Estimating the FOMC's Interest Rate Rule: A Markov-Switching Stochastic Search Variable Selection Approach

Adam Check*

February 28, 2017

Abstract

Many researchers and economic commentators believe that the Federal Open Market Committee (FOMC) has changed interest rate policy over time. Mathematically, FOMC policy is typically described as a linear interest rate rule, in which the FOMC changes interest rates based on measure of the inflation gap and a measure of the output gap. In this context, a change in policy is associated with a change in the coefficients of this rule. I introduce a new econometric model in order to determine if there have been changes in these coefficients. I call this model a Markov Switching Stochastic Search Variable Selection (MS-SSVS) model, as it builds on the work of George and McCulloch (1993) who introduced SSVS in linear models, and also on the work of Hamilton (1989), Kim and Nelson (1999), Früwirth-Schnatter (2006) and others who have popularized the use of Markov-Switching models. This model embeds linear regression and Markov-Switching models as special cases, as it allows regression coefficients to be freely estimated, restricted to be identical across regimes, or restricted to be equal to zero. Because it builds these features into the likelihood function of a single model, estimation is fast and does not involve the approximation of marginal likelihoods. I find that there is very little evidence of coefficient change, with the data suggesting only a change in response to the unemployment gap.

JEL Codes:

*I can be contacted at ajc@uoregon.edu

1 Introduction

The traditional view of monetary policy is that the Federal Federal Open Market Committee (FOMC) adjusts the nominal Federal Funds rate based on measures of economic performance. Mathematically, this is typically formulated as a version of the the Taylor Rule, first described in Taylor (1993), in which the target nominal Federal Funds rate is a linear function of output and inflation. This policy rule, and others taking very similar forms, are the foundation of empirical analysis of FOMC behavior over the past two decades.

Given the episode of "Great Inflation" that occurred in the 1970s, many authors have suggested the policy rule followed by the FOMC has changed over time, with the FOMC being less proactive against inflation in the 1970s, and more proactive since the early 1980s. This view was popularized by Clarida et al. (2000), who used a split-sample regression approach to show that the inflation response coefficient was smaller before the the appointment of Paul Volcker as FOMC chairman in 1979. However, later studies cast doubt on this finding. Orphanides (2004) finds evidence in favor of a change in output response, but not in inflation response, and Sims and Zha (2006) find no evidence of change in either the output or the inflation response. In a similar vein, Boivin (2006) finds that the point estimate of the inflation response coefficient has changed over time, but his time-varying estimates have a very high degree of uncertainty associated with them. However, other authors have continued finding changes in the inflation response coefficient. For example, Taylor (2013) and Kahn (2010) have recently argued not only that the inflation response was relatively low in the 1970s, but also that the FOMC reverted to a weak inflation response during the lead-up to the Great Recession in 2008.

Addressing the question of structural change in FOMC behavior is of importance to both academics and policymakers. As has been demonstrated in both small and medium sized DSGE models by numerous authors, different monetary policies can lead to differences in inflation rates and volatility, inflation persistence, and short-term output growth rates and volatility. If monetary policy in the United States has changed, it is crucial that we document this fact, as it will eventually allow us to attribute changes in economic performance to changes in policy. This is especially true in light of Taylor's (2013) claim that a weak inflation response returned in the mid-2000s and engendered a financial bubble.

In this paper, I introduce a new econometric model in order to address this question. This Markov-Switching Stochastic Search Variable Selection (MS-SSVS) model nests both a constant coefficient model, consistent with the findings of Sims and Zha (2006), and a Markov-Switching model, consistent with Taylor's hypothesis, as special cases. In addition, the MS-SSVS model can probabilistically restrict coefficients in either one or both regimes to be zero, so that a variable may be completely excluded from the regression in either one regime or in both regimes. In short, for each coefficient in each regime, there are three possibilities: (1) the coefficient is restricted to zero; (2) the coefficient is restricted to be the same as the coefficient in the other regime; (3) the coefficient is freely estimated independently of the coefficient in the other regime.

My MS-SSVS model builds on the work of George and McCulloch (1993) and George et al. (2008), who developed Stochastic Search Variable Selection (SSVS) in order to perform variable selection in linear regression models and linear VARs. SSVS has some differences with competing methodologies such as Bayesian Model Averaging (BMA) that make it especially attractive in a Markov-Switching environment. One major advantage of SSVS is that it is not necessary to directly compute or approximate the marginal likelihood, which is a computationally intensive task in Markov-Switching models. Instead, the uncertainty associated with variable inclusion and variable switching is nested within a unified hierarchical model.

This type of variable selection and coefficient restriction has been found to have good small sample properties. In a linear regression framework, as the number of variables grows relative to the sample size, estimators lose power, and regression coefficient estimates become more imprecise. Coefficients estimates which might get sent to zero in larger sample sizes may instead be relatively large and appear to be economically and statistically significant in smaller samples. This problem becomes more acute in Markov-Switching models, since the numbers of parameters is more than doubled in these models compared to their linear counterparts. Shrinkage-type estimators such as SSVS have been shown to alleviate these problems, leading to more accurate estimates and better out-of-sample forecasting performance.

For example, in an application to recession forecasting, Owyang et al. (2015) show that using BMA improves upon the full sample forecasts, allowing them to identify the onset of recessions more quickly. In addition, when considering models in which switching is possible, accounting for potential model uncertainty can help to identify which features of the model are actually changing. For example, if volatility is changing over time in the data generating process, but is unaccounted for in estimation, it may appear that the coefficients of the model are changing. Therefore, it is important to consider more than one type of model, or to consider a model that can endogenously determine which restrictions are appropriate.

Through Monte-Carlo exercises with simulated data, I show that my methodology is able to correctly identify parameter restrictions - this is true both when the actual coefficients are near zero, and when the coefficients are the same across regimes. As expected, my MS-SSVS model is better able to identify coefficient restrictions as the signal to noise ratio increases. This increase in signal is modeled in two ways, either as a reduction in error volatility or as an increase in sample size. The MS-SSVS model is particularly adept at identifying zero-restrictions with a high degree of accuracy, even as the amount of noise increases.

When I apply this new methodology to Federal Funds Rate data from 1970-2008, I find two things. First, contrary to Clarida et al. (2000), I find very little evidence that the FOMC's inflation response differs across two regimes. Second, contrary Sims and Zha (2006) I do find evidence for two distinct regimes, with the unemployment gap response coefficient differing across the regimes. I find that there was a relatively strong unemployment response coefficient in the mid 1970s, the late 1980s and early 1990s, and between 2004-2006. This strong unemployment response does correspond to a heightened probability that the inflation response coefficient was relatively low; however, the mean inflation response coefficient in this regime is only very slightly smaller than it is in the weak unemployment response regime. There is little evidence that there was a noticeable reduction in the inflation response during the run-up of the housing bubble in the mid 2000s, or that a weak inflation response in the 1970s caused the "Great Inflation".

2 Econometric Model

The model I introduce is based on a Markov-Switching model with switching in coefficients:

$$
\begin{aligned}
& y_t = X_t \beta_{St} + \varepsilon_t \\
& \varepsilon_t \sim N(0, \sigma^2) \\
& S_t \in \{0, 1\}
\end{aligned}
$$

The regime, S_t follows a first order Markov process:

$$
\begin{aligned}
Pr(S_t = j | S_{t-1} = i) &= p_{ij} \\
& i, j \in \{0, 1\}
\end{aligned}
$$

The model as detailed above has been well studied, and there exist well known frequentist and Bayesian procedures to estimate it. These methods are described in Hamilton (1989), Kim and Nelson (1999), and Früwirth-Schnatter (2006), among others. While these techniques make it feasible to estimate the model, model comparison remains relatively cumbersome. The estimation process itself can be time consuming, making the estimation of more than a handful of models potentially infeasible. In addition, performing Bayesian Model Averaging requires estimation of the marginal likelihood of each model - this is an extra, complicated, and time consuming step that needs to be undertaken after estimation of the model.

Model comparison in this class of models, however, remains important. In linear models, there are only two possibilities for each regressor - either it belongs in the model or it does not. However, in a regime switching model with two regimes, there are four possibilities for each regressor: (1) it does not belong in either regime, (2) it belongs in each regime, and is the true effect is distinct under each regime, (3) it belongs in each regime but the true effect is the same regardless of regime, (4) the variable belongs in only one of the two regimes. This only increases the burden of model comparison, as the number of models to consider expands more rapidly when the possibility of switching is properly accounted for.

In this paper, I build an econometric model that allows for the four possibilities elicited above. It does so in a computationally feasible manner by utilizing a hierarchical prior to nest all of these possibilities within a single model. To do this, I build on the Stochastic Search Variable Selection (SSVS) technique that was developed in George and McCulloch (1993) and further studied and implemented in George et al. (2008) and Koop and Korobilis (2010). In the SSVS framework, the possibility that some coefficients do not enter the model is built into the model likelihood function, so that only one model needs to be estimated, and the marginal likelihood does not need to be computed. This framework is therefore much simpler and more time effective than performing Bayesian model averaging by estimating hundreds, thousands, or more models and comparing each based on its marginal likelihood.

Let $\beta^k = [\beta_i^k \ \beta_j^k]$ be a vector that contains coefficient k in state i and state j. Under my

MS-SSVS model, β^k is assumed to come from the following prior mixture distribution:

$$\beta^k \sim \gamma_1^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_0\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + \gamma_2^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_1\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + \gamma_3^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_1\begin{bmatrix}1 & 1-\epsilon\\1-\epsilon & 1\end{bmatrix}\right)$$
$$+ \gamma_4^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_1 & 0\\0 & \tau_0\end{bmatrix}\right) + \gamma_5^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_0 & 0\\0 & \tau_1\end{bmatrix}\right)$$
$$\gamma^k = (\gamma_1^k, \gamma_2^k, \gamma_3^k, \gamma_4^k, \gamma_5^k) \in \{(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)\}$$
$$\tau_0 = 0.1 \times \sqrt{\widehat{Var(b_k)}}$$
$$\tau_1 = 15 \times \sqrt{\widehat{Var(b_k)}}$$
$$k \in \{1, 2, 3, 4, 5\}$$

where $\widehat{Var(b_k)}$ is the OLS estimate of the variance of β under the assumption that there is no regime switching, and τ_0, τ_1, and ϵ are parameters chosen by the researcher that control the variances of each distribution in the prior for β^k.[1]

It is important to note that in this model, the posterior probability of each restriction is proportional to the prior probability of the restriction times the prior density for β^k under that restriction evaluated at the posterior draw of β^k. Therefore, the "restrictions" are only enforced approximately. To see why this is necessary, consider a prior density for the zero-restriction in which $\tau_0 = 0$, so that the coefficient was literally restricted to equal zero. Unless this restriction was chosen, our estimate of β^k will almost surely never be exactly zero. Therefore, the posterior density under this restriction will always be zero, since $\beta^k \neq 0$; therefore this restriction will never be enforced. Our goal when choosing τ_0 is to choose a sensible value that will enforce this zero-restriction when appropriate, while keeping the estimate of β^k near zero in the event that this restriction is chosen. Our goal when choosing ϵ is similar. We want to choose a number small enough that the prior density

[1] This is very similar to Bayesian linear regression, where the researcher typically chooses the variance of the prior distribution for the regression coefficients.

under this restriction will be high when both values of β^k are approximately equal, but we need to be careful to not choose a value for ϵ that is so small that the restriction will never be enforced.

The prior laid out above is data dependent since $\widehat{Var(b_k)}$ depends on the dependent variable. Therefore, it does not adhere to the requirement, in a Bayesian approach, that the prior be independent of the observed dependent data. However, as discussed in George and McCulloch (1993), the zero-restriction region depends on the values for both τ_0 and τ_1. George and McCulloch (1993) find that using $\widehat{Var(b_k)}$ in the choices of τ_0 and τ_1 helps to ensure that this zero-restriction region lies over a sensible space so that the coefficients are restricted to be close to zero only where appropriate. This prior, although not technically valid due to its dependence on the observed data, remains popular in the literature, as evidenced by its use in Koop and Korobilis (2010).[2]

The mixture of normal distributions described above represents five possibilities:

1. The coefficient is restricted to be near zero in each state, i.e. the variable is excluded in both states.

2. The coefficient estimates are freely estimated, independently of each other, i.e. the variable is included in each state, and the effect in each state is different.

3. The coefficient estimates are freely estimated, but identical, i.e. the variable is included in each state, and the effect in each state is the same.

4. The coefficient in state 0 is freely estimated, but the coefficient in state 1 is restricted to be near zero, i.e. the variable is excluded from state 1.

5. The coefficient in state 1 is freely estimated, but the coefficient in state 0 is restricted to be near zero, i.e. the variable is excluded from state 0.

[2] Priors of this form are sometimes called "empirical Bayes" methods. One argument for their use, although not mathematically rigorous, is that the goal of empirics is to discover features of the data. Priors of this form should be considered if they can be shown to be well behaved and able to uncover features of the data, even if they do not technically adhere to proper Bayesian theory.

I assume that each indicator vector γ^k comes from the following prior distribution:

$$\gamma_j^k = \begin{cases} (1,0,0,0,0) & \text{with probability } p_1 \\ (0,1,0,0,0) & \text{with probability } p_2 \\ (0,0,1,0,0) & \text{with probability } p_3 \\ (0,0,0,1,0) & \text{with probability } p_4 \\ (0,0,0,0,1) & \text{with probability } p_5 \end{cases}$$

$$\sum_{i=1}^{5} p_i = 1$$

$$0 < p_i < 1 \ \forall \ i \in \{1,2,3,4,5\}$$

3 Estimation Procedure

I set independent priors across the hierarchical parameters:

$$p(p_{00}, p_{11}, \sigma^2, \tau_0, \tau_1, p_1, p_2, p_3, p_4, p_5, \epsilon) = \\ p(p_{00})p(p_{11})p(\sigma^2)p(\tau_0)p(\tau_1)p(p_1)p(p_2)p(p_3)p(p_4)p(p_5)p(\epsilon)$$

I assume that the prior parameters $\tau_0, \tau_1, \epsilon, p_1, p_2, p_3, p_4, p_5$ are each set by the researcher, i.e. their prior is a point-mass at a particular value. This is a common assumption in the SSVS literature. The parameters τ_0, τ_1, and ϵ control the variance of the each prior mixture distribution. The probabilities, p_1, p_2, p_3, p_4, and p_5, control the weights for each prior distribution.

For the other three hyper-parameters, p_{00}, p_{11}, and σ^2, I set prior distributions:

$$p(p_{00}) = \text{Beta}(a_0, b_0)$$

$$p(p_{11}) = \text{Beta}(a_1, b_1)$$

$$p(\sigma^2) = \text{InverseGamma}(\alpha_Q, \beta_Q)$$

Drawing from the full posterior directly is intractable. Instead, I draw from each of the conditional posteriors. This is called the Gibbs sampler. Let $\beta = [\beta_0, \beta_1]'$, $P_s = [p_{00}\ p_{11}]'$, $\tau = [\tau_0, \tau_1]'$, $P_\gamma = [p_1, p_2, p_3, p_4, p_5]'$, $\Gamma = \gamma^K$. The process is as follows:

1. Sample the indicators for the mixture of normals prior each variable:

$$
\begin{aligned}
p(\Gamma^{(z)}|Y, \beta^{(z-1)}, P_s^{(z-1)}, \sigma^{2,(z-1)}, S_T^{(z-1)}, \tau, P_\gamma) &= p(\Gamma^{(z)}|Y, \beta^{(z-1)}, S_T^{(z-1)}, \tau, P_\gamma) \\
p(\Gamma^{(z)}|Y, \beta^{(z-1)}, S_T^{(z-1)}, \tau, P_\gamma) &= \text{ Categorical} \\
\Gamma_k^{(z)} &= \text{ Categorical} \begin{pmatrix} \frac{p_1 f(N(0,\Sigma_1)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_2 f(N(0,\Sigma_2)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_3 f(N(0,\Sigma_3)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_4 f(N(0,\Sigma_4)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_5 f(N(0,\Sigma_5)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \end{pmatrix}
\end{aligned}
$$

This procedure is based on George and McCulloch (1993). My procedure is slightly modified because I have a mixture of five normal distributions rather than two. Once the prior mixture distributions are selected, form the prior variance for β as:

$$
D = \begin{bmatrix} \Sigma^{k=1} & 0 & \cdots & 0 & 0 \\ 0 & \Sigma^{k=2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \Sigma^{k=K} \end{bmatrix}
$$

D is block diagonal, with the Cholesky decomposition of the two-by-two mixture variance for each pair of coefficients, k, Σ^k along the diagonals, with zeros everywhere else.

2. Sample the regression coefficients:

$$
\begin{aligned}
p(\beta^{(z)}|Y,\Gamma^{(z)},P_s^{(z-1)},\sigma^{2,(z-1)},S_T^{(z-1)},\tau,P_\gamma) &= p(\beta^{(z)}|Y,\Gamma^{(z)},\sigma^{2,(z-1)},S_T^{(z-1)},\tau) \\
p(\beta^{(z)}|Y,\Gamma^{(z)},\sigma^{2,(z-1)},S_T^{(z-1)},\tau) &\sim \text{ Normal} \\
\beta^{(z)} &\sim N\left(\hat{\beta},V\right) \\
V &= ((DRD')^{-1}+X'X)^{-1} \\
\hat{\beta} &= VX'Y
\end{aligned}
$$

where R is a prior correlation matrix, typically set to the identity matrix.

3. Sample the variance of the regression error:

$$
\begin{aligned}
p(\sigma^{2,(z)}|Y,\Gamma^{(z)},\beta^{(z)},P_s^{(z-1)},S_T^{(z-1)},\tau,P_\gamma) &= p(\sigma^{2,(z)}|Y,\beta^{(z)},S_T^{(z-1)}) \\
p(\sigma^{2,(z)}|Y,\beta^{(z)},S_T^{(z-1)}) &= \text{ Inverse Gamma} \\
\sigma^{2,(z)} &\sim \text{ IG}\left(a_Q+\frac{T}{2},\beta_Q+\frac{SSE}{2}\right)
\end{aligned}
$$

where T is the sample size and $SSE=(Y-X\beta)'(Y-X\beta)$

4. Sample the Markov State indicators:

$$
p(S_T^{(z)}|Y,\Gamma^{(z)},\beta^{(z)},\sigma^{2,(z)},,P_s^{(z-1)},\tau,P_\gamma) = p(S_T^{(z)}|Y,\beta^{(z)},\sigma^{2,(z)},P_s^{(z-1)})
$$

using the procedure described in Kim and Nelson (1999).

5. Sample the Markov transition probabilities:

$$
\begin{aligned}
p(P_s^{(z)}|Y,\Gamma^{(z)},\beta^{(z)},\sigma^{2,(z)},S_T^{(z)},\tau,P_\gamma) &= p(P_s^{(z)}|Y,S_T^{(z)}) \\
p(P_s^{(z)}|Y,S_T^{(z)}) &= \text{ Beta} \\
P_s^{ii,(z)} &= \text{ Beta}(a_i+N_{ii},b_i+N_{ij})
\end{aligned}
$$

where N_{ij} is the number of times that the regime transitioned from regime i to regime j in $S_T^{(z)}$.

3.1 Helicopter Tour of Prior for β^k

Recall that the prior for $\beta^k = [\beta_0^k \ \beta_1^k]'$ is given by a mixture of five Normal distributions:

$$\beta^k \sim \gamma_1^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_0\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + \gamma_2^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_1\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + \gamma_3^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \tau_1\begin{bmatrix}1 & 1-\epsilon\\1-\epsilon & 1\end{bmatrix}\right)$$
$$+ \gamma_4^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_1 & 0\\0 & \tau_0\end{bmatrix}\right) + \gamma_5^k N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_0 & 0\\0 & \tau_1\end{bmatrix}\right)$$

In my application, I choose:

$$Pr(\gamma_1^k = 1) = 0.4$$
$$Pr(\gamma_2^k = 1) = 0.2$$
$$Pr(\gamma_3^k = 1) = 0.2$$
$$Pr(\gamma_4^k = 1) = 0.1$$
$$Pr(\gamma_5^k = 1) = 0.1$$

This imposes a prior belief that there is a 50% probability that β_i^k is restricted to be near 0. Given that β_i^k is not restricted to be near 0, there is a 40% prior probability that $\beta_i^k \approx \beta_j^k$, and a 60% probability that β_i^k is independent of β_j^k.

In addition, I choose:

$$
\begin{aligned}
\tau_0^k &= c_0\sqrt{\widehat{var(\beta^k)}}\\
\tau_1^k &= c_1\sqrt{\widehat{var(\beta^k)}}\\
c_0 &= 0.1\\
c_1 &= 15.0
\end{aligned}
$$

where $\widehat{var(\beta^k)}$ is the OLS estimate of the variance of β^k under a no regime switching assumption. These priors are similar to ones suggested in George and McCulloch (1993) and Koop and Korobilis (2010). Finally, for the case of parameters restricted to be equal across regimes, I set $\epsilon = 1.0 - 0.99999$.

In figures (1)-(3), I plot the prior probability density function of β_0 and β_1. This prior density function has some striking features. It is strongly peaked near $\beta_0 = 0$ and $\beta_1 = 0$, so there is a relatively high prior probability that both coefficients are restricted to zero. If the estimated coefficients land in the orange region of figure 2 (or the yellow region of figure 3), it is almost a certainty that the priors for β_0 and β_1 will be centered on zero with a very tight prior variance. Additionally, there are three other regions which receive relatively large prior mass: both regions where one of the coefficients is restricted to be near zero, and the diagonal region representing coefficients that are (roughly) identical under each regime. Outside of these four relatively narrow but sharply peaked regions, the Normal distribution with the highest probability density function corresponds to both regimes being freely estimated.

Figure 1: Prior Probability Density Function for Different Values of β_0 and β_1

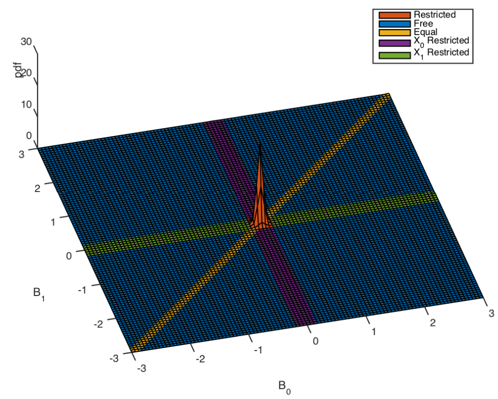


Figure 2: Prior Probability Density Function for Different Values of β_0 and β_1: View from Above

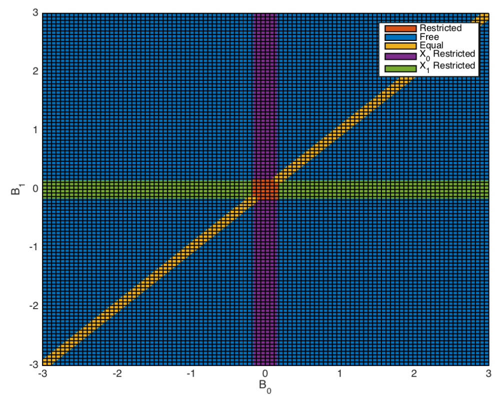
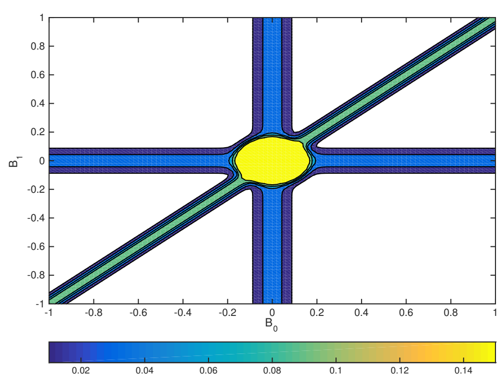


Figure 3: Contour Plot of the Prior Probability Density Function for Different Values of β_0 and β_1



4 Monte-Carlo Analysis

With the estimation procedure in hand, I turn to analyzing how effective the MS-SSVS procedure is at identifying features of the data. I am most interested in the ability of this model to identify the various restrictions that are built into it, when they are actually present in the data. I consider two cases: one in which data is generated from a process that has several types of restrictions, and another that corresponds to linear regression. However, before turning to data analysis, I investigate a particular choice of prior parameters to see what the prior mixture distribution for β^k looks like.

4.1 Monte-Carlo Exercise

To investigate how well this procedure does in identifying features of restricted Markov-Switching models, I run a series of simulations. Under the true model:

$$\begin{aligned} y_t &= X_t \beta_{St} + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma^2) \\ Pr(S_t = j | S_{t-1} = i) &= p_{ij} \end{aligned}$$

Where $X_t = \begin{bmatrix} 1 & X_{1,t} & X_{2,t} \end{bmatrix}$ and

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} \sim MvN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

i.e. there are three elements in each X_t: an intercept term and two randomly generated and independent regressors. Since there are three columns in X_t, $K = 3$. Recall that $\beta^k = [\beta_0^k \ \beta_1^k]'$ for $k \in \{1, \cdots, K\}$. In words this vector β^k contains the coefficients on regressor k in each state. In this exercise, I chose:

$$\beta^1 = \begin{bmatrix} 1.0 \\ -0.5 \end{bmatrix} \quad \beta^2 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \quad \beta^3 = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$$

An alternative way of viewing the coefficients is to define the coefficients separately in each state:

$$\beta_0 = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.0 \end{bmatrix} \quad \beta_1 = \begin{bmatrix} -0.5 \\ 1.0 \\ 0.0 \end{bmatrix}$$

Finally, the transition probabilities for each state are given by:

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

I will discuss the choices of σ^2 later.

4.1.1 Priors

In this exercise, I assume that the variance term is constant. Therefore, I set an inverse-Gamma prior on the variance term. My priors are as follows:

Table 1: Monte-Carlo Priors

Parameter	Prior Mean	Prior S.D.
c_0	0.1	0.0
c_1	15	0.0
τ_0^k	$c_0\sqrt{\widehat{Var(b^k)}}$	0.0
τ_1^k	$c_1\sqrt{\widehat{Var(b^k)}}$	0.0
p_1	0.4	0.0
p_2	0.2	0.0
p_3	0.2	0.0
p_4	0.1	0.0
p_5	0.1	0.0
p_{00}	0.8	0.16
p_{11}	0.8	0.16
σ^2	1.0	0.58

4.1.2 Results

In the Monte-Carlo exercise, I vary both the number of observations, T, and the standard deviation of the error term, σ. I consider $T \in \{50, 100, 150, 200, 250\}$ and $\sigma \in \{0.1, 0.5, 1.0, 2.0\}$. I find that the models with large T and small σ are most able to pick out the correct restrictions. Intuitively, this makes sense, since these models have the largest sample size and smallest variance, allowing the true features of the model to shine through.

In the tables below, I present the average accuracy of identification of the correct restriction by the estimation procedure. This number has been averaged over the results across 200 separate data generation and estimation procedures. For example, for the first column of table (2), I set $\sigma = 0.1$ and $T = 50$. I then generate 200 data sets and run the estimation procedure on each. For each data set, I calculate the percentage of the time the appropriate restriction was chosen, and I average this percentage across all 200 data sets. For my estimation procedure, I use 15,000 burn-in draws and 20,000 posterior draws. For ease of notation, let the regression coefficients on the intercept term be rewritten as $\begin{bmatrix} \beta_0^1 \\ \beta_1^1 \end{bmatrix} \equiv \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}$. Finally, in the tables below, I have abused notation in the first two rows. For $\mu_0 \neq 0$ and $\mu_1 \neq 0$, I mean that each is freely estimated, so that they are not restricted to be equal to zero and not restricted to be identical to each other.

Table 2: $\sigma = 0.1$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	98.6%	100%	100%	100%	100%
$\mu_1 \neq 0$	97.5%	100%	100%	100%	100%
$\beta_0^1 = \beta_1^1$	92.5%	96.2%	97.3%	97.9%	98.4%
$\beta_0^2 = 0$	96.4%	97.6%	97.9%	98.1%	98.2%
$\beta_1^2 = 0$	95.8%	97.4%	97.9%	97.9%	98.1%

Table 3: $\sigma = 0.5$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	71.7%	96.9%	99.8%	100%	100%
$\mu_1 \neq 0$	39.4%	81.8%	96.1%	99.3%	100%
$\beta_0^1 = \beta_1^1$	71.1%	84.7%	89.2%	90.4%	91.6%
$\beta_0^2 = 0$	88.8%	91.9%	91.8%	93.2%	92.6%
$\beta_1^2 = 0$	87.5%	90.8%	90.3%	92.2%	91.5%

Table 4: $\sigma = 1.0$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	51.5%	72.3%	85.5%	91.1%	95.0%
$\mu_1 \neq 0$	22.3%	36.3%	47.3%	54.6%	64.7%
$\beta_0^1 = \beta_1^1$	54.4%	65.7%	72.4%	75.1%	77.3%
$\beta_0^2 = 0$	84.5%	87.4%	88.1%	88.0%	88.9%
$\beta_1^2 = 0$	82.1%	86.6%	86.5%	86.7%	87.2%

Table 5: $\sigma = 2.0$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	35.3%	44.6%	56.6%	62.2%	65.3%
$\mu_1 \neq 0$	19.3%	20.2%	23.8%	27.9%	30.6%
$\beta_0^1 = \beta_1^1$	34.4%	45.5%	50.0%	53.3%	54.5%
$\beta_0^2 = 0$	82.9%	86.0%	84.0%	83.6%	85.1%
$\beta_1^2 = 0$	82.7%	85.5%	83.6%	83.7%	84.8%

Three things become apparent when looking at these tables. First, the MS-SSVS model is able to correctly identify all types restrictions when the data has a high signal to noise ratio. This is a sign that our estimation procedure is well-behaved when the amount of noise in the data generating process is relatively small. Second, the model performs very well at detecting true zero restrictions, but as the noise rises, correct selection of the "identical" restriction declines. I find that even under the noisiest condition tested, $T = 50$ and $\sigma = 2.0$, the model still detects the zero restrictions with about 83% accuracy. Third, as the noise increases, the model has a relatively more difficult time detecting that μ_1 is actually different than zero compared to μ_0. This is due to the fact that the absolute value of μ_1 is smaller than the absolute value of μ_1. Since μ_1 is closer to zero, the model has a harder time distinguishing μ_1 from 0 than it does with μ_0. Note that this has some carryover effect on the accuracy of the other restrictions, causing $\beta_1^2 = 0$ to be slightly less accurately detected than $\beta_0^2 = 0$ (about one percentage point in most simulations).

4.1.3 Results: Linear Model

I repeat the same exercise as above, except with the following:

$$\beta^1 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \quad \beta^2 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \quad \beta^3 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

This case corresponds to linear regression. I leave everything else, including my priors, unchanged and conduct the same analysis as above. I present my results in the tables below.

Table 6: $\sigma = 0.1$, Linear Regression

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 = \mu_1$	93.2%	100%	100%	100%	100%
$\beta_0^1 = \beta_1^1$	91.6%	100%	100%	100%	100%
$\beta_0^2 = \beta_1^2$	92.0%	100%	100%	100%	100%

In the linear regression case, there is a rapid deterioration in performance of identifying $\beta_0^k = \beta_1^k$ as the amount of noise in the data generating process increases. This may be partially driven by the fact that there is only a 20% prior probability placed on each of these

Table 7: $\sigma = 0.5$, Linear Regression

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 = \mu_1$	58.3%	65.1%	68.9%	72.9%	75.3%
$\beta_0^1 = \beta_1^1$	56.7%	64.3%	68.4%	72.0%	74.8%
$\beta_0^2 = \beta_1^2$	56.6%	64.2%	68.2%	72.8%	74.7%

Table 8: $\sigma = 1.0$, Linear Regression

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 = \mu_1$	49.9%	53.8%	54.1%	57.3%	57.9%
$\beta_0^1 = \beta_1^1$	52.7%	56.7%	59.2%	60.8%	62.0%
$\beta_0^2 = \beta_1^2$	52.6%	55.5%	58.7%	59.1%	62.2%

Table 9: $\sigma = 2.0$, Linear Regression

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 = \mu_1$	25.2%	35.5%	37.3%	40.4%	42.8%
$\beta_0^1 = \beta_1^1$	38.1%	48.8%	52.7%	54.2%	53.3%
$\beta_0^2 = \beta_1^2$	36.6%	48.0%	51.9%	53.7%	55.9%

coefficients being identical. It is important to note that while I have fixed the prior mixture probabilities to be identical for all sets of parameters, in general a researcher could relax this assumption, placing different prior mixture probabilities on each pair of coefficients. If a researcher suspected that one pair of coefficients would be identical in each regime, she could increase the prior probability of that restriction holding.

Additionally, the estimation procedure seems to have more trouble identifying that the mean parameters are identical than it does identifying the other regression coefficients are identical. In linear regression, it is often the case that the mean is the least precisely estimated coefficient, since there is no variation along that dimension of X. Therefore, a researcher should expect that the densities of the mean coefficients will have a greater spread than the other coefficients. Therefore, it is relatively unsurprising that the model has a hard time detecting that the mean coefficients are actually identical.

Finally, there is no reason a researcher needs to stop their analysis after conducting this estimation procedure. If they find relatively strong evidence for coefficient restrictions, it might make sense to enforce those restrictions exactly. For example, they could use this model to find the mode of the mixture distribution, and then estimate a standard Markov-Switching model at that mode, allowing only some coefficients to switch and discarding regressors that are restricted to zero.

5 Application: Interest Rate Rules

Now that we understand more about the properties and accuracy of this estimator, I apply it to monetary policy rule estimation. I use a data set compiled in Check (2016) that uses the official forecasts prepared for the FOMC by their staff. These forecasts are contained in what is known as the "Greenbook", which is published with a six year lag. I follow many papers in this literature, and estimate a rule of the form:

$$i_t = \mu_{s_t} + \rho_{s_t} i_{t-1} + \phi_{\pi_{s_t}}(\pi_t^e - \pi^T) + \phi_{u_{s_t}} u_t^e + \phi_{\Delta u_{s_t}} \Delta u_t^e + \sigma_t \varepsilon_t$$

$$s_t = j$$

$$j \in \{0, 1\}$$

$$\varepsilon_t \sim N(0, 1)$$

Finally, I assume that the volatility of the error term follows a random walk. Let $\sigma_t = \exp(\frac{h_t}{2})$. Then:

$$h_t = h_{t-1} + v_t$$

$$v_t \sim N(0, Q)$$

This rule is fairly standard, with four exceptions. First, it allows for the possibility of Markov-Switching in the coefficients. Second, it is estimated using "meeting-based timing", with the Federal Funds rate on the left-hand side being the average Federal Funds rate between meeting dates rather than between months or quarters. This helps to ensure that the left hand side truly is the nominal federal funds rate target that is implemented by the central bank. Third, when considering the employment response, it includes the expected future change in the unemployment rate in addition to the unemployment gap to account for possible asymmetric unemployment responses over the business cycle. In Check (2016), I found that inclusion of this variable allowed me to better explain past FOMC behavior, and better forecast future FOMC behavior. Fourth, and finally, it includes stochastic volatility. The use of stochastic volatility is relatively rare in this literature, but it is not unique to this study. Sims and Zha (2006) included stochastic volatility and found strong evidence for its inclusion.[3]

In Markov-Switching models, the likelihood function is bi-modal, with two peaks of identical height. This is due to the fact that the model is symmetric to relabeling, so the value of the likelihood function would be identical if the labeling of the regimes were switched. Because the Gibbs-sampler wanders around the posterior density, if the peaks of the likelihood function are close enough, then a researcher can encounter a "label-switching" problem, where the sampler will switch between the two peaks of the likelihood function, and the regimes will flip. This causes bi-modal densities for the regression coefficients in each regime, each spanning the same space. In addition, both regime probabilities at each point in time get pushed towards 50%, since the sampler is switching the labeling of the regimes.

One way to circumvent this problem is to normalize the model, and this is the strategy that I employ in this paper. In general, a researcher typically selects one (or more) coefficients on which to include inequality restrictions across regimes. For example, when estimating a Markov-Switching model on U.S. GDP data, the researcher typically restricts the mean coefficients so that in one regime the mean coefficient is always greater than the mean coefficient in the other regime. This type of normalization allows the sampler to converge to a distribution around one of the two peaks of the likelihood function, rather than switching back and forth between peaks.

[3] Allowing for stochastic volatility requires replacing step (3) on page 11 with the procedure described in Kim et al. (1998).

Following the work of Clarida et al. (2000), among others, I first tried to normalize the model by restricting the inflation response in one regime to be greater than the inflation response in the other. Implementing this restriction led to extremely poor performance of the sampler. After inspecting the histograms of the regression coefficients, it became clear that the model preferred the inflation coefficients to be approximately. Therefore, normalization using the inflation coefficient was ineffective. Under this normalization, the while the coefficients on inflation were approximate equal, the coefficients on the unemployment gap were both bi-modal, spanning the same two modes. This indicated that label-switching was occurring. This was also clear from inspection of the plot of regimes over time, as the probability of both regimes was near 50% throughout the sample.

Due to the evidence that normalizing on the inflation response coefficient failed to properly normalize the model, and that the only coefficient where differences were pronounced between regimes was the coefficient on the unemployment gap, I instead normalized the model on the unemployment gap response coefficient. After doing so, I found that the sampler was much better behaved, with the densities on the unemployment coefficient distinct and unimodal. I still find that the coefficients on all other variables, including the inflation gap, are unimodal and nearly identical. This evidence contradicts the findings by Clarida et al. (2000), but is consistent with evidence presented in Orphanides (2004) and Sims and Zha (2006). Orphanides (2004) finds that the inflation response has been relatively unchanged over time, but that there has been variation in the response to a measure of the real-time output gap. Sims and Zha (2006) find that the introduction of stochastic volatility implies that the coefficients in the interest rate rule have remained constant over time.

5.1 Convergence Diagnostics

Below, I present evidence that the estimator converges to a unique stationary distribution. I first present running mean plots throughout the samples that are discarded. If the sampler is converging to a stationary distribution, then the means of all of the parameters of the model should converge to their means in the stationary distribution. If it is not, then these means will be trending up, down, or bouncing around. Next, I present the autocorrelation functions for the parameters of the model. These functions show the correlation between the draw of the parameter at one iteration and the draw of the same parameter t iterations later. If the sampler is well-behaved, then the autocorrelation functions should fall towards zero as the number of iterations increases. A simple rule-of-thumb is that the number of discarded "burn-in" draws should be at least ten times larger than the maximum number of iterations that it takes the autocorrelation of any parameter to drop to zero.

5.1.1 Running Mean Plots

Figure 4: Regression Parameters

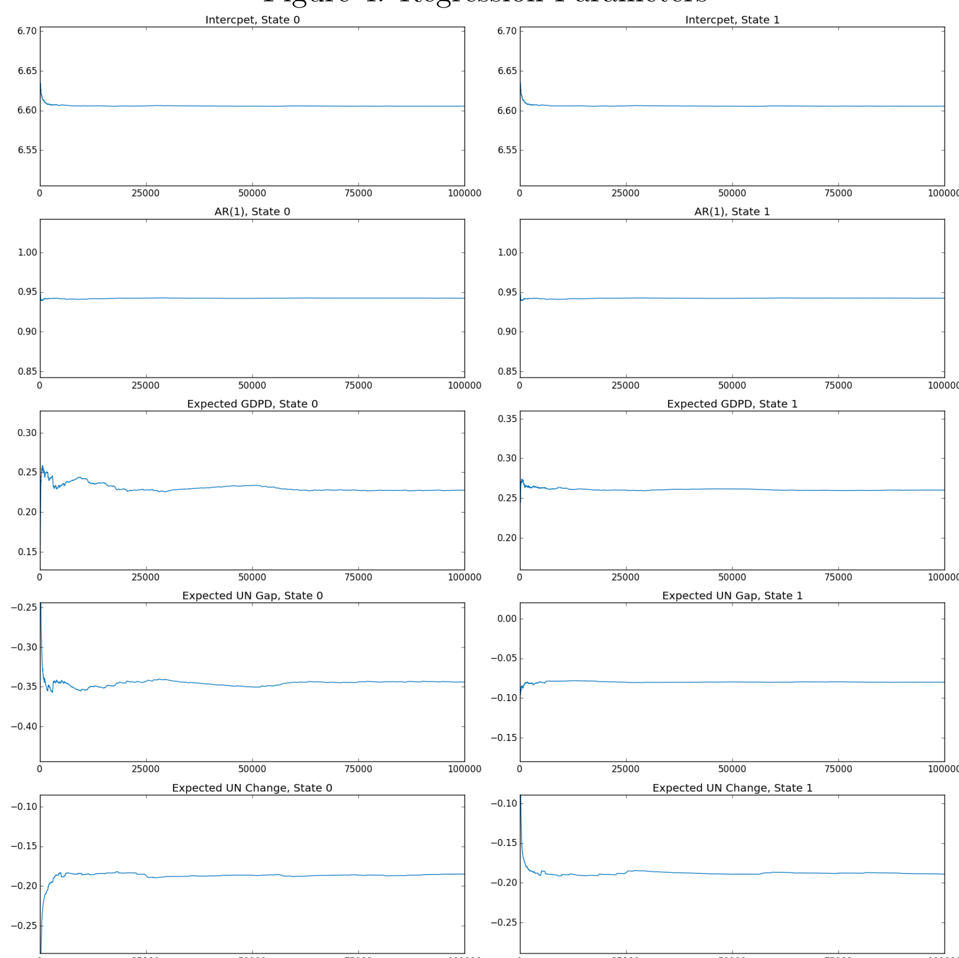
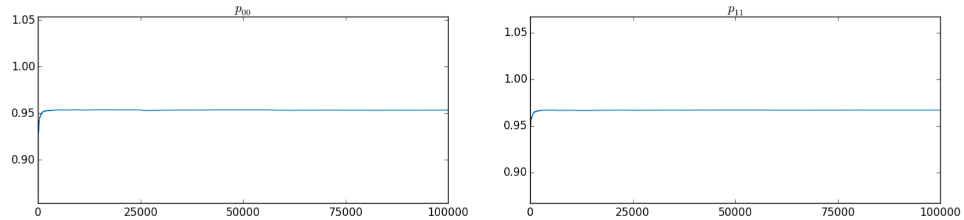


Figure 5: Regime Probabilities



5.1.2 Autocorrelation Functions

Figure 6: Regression Parameters

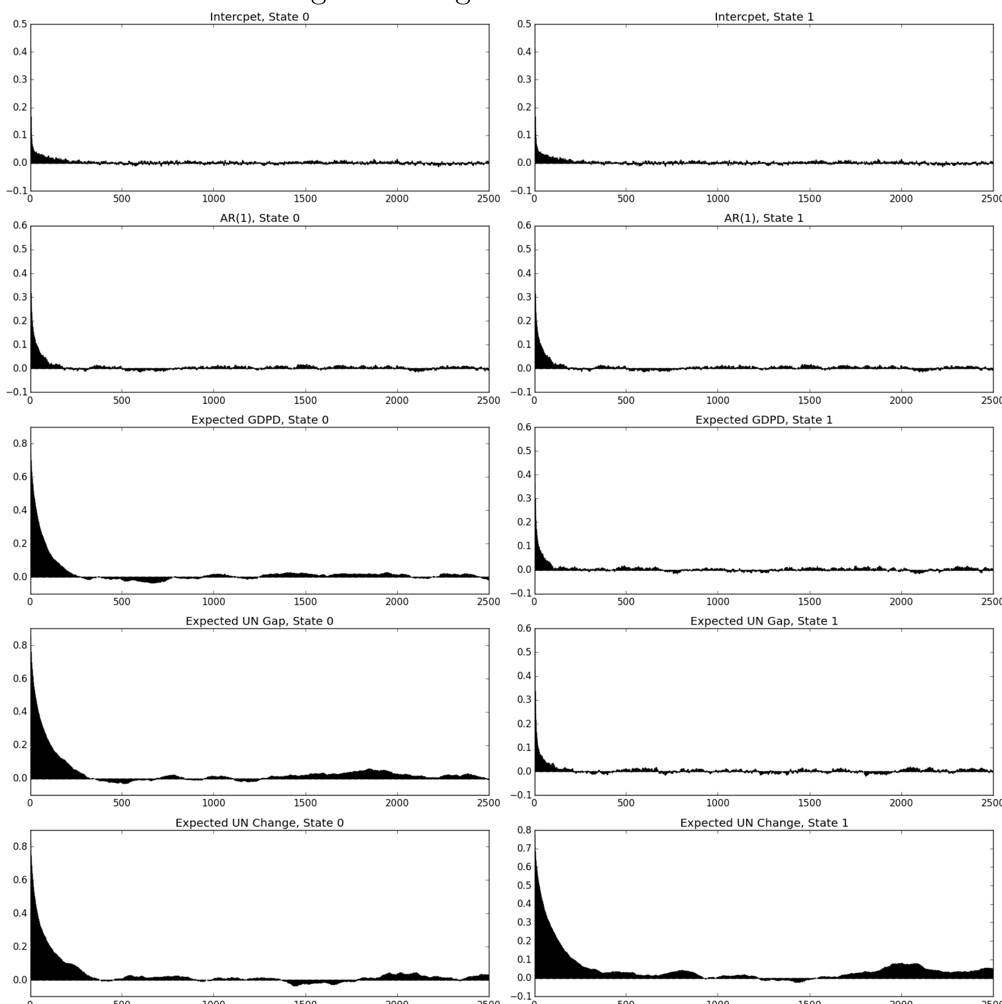
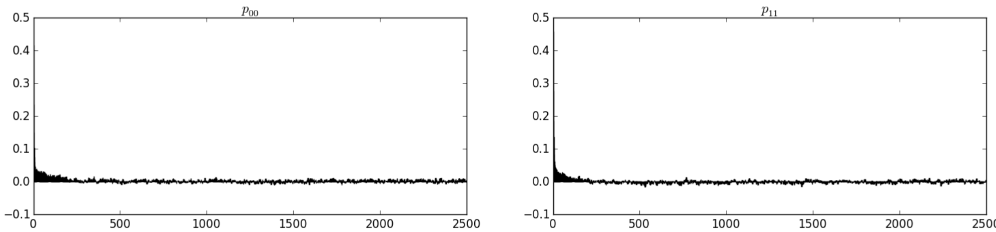


Figure 7: Regime Probabilities



Both of these metrics suggest that the sampler is well-behaved. The running mean plots become flat towards the end of the discarded draws, suggesting that the sampler has converged to a stationary distribution. In addition, the autocorrelation plots settle down to zero at roughly 500-1000 draws, suggesting that only 5,000-10,000 burn-in draws are needed. I perform 100,000 burn-in draws in an abundance of caution. I keep the next 150,000 draws and use them to form posterior inference.

5.2 Results

Next, I present my results. First, I present estimates of the restriction probabilities and regression coefficients in each regime. Second, I plot the coefficient densities for each parameter of the model. For the regression coefficients, I display the densities under each regime. Third, I plot the estimates of the regimes. Because the only major difference between the two regimes is the unemployment gap response, I name one regime the "weak unemployment response regime" and the other the "strong unemployment response regime". Finally, I display the estimate of the volatility term over time, along with the uncertainty associated with it.

5.2.1 Restriction Probabilities and Mean Regression Coefficient Estimates

Table 10: Estimated Restrictions in the “Strong” Unemployment Response Regime

	Zero-Restriction	Freely Estimated	“Identical” Restriction
μ_0	0.0%	0.0%	100%
ρ_0	0.0%	0.0%	100%
$\phi_{\pi,0}$	15.4%	14.7%	69.9%
$\phi_{UN,0}$	1.2%	94.3%	4.5%
$\phi_{\Delta UN,0}$	5.3%	15.6%	79.1%

Table 11: Estimated Restrictions in the “Weak” Unemployment Response Regime

	Zero-Restriction	Freely Estimated	“Identical” Restriction
μ_1	0.0%	0.0%	100%
ρ_1	0.0%	0.0%	100%
$\phi_{\pi,1}$	4.6%	25.5%	69.9%
$\phi_{UN,1}$	49.9%	45.6%	4.5%
$\phi_{\Delta UN,1}$	2.3%	18.6%	79.1%

Table 12: Mean Coefficient Values in Each Regime

	“Weak” Regime	“Strong” Regime
μ	6.61	6.61
ρ	0.94	0.94
ϕ_{π}	0.26	0.22
ϕ_{UN}	-0.08	-0.34
$\phi_{\Delta UN}$	-0.19	-0.18

5.2.2 Coefficient Densities

Figure 8: Regression Parameters

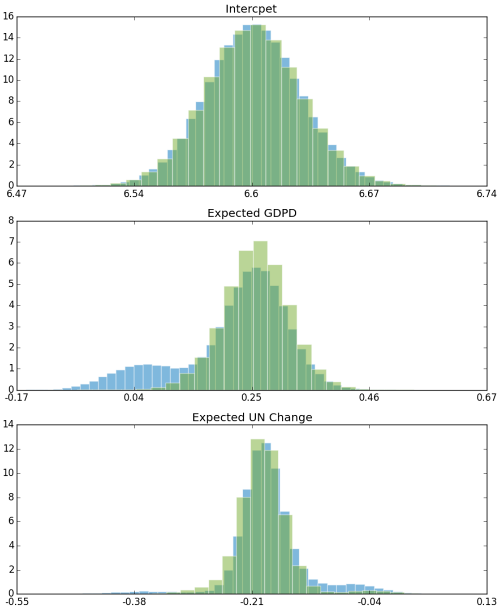
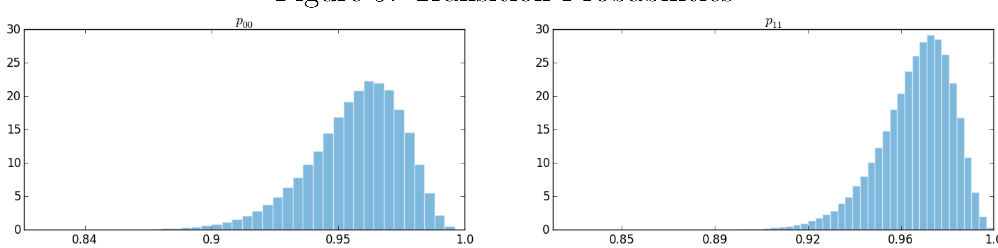


Figure 9: Transition Probabilities



5.2.3 Regime Estimation

Figure 10: Probability of Weak Unemployment Response Regime

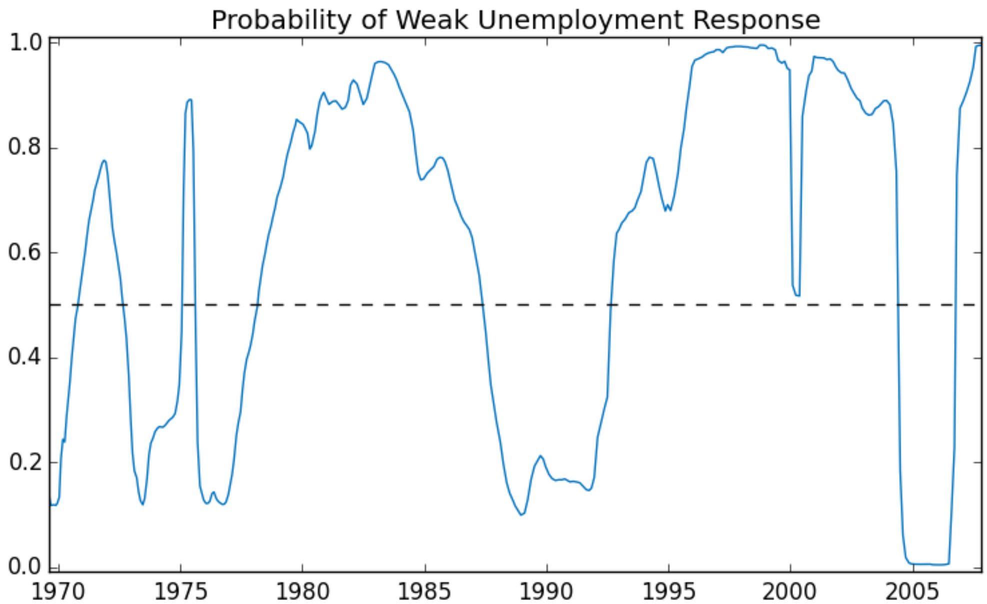
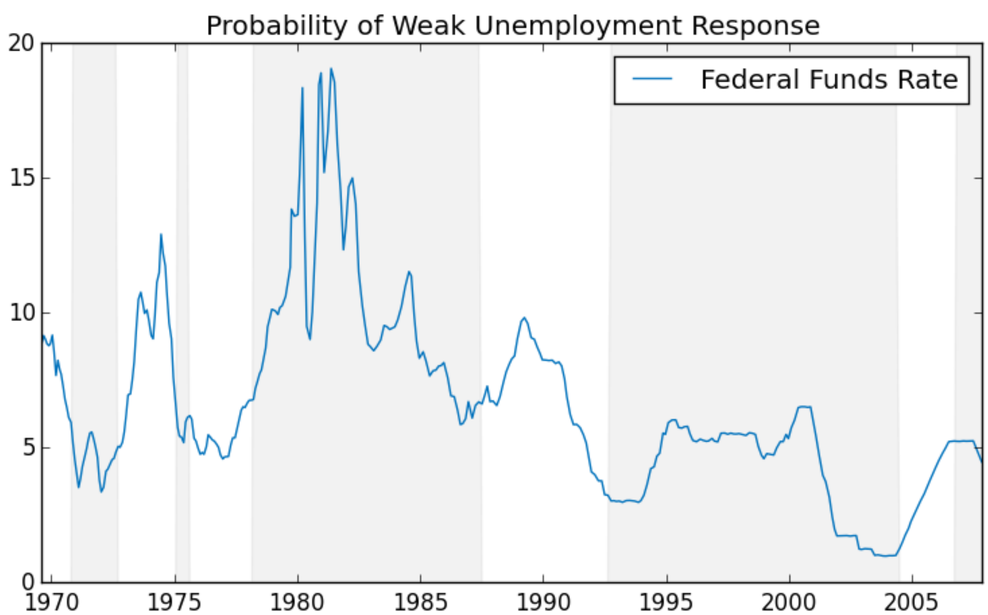


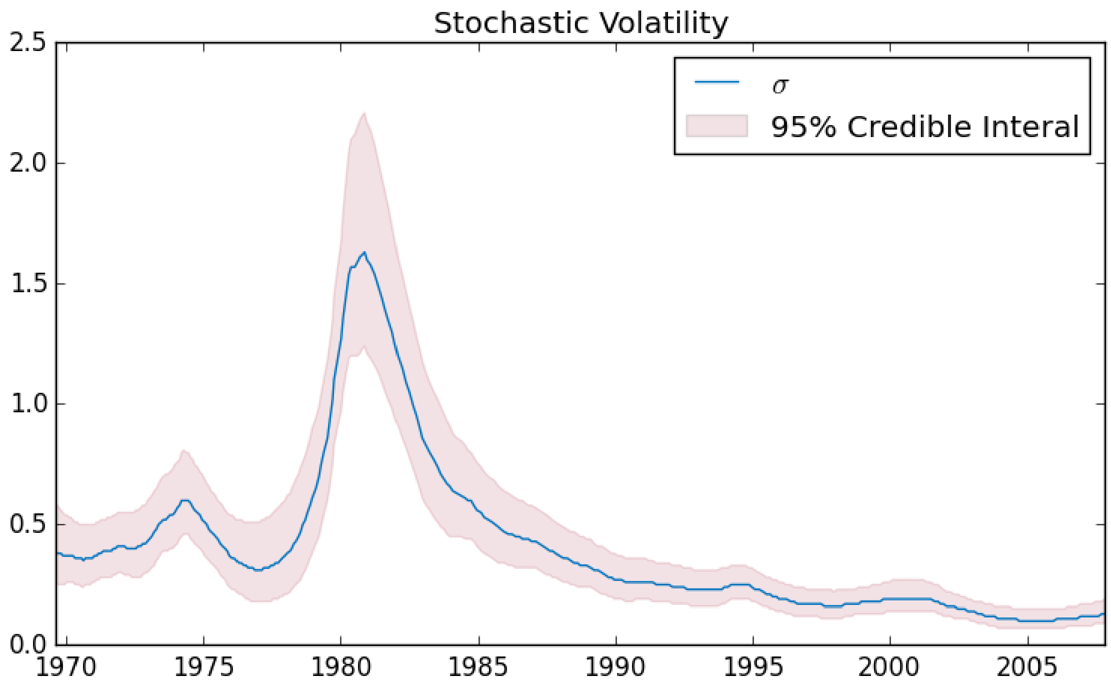
Figure 11: Federal Funds Rate



Note: Gray areas correspond to weak unemployment response regime probability ≥ 0.5

5.2.4 Stochastic Volatility

Figure 12:



These results add to the evidence found by both Sims and Zha (2006) and Orphanides (2004). First, the stochastic volatility term seems very important, as it fluctuates greatly over time. The period from 1979-1983 has the highest volatility. This should be expected, since the FOMC was on record as targeting the money supply rather than the Federal Funds rate during this period. Therefore, it stands to reason that the error term of a linear rule which suggests the FOMC has targeted the Federal Funds rate would be much higher during this period. Next, as found by Orphanides (2004) when using real time data, I find evidence that the only change that occurred in the FOMC's reaction function was its response to the output gap. The FOMC seems to have responded very strongly to the forward-looking unemployment gap in the 1970s. If their forecasts about this gap were incorrect, which Orphanides (2004) suggests, then they may have engaged in overly loose policy during part of this time period. In other words, the FOMC may have allowed interest rates to be low due to an anticipated high level of unemployment that never materialized.

These results also stand in contrast to the recent work of Taylor (2013) and Kahn (2010). Particularly, it does not appear that the FOMC became lax against fighting inflation in the mid 2000s, or even that it had responded weakly to inflation in the mid to late 1970s. The probability of being in the "strong unemployment response" regime was nearly one between late 2004 and 2006, but that would only imply overly loose policy if the FOMC had believed at each meeting that the unemployment rate was going to increase over the following year.

Finally, my results stand in contrast to a recent application of a Markov-Switching model to interest rate rules in Murray et al. (2015). These authors find that the inflation response in one regime was much lower than in the other regime, and that it failed to satisfy the "Taylor Principle", i.e. the long-run response to a one percentage point increase in the inflation gap was a less than one percentage point increase in the Federal Funds rate. They find that this weak inflationary response occurred between roughly 1973-1975 and again during the Volcker years, 1979-1985. This second period seems highly counterfactual, since most economists, and previous studies such as Clarida et al. (2000), attribute the fall in inflation after 1980 to the *strong* inflation response during Volcker's tenure.

There are three major differences between my estimation procedure and the procedure used in Murray et al. (2015). First, I use meeting-based timing, and they use quarterly timing. Second, and very importantly, I allow the volatility associated with the interest rate rule to evolve according to a separate process than the regression coefficients. This is crucial, since their regression coefficient results could be driven entirely by regime-switching in the variance parameter. Indeed, their estimated regimes appear to be highly correlated to periods where I find that volatility was relatively high. Third, I use my newly developed MS-SSVS procedure, where Murray et al. (2015) use an unrestricted Markov-Switching model.[4]

[4]In separate, unpublished, analysis, I find that their main result - regime switching between one regime that satisfies the Taylor principle and another regime that doesn't - falls apart when a standard Markov-Switching model is estimated, but using stochastic volatility instead of forcing the switch in both the regression coefficients and the volatility term to occur at the same time.

6 Conclusion

Over the past 15 years, there has been considerable disagreement about the existence of changes in the response coefficients in the FOMC's interest rate rule. In order to address this question, I build a Markov-Switching model that can endogenously determine the existence of two types of restrictions: (1) zero-restrictions, in which a variable may not be included in one or all of the regimes and (2) identity-restrictions, in which the regression coefficient on the same variable may be restricted to be identical across all regimes. My estimation procedure blends and extends the Gibbs samplers that were previously derived for estimation of Markov-Switching models and Stochastic Search Variable Selection models. I call this unified model an MS-SSVS model.

I find that the MS-SSVS model performs well at identifying true restrictions in a Monte-Carlo exercise using simulated data. In general, the MS-SSVS model performs best in data-sets that have a relatively small amount of noise. In these data sets, it is able to detect zero-restrictions, "identical" restrictions, and switching in the coefficients with high probability. The MS-SSVS is still able to identify these restrictions as the amount of noise grows, and it is able to detect zero-restrictions with a surprisingly high degree of accuracy in even the noisiest data sets that I generated.

When I apply this model to Federal Funds rate data I find three major things. First, there is relatively little evidence that there have been economically significant shifts in inflation response over the period 1970-2007. Second, there is substantially more evidence that there has been a shift in the unemployment gap coefficient, between strong and weak responses to the unemployment rate. I find that the periods most likely to have had a weak response are the early to late 1980s, and roughly 1995-2004. The first period corresponds to the chairmanship of Paul Volcker, suggesting that the FOMC focused relatively less on responding to changes in output or unemployment under his leadership. The second period corresponds to the middle of Greenspan's tenure as chairman, however both the beginning and end of his leadership are characterized by a strong unemployment response. As the estimated volatility of the interest rate rule declines after 1980, the distinction between regimes grows. Finally, I find strong evidence that there have been changes in the volatility of interest rate rule. This adds to a relatively strong body of existing evidence, as Sims and Zha (2006), Check (2016), and Murray et al. (2015) all find that models that allow for a change in variance out-perform models with constant variance.

These findings add to the growing body of literature that the FOMC has not drastically changed policy over the past 45 years. After allowing for switches in mean, persistence, inflation response, unemployment gap response, and the response to the change in the unemployment rate, I find evidence that only the unemployment gap response has changed over time. One potential explanation for this is that the staff at the FOMC is well respected and holds great influence in policy-making decisions. Additionally, because the FOMC makes decisions by committee, any change in chairman may have a limited influence on policy. While the chairs of the FOMC may have strong personal beliefs about how to best respond to changes in the economy, their actions can be fairly well characterized by an interest rate rule that has changed only slightly over time.

References

Boivin, J. (2006). Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data. *Journal of Money, Credit and Banking*, 38(5):1149–1173.

Check, A. (2016). Interest rate rules in practice - the taylor rule or a tailor-made rule? Working paper, University of Oregon.

Clarida, R., Galí, J., and Gertler, M. (2000). Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory. *The Quarterly Journal of Economics*, 115(1):147–180.

Früwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

George, E., Sun, D., and Ni, S. (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics*, 142:553–580.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–84.

Kahn, G. A. (2010). Taylor rule deviations and financial imbalances. *Economic Review*, (Q II):63–99.

Kim, C.-J. and Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, volume 1 of *MIT Press Books*. The MIT Press.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–93.

Koop, G. and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358.

Murray, C. J., Nikolsko-Rzhevskyy, A., and Papell, D. H. (2015). Markov Switching And The Taylor Principle. *Macroeconomic Dynamics*, 19(04):913–930.

Orphanides, A. (2004). Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches. *Journal of Money, Credit and Banking*, 36(2):151–75.

Owyang, M. T., Piger, J., and Wall, H. J. (2015). Forecasting National Recessions Using State?Level Data. *Journal of Money, Credit and Banking*, 47(5):847–866.

Sims, C. A. and Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96(1):54–81.

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

Taylor, J. B. (2013). A Review of Recent Monetary Policy. Economics Working Papers 13103, Hoover Institution, Stanford University.