REGIME SWITCHING AND THE MONETARY ECONOMY

by

ADAM J. CHECK

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2016

DISSERTATION APPROVAL PAGE

Student: Adam J. Check

Title: Regime Switching and the Monetary Economy

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Jeremy Piger	Chair
Bruce McGough	Core Member
George Evans	Core Member
Craig Rolling	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2016

DISSERTATION ABSTRACT

Adam J. Check

Doctor of Philosophy

Department of Economics

June 2016

Title: Regime Switching and the Monetary Economy

For the empirical macroeconomist, accounting for nonlinearities in data series by using regime switching techniques has a long history. Over the past 25 years, there have been tremendous advances in both the estimation of regime switching and the incorporation of regime switching into macroeconomic models. In this dissertation, I apply techniques from this literature to study two topics that are of particular relevance to the conduct of monetary policy: asset bubbles and the Federal Reserve's policy reaction function.

My first chapter utilizes a recently developed Markov-Switching model in order to test for asset bubbles in simulated data. I find that this flexible model is able to detect asset bubbles in about 75% of simulations. In my second and third chapters, I focus on the Federal Reserve's policy reaction function. My second chapter advances the literature in two important directions. First, it uses meeting-based timing to more properly account for the target Federal Funds rate; second, it allows for the inclusion of up to 14 economic variables. I find that the long-run inflation response coefficient is larger than had been found in previous studies, and that increasing the number of economic variables that can enter the model improves both in-sample fit and out-of-sample forecasting ability. In my third chapter, I introduce a new econometric model that allows for Markov-Switching, but can also remove variables from the model, or enforce a restriction that there is no regime switching. My findings indicate that the majority of coefficients in the Federal Reserve's policy reaction function have not changed over time.

CURRICULUM VITAE

NAME OF AUTHOR: Adam J. Check

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Temple University, Philadelphia, PA

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2016, University of Oregon
Master of Science, Economics, 2012, University of Oregon
Bachelor of Business Administration, Economics, 2010, Temple University

AREAS OF SPECIAL INTEREST:

Bayesian Econometrics
Monetary Economics

GRANTS, AWARDS, AND HONORS:

Kleinsorge Summer Research Award, University of Oregon, 2015

Graduate Teaching Fellow Outstanding Teaching Award, 2015

Economics Award, Temple University 2010

ACKNOWLEDGEMENTS

I would like to thank all those who have had a role in shaping the course of my academic life. I could not have made it this far without the support of all of my teachers, instructors, and professors at every level of education. I would especially like to thank Ms. Britt, who fostered my love of reading; Mr. Whelan, who vastly improved my writing; Professor Diamantaras, who gave me the confidence that I could complete a Ph.D. and encouraged me to do so; and Professor Piger for the countless hours of conversation, advice, and support he gave me during this process. I would also like to thank my peers, especially Sacha Gelfer, Chris Gibbs, Tyler Schipper, and Jeff Allen, for being willing to listen to my ideas and being honest enough to tell me which were worthwhile. Finally, I would like to thank Georgina Santos for her constant encouragement and editing assistance.

For my Mom, who sacrificed so much in order for me to have this opportunity.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure | Page

LIST OF TABLES

CHAPTER I

A NEW TEST FOR ASSET BUBBLES

Introduction

Policymakers at the Federal Reserve believe that it is vital to determine whether a price bubble exists in important asset markets.[1] While the appropriate steps to take after recognizing the existence of the bubble are debatable, in order to take any action it would first be necessary to know that the bubble existed. However, as the transcript of the Federal Open Market Committee meeting from June 29 to June 30, 2005 indicates, even during the peak of the massive U.S. housing bubble, policymakers were in disagreement over whether the housing market was in a bubble. This disagreement highlights an important and surprising deficiency in the asset bubble literature - the lack of existence of a powerful and broadly agreed upon test for asset bubbles.

Due to this deficiency, I propose a new test that generalizes the Markov-Switching test for explosive roots in Hall et al. (1999). Specifically, I allow the parameters of Hall et al.'s (1999) model to vary over time. In principle, this should allow the detection of multiple bubbles in the same sample, even if the growth rates of the bubbles differ. At the same time, I introduce Bayesian estimation of this model and use Bayesian model comparison to decide between competing models, rather than using the classical confidence interval-based inference used in Hall et al. (1999). This Bayesian perspective allows me to easily test jointly for both switching

[1]This is evident from a speech given on February 7, 2013 by Federal Reserve Governor Jeremy Stein entitled "Overheating in Credit Markets: Origins, Measurement, and Policy Responses". A transcript of this speech available at: http://www.federalreserve.gov/newsevents/speech/stein20130207a.pdf

in the price dynamics and an explosive root, whereas Hall et al. (1999) tested only for the presence of an explosive root.

To investigate the power of my proposed test, I use artificially generated price series that contain periodically collapsing bubbles. I first estimate the constant parameter Hall et al. (1999) model using Bayesian methods and altering Hall et al.'s (1999) testing procedure slightly to jointly test for both an explosive root and Markov-Switching. Next, I estimate my proposed time-varying parameter generalization. I find that for a test with 5% size, Bayesian estimation and testing of both the constant parameter Hall et al. (1999) model and the more general time-varying parameter model are able to detect these periodically collapsing bubbles nearly 80% of the time.

The idea to econometrically test for bubbles in asset markets has been around for decades, originating shortly after the variance bound tests that were proposed contemporaneously by Shiller (1981) and LeRoy and Porter (1981). These tests attempt to determine whether the observed variance of actual asset prices exceeds the variance bound implied by the frequently used risk-neutral asset pricing equation. Tirole (1985) and Blanchard and Watson (1982) suggest that these variance bounds tests be used to detect bubbles, but Flood et al. (1994) eventually showed that variance bounds tests were actually very poorly suited to test for bubbles, since in the presence of a bubble, the variance may not exceed the bound implied by the test.

Another bubble testing procedure was put forth by West (1987), who suggests the use of a two-step test which estimates the relationship between dividends and stock prices both directly and indirectly. If the estimated relationship differs between the two methods, and the researcher is reasonably certain that they have specified the model correctly, then there exists evidence in favor of a bubble. The major problem with this test is that under different model specifications, researchers have come to different conclusions about the existence of bubbles in U.S. stock prices.

Yet another take on bubble testing was put forward by Diba and Grossman (1988), who use the integration/cointegration properties of dividends and prices to test whether prices take on the integration properties of the dividend process. If they do then there is no bubble, since as suggested by theory, the dynamic properties of the price series depends only on the process followed by the dividends. However, if the price series displays integration patterns that are not shared with the dividend process, then this would suggest that something else is also driving asset prices. If we are sure that dividends are the only relevant fundamental, then we would conclude that there is a bubble in asset prices.

More specifically, Diba and Grossman (1988) use an Augmented Dickey-Fuller (ADF) test in order to test both first differenced prices and first differenced dividends for a unit root. If first differenced prices display a unit root, but first differenced dividends do not, then the price series is consistent with a bubble. Using this test on generated price series data using a very simple bubble process that grows exponentially in every period, they find that their test can detect a bubble 95% of the time.

However, Evans (1991) points out that although the Diba and Grossman (1988) test is appealing and works well on data generated with a relatively simple bubble process, when faced with bubbles that grow at different rates in different time periods it loses almost all of its power, and detects only a handful of bubbles. In order to build on Diba and Grossman's (1988) intuitive and appealing idea for a bubble test, Hall et al. (1999) generalize the Diba and Grossman (1988) test. First, they test for an explosive root in the levels of prices and dividends, rather than testing for a unit root in the first differences. Next, they allow the parameters of this test to alternate between two regimes according to a Markov-Switching process. This allows the test to capture the fact that the bubble is growing much faster in some periods than in others, and will allow at least a subset of periods to be consistent with a bubble. In practice, Hall et al. (1999) find that their test is capable of detecting the Evans (1991) style bubbles about 60% of the time. While this is great improvement over the Diba and Grossman (1988) test, to the dismay of policymakers who wish to determine whether a particular asset is in a bubble, it will still miss the presence of a bubble nearly 40% of the time.

It is through this lens that I introduce a generalization to Hall et al.'s (1999) test for asset bubbles. First, I bring a Bayesian perspective to the test. Second, I allow for the growth rates of bubbles to change upon each episode of bubble, so that the test does not restrict all bubbles in the sample to have the same growth rate. Third, I test jointly for Markov-Switching and an explosive root in the price series, while Hall et al. (1999) assume Markov-Switching under the null, and test only for an explosive root. I find that jointly testing proves important for the detection of bubbles that follow Evans' (1991) bubble generation procedure, as it improves the detection rate to nearly 80%. Allowing for differing growth rates does not meaningfully alter the ability of the test to detect for bubbles in this environment. This may occur because all bubbles have the same expected growth rate under this process.

Rational Bubbles, Testing, and Periodically Collapsing Bubbles

Rational Bubbles

Historically, many bubble tests are designed to detect "rational" bubbles. These include the early variance bounds tests developed by Shiller (1981) and LeRoy and Porter (1981), and implemented by Cochrane (1992); the two-step tests developed in West (1987); and the integration/cointegration test developed in Diba and Grossman (1988).[2] However, before we can properly discuss bubble testing, we must specify what we mean by a "rational" bubble.

Rational bubbles are periods during which agents are willing to pay more for an asset than the asset's fundamental value, which is the value implied by the present value of the expected future dividend stream. The reason that these agents are willing pay a premium during a rational bubble is that they anticipate being able to sell the asset for more than the fundamental value at a later date. This bubble is rational in the sense that if everyone shares this belief, then the asset is priced correctly despite the fact that it trades for more than its fundamental value.

We can mathematically formalize this intuition in a simple asset pricing framework. First, assume that there is an infinitely-lived representative agent, who seeks to maximize expected lifetime utility in an endowment economy:

$$\max_{c_t} E_t \left[\sum_{i=0}^{\infty} \beta^i u(c_{t+i}) \right]$$

$$\text{s.t. } c_{t+i} = y_{t+i} + (p_{t+i} + d_{t+i}) x_{t+i-1} - p_{t+i} x_{t+i}$$

[2]See Gürkaynak (2008) for more detail.

Where y_t is the income of the agent at time t, x_t is the number of units of the asset held by the agent in period t, p_t is the price of the asset in period t, d_t is the dividend paid to those holding the asset at the beginning of period t, and $0 < \beta < 1$ is the discount rate of the representative agent.

We can derive the Euler equation using a variational argument. Using the consumption good, c_t, as the numeraire, we consider the gains and losses from giving up a unit of consumption in order to buy the asset today, and we define the net one-period rate of return from holding an asset as:

$$r_{t+1} = \frac{p_{t+1} + d_{t+1}}{p_t} - 1$$

Then we can see that:

$$u'(c_t) = E_t \beta (1 + r_{t+1}) u'(c_{t+1})$$

In words, if we give up a unit of consumption today, then we can use the proceeds to buy the asset. However, we lose the marginal utility that the unit of consumption would give us today. Tomorrow we will be able to consume $(1 + r_{t+1})$ units of the consumption good, since we have earned the one-period return given by the asset. Therefore, tomorrow we will gain the marginal utility of consuming $(1 + r_{t+1})$ units of the consumption good. Taking into account the fact that we will not get this utility until tomorrow and that the return is risky, we average across all possible returns by using the expectations operator and discount the future utility using the discount factor β.

Using our definition of the net return, r_{t+1}, we can substitute and solve directly for the price of the asset today:

$$p_t = \beta E_t \left\{ (p_{t+1} + d_{t+1}) \frac{u'(c_{t+1})}{u'(c_t)} \right\}$$

Assuming risk-neutral preferences, we have $u'(c_t) = E_t[u'(c_{t+1})] = k \ \forall \ t$, where k is a constant. We can then rewrite the above equation as:

$$p_t = \beta E_t(p_{t+1} + d_{t+1}) \tag{1.1}$$

We can find one solution to this first-order difference equation by using the law of iterated expectations and iterating on the above equation:

$$p_t = \beta^s E_t p_{t+s} + \sum_{i=1}^{s} \beta^i E_t d_{t+i}$$

Taking the limit as $s \to \infty$, we can see that as long as $\beta^s E_t p_{t+s} \to 0$, we have the solution:

$$p_t = \sum_{i=1}^{\infty} \beta^i E_t d_{t+i} = F_t$$

This solution is called the *fundamental* solution, since the price today relies only on the expected value of the future dividend stream. This solution is ensured by imposing a transversality condition on the value of the agent's savings.

However, there is another solution for the first order difference equation in equation (1.1):

$$p_t = F_t + B_t \tag{1.2}$$

where B_t, the *bubble* component of the solution, is any random variable that satisfies:

$$B_t = \beta E_t B_{t+1} \tag{1.3}$$

We verify that this is a solution to equation (1.1) by a direct proof, given below. First, assume that equations (1.2) and (1.3) constitute a solution to (1.1). Then we have:

$$\begin{aligned}
p_t &= \beta E_t(p_{t+1} + d_{t+1}) \\
F_t + B_t &= \beta E_t(F_{t+1} + B_{t+1} + d_{t+1}) \\
\sum_{i=1}^{\infty} \beta^i E_t d_{t+i} + \beta E_t B_{t+1} &= \beta \left(\sum_{i=2}^{\infty} \beta^i E_t d_{t+i} + \beta E_t B_{t+2} + E_t d_{t+1} \right) \\
\sum_{i=1}^{\infty} \beta^i E_t d_{t+i} + \beta E_t B_{t+1} &= \left(\sum_{i=1}^{\infty} \beta^i E_t d_{t+i} + \beta^2 E_t B_{t+2} \right) \\
\beta E_t B_{t+1} &= \beta^2 E_t B_{t+2} \\
E_t B_{t+1} &= \beta E_t B_{t+2} \qquad (1.4)
\end{aligned}$$

Iterating equation (1.3) forward one period, and using the law of iterated expectations, we have:

$$B_{t+1} = \beta E_{t+1} B_{t+2}$$
$$E_t B_{t+1} = \beta E_t B_{t+2}$$

which shows that the equality in equation (1.4) always holds, and verifies that equation (1.2) nests an entire class of solutions, so long as equation (1.3) also holds. ■

As mentioned previously, note that bubbles of this type are usually ruled out by imposing a transversality condition. In fact, Tirole (1982) argues that bubbles can *always* be ruled out in infinitely lived rational expectations models. However, it is common practice in the bubble testing literature to abstract away from this theoretical argument, and work from the assumption that equations (1.2) and (1.3) define the asset price.[3] This approach is justified by the fact that Tirole (1985) shows that rational bubbles can exist in overlapping generations models. In practice, Kindleberger (2000) finds numerous examples of asset bubbles throughout modern history. As pointed out by Evans (1991), if these observed bubbles are not "rational", then a desirable feature of a bubble test would be that it has power against many different bubble specifications.

Imposing Structure on the General Solution

In order to devise a test for rational asset bubbles, it is necessary to put more structure on the economy described above. It is particularly helpful to define

[3] See Gürkaynak (2008)

dynamic equations for both the dividend and the bubble component. For the dividend component, the typical assumption is that d_t is integrated of order one, i.e. it is $I(1)$.[4] Specifically, d_t is usually assumed to follow a random walk with drift:[5]

$$d_t = \mu + d_{t-1} + \varepsilon_t \tag{1.5}$$

Then there are two possible scenarios: the asset price does not contain a bubble component, B_t, or the asset price does contain a bubble component.

In the absence of a bubble component, B_t, we have $p_t = F_t$, and it can be shown that p_t is also $I(1)$, and that p_t and d_t are cointegrated. Furthermore, we have:

$$E_t F_{t+i} = \frac{\beta}{(1-\beta)}(d_t + \mu i) + \frac{\beta}{1-\beta}\mu$$

which becomes dominated by $\frac{\beta}{(1-\beta)}\mu i$ as i gets large.[6] This implies that the forecast of the fundamental value grows linearly over time, increasing by $\frac{\beta}{(1-\beta)}\mu$ each period, and reflects the unit root in the process for d_t.

In addition, rearranging equation (1.3) and using the law of iterated expectations, we can show that the time t expectation of the bubble component

[4] In a model with dynamic growth, the assumption is instead that $\ln(d_t)$ is $I(1)$. However, we will usually be working in the relatively simpler set-up outlined in the text.

[5] We use a random walk for expositional purposes, but the analysis remains the same for any stationary ARMA process for Δd_t. See Evans (1991) for more details.

[6] In general, from Beveridge and Nelson (1981), we know that if d_t follows a stationary ARMA process, then as j tends to infinity, $E_t F_{t+j} \to C_t + jE(\Delta F_t)$ for some C_t. See also Evans (1991).

at time $t+i$ is given by:

$$E_t B_{t+i} = \frac{1}{\beta^i} B_t$$

Therefore, as pointed out in Evans (1991), the conditional expectation of the bubble component grows at rate $\frac{1}{\beta} > 1$. Combining the two components, we can see that as i gets large we have:

$$E_t p_{t+i} \to \frac{\beta}{(1-\beta)} \mu i + \frac{1}{\beta^i} B_t$$

Provided $B_t > 0$, eventually the exponential growth of the bubble component will overwhelm the linear growth of the fundamental component, and the forecast of the price will explode to infinity at the rate of the growth of the bubble component, $\frac{1}{\beta}$.

With the additional structure we have put on the asset pricing model, we have a testable hypothesis. If the price of the asset grows faster than the underlying fundamentals grow, then there is a bubble in the asset price.[7]

Diba and Grossman (1988) Test

Diba and Grossman's (1988) idea is to exploit the specification of a rational bubble, noting that if the price of the asset contained a bubble component, the asset price would grow at a rate faster than suggested by the growth rate of the fundamental process. Furthermore, Diba and Grossman (1988) provide the additional insight that if the fundamental process, d_t is $I(1)$, then the fundamental is stationary in first differences. Therefore, if there is no bubble component, then the first difference of the asset price, Δp_t, would also be stationary.

[7] Of course, this assumes that all fundamentals are observable.

However, in the presence of an exponentially growing bubble component, differencing prices any finite number of times will not yield a stationary process for Δp_t. Additionally, in the presence of a bubble component, the fundamental and the asset price would not be cointegrated. Therefore, Diba and Grossman (1988) propose the following test:[8]

1. Test p_t and d_t for stationarity.
2. If both p_t and d_t are non-stationary test p_t and d_t for cointegration.
3. If they are cointegrated, conclude that there is no bubble.
4. If they are not, then conclude that we cannot rule out the existence of a bubble.

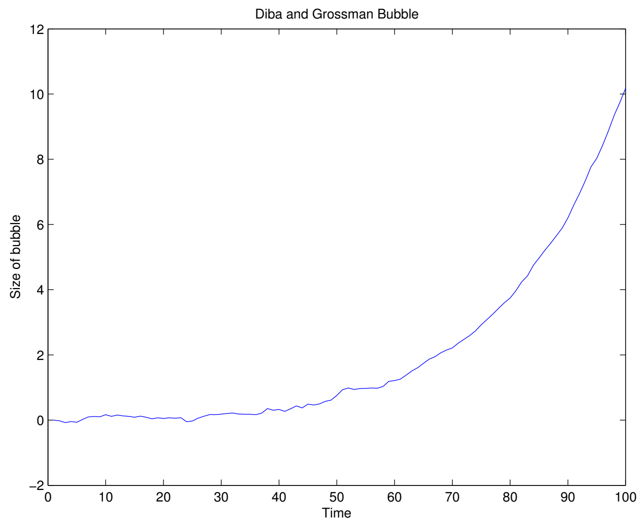
In fact, if we were certain that we had included all relevant measures of fundamentals in d_t, then the lack of stationarity in the price series or the lack of cointegration between the price series and the fundamental series would indicate the presence of a bubble.

Diba and Grossman (1988) use this test on 100 simulated series, each lasting 100 periods and containing an explosive bubble component that evolves according to $B_{t+1} = (1+r)B_t + z_{t+1}$, where $r = 0.05$ and $z_{t+1} \sim$ iid $N(0, \sigma_z)$. They find that their test has high power to detect a bubble, as 95% of their simulated price series are nonstationary in first differences. However, as pointed out in Evans (1991), the process assumed for the evolution of the bubble may be overly simplistic, since it assumes that a bubble will grow at an exponential rate forever. In Figure 1, I show

[8]They also propose other, similar tests, such as testing the first differences of both series for stationarity. All of their proposed tests have approximately the same power, and all were shown in Evans (1991) to have low power against periodically collapsing bubbles.

what a "typical" Diba & Grossman bubble looks like by generating 201 of these bubbles and plotting the bubble with the median variance.

FIGURE 1. Typical Diba and Grossman Bubble



Evans (1991) Bubble Process

Evans (1991) astutely observes that in the real world, bubbles could not possibly have the form hypothesized by Diba and Grossman (1988). That is, no one thinks that an asset bubble could grow unabated forever. In fact, all of the dozens of examples of historical bubbles cited in Kindleberger (2000) eventually collapsed. Therefore, Evans proposes a process for bubbles that allows them to periodically collapse, and shows that the tests suggested by Diba and Grossman (1988) have very little power to detect this more realistic type of bubble.

Evans (1991) retains the same risk neutral asset market set-up considered in Diba and Grossman (1988), and assumes that dividends evolve according to equation (1.5). However, he proposes the following formulation for the bubble component:

$$B_{t+1} = \begin{cases} (1+r)B_t u_{t+1} & \text{if } B_t \leq \alpha \\ \left[\delta + \frac{1+r}{\pi}\left(B_t - \frac{\delta}{1+r}\right)\xi_{t+1}\right] u_{t+1} & \text{if } B_t > \alpha \end{cases}$$

Where δ and α are scalars that satisfy $0 < \delta < (1+r)\alpha$, u_t is a sequence of i.i.d. random variables with $E_t u_{t+1} = 1$:

$$u_t = \exp\left(z_t - \frac{\sigma_z^2}{2}\right)$$

$$z_t \sim N(0, \sigma_z^2)$$

and ξ_t is an exogenous i.i.d. Bernoulli process such that:

$$Pr(\xi_t = 0) = 1 - \pi$$

$$Pr(\xi_t = 1) = \pi$$

In words, the bubble process follows a linear switching process. Recall from equation (1.3) that a rational bubble component must satisfy:

$$B_t = \beta E_t B_{t+1}$$

Therefore, we verify that the above process satisfies this requirement. When $B_t \leq \alpha$, we have:

$$
\begin{aligned}
E_t B_{t+1} &= E_t(1+r)B_t u_{t+1} \\
E_t B_{t+1} &= (1+r)B_t E_t u_{t+1} \\
E_t B_{t+1} &= (1+r)B_t \\
B_t &= \beta E_t B_{t+1}
\end{aligned}
$$

where we have used the fact that in equilibrium, $\frac{1}{\beta} = (1+r)$.

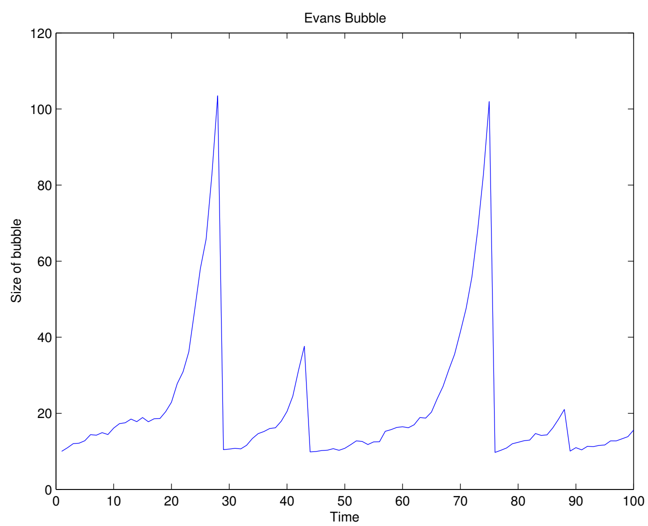
If $B_t > \alpha$, we have:

$$
\begin{aligned}
E_t B_{t+1} &= E_t \left\{ \left[\delta + \frac{1+r}{\pi} \left(B_t - \frac{\delta}{1+r} \right) \xi_{t+1} \right] u_{t+1} \right\} \\
E_t B_{t+1} &= E_t \left\{ \delta u_{t+1} + \frac{1+r}{\pi} \left(B_t - \frac{\delta}{1+r} \right) \xi_{t+1} u_{t+1} \right\} \\
E_t B_{t+1} &= \delta E_t u_{t+1} + \frac{1+r}{\pi} \left(B_t - \frac{\delta}{1+r} \right) E_t \xi_{t+1} u_{t+1} \\
E_t B_{t+1} &= \delta + \frac{1+r}{\pi} \left(B_t - \frac{\delta}{1+r} \right) \pi \\
E_t B_{t+1} &= \delta + (1+r) \left(B_t - \frac{\delta}{1+r} \right) \\
E_t B_{t+1} &= \delta + (1+r)B_t - \delta \\
E_t B_{t+1} &= (1+r)B_t \\
B_t &= \beta E_t B_{t+1}
\end{aligned}
$$

where, in order to go from line 3 to line 4 we have used the fact that ξ_{t+1} and u_{t+1} are each i.i.d. random variables, with $E(\xi_{t+1}) = \pi$ and $E(u_{t+1}) = 1$.

Now that we know that this bubble process conforms to the requirement for a rational bubble, we can analyze some of its properties. If $B_t \leq \alpha$, then the bubble grows at rate $\frac{1}{\beta}$. However, once the size of the bubble exceeds the predetermined level α, then it will grow at the faster rate, $\frac{B_t}{\beta\pi}$ if $\xi_{t+1} = 1$. However, if $\xi_{t+1} = 0$, then the bubble collapses to δu_{t+1}. Therefore, $1 - \pi$ is the probability of the bubble collapsing each period. Once the bubble collapses, it will return to growing at the slower rate, $\frac{1}{\beta}$, until it eventually exceeds the exogenously given scalar α. In Figure 2, I show what a "typical" Evans bubble looks like by generating 201 of these bubbles and plotting the one with the median variance.

FIGURE 2. Typical Evans Bubble



Evans generates 200 price series, using the same fundamentals process as Diba and Grossman (1988), and the periodically collapsing bubbles for the bubble component. Evans then tests these series for the presence of a bubble by using Diba and Grossman's (1988) proposed unit root and cointegration tests.[9] He finds that these tests perform extremely poorly, and detect only a handful of bubbles in these generated price series.[10] Since these collapsing bubbles are a much

[9]See the appendix for full detail and the particular parameter values chosen in Evans (1991).

[10]It is hard to tell from the table presented in Evans (1991). However, it appears that at most three of the 200 bubbles were detected.

more plausible bubble generating process than the process set forth in Diba and Grossman (1988), Evans (1991) concludes that the Diba and Grossman (1988) test for asset bubbles is insufficient, and that work should be done to devise a more powerful and flexible bubble test.

Hall et al.'s (1999) Test for bubbles

Hall et al. (1999) attempt to design a bubble test that is better able to detect the presence of periodically collapsing bubbles than the tests presented in Diba and Grossman (1988). To do so, they return to the Augmented Dickey Fuller (ADF) test that was used in Diba and Grossman (1988), with two main differences. In Diba and Grossman (1988), this ADF test was used to test for the presence of a unit root in p_t or Δp_t. In Hall et al. (1999), the authors instead modify this test to test for an *explosive* root in p_t. Furthermore, Hall et al. (1999) allow the parameters of the test to switch between two regimes: a low growth regime and a high growth regime.

The equation that Hall et al. (1999) estimate to conduct their test for an explosive root in the level of the price series, p_t, is given below:

$$\Delta p_t = \mu_0(1 - S_t) + \mu_1 S_t + [\phi_0(1 - S_t) + \phi_1 S_t]p_{t-1} + \sum_{j=1}^{k}[\psi_{0,j}(1 - S_t) + \psi_{1,j} S_t]\Delta p_{t-j} + \sigma_e e_t$$

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

Here, S_t is an indicator variable, indicating whether we are in a low growth regime ($S_t = 1$), or a high growth regime (S_t=0). If these regimes were directly observable, then we could estimate the above equation by treating S_t as a dummy variable. An alternate way to write this process is as a piecewise function:

$$\Delta p_t = \begin{cases} \mu_0 + \phi_0 p_{t-1} + \sum_{j=1}^{k} \psi_{0,j} \Delta p_{t-j} + \sigma_e e_t & \text{if } S_t = 0 \\ \mu_1 + \phi_1 p_{t-1} + \sum_{j=1}^{k} \psi_{1,j} \Delta p_{t-j} + \sigma_e e_t & \text{if } S_t = 1 \end{cases}$$

Here, p_t is the price series of interest,[11] ϕ_i is the AR(1) parameter determining the impact of p_{t-1} on p_t. $\psi_{i,j}$ is the coefficient on the j^{th} lag of the price series, for $j \geq 2$, which determines how p_{t-j} impacts p_t.

Finally, the interpretation of μ_i depends on the estimation of ϕ_i. If $\phi_i < 1$, $\frac{\mu_i}{1-\phi_i}$ is the mean of the price series in regime i. If $\phi_i = 1$, μ_i is the drift (i.e. time trend) of the random walk process for the price series. If $\phi_i > 1$, then μ_i partially determines whether the explosive price process is exploding to negative or positive infinity.

A problem with the procedure outlined above is that the researcher will not know which periods constitute a high return or low growth regime, so the regime dummy variable, S_t is unobserved. Therefore, Hall et al. (1999) estimate these regimes using a Markov-Switching model. In this model, the researcher assumes that the probability of moving from regime i in period $t-1$ to regime j in period t depends only on what regime the price process was in in period $t-1$. Since there are two regimes in the model, then there are four possible transitions that occur, with probabilities given by p_{ij} for $i, j \in \{0, 1\}$, where p_{ij} is the probability of switching from regime i in period $t-1$ to regime j in period t.

The testing procedure is as follows:

1. Estimate the Markov-Switching model via Maximum Likelihood estimation.

[11] In applied empirical work (looking at S&P data, for instance), this will actually be the log of the price series. In my simple price generation equations, I can simply use the level of prices.

2. Use bootstrapping to find a one-sided 95% confidence interval against which to test the presence of an explosive root in the high growth regime.

3. If this test confirms that the price series, p_t, has an explosive root, but the fundamental series does not, then the price series is consistent with the presence of a bubble.

4. However, if the price series does not have an explosive root, or if both the price series and its corresponding fundamental have an explosive root, then conclude that there is not a bubble.

This generalization of the ADF test allows Hall et al. (1999) to more accurately detect the presence of bubbles. Using the equations above, and performing maximum likelihood estimation with bootstrapped errors, Hall et al. (1999) find that the high growth regime has an explosive root over 75% of the time. However, because the existence of a bubble is only confirmed when the price has an explosive root but dividends do not, Hall et al. (1999) are only able to classify about 60% of their price series as containing a bubble. Compared to the results in Evans (1991), which detected about three bubbles out of 200, this is a great improvement. However, the results may be disappointing to policymakers, as this test still misses about 40% of these bubbles.

Since Hall et al. (1999), there have been more attempts to improve bubble testing. Some recently developed tests, such as Phillips et al. (2011), use recursive tests to test for bubbles and remain agnostic about the structural form of the regime. Using Phillips et al.'s (2011) estimation technique has two desirable features. First, Phillips and Magdalinos (2007) have worked out limit theory for mildly explosive processes, and Phillips et al. (2011) are able to apply that theory to their bubble testing procedure to test directly for explosive roots without the need for bootstrapping, a computationally intensive procedure that Hall et al. (1999) needed to undertake to estimate the distribution of the AR(1) parameter under an explosive processes. Second, it allows the researcher to date-stamp the beginning and ending dates of bubbles, although this can also be done in the Markov-Switching framework by using the estimated regime probabilities. However, under the feasible parameters used by Evans (1991) and Hall et al. (1999), Phillips et al.'s (2011) test detects the presence of a bubble only 43.2% of the time.[12]

MS-TVP Model

Due to the relatively low power of existing bubble tests, there remains a deficiency in the bubble testing literature. In order to try to increase the power of existing bubble tests, I generalize Hall et al.'s (1999) test by allowing the AR parameters in their MS model to evolve according to a random walk each time the price series enters a high return or low growth regime. My test uses Bayesian inference, and therefore admits the use of hierarchical priors, as suggested by Koop and Potter (2007), which makes estimation of this test feasible via Gibbs sampling.

To see where the time-varying nature of my test may be particularly helpful, consider the following example. Suppose that there are three different bubbles in our sample, which is a reasonable number for both Evans' (1991) generated price series and for the entire history of S&P 500 stock data. Assume that two of the three bubbles are very large, with very high growth rates, while the third is relatively small, with a slower growth rate. Then the fixed parameter Hall et al. (1999) test may find the two large bubbles, and estimate a large value for ϕ_0, the

[12]This test performs much better for slightly different parameterizations of Evans (1991) bubble generation process that could still be considered realistic.

AR(1) coefficient in the high growth regime. Since the estimated ϕ_0 is so large, the Hall et al. (1999) test may not detect the third bubble, since it will be more qualitatively more similar to the slow growth regime than the bubble regime. However, the new model with time varying parameters should be able to detect this third bubble. Since the AR(1) coefficient in the high growth regime, $\phi_{0,t}$ can change over time, it will be higher during the two large bubbles, and lower, but still explosive, during the third bubble. With this intuition in mind, I present my generalization of the Hall et al. (1999) test below.

Eo & Kim's Model Applied to the ADF Test

In Eo and Kim (2012), the authors present a newly developed MS-TVP model in the context of GDP growth and identification of recessions. Eo and Kim (2012) were trying to overcome a similar problem to the example presented above - that Hamilton's (1989) Markov-Switching model with constant coefficients did a poor job at identifying relatively mild recessions, like the one in 2001. While Eo and Kim (2012) only consider a model that has no autoregressive components, so that it is only a mean switching model, it is easy to generalize their model to one with lags. The MS-TVP Augmented Dickey fuller test can be written as:

$$\Delta p_t = \mu_{0,\tau}(1 - S_t) + \mu_{1,\tau} S_t + [\phi_{0,\tau}(1 - S_t) + \phi_{1,\tau} S_t] p_{t-1} + \sum_{j=1}^{k} [\psi_{0,\tau,j}(1 - S_t) + \psi_{1,\tau,j} S_t] \Delta p_{t-j} + \sigma_e e_t$$

Let $\beta_{0,\tau} = [\mu_{0,\tau}\ \phi_{0,\tau}\ \psi_{0,\tau,1}\ \ldots\ \psi_{0,\tau,k}]'$ and $\beta_{1,\tau} = [\mu_{1,\tau}\ \phi_{1,\tau}\ \psi_{1,\tau,1}\ \ldots\ \psi_{1,\tau,k}]'$ be the vectors of time varying parameters in regime 0 and regime 1 in a model with

$k + 1$ lags. Then the parameters transition according to:

$$\begin{bmatrix} \beta_{0,\tau} \\ \beta_{1,\tau} \end{bmatrix} = \begin{bmatrix} \beta_{0,\tau-1} \\ \beta_{1,\tau-1} \end{bmatrix} + \begin{bmatrix} \omega_{\beta_0,\tau} \\ \omega_{\beta_1,\tau} \end{bmatrix}$$

where $\tau = 1, 2, \cdots, N_0 + N_1$, and the $\omega_{x,\tau}$ are white noise shocks particular to each parameter. In words, τ is the number of the particular realization of the regime. For example, if $\tau = 1$ is the first realization of the high growth regime, then $\tau = 2$ is the first realization of the low growth regime, $\tau = 3$ is the second realization of the high growth regime, $\tau = 4$ is the second realization of the low growth regime, etc. Therefore N_0 is the number of times the price series has been in the high growth regime, N_1 is the number of times the price series has been in the low growth regime, and $N_0 + N_1$ is the total number of realizations of all regimes.

For estimation, it will be helpful to collect all parameters in a vector and rewrite the system in terms of time t instead of τ. First, let $\beta_t = [\beta'_{0,t}\ \beta'_{1,t}]'$ be the vector of all estimated coefficients at time t. Then the system can be written in state space form. First the Measurement Equation (ME):

$$\Delta p_t = H_t \beta_t + e_t$$

$$H_t = [(1 - S_t)p_{t-1}\ S_t p_{t-1}\ (1 - S_t)\Delta p_{t-1} \ldots S_t p_{t-k}]$$

$$e_t \sim NID(0, \sigma_e^2)$$

Likewise, we can write the State Equation (SE):

$$\beta_t = F\beta_{t-1} + \omega_t$$

$$F = I_{\text{length}(\beta)}$$

$$\omega_t \sim MVN\left(\begin{bmatrix}0\\0\\\vdots\\\vdots\\0\\0\end{bmatrix}, \begin{bmatrix} d_{10,t}\sigma^2_{\mu_0} & 0 & 0 & 0 & \dots & 0 \\ 0 & d_{01,t}\sigma^2_{\mu_1} & 0 & 0 & \dots & 0 \\ 0 & 0 & d_{10,t}\sigma^2_{\phi_0} & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & d_{01,t}\sigma^2_{\psi_{1,k}} \end{bmatrix}\right)$$

Here, $d_{ij,t}$ is a dummy variable which equals one when $t - 1 = i$ and $t = j$. Therefore, the shocks, ω_t, are heteroscedastic - a shock to the regime i parameters only occurs when the price series enters an episode of regime i, and equals 0 otherwise. However, because the disturbance term in the ME is conditionally Gaussian, we are still able to use the Kalman Filter to estimate the parameters of this state space model.

Estimation

I use Markov Chain Monte Carlo (MCMC) Bayesian estimation techniques to estimate both Hall et al.'s (1999) model and the time-varying generalization outlined above. In both estimations, I use a Normal prior on the regression coefficients, a Gamma prior on the inverse of the variance parameter, and a Beta prior on the regime transition probabilities. These priors are conditionally conjugate, so they admit the use of the Gibbs sampler.

As stated above, in my estimation I use a Normal prior on all of the regression coefficients, including the AR(1) coefficient, $\phi_{i,t}$, at all points in time. There is a vast literature outlining the sensitivity with respect to the choice of prior in testing the root of an AR process in a Bayesian framework, with primary contributions coming from Sims (1988), Berger and Yang (1994), and Lubrano (1995). Xia and Griffiths (2012) demonstrate that for Bayesian posterior confidence interval based tests, a uniform prior over the AR(1) coefficient rejects the null hypothesis of a unit root too infrequently, but that this prior performs much better when using Bayesian model comparison. Of all the priors considered by Xia and Griffiths (2012), this uniform prior is most similar to the Normal prior used in this paper.

Estimation of Hall et al.'s Model

In order to estimate the model presented in Hall et al. (1999), I use the Gibbs sampler with relatively tight priors on the AR(1) coefficients, but diffuse priors on μ_i, the variance parameter, and the transition probabilities. The Gibbs sampler for this model is standard for a two-state Markov-Switching regression with autoregressive parameters. I omit a detailed description of the sampler here, but the interested reader can find a detailed exposition of the Gibbs sampler for this model in Kim and Nelson (1999).

Recall that the Hall et al. (1999) model is given by:

$$\Delta p_t = \begin{cases} \mu_0 + \phi_0 p_{t-1} + \sum_{j=1}^{k} \psi_{0,j} \Delta p_{t-j} + \sigma_e e_t & \text{if } S_t = 0 \\ \mu_1 + \phi_1 p_{t-1} + \sum_{j=1}^{k} \psi_{1,j} \Delta p_{t-j} + \sigma_e e_t & \text{if } S_t = 1 \end{cases}$$

Since I will be testing for an explosive root, the prior on the AR(1) coefficient is of extreme importance. I place a tight prior on the AR(1) coefficient in both regimes. In the high growth regime (regime 0), the prior on ϕ_0 is a Normal distribution with mean zero and standard deviation set to 0.05 that is truncated to lie above zero (the AR(1) parameter in the high growth regime is restricted to be consistent with explosive growth). This relatively small standard deviation reflects our prior knowledge that even when a root is only very slightly explosive, the process blows up quickly. Therefore, I believe that even in the presence of a bubble, the explosive root will not be very large. This is reflected by my prior, which holds that I have roughly 95% confidence that the AR(1) parameter is less than 1.1.

In the low growth regime (regime 1), the prior on the AR(1) parameter, ϕ_1, is also a Normal distribution with mean zero and standard deviation set to 0.05. However, this distribution is truncated from above at ϕ_0, in order ensure uniqueness of the likelihood function of the MS model. This prior again reflects the fact that I believe that the AR(1) parameter is near 0, and if it is explosive, it is probably only very slightly explosive.

Estimation of MS-TVP Model

As shown in Eo and Kim (2012), with standard prior distributions on the parameters that are being estimated, the conditional posterior distributions can each be derived analytically, so we can use the Gibbs sampler to estimate this model. Since this model is less well known, I will present an overview of the steps involved in the Gibbs sampler below. However, since my estimation procedure is nearly identical to the procedure discussed at length in Eo and Kim (2012), I will keep this overview relatively brief.

As in Hall et al.'s (1999) model, the priors on the AR(1) coefficients are very important. In order to facilitate estimation of the more flexible switching model, I do not center the prior distributions for the initial conditions for the AR coefficients at zero.[13] Instead, for the AR(1) parameter in the high growth regime, I set the prior for the initial condition to 0.05, and in the low growth regime, I set the prior for the initial condition at -0.05, each with a small variance.

The steps for the Gibbs sampler are as follows:

Step 0:

Initialize the hyperparameters of the model, $\tilde{\Omega} = [\sigma_e^2 \; \sigma_{\mu_0}^2 \; \sigma_{\mu_1}^2 \ldots \; \sigma_{\psi_{1,k}}^2]'$, the time-varying parameters, $\tilde{\beta_T} = [\beta_1 \; \beta_2 \; \ldots \; \beta_T]'$, and transition matrix

$$\tilde{P} = \begin{bmatrix} p_{00} & 1-p_{00} \\ 1-p_{11} & p_{11} \end{bmatrix}$$

Step 1:

Generate the regime for each time period, $\tilde{S}_T = [S_1, S_2, \ldots, S_T]'$, conditional on $\tilde{\beta_T}$, $\tilde{\Omega}$, $\tilde{P}$, and data $\tilde{Y_T}$. This is based on the multi-move sampler developed by Carter and Kohn (1994) and explained in Kim and Nelson (1999).

Step 2:

Based on the state space model, generate the time-varying parameters:

$$\tilde{\beta_T} = [\mu_{0,T} \; \mu_{1,T} \; \phi_{0,T} \; \phi_{1,T} \; \psi_{0,1,T} \; \psi_{1,1,T} \; \ldots \; \psi_{1,k,T}]'$$

[13]This issue is discussed in more detail in section 6

conditional on $\tilde{\Omega}$, $\tilde{S}_T$, $\tilde{P}$, and the data, $\tilde{Y}_t = [p_1 \ p_2 \ \ldots p_T]'$. This can be done by exploiting the state space form of the model to run a Carter and Kohn (1994) algorithm utilizing the Kalman filter.

Step 3:

Generate the hyperparameters of the model, $\tilde{\Omega}$, conditional on $\tilde{\beta_T}$, $\tilde{P}$, $\tilde{S}_T$ and $\tilde{Y}_T$. This is done by exploiting the fact that conditional on the other parameters of the model, each of the state space equations are line by line OLS, as is the measurement equation.

Step 4:

Generate the matrix of transition probabilities, $\tilde{P}$, conditional on $\tilde{S}_T$.

Testing Procedure

The estimation method laid out above gives us parameter estimates the for MS-TVP model. However, after obtaining these estimates, I need to assess the performance of each test, both in absolute terms and relative to the performance of the Hall et al. (1999) test. To do so, I use two sets of competing models, and use Bayesian model comparison to determine which of the two models is more likely.

To assess the performance of the MS-TVP test, I generate 201 time series each consisting of 100 periods according to Evans' (1991) dynamic price and bubble equations. Then, to test for the presence of an explosive root, I specify two competing models. The first comparison is between the Hall et al. (1999) model and a null model. The second is between the MS-TVP model and a null model. In order to assess the performance of the MS-TVP model relative to the Hall et al. (1999) model, I compare the performance of each test against their corresponding null models.

For the standard Hall et al. (1999) model, I first estimate the model restricting the AR(1) process to be explosive in at least one of the two regimes. Next, I estimate a model that restricts the price series to behave as it would in the absence of a bubble. In the absence of a bubble, the price series would be driven only by the underlying fundamental, so two features would change. First, and most obviously, the explosive root in the price series would be replaced by a unit root. However, it is also the case that there would be no regime switching, since in the Evans's (1991) bubble generation procedure, regime switching is only a feature of the bubble component. Therefore, the second model simplifies to Bayesian linear regression, with the series restricted to be nonexplosive, i.e. I restrict $\phi \leq 0$. Table 1 summarizes these two competing models.

TABLE 1. Assessing Hall et al.'s (1999) Test

	Model Equations	Model Restrictions
Model 1	$\Delta p_t = \mu_0 + \phi_0 p_{t-1} + \sum_{j=1}^{k} \psi_{0,j} \Delta p_{t-j} + \varepsilon_t$	$\phi_0 > 0$
	$\Delta p_t = \mu_1 + \phi_1 p_{t-1} + \sum_{j=1}^{k} \psi_{1,j} \Delta p_{t-j} + \varepsilon_t$	$\phi_1 \leq \phi_0$
Model 2	$\Delta p_t = \mu + \phi p_{t-1} + + \sum_{j=1}^{k} \psi_j \Delta p_{t-j} + \varepsilon_t$	$\phi \leq 0$

Next, I do the same for the MS-TVP model. The only difference is that for this model, I assume that prices follow an AR(1) process in order to reduce the risk of overfitting. Therefore, when assessing the performance of the MS-TVP model, I use the following two models shown in Table 2.

For both assessments, I use Bayesian model comparison to compare model 1 to model 2. In order to do so, I first estimate the marginal density of each model

TABLE 2. Assessing the MS-TVP Test

	Model Equations	Model Restrictions
Model 1	$\Delta p_t = \mu_{0,t} + \phi_{0,t} p_{t-1} + \varepsilon_t$	$\phi_{0,t} > 0$
	$\Delta p_t = \mu_{1,t} + \phi_{1,t} p_{t-1} + \varepsilon_t$	$\phi_{1,t} \leq \phi_{0,t}$
Model 2	$\Delta p_t = \mu + \phi p_{t-1} + \sigma_e e_t$	$\phi \leq 0$

using the decomposition given in Chib (1995):

$$m(Y_T) = \frac{f(Y_T|\tilde{\theta})\pi(\tilde{\theta})}{\pi(\tilde{\theta}|Y_T)} \tag{1.6}$$

where $\theta = [\Omega, P]$, i.e. it is the collection of all of the hyperparameters that are being estimated. In equation (1.6), $f(Y_T|\tilde{\theta})$ is the sampling density, $\pi(\tilde{\theta})$ is the prior density of θ, and $\pi(\tilde{\theta}|Y_T)$ is the posterior density of θ, each evaluated at $\theta = \tilde{\theta}$ where $\tilde{\theta}$ is any fixed value of θ in the posterior distribution. Chib (1995) recommends setting $\tilde{\theta}$ equal to a value that occurs with great frequency, such the posterior mean or median, in order to achieve the most accurate approximation. In my application, I use the posterior mean.

Recall that in the MS-TVP model, the shocks in the transition equation are heteroscedastic, since they only occur when there is a shift in regimes. Therefore, there is not an easily computable analytical expression for the sampling density, $f(Y_T|\tilde{\theta})$, in the MS-TVP model. I estimate it using a simple particle filter, based on Fernández-Villaverde and Rubio-Ramírez (2004), which can be used to approximate the marginal density of any parameterized nonlinear state space model. In order to decrease the impact of numerical estimation error, I set the number of particles high enough so that to the first decimal place, the filter produces identical estimates.[14]

[14]In practice, I used 20,000 particles.

To compute an estimate of $\pi(\tilde{\theta}|Y_T)$, I use the method suggested in Chib (1995), which consists of running the Gibbs sampler successively, each time holding an additional element of θ, the collection of all estimated hyperparameters, at its posterior mean, $\tilde{\theta}$. Once I have the numerically estimated values for both the sampling density and the posterior density, I can compute the marginal density using the formula in equation (1.6).

After computing an estimate of the marginal density for each model, I compare the two models by computing the posterior odds ratio. The posterior odds ratio simply describes how likely one model is relative to another. For example, if the posterior odds ratio for model one compared to model two is 3.0 (so that the odds are 3:1), then I would say, given the data, model one is $\frac{3}{3+1} - \frac{1}{3+1} = 50\%$ more likely to have generated it than model two.

In general, the posterior odds ratio of model one compared to model two can be given as:

$$\frac{pr(m_1|Y_T)}{pr(m_2|Y_T)} = \frac{m_1(Y_T)}{m_2(Y_T)}\frac{pr(m_1)}{pr(m_2)}$$

where m_i denotes model i, and $\frac{pr(m_1)}{pr(m_2)}$ is the prior odds ratio. In my case I assume equal prior probability between the two models, so the latter term drops out and equation (1.1) becomes:

$$\frac{pr(m_1|Y_T)}{pr(m_2|Y_T)} = \frac{m_1(Y_T)}{m_2(Y_T)} = B_{12}$$

This makes clear that once I have numerically computed my estimates of the marginal densities, I have all I need to compare the two models.

Finally, I need to set criteria that determines when I will prefer one model to another. To do this, I calibrate the testing procedure, by first running competing models on all generated series of fundamentals. After running these competing models on data that I know has been generated according to a stationary process, I compute the odds ratios, and find the odds ratio for which I would incorrectly prefer model one five percent of the time. In other words, I calibrate the test such the size of the test is approximately five percent.

For example, when comparing the performance of the MS-TVP model with the non-explosive linear regression model, this occurs at an odds ratio equal to 2.16. Therefore, if a given price series has an odds ratio greater than 2.16, and its corresponding fundamental series has an odds ratio less than 2.16, this suggests the presence of a bubble in the price series. In other words, since the price series is explosive but the fundamental series is nonexplosive, this suggests that there is another component aside from the fundamental that is driving the asset price.

However, if both the price series and the fundamental series had an odds ratio greater than 2.16, this would not suggest the existence of a bubble. Since both series are determined to be explosive, it does not suggest that the something other than the fundamental is driving the movements in the price of the asset.

Results

Estimation of Generated Data

Although this procedure is fairly straightforward, I did experience practical difficulties during estimation of the MS-TVP model. When estimating a Markov-Switching model, we need to enforce some type of inequality restriction on a subset of the parameters to ensure uniqueness of the likelihood. In other words, the likelihood function is symmetric - it makes no difference to the likelihood function whether we label the high growth regime as "regime zero" or "regime one".

Since my procedure requires a forward run of the Kalman filter,[15] if I have an "unlucky" draw of regimes and hyperparameters, the time-varying parameters may wander outside their restricted region. When I attempt to enforce this restriction on the backward draw via rejection sampling, my sampler may have to sample billions (or more) of times in order to find parameters that fit the restriction. Note that Koop and Potter (2011) point out a similar problem in a time-varying parameter vector autoregression. In their estimation, they compare a multi-move algorithm, which is similar to the Carter-Kohn algorithm I use in estimation of my model, and a single-move algorithm. They find that when they use the multi-move algorithm, their rejection rates are as high as 99.97%. Although in theory the single-move algorithm mixes at a slower rate, the rejection rate is substantially lower than the multi-move algorithm, so in practice Koop and Potter (2011) suggest using a single-move algorithm.

However, Koop and Potter (2011) perform their analysis within a Metropolis-Hastings setting on a slightly different estimation procedure, so it is not straightforward to generalize their results to my estimation technique. Therefore, to attempt to solve this problem in my model, I set the variances time-varying parameters equal to a small constant instead of estimating them.[16]

Because estimation along with approximation of the marginal density is relatively computationally intensive, I try to strike a balance between accuracy and

[15] See Step 2 in the previous section.

[16] While this fixes the issue in most cases, the issue remains in samples that have a very high variance (about 5% of all price series). When the issue remains, I count that particular series as "no bubble" series.

timeliness. Auto-correlation functions and running mean plots on randomly selected time series simulations suggested the use of at least 5,000 burn-in draws. To be conservative, I chose to use 10,000 burn-in draws. However, in order to also ensure a feasible speed of estimation,[17] I chose a relatively modest 20,000 post burn-in draws.[18]

Power of MS and MS-TVP Tests

In my estimation, I seek to investigate both the absolute and relative power of the Hall et al. (1999) test and the MS-TVP test. The exact priors used can be found in the appendix. Below, I present the results from the model assessment, first comparing the Hall et al. (1999) model restricted to have an explosive root with a nonexplosive model with no switching, and then comparing the MS-TVP test with an explosive root to a nonexplosive model with no switching. As described in the previous section, I calibrate these tests to have size of .05, i.e. 5% of the time it will incorrectly prefer the model with switching when the true model is the stationary linear process. My results are presented in Table 3.

TABLE 3. Power of Bayesian MS Test with 5% Size

	Calibrated Odds Ratio	% Containing Bubble
Bayesian Hall et al. (1999)	45.40	78.61%
MS-TVP	2.16	79.60%

[17]For each time series, estimation entails the estimation of both competing models, as well as approximating the likelihoods of each model. This takes about 30 minutes in Matlab on a 2013 Macbook Pro with a 2.7 Ghz. Intel i7 processor.

[18]Experimenting on a few randomly selected time series, my results were not very sensitive the increasing the number of post burn-in draws.

First, compared to the results in Hall et al. (1999), the Bayesian implementation of the test which jointly tests for both switching and an explosive root displays higher power to detect the presence of a bubble. Hall et al. (1999) are only able to detect a bubble in approximately 60% of the price series. Second, for this particular specification, the more general MS-TVP model does not add much power to the detect a bubble.

The first result, that the Bayesian implementation has more power than the classical estimation in Hall et al. (1999), may be partially driven by the fact that I am jointly testing for both switching and an explosive root, while Hall et al. (1999) test only for an explosive root. A model with an explosive root and switching may represent the price process better than a linear model, even if a switching model with a nonexplosive root might provide a fit of the price series data that is superior to both of these. This intuition is supported by my initial attempt at implementing Bayesian testing in the Hall et al. (1999) model, which found that a version of the test that tested only for an explosive root only detected a bubble about 40% of the time.[19] However, since the underlying dividend series does not have switching, I believe that testing only for an explosive root would be incorrect, and that the test presented in this paper fully exploits the specification of the periodically collapsing bubble process found in Evans (1991).

The second result, that the MS-TVP model does not seem to provide superior bubble detection in this model, is actually quite intuitive. Since the growth rates of all of the bubbles in this model are identical conditional on the realized values of the shocks, the standard Hall et al. (1999) test should be expected to perform relatively well compared to the more general model. Under an alternate time-

[19]However, these estimations were conducted with slightly different priors. To the extent that the results are sensitive to the priors, this would fail to be an apples-to-apples comparison.

varying specification of the bubble growth rate, or even in an application to real world price series that may contain several bubbles, my priors are that the MS-TVP model would perform better relative to the Hall et al. (1999) test.

Finally, for purposes of intuition, I believe it is helpful to see what one of my generated and estimated price series actually looks like. Following Evans (1991), I chose the price series with median variance. The first image in Figure 3 plots the actual asset return and the estimated asset return for all periods. The second image plots the probability of being in the low growth regime for all periods. The fact that this probability is virtually zero except for four brief periods suggests that this price series is in a bubble in almost all time periods. For this price series, model comparison suggests that this price series contains a bubble.

In Figure 4, I present the same graphs, estimated instead in the MS-TVP model. In addition, I present the time path of the explosive root, $\phi_{0,t}$, which shows that following the first large bubble collapse, the next bubble contains a more explosive root. Following the collapse of the second bubble, the explosive root is smaller.

FIGURE 3. Median Bubble (Standard MS Model)

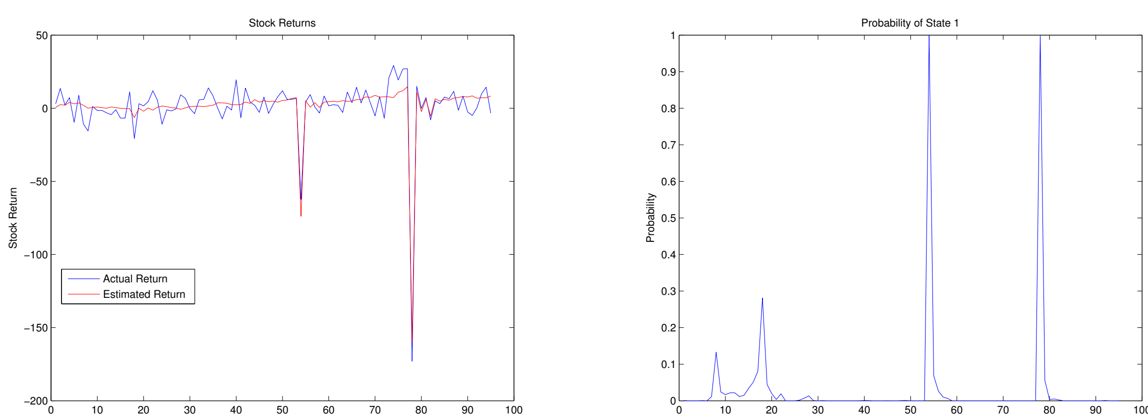
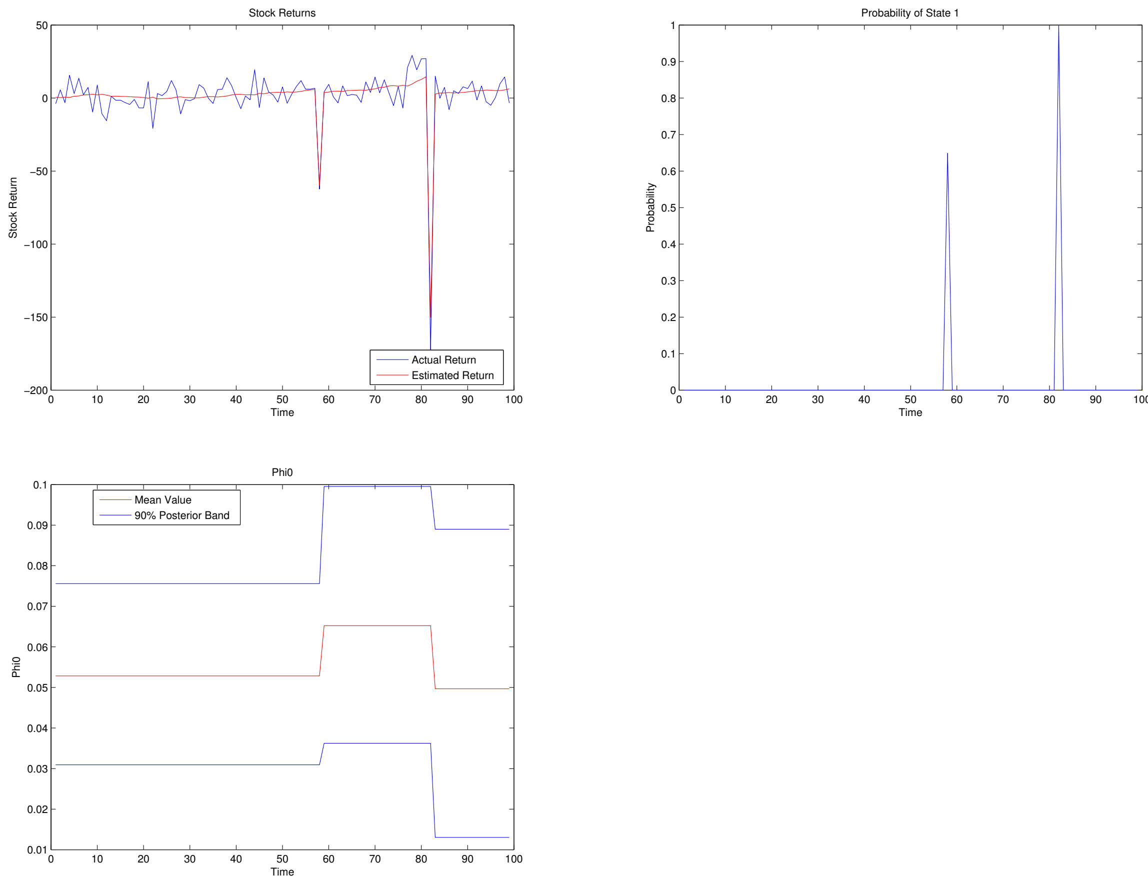


FIGURE 4. Median Bubble (MS-TVP Model)



Conclusion

Detection of asset bubbles would allow policymakers to implement policies to either prevent future bubbles or to take action against an ongoing bubble. In order to aid this detection, I have introduced Bayesian estimation of Hall et al.'s (1999) Markov-Switching test for asset bubbles, and expanded on it to test for both regime switching and an explosive root simultaneously. This leads to an increase in power over the test presented in Hall et al. (1999). In addition, I modify a Markov-Switching time-varying parameter (MS-TVP) model developed in Eo and Kim (2012), and I use this MS-TVP model to test for asset bubbles. In theory, this test is more flexible, since it allows for the possibility that the unconditional growth rate of the bubble changes each time the high growth regime is entered.

Using Monte-Carlo analysis, I show that even in the case where the unconditional growth rate of the bubble remains constant over time, this MS-TVP test has slightly more power than the Hall et al. (1999) test. This is a promising sign, since in the real-world it is unclear whether repeated bubbles will have the same growth rate or different growth rates than previous bubbles. The test developed in this paper is more flexible than Hall et al.'s original test, yet retains power against a more restrictive class of bubbles.

CHAPTER II

INTEREST RATE RULES IN PRACTICE - THE TAYLOR RULE OR A TAILOR-MADE RULE?

Introduction

Many studies concerning the conduct of monetary policy in the United States assume the target Federal Funds rate evolves according to a Taylor rule. Under this rule, the target Federal Funds rate depends only on inflation and output, with this assumption justified on both theoretical and empirical grounds. However, there are many different measures of inflation and output, and it is not clear which of these measures should be used to produce the most accurate description of policy. Furthermore, there are a host of sectoral level variables, such as industrial production and commodity price growth, that may be important to the Federal Open Market Committee's (FOMC's) decision making. The primary goal of this study is to determine what variables have been relevant to the FOMC over the past 40 years.

Determining the variables considered by the FOMC should not only be of interest to economic historians or Fed watchers. Many macroeconomists need to specify a policy rule in order to conduct their research, regardless of whether monetary policy is of central importance to their research question. For example, it is necessary to specify a policy rule in all monetary DSGE models. If the researcher's goal is to evaluate forecasting performance or to study other features of observed data, knowing the correct form of the policy rule will be of great importance, and could potentially influence the results.

Given the long history and large volume of monetary policy research, it is surprising that this issue has not been studied in detail. Instead, the profession has largely followed the work of Taylor (1993), which argues that the behavior of the FOMC can be usefully described by an interest rate rule depending only on inflation and the output gap. The original justification for use of this rule was policy arising from the rules vs. discretion literature of the late 1980s. The empirical application in Taylor (1993) showed that this type of rule fit the Federal Funds rate data fairly well from 1987-1992, and this analysis was extended in Taylor (1999) to cover a much longer time frame. By the early 2000s, based on this and other similar research, "Taylor-type" interest rate rules that include one measure of inflation and one measure of the output gap became the default policy rule used in both theoretical and empirical studies of the macroeconomy and monetary policy. This is still the case today, with some authors also including lags of the interest rate to account for interest rate smoothing.[1]

While these Taylor-type rules have clearly become the dominant paradigm for describing monetary policy in the United States, there is no consensus on the actual measures of inflation and output that should be used to describe policy. This is demonstrated in Table 4, which shows the wide range of definitions that are commonly used. Popular measures of inflation include GDP Deflator inflation and CPI inflation, while the unemployment gap and the GDP gap are most commonly used to measure output. Even among studies that include the same variables, there can be uncertainty about timing; this can be seen in the first two rows of the table, as Taylor (1999) assumes the FOMC responds to contemporaneous values

[1]Throughout this paper, policy rules that include lags of the interest rate are referred to as "generalized" Taylor rules.

while Clarida et al. (2000) assume the FOMC is forward looking and responds to forecasts.

TABLE 4. Explanatory Variables Used in Interest Rate Rule Estimation

Study	Inflation Measure	Output Measure	Horizon	Other
Taylor (1999)	GDP Deflator	GDP Gap	Contemp.	-
Clarida et al. (2000)	GDP Deflator	GDP Gap	Forecast	-
Bernanke and Boivin (2003)	CPI	UN Gap	Forecast	Factor
Orphanides (2004)	GDP Deflator	GDP Gap	Contemp.	-
Cogley and Sargent (2005)	CPI	UN Rate	Past	-
Primiceri (2005)	GDP Deflator	UN Rate	Past	-
Schorfheide (2005)	CPI	GDP Gap	Contemp.	-
Boivin (2006)	GDP Deflator	UN Gap	Forecast	-
Sims and Zha (2006)	Core PCE	GDP growth, UN Rate	Past	PCom
Davig and Doh (2008)	GDP Deflator	GDP Gap	Contemp.	-

In addition to disagreement about the precise measures of inflation and output included in the rule, a potential pitfall when using a Taylor rule is that the FOMC may actually respond to more variables than inflation and output. In this case, policy rules that include only inflation and output would suffer from omitted variables bias. In fact, when estimating the Taylor rule using historical data, the residuals are highly autocorrelated. Therefore, many authors already include a third variable in their Taylor-type rule - the first lag of the Federal Funds rate. While this results in residuals that are substantially less autocorrelated, failure to include an even greater number of relevant variables could still bias coefficient estimates. Finally, if one goal of a study is to be able to best predict the Federal Funds rate in the future, failure to include relevant variables will likely result in predictions that are not as accurate as they could be.

One potential solution would be to include all possibly relevant variables in a regression model, but this solution has several drawbacks. First, including all variables implicitly assumes they are all relevant, but it is not necessarily realistic that the FOMC adjusts its Federal Funds rate target every time one of a large number of variables changes. Second, forcing the inclusion of all variables will reduce the degrees of freedom, leading to less precise estimation of regression coefficients. While this loss of precision would be justified if all variables actually belong in the model, it would harm inference if they do not. Similarly, including potentially irrelevant variables could lead to overfitting in-sample.

Due to these problems, I use Bayesian Model Averaging (BMA) to average across a large number of regression models. BMA is naturally suited to the current context in which there is uncertainty about the true underlying model. Under BMA, each regression model receives posterior weight according to how well it fits the data. As is well known in the Bayesian literature, this weight includes a built-in penalty for the number of parameters that the model includes.[2] Therefore, ceteris paribus, more parsimonious models receive higher posterior weight. Since this technique averages across a large number of regression models, coefficients on variables that are deemed unlikely to be included in the FOMC's interest rate rule are shrunk toward zero. This occurs because the marginal coefficient value is determined by a weighted average of zero, when the variable is not included, and the estimated coefficient value when it is included, with the weight on zero being very high. This shrinkage toward zero typically increases out of sample forecasting performance relative to the regression model that simply includes all variables. As a byproduct of this procedure, I get inclusion probabilities for each variable, which are useful in this context, since a main goal of this study is to determine the variables that the FOMC responds to.

[2] See Koop (2003).

After using BMA, I find four interesting features of monetary policy: (1) the FOMC has been forward looking, (2) interest rate rules of the "generalized" Taylor rule form that include only one measure of inflation, one measure of output, and the first lag of the Federal funds rate receive almost no posterior probability, (3) the FOMC is much more likely to respond to employment statistics than GDP, and (4) rules formed using BMA forecast more accurately than generalized Taylor-type rules.

First, the FOMC has been forward looking, responding to forecasts of future inflation rather than past inflation. This is evidenced by the posterior inclusion probabilities on inflation measures where, for example, expected future GDP Deflator inflation is included with 95.7% probability, while lagged GDP Deflator inflation is included with only 10.3% probability. Second, "generalized" Taylor-type interest rate rules, rules that include only one measure of inflation, one measure of output, and the first lag of the Federal Funds rate, receive almost no posterior probability. This is true under all three different versions of model priors considered in this paper, each of which imply very different things about the variables included in the interest rate rule. A low posterior probability for generalized Taylor-type rules is consistent with the results of Bernanke and Boivin (2003) and Cúrdia et al. (2011), who find that standard formulations of the Taylor rule do a relatively poor job of explaining historical policy responses. Third, the FOMC is much more likely to respond to the unemployment gap and the change in the unemployment rate than to the growth rate of GDP. This result aligns with the mandate of the Federal Reserve, which tasks it with maintaining full employment. Finally, one-step ahead forecasts formed using BMA are more accurate than those formed using generalized Taylor-type rules. As judged by Root Mean Squared Forecasting Error (RMSFE), a commonly used forecast evaluation metric, the forecasts from BMA are on average about 20% more accurate than forecasts formed using generalized Taylor-type rules.

Data and Interest Rate Rule Specification

In my analysis, I consider a total of 14 regressors: one lag of the Federal Funds rate, CPI inflation, past GDP deflator inflation, expected future GDP deflator inflation, past real GDP growth, expected future real GDP growth, the unemployment gap, the change in the unemployment rate, industrial production, housing starts, real PCE growth, payroll employment growth, commodity price growth, and oil price growth. For the forward looking variables, I use Greenbook forecasts, which are available over the entire sample. For all other variables, including variables for which Greenbook forecasts become available later, but are not available over the entire sample, I use lagged values over the entire sample. For these lagged values, I use the last available real time data release occurring on or before the corresponding FOMC meeting date. A description of the variables I use is presented in Table 5.

TABLE 5. Variables Included in BMA Exercise

Variable	Measure	Horizon	Source
CPI	YoY growth	Past	ALFRED
GDP Deflator	YoY growth	Past	ALFRED
GDP Deflator	Mean QoQ growth, 3 Quarters	Future	Greenbook
RGDP	QoQ growth	Past	ALFRED
RGDP	Mean QoQ growth, 3 Quarters	Future	Greenbook
Unemployment Rate	Gap	Future	Greenbook
Unemployment Rate	Change	Future	Greenbook
Industrial Production	Mean QoQ growth, 3 Quarters	Future	Greenbook
Housing Starts	Units	Past	ALFRED
Real PCE	QoQ growth	Past	ALFRED
Commodity Prices	QoQ growth	Past	World Bank
Payroll Employment	QoQ growth	Past	ALFRED
Oil Prices	QoQ growth	Past	ALFRED

As far as the frequency of the data collected, I use FOMC meeting-based timing, which is novel to the Taylor rule literature. That is, for the regressors I assume that the FOMC had the most recent release of the data that was available on the meeting date. For the outcome variable, the Federal Funds (FF) rate, I use the daily Federal Funds rate to construct the average FF rate between meeting dates. For example, the FOMC met on August 7, 2007. I assume that they had the latest release of all of the "past" regressors, and that they used the Greenbook forecast corresponding to the August 7 meeting for all of the "future" regressors. For the Federal funds rate, I assume that they enforce the agreed upon target until the next meeting, which occurred on September 16, 2007, and I use the average of the daily Federal Funds rate between August 7 to September 15 as the outcome variable.[3]

The meeting-based timing solves several issues that arise when using monthly or quarterly averages, which is typically used in studies in FOMC behavior. In these studies, the Federal funds rate is formed using monthly or quarterly averages of the Federal Funds rate. These averages are then matched up with corresponding monthly or quarterly inflation and output data. However, throughout the sample, the FOMC typically meets eight times per year, twice per quarter. This idiosyncrasy creates measurement error when using monthly or quarterly averages. Furthermore, the meeting dates are not necessarily regular throughout the course of each quarter or each month, which only serves to increase the errors introduced by using quarterly or monthly data.

Use of meeting date-based timing does create one complication for data collection, particularly for forecast data found in the FOMC Greenbook. The

[3]Note that, on occasion, the FOMC changes policy in between formal meetings. This appears to have happened seven times in my sample, and is unaccounted for with my methodology.

complication arises from the fact that these Greenbook variables are forecasted at a quarterly horizon, but the meeting dates of the FOMC occur at vastly different stages of the quarter. This causes a problem because if a researcher uses the quarterly forecasts, the meeting date can substantially alter the degree to which the FOMC is forward looking. To illustrate this potential problem more clearly, consider the following example in which the FOMC is forward looking and would like to respond to their "one quarter ahead" GDP growth forecast, presented in Table 6.

TABLE 6. Greenbook Forecasting Example, GDP Growth

Forecast Horizon	Last Day of 2nd Q	First Day of 3rd Q
Current Quarter	1.0%	3.0%
One Quarter Ahead	3.0%	-2.0%
Two Quarters Ahead	-2.0%	-1.0%

If the FOMC meets on the last day of the second quarter, their one quarter ahead forecast will be for the third quarter, at 3.0%. But if the meeting was shifted one day into the future, so that they meet on the first day of the third quarter, their one quarter ahead forecast will be for the fourth quarter, at -2.0%. But since they are meeting on the first day of the quarter, in some sense this -2.0% forecast is really a two-quarter ahead forecast, since it is their best guess of what growth will be throughout the fourth quarter, which doesn't begin for another 90 days. In this case, the forecast for the "current" quarter, 3.0%, more accurately represent beliefs about the one-quarter ahead forecast.

To address this problem in a consistent manner, I use a strategy that weighs future forecasts based on the date of the meeting inside of the current quarter. This weight changes linearly with the timing of the meeting date. Continuing with the above example, if the FOMC truly cared about a "one-quarter ahead" forecast, I assume that they form their forecast in the following way:

$$\text{GDP forecast} = (1-p)\text{GDP}_t^f + p\text{GDP}_{t+1}^f$$
$$p = \frac{\text{days into current quarter}}{\text{total days in current quarter}}$$

Where "GDP forecast" is the forecast that the FOMC will actually respond to, while GDP_t^f is the forecast for the current quarter contained in the Greenbook, and GDP_{t+1}^f is the one-quarter ahead forecast contained in the Greenbook. Applying this formula to the example above, we see that if the meeting falls on the last day of the 2nd quarter, the actionable one quarter ahead GDP forecast would be 2.98%, while if the meeting falls on the first day of the 3rd quarter, it would be 2.95%. Even in the extreme example outlined above, this strategy leads to a sensible and smooth change in the future forecast.

In its most general form, I apply the following formula to get the h quarter ahead forecast of variable x as of the meeting date:

$$x_{t+h} \text{ forecast} = \frac{1}{h}\left[(1-p)x_t^f + \sum_{j=1}^{h-1} x_{t+j}^f + p x_{t+h}^f\right]$$
$$p = \frac{\text{days into current quarter}}{\text{total days in current quarter}}$$
$$h > 1$$

Typically, I am interested in the average of the three quarter ahead growth rates of the variables included in the Greenbook. Therefore, the exact formula is given as:

$$x_{t+3} \text{ forecast} = \frac{1}{3}\left[(1-p)x_t^f + x_{t+1}^f + x_{t+2}^f + p x_{t+3}^f\right]$$

In words, to form the "true" three quarter ahead average forecast, I weight the nowcast for the current quarter and the forecast for the three quarter ahead growth rate according to the time remaining in the current quarter, while the one and two-quarter ahead forecasts receive equal weight. This procedure is necessary to keep the forecast horizon consistent across all observations, since the meeting dates vary substantially within each quarter, and the forecasts contained in the Greenbook are expressed as quarterly forecasts.

With forecasts in hand, I turn to computing measures of the inflation gap and the output gap, which are typically included in Tayor-type rules instead of raw inflation and output. Unfortunately, the FOMC did not announce their inflation target until 2012, and they do not regularly provide estimates of potential output or the natural rate of unemployment. Therefore, I construct these measures using historical data. For simplifying purposes, I assume that the natural rate of unemployment is constant. Because addition or subtraction of a constant from a regressor will not impact inference, I simply leave the unemployment rate unadjusted.[4]

For a measure of the inflation target, I use Matlab code that accompanies the paper by Chan et al. (2013). In that paper, the authors allow for the inflation gap to evolve according to an autoregressive process and probabilistically bound the target inflation rate above at 5%.[5] I believe that the former is both reasonable and realistic as a measure of the inflation target, since if the FOMC misses its target two quarters in a row, it is more likely than not that the misses will be in

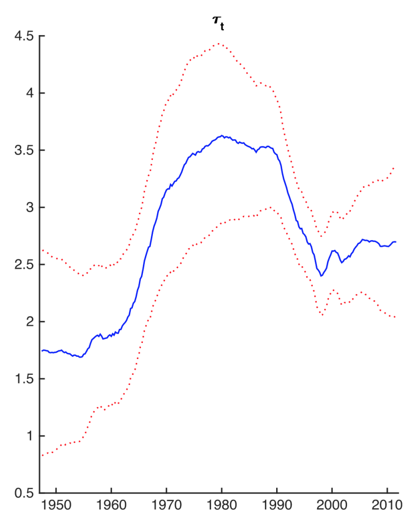
[4]I have experimented with a constant gain learning rule for the natural rate of unemployment, but found that my results do not change substantively for typical values of the gain parameter.

[5]They estimate the upper bound, but set a prior on it that only has support between 0% and 5%.

the same direction. For example, as of writing, GDP Deflator inflation has been below the Fed's stated 2% target for 13 consecutive quarters, Q2 2012-Q2 2015. Additionally, it seems reasonable that the FOMC never desired an inflation rate higher than 5%, even though the inflation rate reached much higher levels in the 1970's. Moreover, in an online appendix, Chan et al. (2013) show that increasing the bound on inflation to 10% has very little influence on their results.

When estimating the inflation target, I use a fully revised measure of the GDP Deflator. Doing so produces the inflation target measure displayed in Figure 5, and we can see that the 5% upper bound of the target is not binding.

FIGURE 5. Estimated Inflation Target



After computing the inflation gap for each measure of inflation, I have the entire set of regressors. I consider interest rate rules of the following form:

$$i_t = X_t \beta_t + \sigma_t \varepsilon_t$$

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

where i_t is the nominal federal funds rate at time t; X_t is a data matrix containing an intercept, the first lag of the nominal federal funds rate, and the exogenous variables; β is the coefficient vector; σ_t is the standard deviation of the monetary policy shock at time t, and ε_t is an i.i.d. error term. Note that the coefficient vector, β_t, and the standard variation of the shock, σ_t, can vary over time. In full sample estimation, I will allow for the possibility of structural changes in the values of these parameters in both 1979 and 1983.[6]

Full Sample Estimation Procedure & Results

Instead of estimating the full model, which implicitly assumes that all of the included variables were relevant to FOMC decision making, I use Bayesian Model Averaging (BMA) to average results over every potential regression model. Essentially, when performing BMA, I run regressions for every possible combination of regressors, and probabilistically average across the results. The major steps of BMA are as follows:

1. Run Bayesian Ordinary Least Squares (BOLS) on all possible models.

2. Based on the posterior marginal likelihood, which takes into account in-sample fit and includes a built in penalty for including more regressors, compute the probability of each model.

3. Using the model probabilities and the posterior statistics of each model, such as the mean coefficient values, compute posterior statistics averaged across the posterior model space.

[6] The choice of these dates is based on the timing of known changes in monetary policy, and is discussed in more detail later.

In order to run BOLS, I need to set priors over all regressors in every model. For full sample estimation, I assume an independent Normal Inverse-Gamma prior. That is, I assume that no matter the model under consideration, the regression coefficients are drawn from a normal distribution, and the variance of the residual is drawn from an Inverse-Gamma Distribution. Because there are 14 potential regressors, there are $2^{14} = 16,384$ models for which priors are needed. Clearly, this task would be infeasible without setting priors in an automatic fashion. In order to set priors for the regression coefficients, I rely on the g-prior suggested in Zellner (1986). Let X_r denote the data vector corresponding to model r, and β_r be the regression coefficients in that model. In each model, I center this prior for β_r on $\beta_r = 0_{p_r}$, where 0_{p_r} is a vector of zeros with length p_r, the number of variables included in model r. For the covariance matrix of the regression coefficients, V_r, I set the following prior:

$$V_{r,pri} = (g_r X_r' X_r)^{-1}$$

The hyperparameter g_r is set to be constant across models, i.e. $g_r = g \; \forall \; r$. It is set according to the recommendations of Fernandez et al. (2001). Since I have 14 potential regressors and my sample size is $T = 351$, I set $g = \frac{1}{T} = \frac{1}{351}$.[7]

I assume two breaks in the variance of the interest rate rule. These breaks are known, and they occur at the October 6th, 1979 meeting and the March 29th, 1983 meeting. These dates were chosen because in the intervening period the FOMC targeted the money supply rather than the nominal interest rate. Since the Federal funds rate was allowed to move freely during this time, it is likely that its behavior

[7] In estimation, I restrict the AR(1) coefficient to be less than one in absolute value. I enforce this restriction via rejection sampling.

was much more volatile. Precise prior statistics are provided in Table 7 below, where σ_1 represents the standard deviation of the error term before October 6, 1979, σ_2 represents the standard deviation of the error term between October 6, 1979 and March 29, 1983, and σ_3 represents the standard deviation of the error term after March 29, 1983. I assume that the prior distribution of the variance terms, σ_1^2, σ_2^2, and σ_3^2, is Inverse-Gamma.

TABLE 7. Prior Distribution of Standard Deviation of Error Terms

Parameter	Mean	S.D
σ_1	1.0	0.60
σ_2	3.0	1.02
σ_3	0.7	0.42

With the priors set, I turn to posterior computation. The independent Normal Inverse-Gamma prior is conditionally conjugate, meaning that I can use the Gibbs sampler to draw from the full posterior distributions. Because I am using BMA, I need to be able to compute the marginal likelihood of each model. To do so, I use an additional simulation step, which is described in Chib (1995). In theory, the accuracy of posterior statistics such as the marginal likelihood increases as the number of simulations increases. In practice, in this relatively simple linear regression framework, a high level of accuracy can be achieved with as few as 500 posterior draws. This relatively low number of draws makes comparing thousands of models relatively easy on a modern computer.[8]

Finally, in addition to the model with breaks only in variance, I estimated a "flexible coefficients" BMA model that allowed both the regression coefficients and the variance to change at the break dates. However, consistent with the results

[8]The full details of this estimation procedure are presented in an online appendix.

of Sims and Zha (2006), these models did not fit the data well, and resulted in marginal likelihoods that were lower than the model with breaks only in variance. In fact, when performing BMA using both the flexible coefficients and the baseline set-up, the entire set of 16,384 flexible coefficients models received posterior weight that was less than 10^{-25}, and therefore would have almost no impact on any posterior feature of interest. For this reason, I drop the flexible coefficients model and focus only on models that have only a change in variance.

After running BMA I find that the FOMC seems to be strongly forward looking, responding to expected future inflation with much greater probability than past inflation. This can be seen in Table 8, where both measures of lagged inflation, CPI and past GDP Deflator inflation each receive less than 16% posterior probability, while expected future GDP Deflator inflation receives over 95% posterior probability. Additionally, it is much more likely that the FOMC responds to the change in the unemployment rate than the percentage change in real GDP. Both expected future and lagged real GDP growth receive less than 15% posterior probability, while the change in the unemployment rate receives an inclusion probability of 95.8%.

Histograms for the conditional posterior distribution of each coefficient are presented in Figure 6. These histograms are formed by resampling the posterior simulations in the following way. First, I draw a model at random, with each model being chosen in accordance with its posterior model probability. Next, once a model is selected, I draw one of the 500 posterior draws at random, with each draw being equally probable. I save this draw, and repeat this process N times to get N draws from the posterior. I choose $N = 3{,}000{,}000$. These histograms plot the value of the coefficient conditional on inclusion in the model, and ignore the

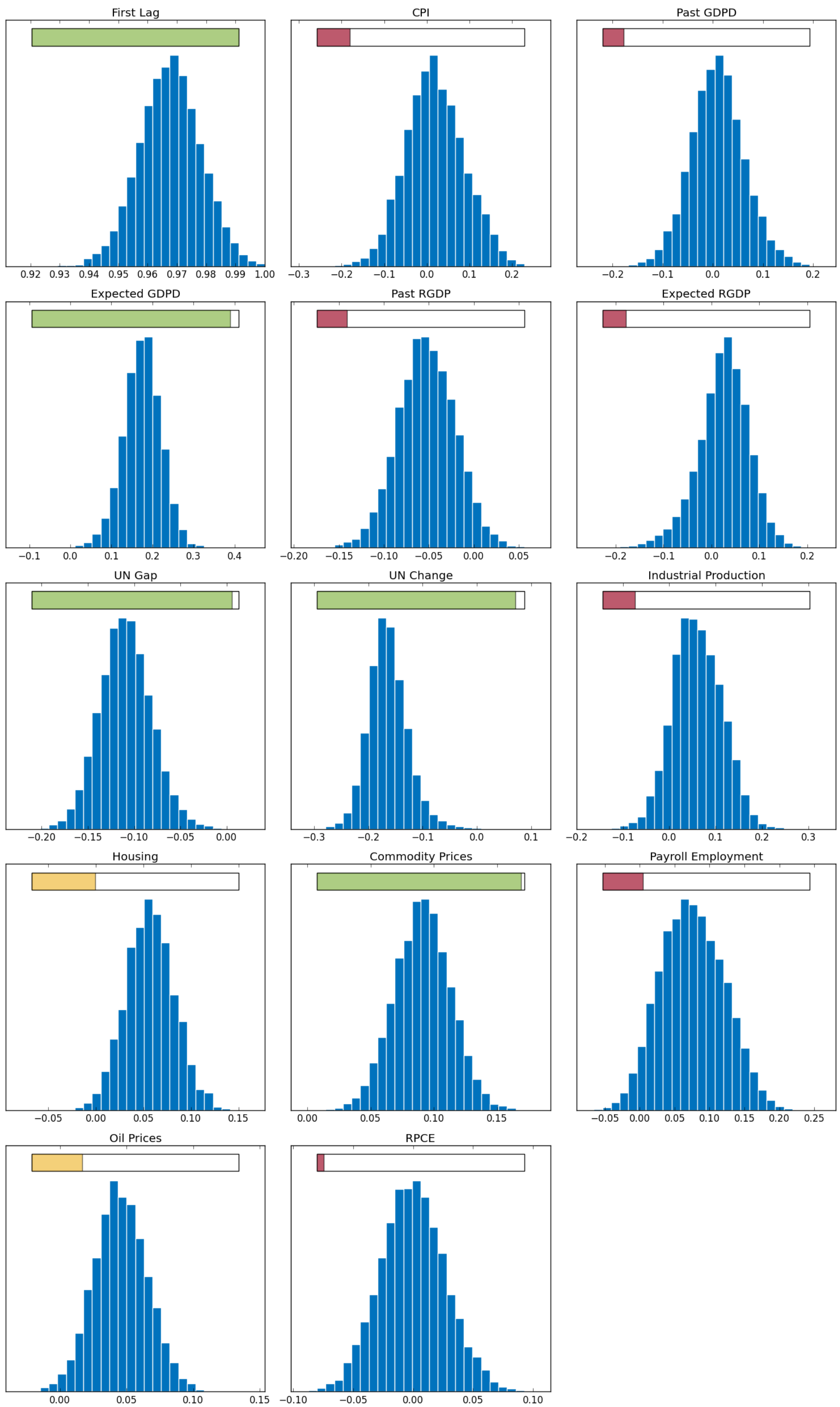
TABLE 8. Inclusion Probabilities from BMA

Variable	Probability
First Lag	100.0%
CPI	15.8%
Past GDPD	10.3%
Expected GDPD	95.7%
Past RGDP	14.3%
Expected RGDP	11.4%
UN Gap	96.8%
UN Change	95.8%
Industrial Production	15.7%
Housing Starts	30.8%
Commodity Prices	98.5%
Payroll Employment	19.5%
Oil Prices	24.5%
RPCE	3.2%

point mass that occurs at zero for variables included with probability less than one. The bars above each histogram represent the inclusion probability, with a full bar representing inclusion with probability one, and an empty bar representing inclusion with probability zero. The bars are also color-coded, with green bars signifying greater than 80% inclusion probability, red bars signifying less than 20% inclusion probability, and yellow bars indicating anything in between.

Aside from the individual inclusion probabilities and coefficients, I group variables by their type and measure the associated inclusion probability. I consider four types: lag of the Federal Funds rate, measures of general inflation, measures of real output, and sectoral measures. The first type corresponds exactly with one variable, the first lag of the Federal Funds rate. The next type, measures of general inflation, includes CPI, past GDPD, and expected GDPD. Real output includes both past and expected RGDP, the UN gap, and UN change. Sectoral measures includes all other variables: industrial production, housing starts, commodity

FIGURE 6. Coefficient Histograms



prices, oil prices, payroll employment growth,f and RPCE. In Table 9, I show the prior and posterior probabilities of rules that include at least one variable of each type. Recall that all models receive equal prior probability. Therefore, categories that include more variables receive a higher prior weight. Turning to the posterior, we see that the inclusion probability of each type of aggregated measure moves towards 100%.

TABLE 9. Inclusion Probability by Variable Type

Rule	Lag FF	Inflation	Real Output	Sectoral
Prior Probability	50%	87.5%	93.8%	98.4%
Posterior Probability	100%	98.5%	99.1%	99.9%

While the posterior inclusion probability of at least one sectoral variable moves towards 100%, the prior inclusion probability was already very high, at 98.4%. Therefore, I conduct a prior robustness check to verify that my result is coming from the information in the data, rather than the information in the prior. I use two alternative model priors. First, instead of equal prior probability across all models, I use equal prior probability across models of different sizes. I call these priors "binomial", and they are popular for model comparison and model averaging since they control for the fact that there are many more medium sized models than either small or large models. For example, in my current case, there are $\binom{14}{7} = 3,432$ models that include seven variables, but only $\binom{14}{2} = 91$ models that include two variables. Continuing with this example, under the binomial prior each model that includes seven variables receives the same weight as all other models with seven variables, and the sum of the weights on all models including seven variables is equal to the sum of the weights on all models including two variables. For my second alternative prior, I use equal prior probability across two sets of models: those that take the form of the generalized Taylor rule, and those that do not. I call these model priors the "50% Taylor" prior, and I set the prior probability that one of the versions of the generalized Taylor rule that has been followed is 50%, with 50% prior probability equally divided across all other models.

The results of this robustness exercise are shown in Table 10. We can see that regardless of the exact prior used, the posterior probability of inclusion of at least one of the sectoral variables remains near 100%. This demonstrates that the high prior inclusion probability for sectoral variables under the baseline prior is not driving my results, but rather the information contained in the data is capable of moving the posterior inclusion probabilities very far from the prior inclusion probabilities. In other words, for all three different versions of model priors, I find that it is very likely that at least one sectoral variable has been included in the policy rule of the FOMC.

TABLE 10. Inclusion Probability by Variable Type - Prior Robustness

Model Prior		Lag FF	Inflation	Real Output	Sectoral
Equal	Prior Probability	50%	87.5%	93.8%	98.4%
	Posterior Probability	100%	98.5%	99.1%	99.9%
Binomial	Prior Probability	53.9%	80.8%	86.2%	89.7%
	Posterior Probability	100%	97.7%	98.5%	99.9%
50% Taylor	Prior Probability	75.0%	93.8%	96.9%	49.2%
	Posterior Probability	100%	98.5%	99.1%	99.9%

I am also interested in the probability that the generalized Taylor-type rule was followed. Under a generalized Taylor-type rule, used in a large number of studies, I assume that the FOMC responds to only the first lag of the Federal

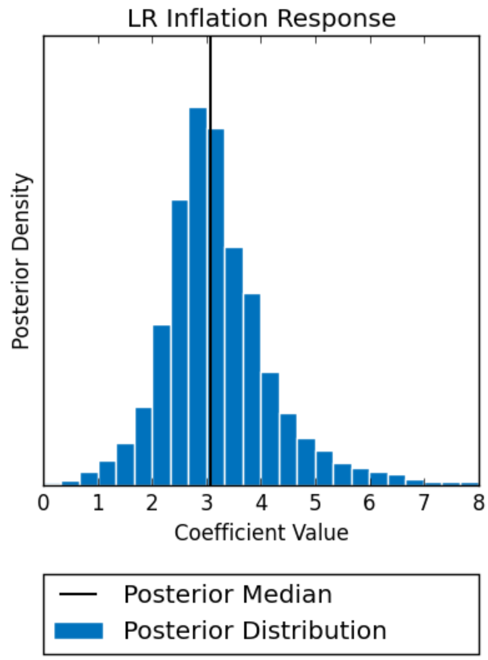
Funds rate, one measure of inflation, one measure of real production, and no sectoral variables. Therefore, they respond to only one of CPI, past GDPD, and expected GDPD; one of past RGDP, expected RGDP, UN gap, and UN change; and none of the other variables. I show the results in Table 11. We can see that for both equal model priors and the 50% Taylor rule priors, the posterior probabilities of the generalized Taylor rule are low. This shows that for sensible, but very different, model priors, there is very little evidence in support of the hypothesis that the FOMC's behavior is best approximated using a generalized Taylor-type rule.

TABLE 11. Probability of "Generalized Taylor Rule"

Model Prior	Prior Probability	Posterior Probability
Equal	7.3×10^{-4}	6.4×10^{-9}
50% Taylor	0.5	8.7×10^{-6}

Finally, a posterior feature of interest is the long run inflation response coefficient. This response coefficient is very important within economic models, as it helps to pin down determinate equilibria. In the most common case, in order for a determinate equilibrium in a simple New Keynesian model, the inflation response coefficient needs to be greater than one. In this paper, since there are several possible inflation measures included, it is necessary to add the coefficients on each in order to determine the total short-run inflation response. Then, in models in which the first lag of the Federal Funds rate is included, I divide this short-run inflation by one minus the AR(1) coefficient on the lag of the Federal Funds rate. Mathematically, $\phi_{\pi,LR} = \frac{\phi_{\pi,SR}}{(1-\rho)}$ where $\phi_{\pi,LR}$ is the long run inflation response, $\phi_{\pi,SR}$ is the short-run inflation response, and ρ is the AR(1) coefficient.

FIGURE 7.



Like the histograms presented earlier, Figure 7 presents the histogram for the long run inflation response conditional on inclusion, so the point mass at zero is ignored, and it is weighted according to the posterior model probabilities.

We can see that the long run inflation response coefficient is unimodal and slightly right-skewed. The unimodal nature of the long-run inflation response coefficient suggests the presence of only one policy regime over the sample. If there had been two policy regimes, one with a weaker response closer to 1.0 and one with a stronger response, as is often hypothesized and has been studied extensively by Clarida et al. (2000), Orphanides (2004), and numerous others, we would expect to see a bi-modal distribution. My result supports the conclusions of Orphanides (2004) and Sims and Zha (2006), who find little evidence of change in the long-run inflation response over time.

In addition to being unimodal, the density lies almost entirely to the right of one, and the posterior median is above three. Roughly 99% of the distribution lies above 1.0; in other words, there is a 99% chance that, conditional on inclusion of at least one measure of inflation, the Taylor principal was satisfied. In addition, the posterior median of the inflation response is relatively high, at 3.0. This is much higher than other authors that use single equation Taylor rule estimation have found. For example, Orphanides (2004) finds that the long run inflation coefficient is about 1.5. After experimenting with different data definitions, I found that my relatively high inflation response is largely driven by my use of meeting-based timing. Performing BMA using quarterly averages for the Federal Funds rate and all regressors yields an estimated long run inflation response coefficient of 1.85, which is much closer to the estimates typically encountered in the policy rule estimation literature.

My estimation procedure has uncovered several features of monetary policy between 1970-2007. First, the generalized Taylor rule does a relatively poor job of describing FOMC behavior. It is much more likely that the FOMC responds to several measures of inflation and output along with at least one additional sectoral variable. Next, the long-run inflation response coefficient is unimodal, suggesting that there has only been one inflation response regime over the sample. The long run inflation coefficient satisfies the Taylor principal with high probability. Finally, the median value of this coefficient is high compared to estimates derived in earlier single equation research. I find that this result is largely driven by my use of meeting-based timing.

Forecasting

In order to further assess the gains made by using BMA, I conduct an out of sample forecasting exercise. In order to avoid potential uncertainty surrounding the break dates in variance in real time, I focus only on the post 1983 sample.[9] I use two types of forecasts, rolling window and recursive. I find that the recursive forecasts are superior to those formed via rolling window estimation, which implies that allowing for structural breaks in a non-parametric fashion by using rolling window estimation does not lead to increased forecasting performance. This suggests that to the extent that there have been changes in FOMC interest rate policy since 1983, they have not been quantitatively important.

When conducting forecasts, I use a slightly different BMA procedure than when performing full sample analysis. Since I am focusing on post 1983 data, I assume a homoskedastic error term, which allows me to use the fully conjugate Normal-Gamma prior. Use of this prior means that the posterior distribution can be described analytically, and posterior simulation is not necessary. In other words, for each possible regression model, I am able to compute the exact posterior distribution, and exact marginal likelihood.[10] Doing so greatly speeds computation, which is important when doing multiple estimations in a recursive exercise. For the priors on the regression coefficients, I use the same prior as in the previous section, $\beta_r \sim N(0_{p_r}, V_{pri,r})$. Because the intercept and variance term are included in all regression models, I set an uninformative prior on each.

I conduct both rolling sample and recursive estimation. Under both techniques, forecasting begins on March 23, 1993, and continues until December

[9] Due to the post 1983 sample, I add PCE inflation to the analysis.

[10] These computations are detailed in chapter 12 of Koop (2003)

11, 2007. With rolling sample estimation, the first observation used in estimation advances as necessary to keep the sample size constant at 80 observations, approximately 10 years. With recursive estimation, the first observation remains fixed at March 23, 1983, and the sample size increases as more observations are added. I only consider the one-meeting-ahead forecast, and this is formed by assuming the FOMC has all of the information that will be available to it at the next meeting.

I conduct rolling sample estimation in order to allow for the possibility of structural change in the behavior of the FOMC in a non-parametric way. While I could perform more advanced estimation, such as estimating potential break dates, or performing a time varying parameter analysis, the relatively simple rolling window approach admits the use of conjugate priors, which greatly speeds estimation and makes re-estimation at each observation feasible. If the rolling window forecasts out-perform the recursive forecasts, this will suggest the presence of a structural break in the FOMC policy rule.[11]

I consider three measures of forecasting performance: Mean Absolute Forecasting Error (MAFE), Root Mean Square Forecasting Error (RMSFE), and the Sum of the Log Predictive Density (SLPD). The first two metrics are common in both Bayesian and frequentist environments, while the latter is gaining traction in Bayesian forecast evaluation. The MAFE measures the mean absolute difference between the forecasted value and the observed value, and is expressed

[11] Use of the parametric, more computationally intensive techniques, may be warranted if rolling window estimation provides evidence of possible structural breaks, but as I will show later, this is not the case here.

mathematically as:

$$\text{MAFE} = \frac{1}{T^f} \sum_{i=1}^{T^f} |y_i^f - y_i|$$

where T^f is the total number of forecasted periods, y_i^f is the forecasted value of the variable of interest at time i, and y_i is the actual observed value of the variable of interest at time i. The spirit of RMSFE is similar, and it is computed by:

$$\text{RMSFE} = \sqrt{\frac{1}{T^f} \sum_{i=1}^{T^f} (y_i^f - y_i)^2}$$

When using MAFE, predictions that are twice as far away are punished exactly twice as much, but when using RMSFE, predictions that are twice as far away are punished more than twice as much. In this sense, RMSFE will punish a prediction model containing a few very bad predictions much more harshly than will MAFE.

The last measure of forecasting performance I use is the sum of the log predictive density. This metric is computed by evaluating the posterior predictive density at the observed value of the variable of interest:

$$\text{SLPL} = \sum_{i=1}^{T^f} \log\left[p(y_i^f = y_i)\right]$$

where $p(y_i^f = y_i)$ is the posterior predictive density evaluated at the point $y_i^f = y_i$. This measure has several nice properties that have led to its increased use as the forecasting metric of choice in forecasts arising from Bayesian methods. First, it is robust to non-normal posterior predictive densities in a way that RMSFE and MAFE are not. For instance, imagine a bi-modal posterior predictive distribution for y_i. In this case the point estimate, y_i^f, will likely have relatively low posterior probability, and lead to RMSFEs and MAFEs that do not do a good job of capturing the predictive accuracy of the model. This problem is avoided when using the SLPL, since it fully captures the asymmetries in the predictive density. Second, as shown in Geweke and Amisano (2011), the log marginal likelihood of a model can be decomposed into the sum of the logs of the one-step ahead predictive likelihood, where prediction of the initial observation is made using only the prior distributions on the parameters in the model. Therefore, the sum of the log predictive likelihoods starting far away from initial observation mirrors the marginal likelihood, but diminishes the impact of the prior.

After conducting the forecasting exercise, I find three main results. First, recursive estimation produces more accurate forecasts than rolling sample estimation. While this does not prove that structural breaks did not occur, it shows that the gains achieved by using a larger sample size outweigh those from allowing for instability. Second, the forecasts produced by BMA generally outperform forecasts produced by any of the generalized Taylor-type rules. Third, the performance of generalized Taylor-type rules that consider output or the unemployment gap deteriorate sharply during the 2001 recession. The performance of the generalized Taylor rule that instead includes the *change* in the unemployment rate greatly improves during this recession, suggesting that during this recession policymakers only reduced interest rates when they expected the unemployment rate to increase.

First I compare the three forecasting metrics for rolling sample vs. recursive estimation, and show the results in Table 12. While I only present the results for the statistics arising from BMA below, the general pattern is true across Taylor rules as well - all three forecasting statistics improve when using recursive estimation. While MAFE and RMSFE improve slightly when using the recursive technique, the sum of the log predictive density increases dramatically. The large increase in the value of the SLPD is most likely due to a predictive density that is more sharply peaked, due to the fact that we are using more information when estimating the parameters of the model.

TABLE 12. Forecast Performance, Rolling Sample vs. Recursive

	MAFE	RMSFE	SLPD
BMA, Rolling	0.1679	0.2098	-163.41
BMA, Recursive	0.1434	0.1963	-130.94

Next, I present forecasting metrics for a variety of Taylor rules, relative to the statistics of BMA, and show the results in Table 13. Here, I have normalized the MAFE and RMSFE by dividing these statistics for each Taylor rule by the values listed in Table 12 above. Therefore, a value greater than 1 indicates larger values of these statistics, which indicates worse forecasting performance. For example, a value of 1.25 indicates forecasting performance that is 25% worse than BMA. For the SLPD, I normalize these statistics by subtracting the SLPD of the Taylor rule from the SLPD from BMA. A positive value indicates worse forecasting performance relative to BMA, while a negative value indicates superior forecasting performance.

We can see that the forecasts formed using BMA are superior to every version of the generalized Taylor rule when forecasting performance is measured by MAFE or RMSFE. With SLDP, the Taylor rule that includes the change in the unemployment rate outperforms BMA, but BMA is superior to the other three measures. In Figure 8, we see that much of the difference in the SLPD's

TABLE 13. Forecasting Performance of Taylor Rules, Relative to BMA

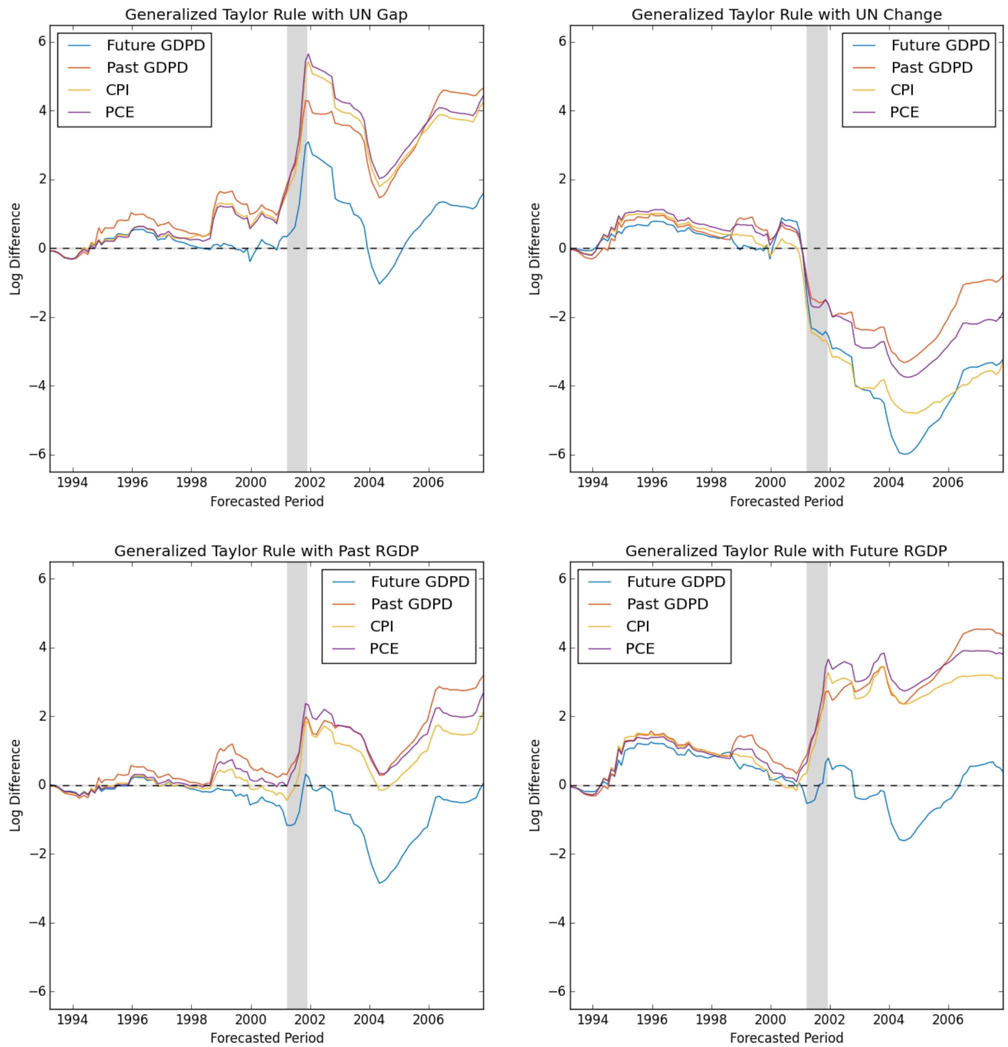
Output Measure	Inflation Measure	MAFE	RMSFE	SLPD
UN gap	CPI	1.25	1.28	4.23
	PCE	1.22	1.27	4.41
	Past GDPD	1.26	1.28	4.65
	Future GDPD	1.14	1.22	1.58
UN change	CPI	1.12	1.06	**-3.29**
	PCE	1.14	1.09	-1.87
	Past GDPD	1.17	1.12	-0.81
	Future GDPD	1.14	1.06	-3.24
Past RGDP	CPI	1.22	1.21	2.10
	PCE	1.22	1.21	2.65
	Past GDPD	1.24	1.22	3.16
	Future GDPD	1.17	1.15	0.03
Future RGDP	CPI	1.26	1.24	3.08
	PCE	1.26	1.25	3.80
	Past GDPD	1.29	1.26	4.33
	Future GDPD	1.21	1.16	0.39
BMA, Recursive		**1.00**	**1.00**	0.00

is driven by the performance of these rules during the 2001 recession, with the Taylor rule including only the change in the unemployment rate the only one that sees performance increase relative to BMA during this recession. The superior performance of this rule is interesting, as the change in the unemployment rate is rarely included in other studies estimating the FOMC's policy reaction function.

Conclusion

The Taylor rule, which has been justified by both its theoretical elegance and empirical success, is the standard way to formulate monetary policy in macroeconomic models. However, using Bayesian Model Averaging (BMA) with many potential variables, I have shown that virtually no posterior probability is

FIGURE 8. Cumulative Sum of Log Predictive Density Relative to BMA



assigned to generalized Taylor-type rules that include one lag of the Federal Funds rate, one measure of inflation, and one measure of either the output gap or output growth. In addition, I find that in a forecasting exercise rules formed using BMA outperform all generalized Taylor-type rules when forecasting performance is judged by either Root Mean Squared Forecast Error (RMSFE) or Mean Absolute Forecast Error (MAFE). Both of these results suggest that most policy rules considered in empirical and theoretical settings are misspecified, and that is is important to model the FOMC as responding to many variables.

My analysis also reveals that the FOMC focuses more on the change in employment than the change in output, and that the FOMC is forward looking. The former result makes intuitive sense, because the Federal Reserve is mandated with promoting maximum sustainable employment, not maximum sustainable output.[12] The latter result, that the FOMC is forward looking, also aligns with the commonly held view that the FOMC should be proactive rather than reactive in an effort to smooth business cycles and prevent inflation before it happens. However, this view is not yet ubiquitous in the profession, as many empirical studies of the Taylor rule and many theoretical models use a backward looking policy rule.

Finally, I find that the point estimate of the long-run inflation response coefficient is about twice as large than in comparable studies, and that the Taylor principle has been satisfied throughout the entire 1970-2007 sample. I find that the relatively large inflation response coefficient is mostly driven by my use of meeting-based timing. This indicates that monthly or quarterly averages of the Federal Funds rate introduce measurement error and dampen the observed inflation response coefficient. The fact that I find that the Taylor principle has been satisfied over the full sample adds to the growing body of research (e.g. Orphanides (2004)) that the high inflation of the 1970s was not driven by a weak inflation response.

These findings are important for economic historians, macroeconomists studying policy in theoretical models, and policymakers. For economic historians, it is useful and interesting to know how the FOMC has set policy in the past.

[12] Of course, these two statistics are linked, so in theory the FOMC may respond to changes in GDP since changes in GDP may lead to changes in employment. In practice, I find this is not the case.

For macroeconomists studying policy in theoretical models, it is important to know what form the interest rate rule has, so that they can accurately represent it in their model. Changing the form of the interest rate rule could impact policy analysis, and models with a misspecified interest rate rule may fail to deliver accurate results. Finally, for policymakers, it is important to know how policy has been set in the past, as that often serves as a guide for what to do in the future.

CHAPTER III

THE FOMC'S INTEREST RATE RULE: AN MS-SSVS APPROACH

Introduction

The traditional view of monetary policy is that the 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, has served as 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 the chair of the FOMC in 1979. However, later studies cast doubt on this finding. Orphanides (2004) finds evidence in favor of a change in the output response, but not in the 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 to find 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 policy rules can lead to differences in inflation rates, volatility, and 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.

This newly developed 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 in 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 using 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 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. Coefficient 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 using 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: 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 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 corresponds 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".

Econometric Model

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

$$y_t = X_t \beta_{St} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

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

The regime, S_t follows a first order Markov process:

$$Pr(S_t = j | S_{t-1} = i) = p_{ij}$$

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

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 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 in regime, (3) it belongs in each regime, but the true effect is the same regardless of regime, (4) it belongs in only one of the two regimes. This only increases the burden of model comparison, as the number of models to consider expands 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 that nests these four possibilities within a single model. To do this, I build on the 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 the 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 $\sqrt{\widehat{Var(b_k)}}$ is the standard error of β under the assumption that there is no regime switching, and τ_0, τ_1, and ϵ are hyper-parameters chosen by the researcher that control the variance of each distribution in the prior for β^k.[1]

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.

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

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.

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

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

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

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

It is important to note that 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 $\beta^k \neq 0$, and 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 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]

[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.

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:

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

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

$$\Gamma_k^{(z)} = \text{ Categorical} \begin{pmatrix} \frac{p_1 f(N(0,\Sigma_1)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_2 f(N(0,\Sigma_2)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_3 f(N(0,\Sigma_3)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_4 f(N(0,\Sigma_4)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \\ \frac{p_5 f(N(0,\Sigma_5)|\beta^k)}{\sum_{i=1}^5 p_i f(N(0,\Sigma_i)|\beta^k)} \end{pmatrix}$$

where Σ_i is the variance/covariance matrix of the i^{th} prior mixture distribution, described on page 75. 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, from 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:

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

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

3. Sample the variance of the regression error:

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

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 $S_T^{(z)}$.

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:

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

In my application, I set the following prior probabilities on the mixture distributions:

$$Pr(\gamma_1^k = 1) = 0.4$$

$$Pr(\gamma_2^k = 1) = 0.2$$

$$Pr(\gamma_3^k = 1) = 0.2$$

$$Pr(\gamma_4^k = 1) = 0.1$$

$$Pr(\gamma_5^k = 1) = 0.1$$

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:

$$\tau_0^k = 0.1\sqrt{\widehat{var(\beta^k)}}$$

$$\tau_1^k = 15.0\sqrt{\widehat{var(\beta^k)}}$$

where $\widehat{Var(b^k)}$ is the uncertainty associated with the OLS estimate 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 (9)-(11), 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 9. Prior Probability Density Function for Different Values of β_0 and β_1

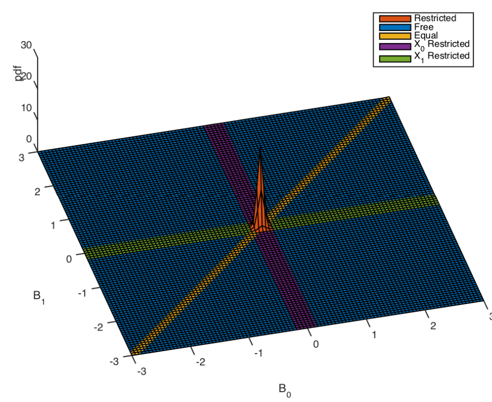


FIGURE 10. Prior Probability Density Function for Different Values of β_0 and β_1: View from Above

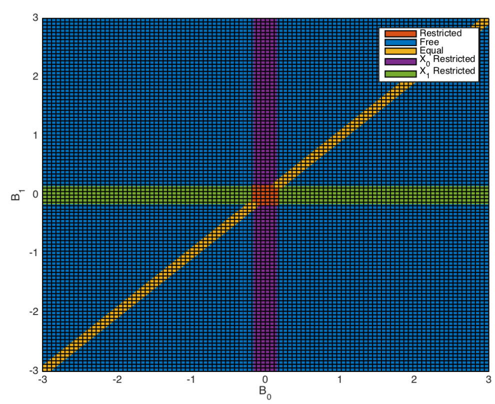
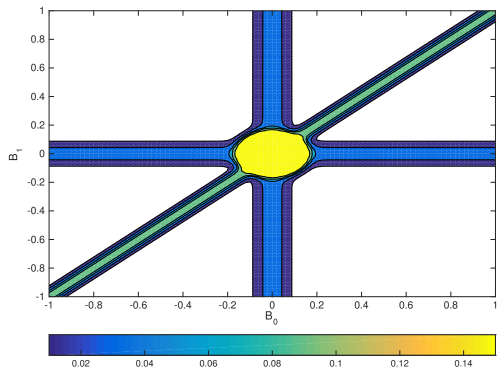


FIGURE 11. Contour Plot of the Prior Probability Density Function for Different Values of β_0 and β_1



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.

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:

$$y_t = X_t \beta_{St} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$Pr(S_t = j | S_{t-1} = i) = p_{ij}$$

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, the 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, since σ^2 will vary across simulations.

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 described in Table 14.

TABLE 14. Monte-Carlo Priors

Parameter	Prior Mean	Prior S.D.
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

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 Tables 15, 16, 17, 18, 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, in the first column of Table 15, 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 "correct" 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, define the regression intercept 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 also not restricted to be identical to each other.

TABLE 15. $\sigma = 0.1$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	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 16. $\sigma = 0.5$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	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 17. $\sigma = 1.0$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	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 18. $\sigma = 2.0$

	$T = 50$	$T = 100$	$T = 150$	$T = 200$	$T = 250$
$\mu_0 \neq 0$	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. 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 less accurate in most simulations).

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 19, 20, 21, 22.

TABLE 19. $\sigma = 0.1$, 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%

TABLE 20. $\sigma = 0.5$, 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 21. $\sigma = 1.0$, 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%

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.

TABLE 22. $\sigma = 2.0$, 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%

This may be partially driven by the fact that there is only a 20% prior probability placed on each of these 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. Given this, 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, in some circumstances 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 the regressors that were restricted to zero at the mode.

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 the policy reaction function literature, and estimate a rule of the form:

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

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

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

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 interest rate srule 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 FOMC. 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 was important for both in-sample fit and for out-of-sample forecasting. 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 when estimating a similar interest rate rule inside of a VAR, 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.

[3] Allowing for stochastic volatility requires replacing step (3) on page 80 with the procedure described in Kim et al. (1998).

For example, when estimating a Markov-Switching model on U.S. GDP data, the researcher typically restricts the mean growth rate so that in one regime the mean 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.

Due to 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 equal. 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 found that the coefficients on all other variables, including the inflation gap, were unimodal and nearly identical. This evidence contradicts the findings of 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.

Convergence Diagnostics

Below, I present evidence that the estimator converges to a unique stationary distribution. I first present running mean plots throughout the burn-in samples in Figures 12 and 13. 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 in Figures 14 and 15. 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.

Running Mean Plots

FIGURE 12. Regime Probabilities

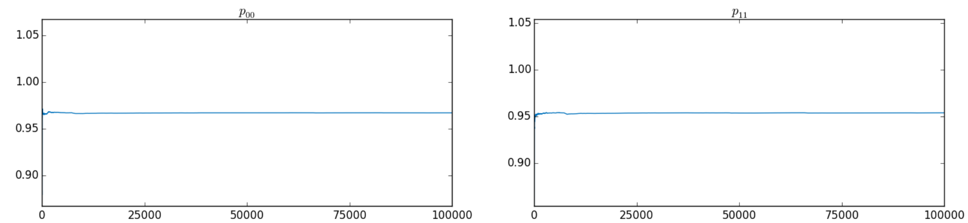
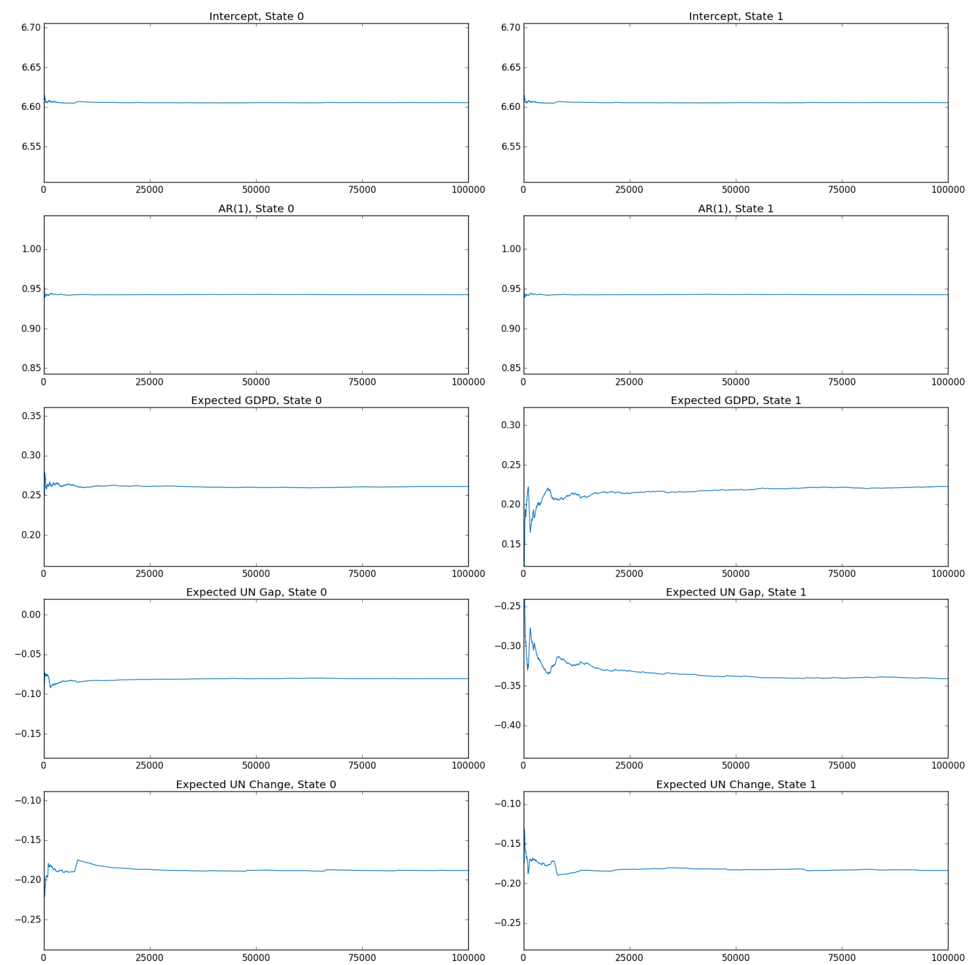


FIGURE 13. Regression Coefficients



Autocorrelation Functions

FIGURE 14. Regime Probabilities

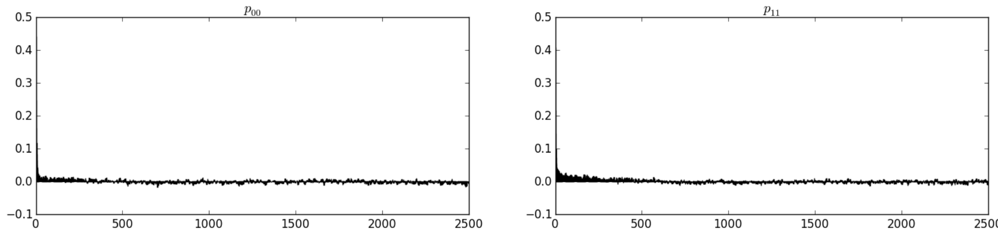
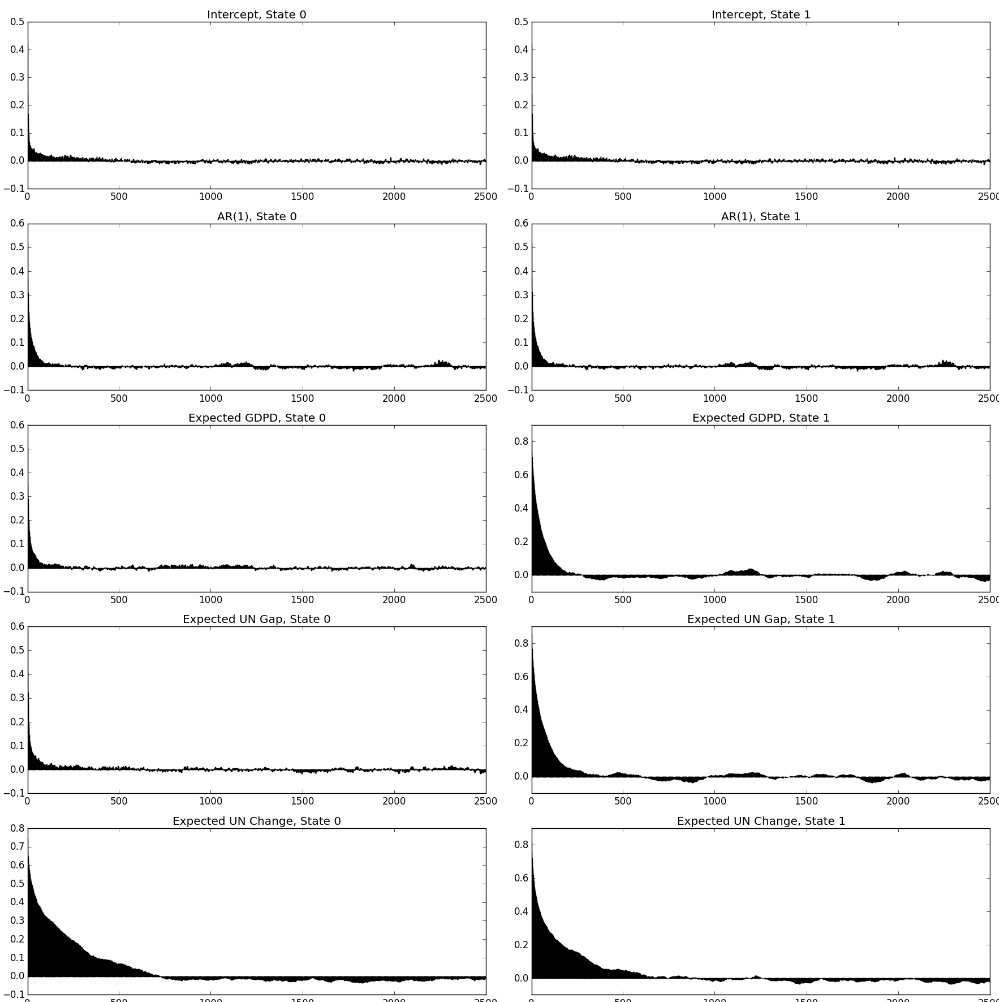


FIGURE 15. Regression Coefficients



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.[4]

Results

Next, I present my results. First, in Tables 23, 24, and 25, I present estimates of the restriction probabilities and regression coefficients in each regime. Second, in Figures 16 and 17, I plot the coefficient densities for each parameter of the model. For the regression coefficients, I display the densities under each regime on the same plot. Third, in Figures 18 and 19, 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, in Figure 20, I display the estimate of the volatility term over time, along with the uncertainty associated with it.

<u>Restriction Probabilities and Mean Regression Coefficient Estimates</u>

[4] I have repeated this exercise with $p_1 = p_2 = p_3 = p_4 = p_5 = 0.2$ and found nearly identical results to what is presented in this section and in the Results section below.

TABLE 23. 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 24. 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 25. 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

Coefficient Densities

FIGURE 16. Regression Coefficients

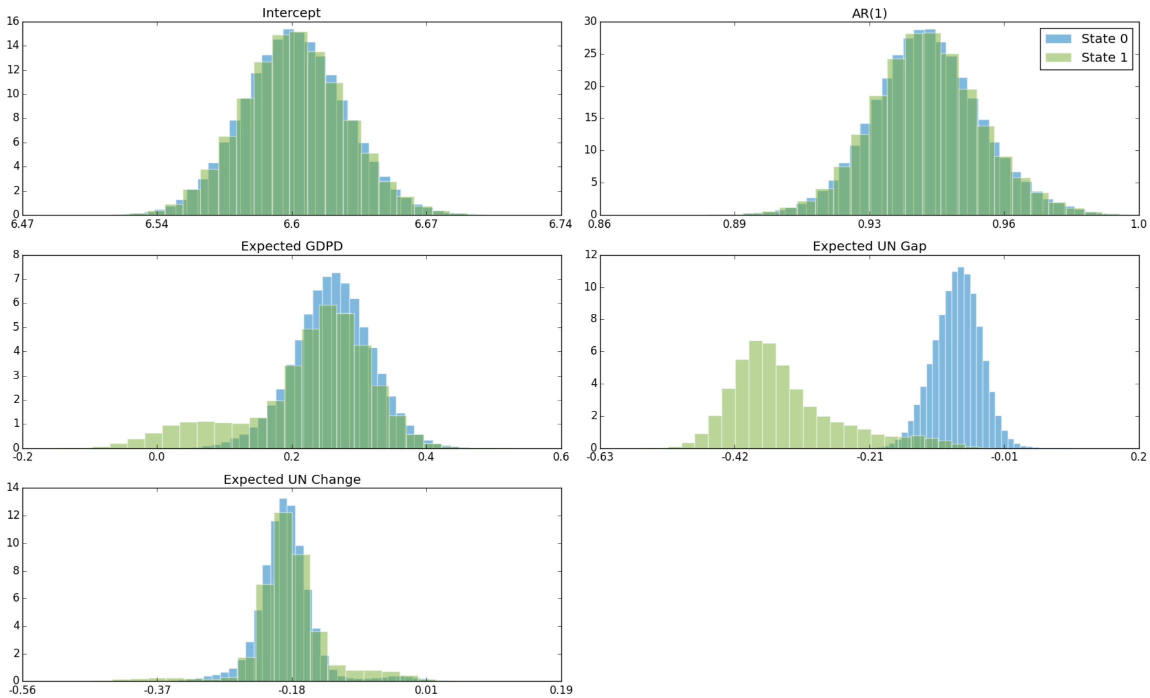
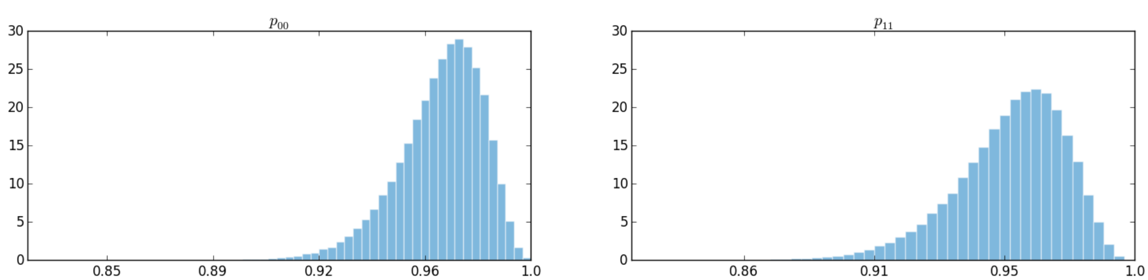


FIGURE 17. Transition Probabilities



Regime Estimation

FIGURE 18. Probability of Weak Unemployment Response Regime

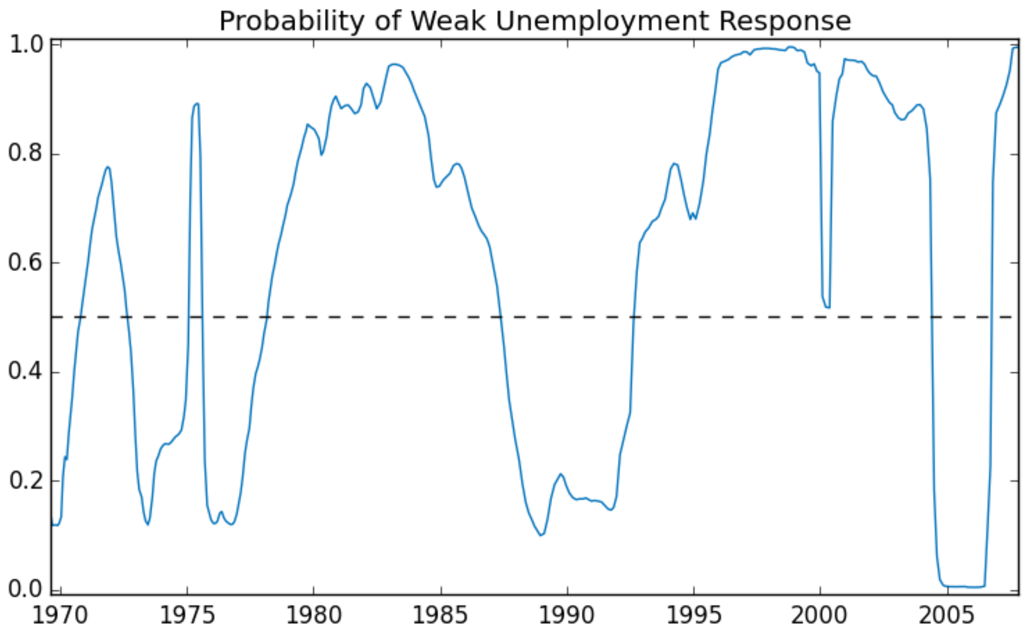
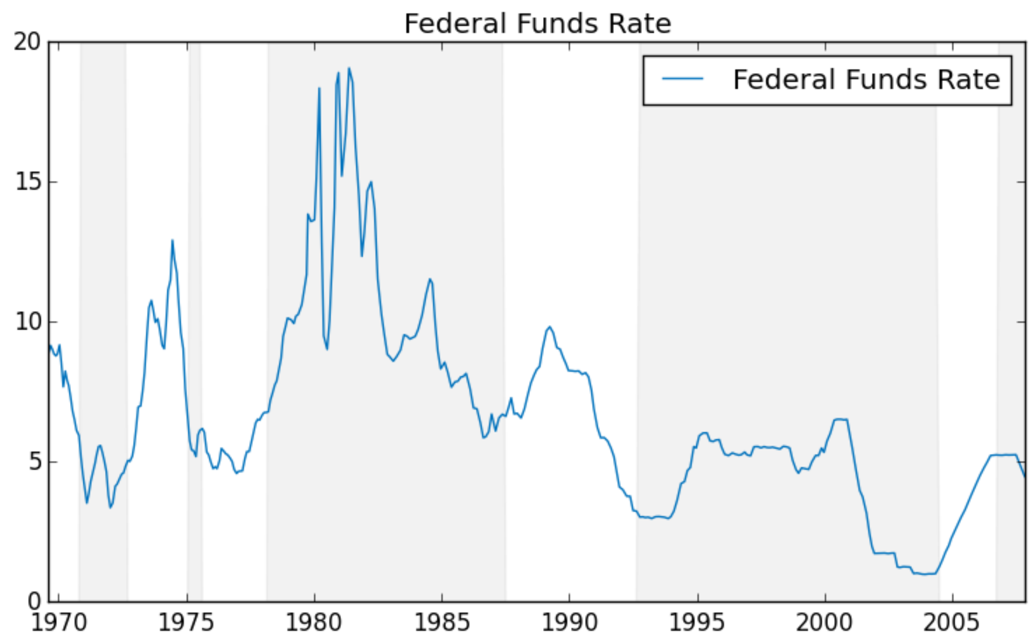


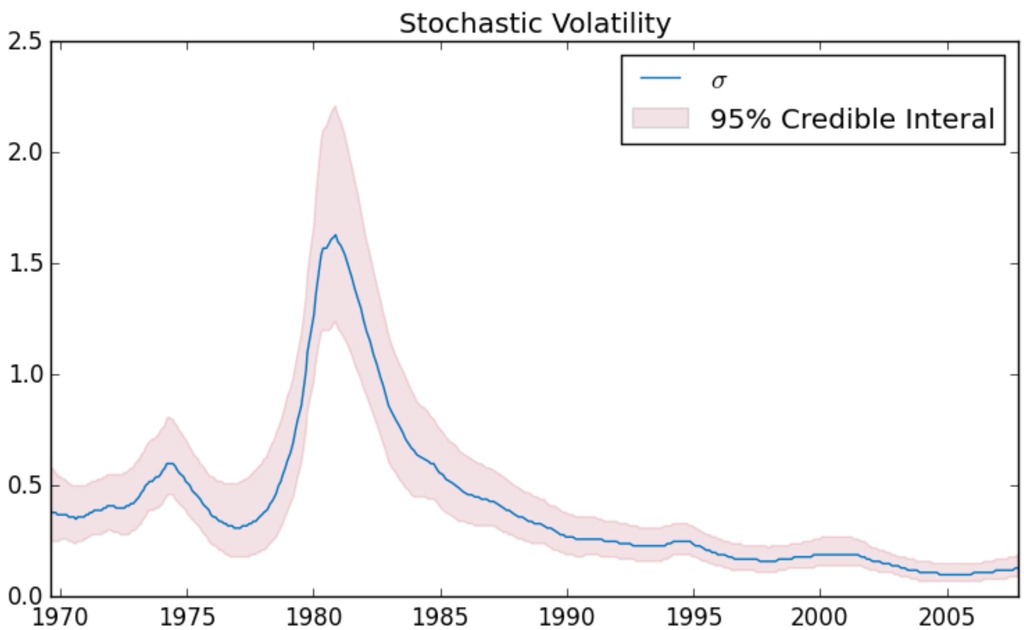
FIGURE 19. Federal Funds Rate



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

Stochastic Volatility

FIGURE 20.



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 targeted the money supply rather than the Federal Funds rate during this period. Therefore, we should expect the error term of a linear rule that 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 unemployment 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 accommodative policy during 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 decline 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.[5]

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 procedure is still able to identify these restrictions as the amount of noise grows, and it is able to

[5] 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.

detect zero-restrictions with a very 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 CITED

Berger, J. O. and Yang, R. Y. (1994). Non informative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, 10:461–482.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Beveridge, S. and Nelson, C. R. (1981). A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle. *Journal of Monetary Economics*, 7:151–174.

Blanchard, O. and Watson, M. (1982). Bubbles, rational expectations, and financial markets. In Wachter, P., editor, *Crises in the Economic and Financial Structure*, pages 295–315. Lexington Books, Lexington, MA.

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.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Chan, J. C. C., Koop, G., and Potter, S. M. (2013). A New Model of Trend Inflation. *Journal of Business & Economic Statistics*, 31(1):94–106.

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

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90:1313–1321.

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.

Cochrane, J. H. (1992). Explaining the Variance of Price-Dividend Ratios. *Review of Financial Studies*, 5(2):243–80.

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

Coibion, O. and Gorodnichenko, Y. (2011). Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation. *American Economic Review*, 101(1):341–70.

Cúrdia, V., Ferrero, A., Ng, G. C., and Tambalotti, A. (2011). Evaluating interest rate rules in an estimated DSGE model. Staff Reports 510, Federal Reserve Bank of New York.

Davig, T. and Doh, T. (2008). Monetary policy regime shifts and inflation persistence. Research Working Paper RWP 08-16, Federal Reserve Bank of Kansas City.

Diba, B. T. and Grossman, H. I. (1988). Explosive Rational Bubbles in Stock Prices? *American Economic Review*, 78(3):520–30.

Eo, Y. and Kim, C.-J. (2012). Markov-Switching Models with Evolving Regime-Specific Parameters: Are Post-War Booms or Recessions All Alike? Working Papers 2012-04, University of Sydney, School of Economics.

Evans, G. W. (1991). Pitfalls in Testing for Explosive Bubbles in Asset Prices. *American Economic Review*, 81(4):922–30.

Fernandez, C., Ley, E., and Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100(2):381–427.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2004). Sequential Monte Carlo Filtering: An Example. Note, University Pennsylvania.

Flood, R., Hodrick, R., and Kaplan, P. (1994). An evaluation of recent evidence on stock price bubbles. In Flood, R. and Garber, P., editors, *Speculative Bubbles, Speculative Attacks, and Policy Switching*, pages 105–133. MIT Press, Cambridge, MA.

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.

Geweke, J. and Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1):1–29.

Gürkaynak, R. S. (2008). Econometric Tests Of Asset Price Bubbles: Taking Stock . *Journal of Economic Surveys*, 22(1):166–186.

Hall, S. G., Psaradakis, Z., and Sola, M. (1999). Detecting Periodically Collapsing Bubbles: A Markov-Switching Unit Root Test. *Journal of Applied Econometrics*, 14(2):143–54.

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.

Kindleberger, C. P. (2000). *Manias, Panics, and Crashes*. John Wiley & Sons, 4 edition.

Koop, G. (2003). *Bayesian Econometrics*. Wiley.

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

Koop, G. and Potter, S. M. (2007). Estimation and Forecasting in Models with Multiple Breaks. *Review of Economic Studies*, 74(3):763–789.

Koop, G. and Potter, S. M. (2011). Time varying VARs with inequality restrictions. *Journal of Economic Dynamics and Control*, 35(7):1126–1138.

LeRoy, S. F. and Porter, R. D. (1981). The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica*, 49(3):555–74.

Lubrano, M. (1995). Testing for unit roots in a Bayesian framework. *Journal of Econometrics*, 69(1):81–109.

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.

Phillips, P. C. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

Phillips, P. C. B., Wu, Y., and Yu, J. (2011). Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values? *International Economic Review*, 52(1):201–226.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*, 72(3):821–852.

Schorfheide, F. (2005). Learning and Monetary Policy Shifts. *Review of Economic Dynamics*, 8(2):392–419.

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review*, 71(3):421–36.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control*, 12(2-3):463–474.

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. (1999). A Historical Analysis of Monetary Policy Rules. In *Monetary Policy Rules*, NBER Chapters, pages 319–348. National Bureau of Economic Research, Inc.

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

Tirole, J. (1982). On the Possibility of Speculation under Rational Expectations. *Econometrica*, 50(5):1163–81.

Tirole, J. (1985). Asset Bubbles and Overlapping Generations. *Econometrica*, 53(6):1499–1528.

West, K. D. (1987). A Specification Test for Speculative Bubbles. *The Quarterly Journal of Economics*, 102(3):553–80.

Xia, C. and Griffiths, W. (2012). Bayesian unit root testing: The effect of choice of prior on test outcomes. In Terrell, D. and Millimet, D., editors, *Advances in Econometrics: 30th Anniversary Edition*, volume 30, pages 27–57. Emerald Group Publishing Limited, Bingley, UK.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In Goel, P. K. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*.