MHGibbs sampler. Firstly, the MHGibbs sampler is not conditioning on the indicator variable, s^T, so it substantially reduces the dependence between the indicator variable and the latent variable, h^T.[6] Secondly, parameters are drawn *without* conditioning on the corresponding latent variable. That is, we draw θ_{ss} unconditional on β^T and θ_{sv} on h^T.

Algorithm 3 MHGibbs sampler for the SS-SV model

1. Draw random sample from $\theta_{ss}, \beta^T | D, \theta_{sv}, h^T$ by drawing
   (a) θ_{ss} from $\theta_{ss}|D, h^T$ and
   (b) β^T from $\beta^T|D, h^T, \theta_{ss}$.
2. Draw random sample from $\theta_{sv}, h^T | D, \theta_{ss}, \beta^T$ by drawing
   (a) θ_{sv} from $\theta_{sv}|D, \beta^T$ and
   (b) h^T from $h^T|D, \beta^T, \theta_{sv}$.

3.2.5 Comparing Some Alternative Samplers

The literature of the stochastic volatility is too enormous to be briefly surveyed in this subsection. We thus only focus on the most relevant approach. Chib et al. [2002] use Metropolis-Hastings algorithm to draw θ_{sv} without conditioning on h^T, but conditioning on s^T. The idea is that conditional on s^T, the likelihood function can be easily evaluated from the output of the Kalman filter recursions, and thus the posterior kernel can be obtained by combining the likelihood with the prior density. From this point of view, the MHGibbs sampler improves upon Chib et al. [2002] in that it samples θ_{sv} without conditioning on *both*

[6] In fact, making inference about the filtered and smoothed probability of s_t is straightforward using MHGibbs sampler. We can do one more step after step 2.(b): draw s^T from $s^T|D, h^T, \beta^T, \theta_{sv}$. The first block concerning β^T and θ_{ss} is not affected relevant to s^T, so this step is redundant unless s_t is itself of interest for research question.

h^T and s^T. Stoffer and Wall [2004] replace the $\ln \chi^2$ errors in the stochastic volatility process by a mixture of two normals. Instead of approximating $\ln \chi^2$, they treat the means and the variances of the mixture of normals as the model parameters, and estimate the model using MLE, where the likelihood function is evaluated using the same filtering technique as in this paper, i.e., simplified Kim's filter.

Several notable alternative methods for the SS-SV model have have been proposed in the literature. Koopman and Bos [2004] consider a combination of the linear Gaussian state space model and a common stochastic volatility model, meaning that the signal/noise ratio is constant over time. They cast the estimation within a Frequentist framework and develop a MLE based on simulation techniques. However, as they recognized, allowing for only one stochastic volatility process might be regarded as a limitation. The SS-SV model considered in this paper covers a wider range of interesting models that can be used in practice in that it allow for multiple stochastic volatilities and time-varying signal/noise ratio.

Within a Bayesian framework, Bos and Shephard [2006] estimate the SS-SV model using the "single-move" Metropolis-Hasting algorithm for drawing h^T and the FFBS for drawing β^T. Since the mixing properties of the MCMC draws obtained using the single-move sampler can be poor, according to Giordani et al. [2011], the FFBS-based samplers introduced in this paper can potentially outperform the single-move sampler. Giordani and Kohn [2008] discuss how to estimate the state space model with the stochastic volatility based on the algorithm originally proposed by Carter and Kohn [1996]. They point out that drawing the indicator variables, s^T, conditional on the states, h^T, can be very inefficient when s^T and h^T are highly correlated. They propose a MCMC approach to draw s^T without conditioning on h^T and find that the efficiency gain is remarkable. The idea of our approach is in line with Giordani and Kohn [2008]'s argument, but we draw parameters and states without conditioning on s^T. We believe that in the cases where the indicator variable, s_t, just serves as a device for approximating, our algorithm might be easier to implement in that it does not involve novel filtering and sampling techniques.

3.3 Simulation Study

3.3.1 Data Generating Process and Priors

In this section, we compare AMS outlined in Algorithms 1 and MHGibbs sampler outlined in Algorithm 2 using simulated data generated by the following process:

$$
\begin{aligned}
y_t &= \mu_t + e_t, \quad e_t \sim N(0, \exp(h_t/2)),\\
\mu_t &= \mu_{t-1} + u_t \quad u_t \sim N(0, \sigma_u^2),\\
h_t &= h_{t-1} + \eta_t \quad \eta \sim N(0, \sigma_\eta^2).
\end{aligned}
$$

This is a local level model with the stochastic volatility. Note that there are only two fixed parameters in this model, σ_u^2 and σ_η^2, so it is useful to compare performance of these two algorithms. The data are generated based on true value $\sigma_u^2 = \sigma_\eta^2 = 0.01$.

We implement two algorithms assuming that prior distributions for σ_u^2 and σ_η^2 are both Inversed-Gamma, $IG(1, 0.05)$[7]. Following suggestions of Durbin and Koopman [2001], we set diffuse priors, N(0,100), on initial values of latent variables, h_0 and μ_0. For the MHGibbs sampler, we generate the candidate draw according to a random walk, $\theta^{(i)} = \theta^{(i-1)} + z$, where $\theta^{(i)}$ is the parameter drawn for i-th iteration of the Gibbs sampler, and $z \sim N(0, c)$. c is chosen to ensure that the average acceptance probability is in the region 0.2 to 0.5, based on suggestion in section 5.5.2 of Koop [2003]. In this simulation, c is set to be 0.0005, and average acceptance probabilities are 0.3270 and 0.4397 for σ_η^2 and σ_u^2, respectively. Each sampler runs 50,000 times and the first 10,000 draws are discarded to eliminate the effects of initial values. This procedure gives us 40,000 posterior samples, upon which the following analyses are based.

3.3.2 Modifying the AMS

We first report the prior and the posterior distributions of σ_u and σ_η (the upper panels) and the 90 % credible set of the estimated latent variables (the lower panels) using the AMS in Figure (3.1). The vertical line represents the true value, 0.1. It is surprising that the

[7]The inverse gamma density $IG(a, b)$ is $f_{IG}(z; a, b) = \frac{b^a}{\Gamma(a)} z^{-(a+1)} exp(-b/z)$.

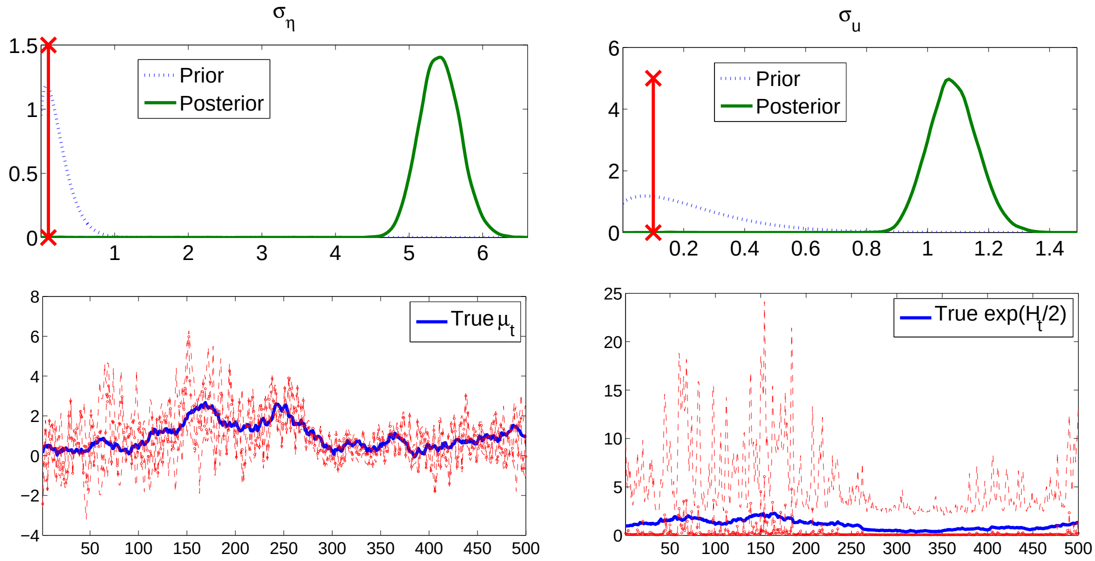


Figure 3.1: Posterior simulation according to the AMS

broadly-used AMS fails to capture the true values in this simple case.[8] In fact, Niemi and West [2010] also find similar results in the supplemental material of their paper, and they argue that the inadequacy of the FFBS-enabling approximation to the distribution of the log chi-squared errors is a key reason. They also note that using normal mixtures with more components can alleviate this problem to some degree, but not remove it. Since Omori et al. [2007] show that the 10-component mixture-normal greatly improves the approximation of log-χ^2 distribution, we believe that there should be other reason for the serious failure of the AMS.

In additional to the approximation error, we find that the mixing indicator variable, s_t, plays a crucial role in causing the failure of the AMS. Because β^T is subject to uncertainty, the left-hand-side variable in equation (3.9) is changing over iterations of the AMS. As the consequence, s^T drawn in previous the iteration might be able to approximate the distribution of errors well, causing a really volatile draw of h^T. Motivated by this conjecture, we modify the generic AMS as follows: for each iteration of the AMS, run the step 3 of

[8]We obtain the same results even with half a million draws and different initial points, so neither the slow convergence nor the bad luck seem to be a plausible reason.

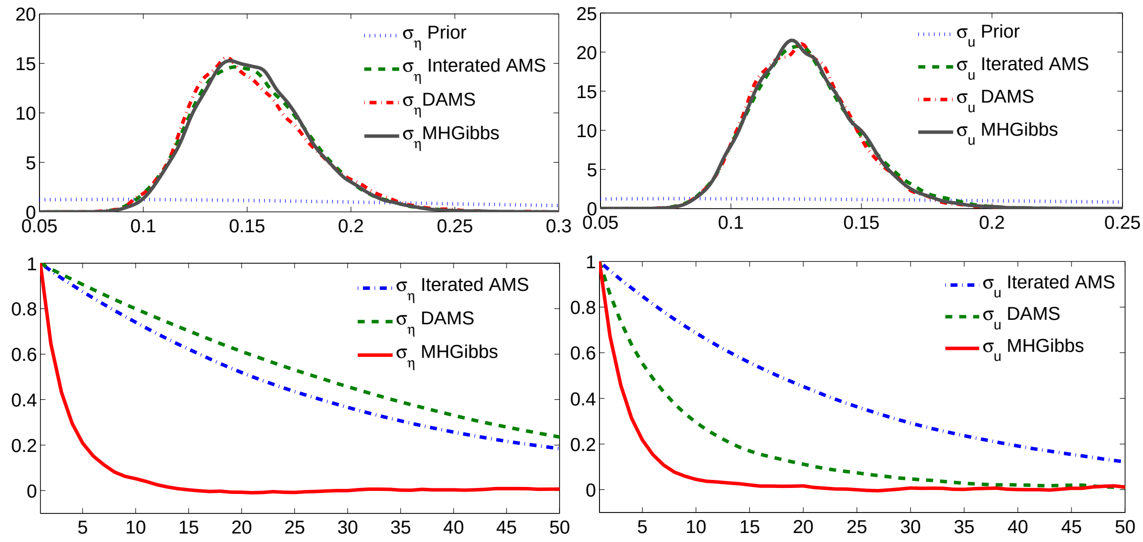


Figure 3.2: Posterior density (upper panels) and autocorrelations (lower panels) of draws of parameters according to different samplers

Algorithm 1 for k times. We call this method as the iterated AMS and use $k = 10$ in what follows.

3.3.3 Performances of Different Samplers

Figure 3.2 shows estimation result using different samplers. The posterior densities of σ_u and σ_η shown in the upper panels cover the true value with positive probability, and all three samplers converge to the same stationary distribution. As slow mixing of MCMC draws implies strong correlation in the chain of sampled parameter vectors, it is instructive to look at the correlation of the individual parameters. The lower panels of Figure 3.2 display the autocorrelation of the samples of σ_η and σ_u using different samplers. The estimated autocorrelation of the draws using the MHGibbs sampler dies out after about 10 lags, while that using the iterated AMS remains positive after 50 lags. The performance of the DAMS is comparable with the iterated AMS for σ_η but is better than the iterated AMS for for σ_u.

Table 3.2: Efficiency Analysis

	Iterated AMS	DAMS	MHGibbs
Inefficiency Factor			
σ_η	45.3694	58.1206	5.8160
σ_u	34.4591	16.8173	4.0324
CPU time for			
100 draws (seconds)	71.9571	15.6046	22.4978

Table 3.2 displays the estimated inefficiency factors defined as

$$\text{Ineff}(\hat{\theta}_G) = \frac{\text{Var}(\hat{\theta}_G)}{s^2/G},$$

where $\hat{\theta}_G = G^{-1}\sum_{g=1}^{G}\theta^{(g)}$ is the sample mean of a generic parameter θ, G is the total number of the MCMC draws, and s^2 is just the sample variance of $[\theta^{(1)}, \ldots, \theta^{(G)}]$. Intuitively, $\text{Var}(\hat{q}_G)$ measures the "long-run" variance of the sequence $[\theta^{(1)}, \ldots, \theta^{(G)}]$ and has to compensate for the fact that this sequence is correlated.[9] Thus the inefficiency factor is the ratio of variance of $\hat{\theta}_G$ to the variance of $\hat{\theta}_G$ relevant for the case of independent draws. A low value of inefficiency factor is preferred, while a value of one indicates that the sampler delivers an uncorrelated set of draws. According to Table 3.2, the MHGibbs sampler outperforms the two versions of the AMS by a factor from 4 to 11. Note that even though the DAMS does not always outperform the iterated AMS, the iterated AMS requires $k \times N$ iterations in total to obtain N posterior draws for SV-block, leading to an increase in computing time by a factor of k. Thus the DAMS is more efficient in terms of the computation time.

Finally, we report the posterior median and the 90 % credible set of MH-Gibbs estimated latent variables in Figure 3.3. Estimates using other samplers are very similar thus we do

[9]We follow Koop [2003] (see chapter 4) and estimate $\text{Var}(\hat{q}_G)$ using the Matlab code from Jim LeSage's Econometrics Toolbox.

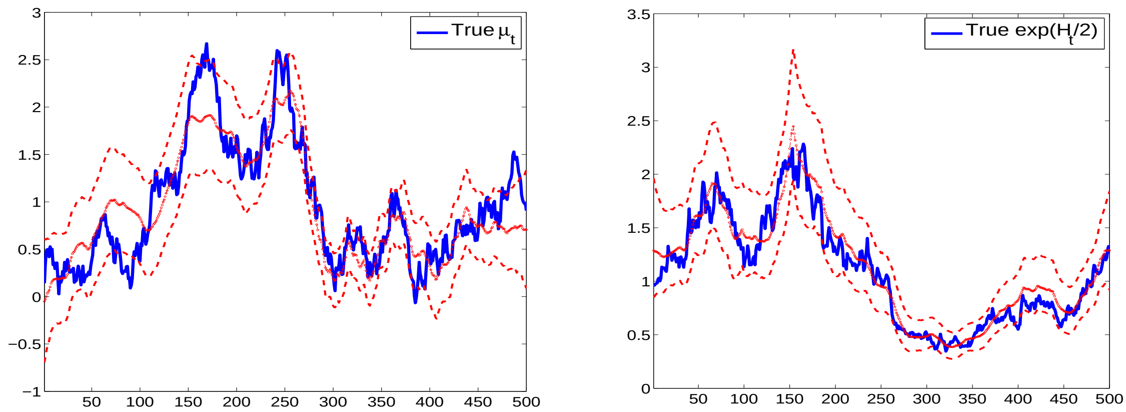


Figure 3.3: Posterior distributions of latent variables

not report here[10]. The MHGibbs sampler works well and captures the true latent variables closely.

3.4 Empirical Application

In this section we revisit the unobserved component stochastic volatility (UCSV hereinafter) model proposed by Stock and Watson [2007] (SW hereinafter). The UCSV model approximates inflation as a sum of a persistent and a transitory component. The persistent component captures the trend in inflation, while the transitory component captures deviations of inflation from its trend value. Furthermore, the variability of both the trend and temporary components are allowed to change over time. The UCSV model assumes that inflation evolves according the the following process:

$$\pi_t = \tau_t + \eta_t, \quad \eta_t = exp(h_{\eta,t}/2)\zeta_{\eta,t} \tag{3.16}$$

$$\tau_t = \tau_{t-1} + \epsilon_t, \quad \epsilon_t = exp(h_{\epsilon,t}/2)\zeta_{\epsilon,t} \tag{3.17}$$

$$h_{\eta,t} = h_{\eta,t-1} + \nu_{\eta,t} \tag{3.18}$$

$$h_{\epsilon,t} = h_{\epsilon,t-1} + \nu_{\epsilon,t} \tag{3.19}$$

[10]These results are not a consequence of using a special sample. We have tried several different sequences generated by the same model, and similar results emerge.

where $\zeta_t = (\zeta_{\eta,t}, \zeta_{\epsilon,t})$ is i.i.d. $N(0, I_2)$, $\nu_t = (\nu_{\eta,t}, \nu_{\epsilon,t})$ is i.i.d. $N(0, diag([\gamma_\eta^2, \gamma_\epsilon^2]))$, ζ_t and ν_t are independently distributed. Note that this model has only two parameters, γ_η^2 and γ_ϵ^2, which control the smoothness of the stochastic volatility process. SW set $\gamma_\eta^2 = \gamma_\epsilon^2 = 0.2$ and find that the 1970s through 1983 was a period of high volatility of the permanent innovation, $\sigma_{\epsilon,t}$, but there is little change in the estimates of the variance of the transitory innovation, $\sigma_{\eta,t}$.

Even though the restriction $\gamma_\eta^2 = \gamma_\epsilon^2 = 0.2$ works well for U.S. inflation, there are two reasons that an attempt to estimate these parameters is important. Firstly, there is no guidance on how to choose reasonable values for parameters while applying the UCSV model to other time series. Furthermore, the number of possible extensions of the UCSV model is unlimited. It is not difficult to incorporate higher order dynamics into the transitory component, η_t, so that the extended model can be used for a broader set of time series, e.g., real GDP. In such cases, it is almost impossible to choose a set of reasonable values for the parameters describing latent variables. Secondly, being able to estimate parameters is important to take advantage of the Bayesian approach, e.g., Model comparison or averaging. The marginal likelihood, by construction, is not conditioning on specific values of the parameters. Therefore, if the model comparison and averaging are of the most importance in a research agenda, one should estimate the model parameters instead of choosing a set of reasonable values.

We apply the MHGibbs sampler to estimate UCSV model using U.S. GDP inflation, from 1953:I-2010:IV. Following SW, we measure inflation as the annualized quarterly change in the GDP price deflator (400 times the log difference of the deflator). We do not impose restriction on γ_η^2 and γ_ϵ^2, so that there are two fixed parameters. Prior distributions for σ_ϵ^2 and σ_η^2 are both Inversed-Gamma, $IG(1, 0.05)$. We set diffuse priors, N(0,100), on initial values of latent variables, $h_{\eta,0}$, $h_{\epsilon,0}$ and μ_0. We generates candidate draws according to a random walk, and c is chosen to be 0.005. The average acceptance probabilities are 0.4975 and 0.4881 for σ_η^2 and σ_ϵ^2, respectively. The sampler runs 15,000 times and the first 5,000 draws are discarded to eliminate the effects of initial values. This procedure gives us 10,000 posterior samples, upon which the following analyses are based.

Figure 3.4 shows the posterior median of latent variables, and the dotted lines represent

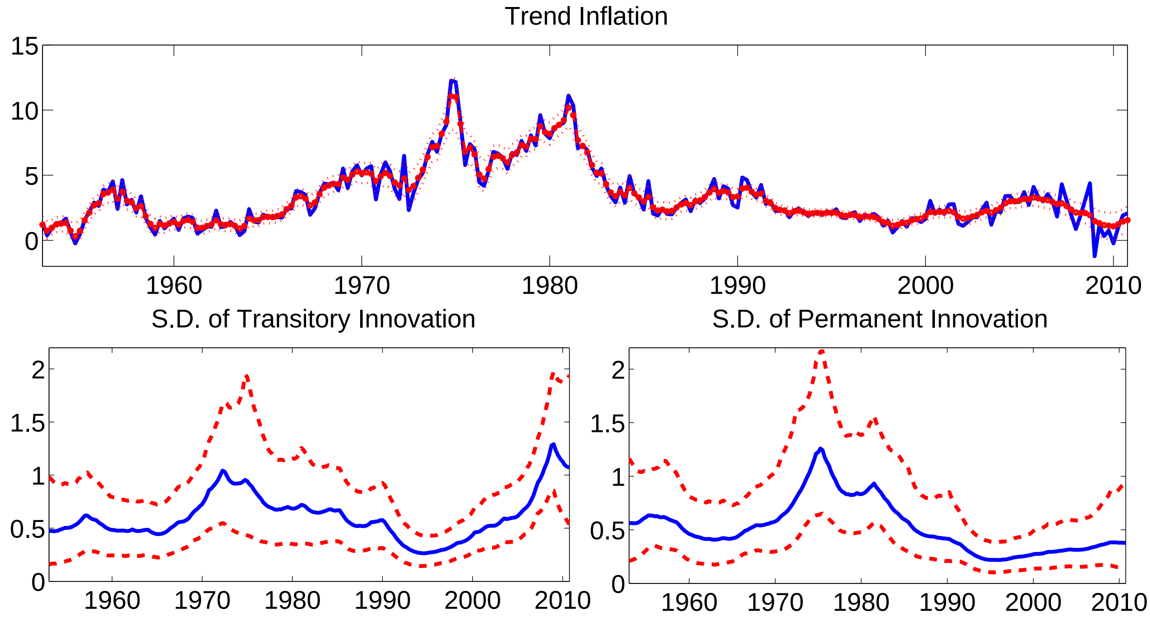


Figure 3.4: Latent Variables in UCSV Model

90 % credible set. The posterior distributions are by and large consistent with those in SW, except that the standard deviation of the transitory component also shifted up during 1970s. Therefore, the standard deviation of the permanent component during 1970s did not increase as much as found in SW. Figure 3.5 shows the posterior distributions of γ_η^2 and γ_ϵ^2, and the vertical lines represent the values chosen by SW. Even though the values chosen by SW have nontrivial posterior density, the uncertainty associated with γ_η^2 and γ_ϵ^2 is great. Because ignoring the uncertainty may lead to biased statistical evidence in favor of the UCSV model, estimating the UCSV model using the MHGibbs sampler is suggested for research concerning the "fitness" of various empirical models for U.S. inflation.

3.5 Conclusion

In this paper we have proposed and compared three methods for the SS-SV model. With the 10-component mixture of normals approximating the log-χ^2 distributions, We have shown that the conventional AMS, which is conditioning on the mixture indicator variable, can fail to capture the true model parameters. Two simple modifications of the AMS

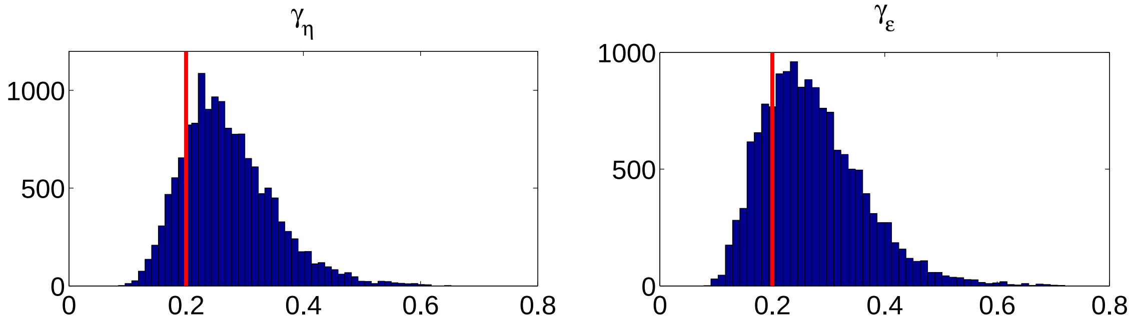


Figure 3.5: Parameters in UCSV Model

have been proposed: iterated AMS and disturbance-based AMS. However, we have shown that drawing the mixture indicator variable is not necessary and proposed a new method, MHGibbs sampler, based on conventional Kalman filter and Kim's filter. We illustrated these methods on simulated and real data.

Using simulation data, we have shown that the MHGibbs sampler outperforms significantly both AMSs in that the simulated chain displayed better mixing properties. Furthermore, the disturbance-based AMS reduces the computation time substantially comparing to the iterated AMS. We have applied the MHGibbs sampler to the UCSV model proposed by Stock and Watson [2007], and similar results have been found except that the transitory component of U.S. inflation exhibited greater volatility in 1970s.

BIBLIOGRAPHY

Olivier Coibion and Yuriy Gorodnichenko. Monetary policy, trend inflation, and the great moderation: An alternative interpretation. *American Economic Review*, 101(1):341–70, 2011.

John B. Taylor. Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39(1):195–214, December 1993.

Christopher A. Sims and Tao Zha. Were there regime switches in u.s. monetary policy? *American Economic Review*, 96(1):54–81, March 2006.

Francesco Bianchi. Regime switches, agents' beliefs, and post-world war ii u.s. macroeconomic dynamics. *The Review of Economic Studies*, 80(2):463–490, 2013.

Richard H. Clarida, Jordi Gali, and Mark L. Gertler. Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of economics*, 115(1): 147–180, 2000.

Peter N. Ireland. Changes in the federal reserve's inflation target: Causes and consequences. *Journal of Money Credit and Banking*, 39(8):1851, 2007.

Timothy Cogley, Giorgio E. Primiceri, and Thomas J. Sargent. Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics*, 2(1):43–69, January 2010.

Christopher A. Sims, Daniel F. Waggoner, and Tao Zha. Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, 146(2):255–274, October 2008.

Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1): 1–45, February 2005.

Troy Davig and Eric M. Leeper. Monetary-fiscal policy interactions and fiscal stimulus. *European Economic Review*, 55(2):211–227, February 2011.

Helle Bunzel and Walter Enders. The taylor rule and "opportunistic" monetary policy. *Journal of Money, Credit and Banking*, 42(5):931–949, 08 2010.

Daniel Leigh. Estimating the federal reserve's implicit inflation target: A state space approach. *Journal of Economic Dynamics and Control*, 32(6):2013–2030, June 2008.

Chang-Jin Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22, 1994.

Efrem Castelnuovo, Luciano Greco, and Davide Raggi. Policy rules, regime switches, and trend inflation: An empirical investigation for the u.s. *Macroeconomic Dynamics*, FirstView:1–23, 5 2013. ISSN 1469-8056.

Bruce E. Hansen. The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp. *Journal of Applied Econometrics*, 7(S):S61–82, 1992.

Rene Garcia. Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, 39(3):763–88, August 1998.

Gary Koop and Simon M. Potter. Bayes factors and nonlinearity: Evidence from economic time series1. *Journal of Econometrics*, 88(2):251–281, November 1998.

Thomas Laubach and John C. Williams. Measuring the natural rate of interest. *The Review of Economics and Statistics*, 85(4):1063–1070, November 2003.

Timothy Cogley and Argia M. Sbordone. Trend inflation, indexation, and inflation persistence in the new keynesian phillips curve. *The American Economic Review*, 98(5): 2101–2126, 2008.

Timothy Cogley and Thomas J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics*, 8(2):262–302, 2005.

Gary Koop and Dimitris Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. 3(4):267–358, 2010.

David Romer and Christina Romer. The evolution of economic understanding and postwar stabilization policy. In *Federal Reserve Bank of Kansas City Symposium on Rethinking Stabilization Policy*, pages 11–78, 2002.

Olivier Blanchard, Giovanni Dell'Ariccia, and Paolo Maur. Rethinking macroeconomic policy. *Journal of Money, Credit and Banking*, 42:199–215, 2010.

Chang-Jin Kim and Charles R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, volume 1. The MIT Press, 1 edition, 1999.

Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, pages 773–795, 1995.

Alan E. Gelfandand Dipak K. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B*, 56(3):501–514, 1994.

John Geweke. Using simulation methods for bayesian econometric models: inference, development,and communication. *Econometric Reviews*, 18(1):1–73, 1999.

Sharon Kozicki and Peter A. Tinsley. Perhaps the 1970s fomc did what it said it did. *Journal of Monetary Economics*, 56(6):842–855, September 2009.

Athanasios Orphanides and David W. Wilcox. The opportunistic approach to disinflation. *International Finance*, 5(1):47–71, Spring 2002.

John B. Taylor. Macroeconomic lessons from the great deviation. In *NBER Macroeconomics Annual 2010*, volume 25, pages 387–395. National Bureau of Economic Research, Inc, 2011.

Ben S. Bernanke. Reflections on the yield curve and monetary policy: a speech before the economic club of new york, march 20, 2006. *Speech*, 2006.

Frecdecrique Bec, Meclika Ben Salem, and Fabrice Collard. Asymmetries in monetary policy reaction function: Evidence for u.s. french and german central banks. *Studies in Nonlinear Dynamics & Econometrics*, 6(2):3, 2002.

Timothy Cogley and Richard Startz. Bayesian iv: the normal case with multiple endogenous variables. 2012.

Arabinda Basistha and Richard Startz. Why were changes in the federal funds rate smaller in the 1990s? *Journal of Applied Econometrics*, 19(3):339–354, 2004.

Chang-Jin Kim and Charles R. Nelson. Estimation of a forward-looking monetary policy rule: A time-varying parameter model using ex post data. *Journal of Monetary Economics*, 53(8):1949–1966, 2006.

Denise R. Osborn, Dong-Heon Kim, and Marianne Sensier. Nonlinearity in the fed's monetary policy rule. *Journal of Applied Econometrics*, 20(5):621–639, 2005.

Mark Gertler, Jordi Gali, and Richard Clarida. The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature*, 37(4):1661–1707, December 1999.

Yunus Aksoy, Athanasios Orphanides, David Small, Volker Wieland, and David Wilcox. A quantitative exploration of the opportunistic approach to disinflation. *Journal of Monetary Economics*, 53(8):1877–1893, November 2006.

M. Hashem Pesaran and Melvyn Weeks. *Nonnested Hypothesis Testing: An Overview*, pages 279–309. Blackwell Publishing Ltd, 2007.

Russell Davidson and James G. MacKinnon. Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 49(3):781–93, May 1981.

Neil R. Ericsson. Asymptotic properties of instrumental variables statistics for testing nonnested hypotheses. *Review of Economic Studies*, 50(2):287–304, April 1983.

Hwan-Sik Choi and Nicholas M. Kiefer. Robust nonnested testing and the demand for money. *Journal of Business & Economic Statistics*, 26(1):9–17, January 2008.

Leslie G. Godfrey. Robust non-nested testing for ordinary least squares regression when some of the regressors are lagged dependent variables. *Oxford Bulletin of Economics and Statistics*, 73(5), 2011.

Alastair R. Hall and Denis Pelletier. Non-nested testing in models estimated via generalized method of moments. *Econometric Theory*, 27:443–456, 3 2011. ISSN 1469-4360.

Russell Davidson and James G. MacKinnon. *Econometric Theory and Methods*. Oxford University Press, 2004.

Cgang-Jin Kim. Dealing with endogeneity in regression models with dynamic coefficients. *Foundations and Trends® in Econometrics*, 3(3):165–266, 2009.

Siddhartha Chibs. Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–21, 1995.

Dean Croushore and Tom Stark. A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130, November 2001.

Christopher Martin and Costas Milas. Testing the opportunistic approach to monetary policy. *Manchester School*, 78(2):110–125, 03 2010.

Carlo A. Favero. Taylor rules and the term structure. *Journal of Monetary Economics*, 53 (7):1377–1393, October 2006.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 65(3):361–93, July 1998.

James H. Stock and Mark W. Watson. Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33, 2007.

Siem Jan Koopman and Charles S. Bos. State space models with a common stochastic variance. *Journal of Business & Economic Statistics*, 22(3):346–357, July 2004.

Drew Creal. A survey of sequential monte carlo methods for economics and finance. *Econometric Reviews*, 31(3):245–296, 2012.

Paolo Giordani, Michael Pitt, and Robert Kohn. Bayesian inference for time series state space models. In John Geweke, Gary Koop, and Herman van Dijk, editors, *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press, 2011.

Charles S. Bos and Neil Shephard. Inference for adaptive time series models: Stochastic volatility and conditionally gaussian state space form. *Econometric Reviews*, 25(2-3): 219–244, 2006.

Christopher K. Carter and Robert J. Kohn. Markov chain monte carlo in conditionally gaussian state space models. *Biometrika*, 83(3):589–601, 1996.

Marco Del Negro and Frank Schorfheide. Bayesian macroeconometrics. *The Oxford Handbook of Bayesian Econometrics*, pages 293–389, 2011.

Christopher K. Carter and Robert J. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541, 1994.

Andrew Harvey, Esther Ruiz, and Neil Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–64, April 1994.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140 (2):425–449, October 2007.

Sylvia Fruhwirth-Schnatter. *Finite mixture and Markov switching models*. Springer series in statistics. Springer, 2006.

Siddhartha Chib, Federico Nardari, and Neil Shephard. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316, June 2002.

David S. Stoffer and Kent D. Wall. Resampling in state space models. In *State Space and Unobserved Component Models Theory and Applications*, pages 227–258. Cambridge University Press Cambridge, 2004.

Paolo Giordani and Robert Kohn. Efficient bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics*, 26(1):66–77, January 2008.

James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.

Gary Koop. *Bayesian Econometrics*. John Wiley & Son, 2003.

Jarad Niemi and Mike West. Adaptive mixture modeling metropolis methods for bayesian analysis of nonlinear state-space models. *Journal of Computational and Graphical Statistics*, 19(2):260–280, 2010.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84, March 1989.

Appendix A

THE GIBBS SAMPLER FOR CHAPTER 2

Choose a starting values, for $g = 1, \ldots, G$:

(1) Generate $\mathbf{\Phi}^{(g)}$ from $f(\mathbf{\Phi}|Data, \boldsymbol{\sigma}^{(g-1)}, \mathbf{S_c}^{(g-1)}, \mathbf{S_v}^{(g-1)}, \mathbf{p}^{(g-1)})$.

Given a random draw of state variable, the MS-VAR model can be rewritten as three regression equations, then the posterior distribution of $\mathbf{\Phi}^{(g)}$ can be drawn from a normal distribution with specific mean and variance. We will use the following fact: Let $Y = X\beta + e$ be a generic regression, where Y and e are $T \times 1$ vectors, and X is $T \times K$ matrix of rank K. $e \sim N(0, \sigma^2 I)$. Assume a normal prior for $\beta \sim N(\beta_0 \Sigma_0)$, the posterior distribution of β is $N(\beta_1, \Sigma_1)$, where

$$\beta_1 = (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'Y), \tag{A.1}$$

$$\Sigma_1 = (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1}. \tag{A.2}$$

Due to the recursive structure of the MS-VAR, drawing sample from posterior distributions of $\mathbf{\Phi}^{(g)}$ can be done equation-by-equation. For the output growth and inflation equations, results (A.1) and (A.2) can be directly applied. For example, the output growth equation is

$$y_t = [1, y_{t-1}, \pi_{t-1}, i_{t-1}, \ldots, y_{t-p}, \pi_{t-p}, i_{t-p}]\phi_y + \epsilon_{y,t}$$
$$\epsilon_{y,t} \sim N(0, h_{y,t}^2), \quad h_{y,t}^2 = (1 - S_{v,t})\sigma_{y,0}^2 + S_{v,t}\sigma_{y,1}^2,$$

where ϕ_y is a $3 * p + 1 \times 1$ vector of parameters. Given the knowledge of $S_{v,t}$ and $\sigma_{y,0}$, the stochastic variance $h_{y,t}^2$ is known. By dividing both side by $h_{y,t}$ we have

$$y_t^* \equiv y_t/h_{y,t} = [1/h_{y,t}, y_{t-1}^*, \pi_{t-1}^*, i_{t-1}, \ldots, y_{t-p}^*, \pi_{t-p}^*, i_{t-p}^*]\phi_y + \epsilon_{y,t}^*$$
$$\epsilon_{y,t} \sim N(0, 1).$$

Stack all observations into a matrix then draw ϕ_y using (A.1) and (A.2). The coefficients in the inflation equation can be drawn using the same procedure, but note that regressors include contemporaneous output growth y_t.

The interest rate equation contains regime-switching coefficients, but it becomes a common linear regression given the knowledge of $S_{c,t}$. We take the following two steps to draw sample from posterior distributions of coefficients in the interest rate equation. Let $\boldsymbol{\alpha} = [\alpha_{0,0}, \alpha_{\pi,0}, \alpha_{y,0}, \alpha_{0,1}, \alpha_{\pi,1}, \alpha_{y,1}]'$ and $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_p]'$.

- Draw $\boldsymbol{\alpha}|Data, \boldsymbol{\rho}, \boldsymbol{\sigma}, \mathbf{S_c}, \mathbf{S_v}$ Subtract both side by $\rho(L)i_{t-1}$:

$$\begin{aligned} i_t - \rho(L)i_{t-1} = (1-\rho(1))[&\alpha_{0,0}(1-S_{c,t}) + \alpha_{\pi,0}(1-S_{c,t})\pi_t + \alpha_{y,0}(1-S_{c,t})y_t \\ &+ \alpha_{0,1}S_{c,t} + \alpha_{\pi,1}S_{c,t}\pi_t + \alpha_{y,1}S_{c,t}y_t] + e_{i,t}. \end{aligned}$$

  where $e_{i,t} \sim N(0, h_{i,t}^2)$ and $h_{i,t}^2 = (1-S_{v,t})\sigma_{i,0}^2 + S_{v,t}\sigma_{i,1}^2$. Divide both side by $h_{i,t}$, we have

$$\begin{aligned} a_t &\equiv \frac{i_t - \rho(L)i_{t-1}}{h_{i,t}} \\ &= (1-\rho(1))\left[\frac{(1-S_{c,t})}{h_{i,t}}, \frac{(1-S_{c,t})\pi_t}{h_{i,t}}, \frac{(1-S_{c,t})y_t}{h_{i,t}}, \frac{S_{c,t}}{h_{i,t}}, \frac{S_{c,t}\pi_t}{h_{i,t}}, \frac{S_{c,t}y_t}{h_{i,t}}\right] \times \\ &\left[\ \alpha_{0,0}, \alpha_{\pi,0}, \alpha_{y,0}, \alpha_{0,1}, \alpha_{\pi,1}, \alpha_{y,1}\ \right]' + e_{i,t}^* \end{aligned}$$

  where $e_{i,t}^* \sim N(0,1)$. Stack all observations into a matrix then draw $\boldsymbol{\alpha}$ using (A.1) and (A.2).

- Draw $\boldsymbol{\rho}|Data, \boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{S_c}, \mathbf{S_v}$

  Let $\hat{i}_t = \alpha_0(S_{c,t}) + \alpha_\pi(S_{c,t})\pi_t + \alpha_y(S_{c,t})y_t$. Given the knowledge of $\boldsymbol{\alpha}$, the interest rate equation can be rewritten as follows:

$$i_t - \hat{i}_t = \rho_1(i_{t-1} - \hat{i}_t) + \ldots + \rho_p(i_{t-p} - \hat{i}_t) + \epsilon_{i,t}$$

  Divide both side by $h_{i,t}$, then we have

$$b_t \equiv \frac{i_t - \hat{i}_t}{h_{i,t}} = \left[\frac{i_{t-1} - \hat{i}_t}{h_{i,t}}, \ldots, \frac{i_{t-1} - \hat{i}_t}{h_{i,t}}\right] \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix} + e_{i,t}^*$$

where $e^*_{i,t} \sim N(0,1)$. Stack all observations into a matrix then draw $\boldsymbol{\rho}$ using (A.1) and (A.2).

(2) Generate $\boldsymbol{\sigma}^{(g)}$ from $f(\boldsymbol{\sigma}|Data, \boldsymbol{\Phi}^{(g)}, \mathbf{S_c}^{(g-1)}, \mathbf{S_v}^{(g-1)}, \mathbf{p}^{(g-1)})$.

Given the knowledge of state variables and MS-VAR coefficients, the error term in each equation is observed. The sampling scheme of volatilities $\boldsymbol{\sigma}^{(g)}$ is then based on the "observed" error terms. Note that the sample scheme is identical across equations.

Let e denote a $T \times 1$ vector of generic error term, and

$$e_t \sim N(0, h_t^2), \quad h_t^2 = (1 - S_{v,t})\sigma_0^2 + S_{v,t}\sigma_1^2$$

Assume $\sigma_1^2 = (1 + f)\sigma_1^2$, and $f > -1$, we have

$$h_t^2 = \sigma_0^2(1 + S_{v,t}f)$$

The sampling scheme consists of two steps:

- Generate $\sigma_1|Data, f, \boldsymbol{\Phi}, \mathbf{S_c}, \mathbf{S_v}$.

  Divide the e_t by $\sqrt{1 + S_{v,t}f}$, we have

$$e_t^* \equiv \frac{e_t}{\sqrt{1 + S_{v,t}f}} \sim N(0, \sigma_0^2).$$

  Given an inverse Gamma distribution, $IG(\nu_0/2, \delta_0/2)$, as a conjugate prior for σ_0^2, the posterior distribution of σ_0^2 is $IG(\nu_1/2, \delta_1/2)$ where

$$\nu_1 = \nu_0 + T, \quad \delta_1 = \delta_0 + \sum_{t=1}^{T} e_t^{*2}$$

- Generate $f|Data, \sigma_0, \boldsymbol{\Phi}, \mathbf{S_c}, \mathbf{S_v}$.

  Divide e_t by σ_0, we have

$$e_t^{**} \equiv \frac{e_t}{\sigma_0} \sim N(0, 1 + S_{v,t}f).$$

  Note that the likelihood function of f depends only on the values of e_t for which $S_{v,t} = 1$. Therefore, given an inverse Gamma distribution as a conjugate prior for $\bar{f} = (1 + f)$, the posterior distribution of $\bar{f}$ is $IG(\nu_2/2, \delta_2/2)$ where

$$\nu_2 = \nu_0 + T_1, \quad \delta_2 = \delta_0 + \sum^{N_1} e_t^{**2},$$

where $N_1 = t : S_{v,t} = 1$; T_1 is the cardinalities of N_1; and the sum is over the elements in N_1.

(3) Generate $\mathbf{S_c}^{(g)}$ from $f(\mathbf{S_c}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_v}^{(g-1)}, \mathbf{p}^{(g-1)})$.

Define

$$\hat{i}_{t,j} = (1 - \rho(1))[\alpha_{o,j} + \alpha_{\pi,j}\pi_t + \alpha_{y,j}y_t] + \rho(L)i_{t-1}, \quad j = 0, 1.$$

It can be seen as the "fitted" interest rate under regime 0 and 1. The interest rate equation can be simplified as follows:

$$i_t = (1 - S_{c,t})\hat{i}_{t,0} + S_{c,t}\hat{i}_{t,1} + e_{i,t}, \quad e_{i,t} \sim N(0, h^2_{i,t}). \tag{A.3}$$

$h_{i,t}$ is known conditional on $\boldsymbol{\sigma}$ and $\mathbf{S_v}$. Equation (A.3) is a standard Markov-switching model Bayesian inference of the Markov-switching model has been widely used, and we would outline the sampling scheme in this appendix and refer readers to Kim and Nelson [1999] for the extensive review.

Suppressing the conditioning on the model parameters, it can be shown that

$$g(\mathbf{S_c}|\tilde{i}_T) = g(S_{c,T}|\tilde{i}_T) \prod_{t=1}^{T-1} g(S_{c,t}|S_{c,t+1}, \tilde{i}_t),$$

where $\tilde{i}_t = [i_1, \ldots, i_t]$. It suggests the following two steps:

1. Run Hamilton [1989]'s basic filter to get $g(S_{c,t}|\tilde{i}_t)$, for $t = 1, \ldots, T$. The last iteration of the filter provides us with $g(S_{c,T}|\tilde{i}_T)$, from which $S_{c,T}$ is generated.
2. To generate S_t conditional on $S_{c,t+1}$ and $\tilde{i}_t$, we employ the following results:

$$g(S_{c,t}|S_{c,t+1}, \tilde{i}_t) \propto g(S_{c,t+1}|S_{c,t})g(S_{c,t}|\tilde{i}_t),$$

where $g(S_{c,t+1}|S_{c,t})$ is the transition probability. Then calculate $Pr[S_{c,t} = 1|S_{c,t+1}, \tilde{i}_t]$ in the following way:

$$Pr[S_{c,t} = 1|S_{c,t+1}, \tilde{i}_t] = \frac{g(S_{c,t+1}|S_{c,t} = 1)g(S_{c,t} = 1|\tilde{i}_t)}{\sum_{j=0}^{1} g(S_{c,t+1}|S_{c,t} = j)g(S_{c,t} = j|\tilde{i}_t)}.$$

Then generate $S_{c,t}$ according to the above probability.

(4) Generate $\mathbf{S_v}^{(g)}$ from $f(\mathbf{S_v}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_c}^{(g)}, \mathbf{p}^{(g-1)})$.

Given the knowledge of state variables and MS-VAR coefficients, the error term in each equation is observed. Thus we have

$$\begin{bmatrix} e_{y,t} \\ e_{\pi,t} \\ e_{i,t} \end{bmatrix} \sim N(0, \Sigma_t), \quad \Sigma_t = (1 - S_{v,t})\Sigma_0 + S_{v,t}\Sigma_1 \tag{A.4}$$

where $\Sigma_j = diag([\sigma^2_{y,0}, \sigma^2_{\pi,0}, \sigma^2_{i,0}])$. Notice that equation (A.4) is also a standard Markov-Switching model. Therefore, the same sampling scheme used to draw $\mathbf{S_c}$ can be applied to generate $\mathbf{S_v}$.

(5) Generate $\mathbf{p}^{(g)}$ from $f(\mathbf{p}|Data, \mathbf{\Phi}^{(g)}, \boldsymbol{\sigma}^{(g)}, \mathbf{S_c}^{(g)}, \mathbf{S_v}^{(g)})$.

Conditional on $\mathbf{S_c}$ and $\mathbf{S_v}$, notice that $\mathbf{p}$ are independent of the data set, and the model's other parameters. We use beta distributions as conjugate priors for the transition probabilities. Interested readers are referred to Kim and Nelson [1999] for the review and details.

Appendix B

THE GIBBS SAMPLER FOR CHAPTER 3

The model can be written as follows:

$$i_t^* = \beta' X_t + e_t \tag{B.1}$$

$$X_t = (I_k \otimes z_t')\gamma + \nu_t, \tag{B.2}$$

$$\begin{bmatrix} e_t \\ \nu_t \end{bmatrix} \sim N\left(0, \Sigma\right).$$

where X_t is a $k \times 1$ vector of endogenous regressors, and z_t is a $q \times 1$ vectors of instrumental variables. The Gibbs sampler consists of the following three blocks: For clarity, we define some matrix that will be used in the what follows: $\boldsymbol{X} = [X_1, X_2, \ldots, X_T]'$ and $\boldsymbol{Z} = [z_1, z_2, \ldots, z_T]'$, $\boldsymbol{i} = [i_1^*, i_2^*, \ldots, i_T^*]'$, $\boldsymbol{e} = [e_1, e_2, \ldots, e_T]'$ and $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_T]'$.

- $\beta | \gamma, \Sigma$

  Conditional on γ, the error terms in the IV regressions, ν_t, are "observed". Let

  $$\Sigma = \begin{bmatrix} \sigma_e^2 & \Sigma_{e\nu} \\ \Sigma_{\nu e} & \Sigma_{\nu\nu} \end{bmatrix}$$

  The distribution of $e_t | \nu_t$ is normal $N(c_t, H)$ where

  $$c_t = \Sigma_{e\nu}\Sigma_{\nu\nu}^{-1}\nu_t$$
  $$H = \sigma_e^2 - \Sigma_{e\nu}\Sigma_{\nu\nu}^{-1}\Sigma_{\nu e},$$

  Therefore, we can have the following regression equation:

  $$yb_t \equiv i_t^* - c_t = \beta' X_t + \xi_t$$

where $\xi_t = (e_t - c_t)$. Notice now the error term ξ_t is iid zero mean and known variance H, therefore, with normal conjugate prior of β, $p(\beta) = N(\beta_0, \underline{V}_\beta)$, then the conditional posterior is

$$\begin{aligned} p(\beta|\gamma, \Sigma, D) &= N(\bar{\beta}, \bar{V}_\beta), \\ \bar{V}_\beta &= H\left(H\underline{V}_\beta + \boldsymbol{X'X}\right)^{-1}, \\ \bar{\beta} &= \bar{V}_\beta\left[\underline{V}_\beta^{-1}\beta_0 + \boldsymbol{X'}\left(\boldsymbol{yb} - c\right)\right]. \end{aligned} \tag{B.3}$$

where $\boldsymbol{yb} = [yb_1, yb_2, \ldots, yb_T]'$.

- $\gamma|\beta, \Sigma$.

Conditional of β, equations (B.1) and (B.2) can be written as a seemingly unrelated regression that is linear in the unknown γ parameters,

$$\begin{bmatrix} \boldsymbol{i_t} \\ vec(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} \beta' \otimes Z \\ I_k \otimes Z \end{bmatrix} \gamma + \begin{bmatrix} \boldsymbol{\nu_0} \\ vec(\boldsymbol{\nu}) \end{bmatrix}. \tag{B.4}$$

Note that the error term in the first equation, $\boldsymbol{\nu_0}$ is a composition of $\boldsymbol{e}$ and $\boldsymbol{\nu}$. For example, the first element of $\boldsymbol{\nu_0}$ is $e_1 + \beta'\nu_1$. Let $u_t = [\nu_{t0}, \nu_{t1}, \ldots, \nu_{tk}]'$, given β and Σ, we know that $u_t \sim N(0, \Omega)$, where

$$\Omega = \begin{bmatrix} 1 & \beta' \\ 0 & I_k \end{bmatrix} \Sigma \begin{bmatrix} 1 & \beta' \\ 0 & I_k \end{bmatrix}',$$

Call the left-hand side variables of i-th equation $\tilde{y}_i$ and the right hand side variables f i-th equation $\tilde{X}_i$. Assuming a normal prior $p(\gamma) = N(\gamma_0, \underline{V}_\gamma)$, the conditional posterior $p(\gamma|\beta, \Sigma, D)$ is normal with variance

$$\bar{V}_\gamma = \left(\underline{V}_\gamma^{-1} + \sum_{i=1}^{k+1} \tilde{X}_i'\Sigma^{-1}\tilde{X}_i\right), \tag{B.5}$$

and mean

$$\bar{\gamma} = \bar{V}_\gamma\left(\underline{V}_\gamma^{-1}\gamma_0 + \sum_{i=1}^{k+1} \tilde{X}_i'\Sigma^{-1}\tilde{y}_i\right). \tag{B.6}$$

- $\Sigma|\boldsymbol{\beta}, \boldsymbol{\delta}$

  Conditional on (Γ, β, D), the residuals in equations (B.1) and (B.2) are observable. Let $\boldsymbol{u} = [u_1, u_2, \ldots, u_T]'$. We assume the prior $p(\Sigma)$ is inverse Wishart with scale matrix $\underline{\mathrm{S}}$ and degrees of freedom $df \geq k+1$. Since the conditional likelihood function is Gaussian, the posterior is also inverse Wishart with scale matrix $\bar{S} = \underline{\mathrm{S}} + \boldsymbol{u}'\boldsymbol{u}$ and degrees of freedom $DF = df + N$. Hence Σ can be drawn by sampling from a $IW(\bar{S}, DF)$ distribution.

B.1 Chib's Method

The marginal likelihood of the model can be computed by applying Chib's (1995) method. Let θ be the set of all parameters, i.e., β, Γ, Σ. The "basic marginal likelihood identity"is

$$p(y, X|Z) = \frac{p(y, X|\theta^*, Z)p(\theta^*)}{p(\theta^*|D)}. \tag{B.7}$$

While the identity in equation (B.7) holds for any value of θ^*, we want to be sure that Σ^* is positive definite. As a recommendation, let $\beta^* = \text{median}(\beta^s)$ and $\gamma^* = \text{median}(\gamma^s)$, where the superscript s indicates draws from the MCMC distribution. Then, using the residuals $e_i(\beta^*, \gamma^*)$, let $\Sigma^* = \sum_{i=1}^{N} e_i e_i'/N$.

The values of the likelihood and prior in equation (B.7) can be computed directly. Note that the log likelihood is given by

$$\log p(y, X|\theta^*, Z) = -\frac{N(k+1)}{2}\log(2\pi) - \frac{N}{2}\log\det\Sigma^* - \frac{1}{2}\sum_{i=1}^{N} e_i'(\Sigma^*)^{-1}e_i.$$

The posterior density for θ^* can be decompose as follows:

$$p(\beta^*, \gamma^*, \Sigma^*|D) = p(\gamma^*|D) \times p(\beta^*|D, \gamma^*) \times p(\Sigma^*|D, \beta^*, \gamma^*). \tag{B.8}$$

The order of parameters in equation (B.8) reflects computational efficiency rather than anything more fundamental. The evaluation of (B.8) proceeds in three steps. If the draws from the MCMC are indexed $s = 1, ...S$, then

$$p(vec(\gamma^*)|D) \approx \frac{1}{S}\sum_{s=1}^{S} f_N(\gamma^*; \bar{\gamma}^s, \bar{V}_\gamma^s). \tag{B.9}$$

where $f_N(\cdot)$ is the density of a normal pdf and $\bar{\gamma}^s, \bar{V}_\gamma^s$ have already been calculated in the second block in the original MCMC.

Calculation of $p(\beta^*|D, \gamma^*)$ – the second term in (B.8) – requires a second run of the MCMC procedure outlined above, except that the second block is omitted and γ^* replaces draws of γ. Index the draws of this auxiliary sampler by $s_\beta = 1, ...S_\beta$. Then

$$p(\beta^*|D, \gamma^*) \approx \frac{1}{S_\beta} \sum_{s_\beta=1}^{S_\beta} f_N(\beta^*; \bar{\beta}^{s_\beta}, \bar{V}_\beta^{s_\beta}), \tag{B.10}$$

where $\bar{\beta}^{s_\beta}, \bar{V}_\beta^{s_\beta}$ are calculated as in the first block above.

The conditional posterior $p(\Sigma|D, \beta, \gamma)$ is simulated in the third block above. Hence the final term in (B.8) can be evaluated as

$$p(\Sigma^*|D, \beta^*, \gamma^*) = f_{IW}(\Sigma^*|DF, \bar{S}), \tag{B.11}$$

where the scale matrix $\bar{S}$ is evaluated at β^*, γ^*,as in the third block above.

Collecting the results of (B.9), (B.10), and (B.11) and multiplying them together delivers the denominator of (B.7). Dividing the posterior kernel $p(y, X|\theta^*, Z)p(\theta^*)$ by the result delivers the marginal likelihood $p(y, X|Z)$.

Appendix C
THE AUXILIARY MIXTURE SAMPLER

In this appendix, we illustrate the details of the AMS for the general SS-SV model as follows:

$$y_t = c + H\beta_t + Re_t \tag{C.1}$$

$$\beta_t = d + F\beta_{t-1} + Wu_t, \tag{C.2}$$

$$e_t = \exp(h_{e,t}/2)e_t^*, \tag{C.3}$$

$$u_t = \exp(h_{u,t}/2)u_t^*, \tag{C.4}$$

$$h_{j,t} = h_{j,t-1} + \eta_{j,t} \quad \text{for} \quad j = e, u, \tag{C.5}$$

where c, H, R, d, F, W are possibly time-varying matrices containing unknown parameters. e_t^* and u_t^* are both i.i.d. N(0,1) and $\eta_{j,t} \sim N(0, \sigma^2_{\eta,j})$

1. **Draw random sample from** $\theta_{ss}, \theta_{sv}|D, \beta^T, h^T$.

   Note that conditional on β^T and h^T, θ_{ss} and θ_{sv} are independent, so we can draw a sample from $\theta_{ss}|D, \beta^T, h^T$ and $\theta_{sv}|D, \beta^T, h^T$. These two conditional distributions are usually standard with Normal or inverse-Gamma prior distributions. For example, in the local level model with stochastic volatility used in our simulation study, $u_t = \beta_t - \beta_{t-1}$ is observed conditional on data and β^T. Assume the prior for σ_u^2 is inverse-Gamma[1] with hyperparameters a_0 and b_0, and the posterior distribution of σ_u^2 will be also inverse-Gamma as follows

   $$\sigma_u^2|D, \beta^T, h^T \sim IG(\frac{T}{2} + a_0, \frac{\sum_{t=1}^T u_t^2}{2} + b_0).$$

   See Kim and Nelson [1999] for more examples of the posterior distribution of model parameters.

[1]The inverse gamma density $IG(a, b)$ is $f_{IG}(z; a, b) = \frac{b^a}{\Gamma(a)} z^{-(a+1)} exp(-b/z)$.

2. **Draw random sample from** $\beta^T, |D, \theta_{ss}, \theta_{sv}, h^T$.

   Given the model parameters and h^T, the SS-SV model becomes a standard state space model. Many simulation method can be used to draw random sample from $\beta^T, |D, \theta_{ss}, \theta_{sv}, h^T$, and we adopt the FFBS developed by Carter and Kohn [1994]. First, run the basic filtering as follows:

   Prediction

$$\begin{aligned}
\beta_{t|t-1} &= d + F\beta_{t-1|t-1}, \\
P_{t|t-1} &= FP_{t-1|t-1}F' + W\sigma^2_{u,t}W' \\
\eta_{t|t-1} &= y_t - y_{t|t-1} = y_t - H\beta_{t|t-1} - c \\
f_{t|t-1} &= HP_{t|t-1}H' + R\sigma^2_{e,t}R'
\end{aligned}$$

   Updating

$$\begin{aligned}
\beta_{t|t} &= \beta_{t|t-1} + P_{t|t-1}H'f^{-1}_{t|t-1}\eta_{t|t-1} \\
P_{t|t} &= P_{t|t-1} - P_{t|t-1}H'f^{-1}_{t|t-1}HP_{t|t-1}
\end{aligned}$$

   Save $\beta_{t|t}$ and $P_{t|t}$, and generate β_T from $N(\beta_{T|T}, Pt_{T|T})$. Then do backward sampling by computing

$$\begin{aligned}
\beta_{t|t,\beta_{t+1}} &= \beta_{t|t} + P_{t|t}F'(FP_{t|t}F' + W\sigma^2_{u,t}W')^{-1}(\beta_{t+1} - F\beta_{t|t} - d), \\
P_{t|t,\beta_{t+1}} &= P_{t|t} - P_{t|t}F'(FP_{t|t}F' + W\sigma^2_{u,t}W')^{-1}FP'_{t|t},
\end{aligned}$$

   for $t = T-1, T-2, \ldots, 1$. Finally, for each t, generate β_t from $N(\tau_{t|t,\beta_{t+1}}, Pt_{t|t,\beta_{t+1}})$.

3. **Draw random sample from** $h^T, s^T|D, \theta_{ss}, \theta_{sv}, \beta^T$.

   This block is done by two steps.

   (a) h^T from $h^T|D, \theta_{ss}, \theta_{sv}, \beta^T, s^T$.

   Transform the SS-SV model according to equations (3.9) and (3.10). Conditional on the mixture indicators, s^T, The FFBS sampling scheme summarized above can be used to draw sample from $h^T|D, \theta_{ss}, \theta_{sv}, \beta^T, s^T$ with correct specifications of the coefficient matrices. That is, $H = 1$, $c(s_t) = m(s_t)$, $R = 1$, $d = 0$, $F = 1$, $W = 1$.

(b) s^T from $s^T|D, \theta_{ss}, \theta_{sv}, \beta^T, h^T$.

This step can be done by independently sampling each s_t using the probability mass function:

$$Pr(s_t = i|D, \beta_t, h_t) \propto q_i f_N(\ln e_t^2|h_t + m_i, \nu_i^2), \quad i \leq 10.$$

The same procedure applies for $\ln u_t^2$.

Appendix D

SIMPLIFIED KIM'S FILTER

Let Θ denotes the all constant parameters, and Ψ_t the information available up to time t. The likelihood function can be obtained by

$$\begin{aligned}
LK &= \prod_{t=1}^{T} f(y_t|\Theta, \Psi_{t-1}) = \prod_{t=1}^{T} \sum_{s_t} f(y_t, s_t|\Theta, \Psi_{t-1}) \\
&= \prod_{t=1}^{T} \sum_{s_t} f(y_t|s_t, \Theta, \Psi_{t-1}) pr(s_t|\Theta, \Psi_{t-1}) \\
&= \prod_{t=1}^{T} \sum_{s_t} f(y_t|s_t, \Theta, \Psi_{t-1}) pr(s_t) \equiv \prod_{t=1}^{T} lk_t
\end{aligned}$$

Note that because the mixture indicator, s_t, are not serially correlated, and it is independent with Θ a priori, $pr(s_t = i|\Theta, \Psi_{t-1}) = pr(s_t = i) = q_t$ given in Table 3.1. For notational clarity, we suppress the conditioning on Θ in this appendix.

Let the SV-part of the SS-SV model be represented by the following equations:

$$y_t = m(s_t = i) + \beta_t + \nu(s_t = i) z_t \tag{D.1}$$

$$\beta_t = \beta_{t-1} + \sigma_\eta \eta_t, \tag{D.2}$$

where $[z_t, \eta_t]' \sim N(0, I_2)$. Let $\beta_{t|t} = E(\beta_t|\Psi_t)$, and $P_{t|t}$ is the mean squared error of the forecast. Given the initial $\beta_{0|0}$ and $P_{0,0}$, implement the Kalman filter for each i as follows:

Prediction

$$\beta_{t|t-1} = \beta_{t-1|t-1},$$
$$P_{t|t-1} = P_{t-1|t-1} + \sigma_\eta^2,$$
$$\xi_{t|t-1}^{(i)} = y_t - \beta_{t|t-1} - m(s = i),$$
$$f_{t|t-1}^{(i)} = P_{t|t-1} + \nu^2(s_t = i),$$

Updating

$$\beta_{t|t}^{(i)} = \beta_{t|t-1} + P_{t|t-1}(f_{t|t-1}^{(i)})^{-1}\xi_{t|t-1}^{(i)}$$
$$P_{t|t}^{(i)} = P_{t|t-1} - P_{t|t-1}(f_{t|t-1}^{(i)})^{-1}P_{t|t-1}$$

The conditional density, $f(y_t|\Psi_{t-1})$, is thus computed by

$$lk_t = f(y_t|\Psi_{t-1}) = \sum_{i=1}^{k} f_N(y_t|f_{t|t-1}^{(i)}, \eta_{t|t-1}^{(i)}) \times q_i,$$

where f_N is the normal density.

Following Kim and Nelson [1999], we can collapse terms to obtain $\beta_{t|t}$ and $P_{t|t}$ which are needed for next iteration of the Kalman filter as follows:

$$\begin{aligned}
\beta_{t|t} &= E(\beta_t|\Psi_t) = E_s[E(\beta_t|\Psi_t)|s_t] = \sum_{i=1}^{k} E(\beta_t|s_t = i, \Psi_t)pr(s_t = i|\Psi_t) \\
&= \sum_{i=1}^{k} \beta_{t|t}^{(i)} pr(s_t = i|\Psi_t) \qquad \text{(D.3)} \\
P_{t|t} &= E_s\{E[(\beta_t - \beta_{t|t})(\beta_t - \beta_{t|t})'|\Psi_t]|s_t\} \\
&= \sum_{i=1}^{k} E[(\beta_t - \beta_{t|t})(\beta_t - \beta_{t|t})'|s_t = j, \Psi_t]pr(s_t = j|\Psi_t) \\
&= \sum_{i=1}^{k} E[(\beta_t - \beta_{t|t}^{(i)} + \beta_{t|t}^{(i)} - \beta_{t|t})(\beta_t - \beta_{t|t}^{(i)} + \beta_{t|t}^{(i)} - \beta_{t|t})'|s_t = j, \Psi_t]pr(s_t = j|\Psi_t) \\
&= \sum_{i=1}^{k} \left\{P_{t|t}^{(i)} + (\beta_{t|t}^{(i)} - \beta_{t|t})(\beta_{t|t}^{(i)} - \beta_{t|t})'\right\} pr(s_t = j|\Psi_t) \qquad \text{(D.4)}
\end{aligned}$$

Note that $pr(s_t = i|\Psi_t) \neq q_i$ because y_t^* contains information about s_t. Once y_t is observed at the end of time t, we can update the probability in the following way:

$$\begin{aligned} pr(s_t = i|\Psi_t) = pr(s_t = i|\Psi_{t-1}, y_t) &= \frac{pr(s_t = i, y_t|\Psi_{t-1})}{f(y_t|\Psi_{t-1})} \\ &= \frac{f(y_t|s_t = i, \Psi_{t-1})q_i}{lk_t} \end{aligned} \tag{D.5}$$

To sum up, the likelihood function is evaluated in the following steps:

1. For each t, compute $\eta^j_{t|t-1}$ and $f^j_{t|t-1}$ by the basic Kalman filter.

2. Compute $llk_t = log(lk_t)$.

3. Update $\beta_{t|t}$ and $P_{t|t}$ using (D.3), (D.4) and (D.5).

Once we run through $t = 1, \ldots, T$, the log of the kernel of $f(\Theta|y_1, \ldots, y_T)$ is computed by

$$log(PK) = \ln f(\Theta) + \sum_{t=1}^{T} llk_t$$

where $f(\Theta)$ is the prior density, and PK denotes the "posterior density". Using the Metropolis-Hasting sampler, we can sample from the posterior distribution $f(\Theta|y_1, \ldots, y_T)$.

Given a draw of Θ, we can easily draw sample from $f(\beta|y_1, \ldots, y_T, \Theta)$ using the FFBS proposed by Carter and Kohn [1994]. Run the filter above and save $\beta_{t|t}$ and $P_{t|t}$. Then generate β_T from $N(\beta_{T|T}, P_{T|T})$ and proceed the following for $t = T-1, \ldots, 1$:

$$\begin{aligned} \beta_{t|t,\beta_{t+1}} &= \beta_{t|t} + P_{t|t}(P_{t|t} + \sigma^2_\eta)^{-1}(\beta_{t+1} - \beta_{t|t}), \\ P_{t|t,\beta_{t+1}} &= P_{t|t} - P_{t|t}(P_{t|t} + \sigma^2_\eta)^{-1}P'_{t|t} \end{aligned}$$

For each t, generate β_t from $N(\beta_{t|t,\beta_{t+1}}, Pt_{t|t,\beta_{t+1}})$.