Modeling Monetary Policy Dynamics: A Comparison of Regime Switching and Time Varying Parameter Approaches

Aeimit Lakdawala*
Michigan State University

April 2015

Abstract

Structural VAR models have been widely used to model monetary policy dynamics. Typically, a choice is made between regime-switching models and time-varying parameter models. In this paper we use a canonical model of monetary policy and estimate both types of time variation in monetary policy while also allowing for changing variances. The models are compared using marginal likelihood and forecast performance evaluation. We analyze whether the two frameworks identify similar dynamic patterns in the data and if the decision to choose one method over the other matters for the identification of monetary policy shocks.

Keywords: Structural VAR, monetary policy, regime-switching, time-varying parameter, model comparison

*Email: aeimit@msu.edu, Address: 486 W. Circle Drive, 110 Marshall-Adams Hall, East Lansing, MI 48824
1 Introduction

There is a large literature that studies changes in the behavior of monetary policy setting in the post-war US data. The most common way to model monetary policy, both in structural macro models and reduced form VAR models, is to use a Taylor-type rule. This is a simple linear rule where the short term interest rate is adjusted in response to some measure of inflation and economic activity. To allow for changes in the behavior of monetary policy, the parameters of this simple rule are allowed to change over time. From a modeling perspective, the econometrician has to specify some process for the time variation in the parameters. There are two common approaches that have been used in the literature. The first one, known as the regime switching (rs) approach, assumes that there are discrete changes in the parameters which are governed by a Markov switching variable. This approach has been used in VAR analysis by Sims and Zha (2006) and in structural DSGE models by Bianchi (2013) and Liu et al. (2011) among others. The other approach, known as the time-varying parameter (tvp) approach, assumes that there are gradual changes in the parameters. The time variation is typically specified as an autoregressive process. This has been used in VAR analysis by Cogley and Sargent (2005) and Primiceri (2005) and in structural DSGE models by Fernandez-Villaverde et al. (2010). Sometimes, a theoretical model is used to motivate the choice between the two frameworks, but more often the choice turns out to be driven by convenience and tractability. In this paper we compare the two approaches using a structural VAR model and explore its implications for inference about monetary policy dynamics.

Specifically, we want to study whether researchers that pick one method over the other end up arriving at different answers to important monetary policy related questions. First, there is a debate in the literature regarding changes in the macroeconomy since the 1980s. Was there a change in monetary policy or was it just a change in the volatility of shocks hitting the economy, or perhaps both? One approach has been to estimate models with time variation allowed in both the monetary policy parameters and the shock volatility parameters. Then some form of model comparison is done to pick the specification which best fits the data. In this paper we want to know whether using a tvp or rs approach gives rise to a different conclusion. Second if there has been a change in monetary policy, how does the choice of tvp vs. rs affect the characterization of historical changes in monetary policy? For instance, was there a big dramatic change in the reaction of the Federal Reserve to

\footnote{See Primiceri (2005) and Sargent et al. (2006) for examples of this approach.}
inflation when Paul Volcker was elected or was it gradual and had it begun before the arrival of Volcker? A related issue is whether Federal Reserve policy in the 1970s involved increasing interest rates more than one for one with inflation, satisfying the Taylor principle. Finally, how does this choice affect the evaluation of monetary policy’s effects on the economy? A key component in this evaluation is the identification of monetary policy shocks and we investigate whether the choice of rs vs. tvp matters for what monetary shocks are identified.

We consider a canonical three variable VAR with unemployment, inflation and the federal funds rate. The structural VAR is identified using the recursive (also called triangular) identifying restriction that is most common in the monetary policy literature. Under this assumption, both inflation and unemployment react with a lag to changes in the federal funds rate while the federal funds rate is allowed to react contemporaneously to inflation and unemployment. The equation corresponding to the interest rate is interpreted as the monetary policy reaction function. The coefficients of this policy rule are allowed to change over time, either using a time-varying parameter or regime-switching framework. Correspondingly, the variance of the shocks are also allowed to change over time. The model is estimated using a Bayesian Markov Chain Monte Carlo algorithm. We use both marginal likelihood and forecasting performance to evaluate the fit of the models.

The main results regarding the monetary policy questions are as follows. First, if a researcher were to adhere to either the tvp or rs framework and consider the best fit model within that framework, the specification with changes only in the variance would be picked for both the rs and tvp specifications. Even though there are some differences in the dynamics of the estimated changes in the variances, the monetary policy shocks identified are almost identical implying very similar effects of monetary policy on the economy. But, as motivated in Benati and Surico (2009), it is interesting to look at a framework with just the changes in the monetary policy coefficients. Here we find that the dynamic patterns are different between the tvp and rs cases. Specifically, the changes in the response to inflation are more gradual in the tvp case. Additionally, the Taylor principle is satisfied for essentially the whole sample. On the other hand, the rs specification implies frequent changes in the 1970s between high response to inflation that satisfies Taylor principle and low responses that do not. As a result the monetary policy shocks identified by the rs regime in the 1970s are quite different (especially in the 1970s) from the tvp case.

To the best of our knowledge, this is the first paper that explicitly compares the implications of the choice of time-varying parameter vs. regime switching approaches for monetary
policy dynamics. There is a growing literature that compares the forecasting performance of alternative macroeconomic models. Clark and Ravazzolo (2014) compare the forecasting performance of a variety of AR and VAR macro models using different specifications of changing volatility of the shocks. D’Agostino et al. (2013) use a tvp model with stochastic volatility and compare its forecasting performance with various other models and find that modeling changes in the coefficients of the VAR delivers superior forecasts. Both these papers focus on forecasting, use reduced form models and do not study any regime-switching models. Perhaps the study closest to ours is Koop et al. (2009). They use a mixture innovation model to examine the transmission mechanism of monetary policy. Their model allows for multiple structural breaks where the number of breaks is modeled using a hierarchical prior setup. Their model nests the tvp model but not the regime switching model. Finally, they allow for variation in the structural equations governing inflation and unemployment as well as the monetary policy equation. There is a strand of the literature that tries to build a more flexible approach to modeling time variation, see for example Hamilton (2001), Koop and Potter (2007). Koop and Potter (2010) consider a more general approach using the concepts of hypothetical data re-ordering and distance between observations. But there is no empirical work using these models that studies the dynamics of monetary policy and appears to be a promising area for future research.

The rest of the paper is organized as follows. The next section describes the model and explains the rs and tvp framework. Section 3 discusses the priors and gives an overview of the estimation methodology, including details on the method used to calculate the marginal likelihood. Further estimation details are provided in an online appendix. The results are discussed in section 4 and section 5 provides some concluding remarks.

2 The Model

We consider a simple 3 variable VAR with $p$ lags for $y_t = [u_t, \pi_t, i_t]'$ with unemployment ($u_t$), inflation ($\pi_t$) and the fed funds rate ($i_t$).

$$y_t = a_{0,t} + \sum_{j=1}^{p} b_{t,j} y_{t-j} + e_t, \quad e_t \sim N(0, \Omega_t)$$  \hspace{1cm} (1)$$

$$\Omega_t = A_t^{-1}\Sigma_t \Sigma_t' A_t^{-1'}$$  \hspace{1cm} (2)$$
Equation 2 represents the triangular decomposition of the reduced-form covariance matrix, where $A_t$ is a lower triangular matrix with 1s on the diagonal and $\Sigma_t$ is a diagonal matrix and $b_{i,j}$ is a 3 x 3 matrix of coefficients. This VAR is popular in the literature and has been used both in the regime-switching case (see Sims et al. (2008)) and in the time-varying parameter case (see Primiceri (2005)). We can write each equation of the VAR as

$$y_{nt} = z'_{nt} \beta_n + \epsilon_{nt}$$

for $n = 1, 2, 3$ and $z_{nt} = z_t = [1, y'_{t-1}, ..., y'_{t-p}]$ is 1 x $k$ and $\beta_n$ (which stacks the $n$th rows of the $b_{i,j}$ matrices) is $k$ x 1, with $k = 3p + 1$. We can now stack the $z_{n,t}$ into a 3 x $r$ matrix $Z_t$ and $\beta_n$ into a $r$ x 1 vector ($r = 3k$) to get the following equation

$$y_t = Z_t \beta_t + A_t^{-1} \Sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, I)$$

with

$$Z_t = \begin{bmatrix} z'_{t-1} & 0 & 0 \\ 0 & z'_{t-2} & 0 \\ 0 & 0 & z'_{t-3} \end{bmatrix}, \beta_t = \begin{pmatrix} \beta_{u,t} \\ \beta_{\pi,t} \\ \beta_{i,t} \end{pmatrix}$$

To give this reduced-form VAR a structural interpretation some identification restrictions are required. The most common identification assumption in the monetary policy literature is the recursive (or triangular) identification, see Christiano et al. (1999) for a survey. In this identification scheme the Fed Funds rate is ordered last, with the implied assumption that unemployment and inflation do not react contemporaneously to changes in the fed funds rate but only with a lag. The fed funds rate is allowed to react contemporaneously to inflation and unemployment. With this assumption in place, the matrix governing the contemporaneous relationships between the three variables exactly coincides with the lower-triangular matrix $A_t$ that comes from the triangular-decomposition of the reduced-form covariance matrix $\Omega_t$. Alternative identification approaches have been used in the literature. See Gambetti et al. (2008) for an example of sign restrictions and Sims and Zha (2006) for non-recursive zero restrictions. Arguably, the recursive approach remains the most popular approach and this motivates its use in the current paper.

Notice that the reduced form covariance matrix $\Omega_t$ is indexed by $t$. One component of the time dependency comes from time variation in $A_t$, which represents the time variation in the contemporaneous relationships between the three variables. Additionally the standard
deviation of the structural shocks $\Sigma_t$ is also allowed to depend on $t$. There is a large literature that has documented changes in the variance of the shocks, for example see Sims and Zha (2006) and Primiceri (2005). Thus we will allow for the variance of the shocks to all three variables to change over time. Since the focus of this paper is modeling monetary policy dynamics we will consider time variation in the parameters only in the monetary policy equation, the 3rd equation in the system shown above. This is a reasonable assumption as Sims and Zha (2006) find that the specification with changes only in the monetary policy equation fit the data better than a specification where parameters of the other equations change as well. Thus in this paper we will consider models where there are changes only in the monetary policy equation (referred to with a $b$), only in the variance of the shocks ($v$) and changes in both ($vb$). The parameters that are allowed to change over time have a $t$ subscript. Note that with the structural interpretation the parameters $\alpha_{31,t}$ and $\alpha_{32,t}$ of the $A_t$ matrix represent the contemporaneous response of interest rates to unemployment and inflation respectively. For expositional clarity the equation $y_t = Z_t \beta + A_t^{-1} \Sigma_t \varepsilon_t$ can be expanded as follows

$$
\begin{pmatrix}
  u_t \\
  \pi_t \\
  i_t 
\end{pmatrix} = Z_t \begin{pmatrix}
  \beta_u \\
  \beta_\pi \\
  \beta_{i,t} 
\end{pmatrix} + \begin{pmatrix}
  1 & 0 & 0 \\
  \alpha_{21} & 1 & 0 \\
  \alpha_{31,t} & \alpha_{32,t} & 1 
\end{pmatrix}^{-1} \begin{pmatrix}
  \sigma_{u,t} & 0 & 0 \\
  0 & \sigma_{\pi,t} & 0 \\
  0 & 0 & \sigma_{i,t} 
\end{pmatrix} \varepsilon_t
$$

(6)

2.1 Regime Switching

In the regime switching framework the parameters of the interest rate equation in $\beta_{i,t}$ and $A_i,t = [\alpha_{31,t}, \alpha_{32,t}]$ and all the parameters in $\Sigma_t$ depend on $s_t$, which is an unobserved Markov switching variable that can take on values 1, 2, ... $M$. We can write the model as

$$
\begin{pmatrix}
  u_t \\
  \pi_t \\
  i_t 
\end{pmatrix} = Z_t \begin{pmatrix}
  \beta_u \\
  \beta_\pi \\
  \beta_{i,(s_t)} 
\end{pmatrix} + \begin{pmatrix}
  1 & 0 & 0 \\
  \alpha_{21} & 1 & 0 \\
  \alpha_{31}(s_t) & \alpha_{32}(s_t) & 1 
\end{pmatrix}^{-1} \begin{pmatrix}
  \sigma_{u}(s_t) & 0 & 0 \\
  0 & \sigma_{\pi}(s_t) & 0 \\
  0 & 0 & \sigma_{i}(s_t) 
\end{pmatrix} \varepsilon_t
$$

(7)

$$
p(s_t = j|s_{t-1} = i) = p_{ij}
$$

(8)

Thus the estimation involves estimating $M$ different sets of parameters for $\beta_{i,(s_t)}, A_i(s_t)$ and $\Sigma(s_t)$. In the baseline specification we model the coefficients of the monetary policy equation governed by the Markov switching variable $s_{1,t}$ which is independent from the
variable \( s_{2,t} \) which governs the switching of the variances. Note that we can combine the two processes \( s_{1,t} \) with \( M_1 \) regimes and \( s_{2,t} \) with \( M_2 \) regimes into the process \( s_t \) with \( M_1 \times M_2 \) regimes. We have also tried the specification where both the monetary policy coefficients and the variances depend on the same Markov process, but found that the data prefer the independent specification. This same finding is also reported in Sims and Zha (2006). In the baseline results we consider up to 4 total regimes. The transition matrix for the Markov switching variable is left unrestricted.

### 2.2 Time-varying parameter

In the time-varying parameter framework the parameters are modeled as latent variables where a law of motion is specified for their dynamics. The most common approach in the monetary policy literature is to model the parameters as following random walks.

\[
\begin{pmatrix}
  u_t \\
  \pi_t \\
  i_t
\end{pmatrix} = Z_t \begin{pmatrix}
  \beta_u \\
  \beta_\pi \\
  \beta_{i,t}
\end{pmatrix} + \begin{pmatrix}
  1 & 0 & 0 \\
  \alpha_{21} & 1 & 0 \\
  \alpha_{31} & \alpha_{32} & 1
\end{pmatrix}^{-1} \begin{pmatrix}
  \sigma_{u,t} & 0 & 0 \\
  0 & \sigma_{\pi,t} & 0 \\
  0 & 0 & \sigma_{i,t}
\end{pmatrix} \varepsilon_t
\]  

\( (9) \)

\[
\beta_{i,t} = \beta_{i,t-1} + \nu_t, \quad \nu_t \sim N(0, Q) \tag{10}
\]

\[
A_{i,t} = A_{i,t-1} + \zeta_t, \quad \zeta_t \sim N(0, S) \tag{11}
\]

\[
\log(\sigma_{j,t}) = \log(\sigma_{j,t-1}) + \eta_{j,t}, \quad \eta_t \sim N(0, W), j \in \{u, \pi, i\} \tag{12}
\]

The covariance matrix \( W \) of the innovations to the log volatility process is assumed to be diagonal following Cogley and Sargent (2005). Thus the tvp setup allows for a break in the parameters in every time period, where the estimates of \( Q, S \) and \( W \) will govern the amount of estimated time variation.
3 Priors and Estimation

3.1 Priors

For the rs case we use symmetric priors, so that regardless of the regime the prior distribution is the same. For the coefficients we use the so-called "Minnesota" style prior, which provides shrinkage towards a random walk. Specifically the coefficients of the VAR are assumed to follow a normal prior $\beta \sim N(m_\beta, M_\beta)$. Note that the $\beta$ includes the parameters of both the monetary and non-monetary equations. $m_\beta$ is set to 1 for the parameters corresponding to the first own lag and the rest are set to zero. The prior variance for coefficients of each equation is set in the following way. $\sqrt{M_\beta} = \frac{\lambda_0 \lambda_1}{\delta_j L(j)^3}$ where $\delta_j$ is the standard deviation of a univariate autoregression of equation $j$ and $L(j)$ represents the order of the lag in equation $j$. This prior variance embodies the idea that own lags are more likely to be important predictors than lags of other variables and the predictive power diminishes as the lag length increases. Following Sims and Zha (1998) we set $\lambda_0 = 1$, $\lambda_1 = 0.2$ and $\lambda_2 = \lambda_3 = \lambda_4 = 1$. The parameters of $A$ matrix are assumed to have a normal prior with mean zero and a large variance. The inverse of the standard deviations $\sigma_j^{-1}$ are assumed to have a Gamma prior, again with priors being symmetric across regimes. Finally for the transition matrix $P$ which contains the regime probabilities $p_{i,j}$ we use a Beta prior when there are two regimes and a Dirichlet prior when there are more than two regimes. The prior parameter vector $\tilde{\alpha}$ of the Beta and Dirichlet distributions are set so that they imply a probability of 0.85 of staying in the same regime and an equal probability of $\frac{1-0.85}{M-1}$ of moving to any another regime, where $M$ is the total number of regimes. The priors are thus summarized as follows with $V_A = 10,000$, $k = \theta = 1$ and $m_\beta$, $M_\beta$ and $\tilde{\alpha}$ set as described above.

$$A \sim N(0, V_A)$$  \hfill (13)

$$\beta \sim N(m_\beta, M_\beta)$$  \hfill (14)

$$\sigma_j^{-1} \sim G(k, \theta)$$  \hfill (15)

$$P \sim Dir(\tilde{\alpha})$$  \hfill (16)

In the tvp case, to keep the priors similar to the rs case, we use a Minnesota-style prior for the constant parameters in $\beta$ with the same prior parameters. For the priors regarding the

\footnote{As a robustness test, we have also used a training sample prior for all the coefficients as explained below.}
latent variables, we follow Primiceri (2005) and use a training sample approach. The first 10 years of the sample is used to estimate a constant parameter VAR by OLS. Then the priors for the initial values of the latent variables are based on OLS estimates \( \log(\hat{\sigma}_{OLS}) \), \( \hat{A}_{i,OLS} \) and \( \beta_{i,OLS} \). The variance of the shocks in the random walk processes are assumed to follow inverse-Wishart distributions with the scale matrix that depends on the constant parameters, \( k_Q \), \( k_W \) and \( k_S \). For the benchmark results we use the values \( k_Q = 0.01 \), \( k_S = 0.1 \) and \( k_W = 0.01 \) that are very common in literature, see Primiceri (2005), D’Agostino et al. (2013) and Cogley and Sargent (2005) among others. Thus the prior setup is summarized as follows

\[
\beta_{u\pi} \sim N(m_{\beta_{u\pi}}, M_{\beta_{u\pi}}) \quad (17)
\]
\[
\alpha_{21} \sim N(0, 10,000) \quad (18)
\]
\[
\log(\hat{\sigma}_0) \sim N(\log(\hat{\sigma}_{OLS}), I) \quad (19)
\]
\[
\beta_{0,i} \sim N(\hat{\beta}_{i,OLS}, 4.V(\hat{\beta}_{i,OLS})) \quad (20)
\]
\[
A_{0,i} \sim N(\hat{A}_{i,OLS}, 4.V(\hat{A}_{i,OLS})) \quad (21)
\]
\[
Q \sim IW(k_Q^2, 40.V(\hat{\beta}_{i,OLS}), 40) \quad (22)
\]
\[
W \sim IW(k_W^2, 4.I_n, 4) \quad (23)
\]
\[
S \sim IW(k_S^2, 3.V(\hat{A}_{i,OLS}), 3) \quad (24)
\]

### 3.2 Estimation

For the tvp specification we use the algorithm outlined in Primiceri (2005). This is a Gibbs Sampler which uses the simulation smoother of Carter and Kohn (1994) and a mixture of normal approximation for the stochastic volatility based on Kim et al. (1998). We make sure to use the correct ordering as pointed out by Del Negro and Primiceri (2013). For the rs case we use a single block random-walk Metropolis-Hastings (M-H) algorithm. Although a Gibbs Sampler can also be constructed for the models used here, we found that the Metropolis-Hastings algorithm had good convergence properties and was much more convenient to use.\(^3\) All the estimation details are provided in an online appendix.

\(^3\)The basic random-walk M-H algorithm is the same for the various rs specifications and requires just changing the log-likelihood file while a Gibbs Sampler would have to be specifically tailored for each different specification.
3.3 Model Comparison

With a Bayesian estimation framework, a natural way to perform model comparison is to calculate the posterior odds. With a priori equal weight associated to each model the Bayesian posterior odds ratio boils down to comparing the marginal likelihood. Let \( \theta \) denote all the parameters and \( \xi^T \) the latent variables. Gathering all the parameters and latent variables in \( \Theta = \{ \theta, \xi^T \} \) the marginal likelihood is written as

\[
P(Y) = \int p(Y|\Theta)\pi(\Theta)d\Theta
\]  

(25)

where \( p(Y|\Theta) \) is likelihood and \( \pi(\Theta) \) is the prior. We will use the modified harmonic mean estimator of Gelfand and Dey (1994), with the truncated normal weighting function \( f() \) suggested by Geweke (1999).

\[
p(Y)^{-1} = \left[ \frac{1}{D} \sum_{i=1}^{D} \frac{f(\Theta^{(i)})}{p(Y|\Theta^{(i)})\pi(\Theta^{(i)})} \right]^{D}
\]

(26)

\( \Theta^{(i)} \) represents the \( ith \) draw from the posterior distribution, with \( D \) representing the total number of draws. Given the high dimension of latent variables we make some simplifying assumptions to aid in of the computation of the marginal likelihood, following Justiniano and Primiceri (2008). First we we assume an independent structure for both the priors and the weighting function for the parameters and the latent variables, \( \pi(\theta, \xi^T) = \pi(\theta)\pi(\xi^T) \) and \( f(\theta, \xi^T) = f(\theta)f(\xi^T) \). Next we assume that the weighting function of the latent variable is just equal to the prior, \( f(\xi^T) = \pi(\xi^T). \)

Given these assumptions, the marginal likelihood can be calculated using the simpler equation

\[
p(Y)^{-1} = \left[ \frac{1}{D} \sum_{i=1}^{D} \frac{f(\Theta^{(i)})}{p(Y|\Theta^{(i)})\pi(\Theta^{(i)})} \right]
\]

(27)

4 Results

The data used in the estimation are as follows. The unemployment rate is for civilians 16 years and over. Inflation is the annualized percentage change in the GDP Deflator. The

The assumption about the independent structure is fairly innocuous. For a detailed discussion about the second assumption see Sims et al. (2008).
The interest rate is the Federal Funds rate at an annualized rate. The data sample runs from 1954:Q4 to 2008:Q3. We use the first 10 years as a training sample so the effective sample runs from 1964:Q3 to 2008:Q3. The post-financial crisis sample is not used due to the zero lower bound constraint on the fed funds rate.

Table 1 shows the marginal likelihood calculations for the different specifications of the models. The top row shows that for the constant parameter VAR the log marginal likelihood is $-484.85$. This is useful as a benchmark to compare with the other specifications. The highest marginal likelihood is achieved for the regime switching variance only model with 3 regimes (rs v3) with a value of $-393.42$. Looking only at the tvp case, the best fit model is also the one with only variation in the shock variances with a marginal likelihood of $-396.76$. Thus the marginal likelihood suggests that data prefer change in shock variances rather than a setting that has changes in the monetary policy rule, or changes in both the variances and the policy rule. The marginal likelihood for each of the rs specifications is higher than the constant parameter VAR. But for the tvp b and tvp vb specifications, the marginal likelihood is actually lower than the constant parameter case. For the TVP specification the marginal likelihood of the tvp b specification is $-821.35$, which is significantly lower than the value for the tvp v model or the rs b model with 2 or 3 regimes.

What is the reason for such a low value? To better understand this, the next two columns in the table show the value of the log likelihood and the log prior evaluated at the posterior mean of the parameters. The analytical form of the marginal likelihood is not known for these models and thus we cannot exactly formulate the contribution of the prior and likelihood. Nonetheless, this exercise can provide some useful insight. Intuitively, the likelihood is a measure of the fit, while the prior contributes to a penalty for over-parameterization, with the penalty getting higher as the posterior has lesser overlap with the prior. The value of the log prior for the baseline case, for all the rs specifications and for the tvp v case are between $-30$ and $-63$. On the other hand, for the tvp b and tvp vb specification the log prior at the posterior mean is $-375.84$ and $-402.99$ respectively. Thus the prior seems to be lowering the marginal likelihood dramatically for the tvp b and tvp vb cases. We can further analyze the contribution of the various parameter blocks to the log prior, as show in table 2. This table shows that the biggest contribution to the log prior comes from $Q$, the covariance matrix of the shocks to the time-varying coefficients of the monetary policy rule. One concern is that large negative values reflect a situation where the posterior distribution of $Q$ has very little overlap with the prior distribution. To investigate
this, we consider the following exercise. $Q$ has an Inverse-Wishart prior, $Q \sim IW(\nu, Q)$ with shape parameter $\nu = 40$ and scale matrix $Q = k^2 Q.40.V(\beta_{OLS})$ which depends on the variance of coefficients from a training sample regression, as explained in section 3.1. We can evaluate the log density of this prior distribution at the mode which is given by $Q(\nu - 7 - 1)$. The log prior density evaluated at the mode is -327.99, which is not much higher than the -375.84 and -402.99. In other words, the posterior distribution of $Q$ has reasonable overlap with the specified prior distribution. Taking this idea a step further we can consider a hypothetical situation where the prior distribution is the same as the estimated posterior. Even in this case the log prior density at the mode is not high enough to change our conclusion. This suggests that by construction the Inverse-Wishart prior adds a relatively large penalty compared to the other parameter blocks. Since the Inverse-Wishart distribution is very common in the time-varying parameter literature, this is an important consideration for any studies that use the marginal likelihood for model comparison.

To shed more light on the fit of the different models, figure 1 plots the conditional log density of the $t$th observation, evaluated at the posterior mean of the estimates. The axes in each row of the figure are set to the same values for ease of comparison. A common theme is that the conditional log density value is low for all specifications around 1980 with the lowest value reached in 1980:Q4. This coincides with the reserves targeting experiment in the early years of Paul Volcker’s regime. For the tvp models there is also a dip in the conditional log density around 1975 (especially for the tvp v model) while the regime switching models tend to fit well during this period. As will be clear from figures and discussion below, this is because the regime-switching models are good at capturing big abrupt changes while the tvp model is designed to better capture gradual movements.

We now analyze whether the choice of tvp vs. rs has important implications. The typical strategy in the empirical literature is to consider either the rs or the tvp framework. Within each framework, alternative specifications of changes in variance or policy parameters (or both) are considered. The selection of the model tends to be based on marginal likelihood calculations. Armed with the same underlying VAR model, identifying assumptions, data sample and priors (as much as possible) we can now evaluate whether researchers conducting analysis in only the rs framework or tvp framework would arrive at the same answers to important questions of interest.

\footnote{In ongoing work, we are conducting a simulation study where the true data generating process has a time-varying parameter specification to determine whether the marginal likelihood would correctly pick out the right model.}
4.1 Change in Shock Variances

First we consider the characterization of the change in the variance of the shocks. The best fit model in both the rs and tvp framework is the variance only model and we compare the variance dynamics implied by each model. Figure 2 shows the smoothed probabilities of the variance regimes in the rs v3 model, with the figures arranged in ascending order of variance starting with the lowest variance in the top left. We see a pattern that is commonly found in the literature. The lowest variance regime is in place for the Great Moderation period from mid-1980s to around 2008 and also in the late 1960s. The 1970s and 1980s are characterized by switches between the two higher variance regimes. The high variance regime is prevalent for two episodes, one in the mid-1970s and one in the early 1980s. A similar picture emerges from the tvp v model. Figure 3 plots the posterior mean of the time-varying standard deviation \( \sigma_{j,t} \) from the tvp v model. The standard deviation of the shocks in all three equations display a higher level in the 1970s and early 1980s with a significant decline around the mid 1980s. Moreover the two peaks in the standard deviation of the fed funds rate equation (in the late 1970s and early 1980s) correspond to the third volatility regime being in place in the rs v3 model. As one would expect, the regime switching approach suggests more abrupt and frequent changes in the volatilities in the 1970s. But overall, we can conclude that the two best fit models paint a similar picture about the dynamics of shock variances.

Next, we consider an important question that is the focus of a large number of monetary VAR studies: What is the effect of monetary policy on the economy? To get around the endogeneity issue the literature has focused on identifying monetary policy shocks. In the structural VAR framework with the recursive identification, these are the disturbances in the interest rate equation that can be backed out using the estimates of the covariance matrix of the reduced-form VAR. We analyze whether using a regime switching approach or a time-varying parameter approach gives any discernible differences in the identified monetary policy shocks. Figure 4 plots the monetary policy shocks from both the rs v3 model and tvp v model. Unsurprisingly, it is clear from the figure that the two measures of monetary policy shocks are almost identical. The estimates of the coefficients are also very similar in the two specifications.\(^6\) Thus an important conclusion is the following: if the usual process of estimating a flexible model and then picking the best fist model is used then it does not matter whether the time variation in the shock variances is rs or tvp. Both methods deliver

---

\(^6\) The parameter estimates are not show here, but are available upon request.
almost identical monetary policy shocks and imply the same dynamics for the changes in the variance of the shocks.

4.2 Change in Monetary Policy Rule

Next we turn our attention to comparing models that just allow changes in the monetary policy rule. Even though the marginal likelihood for these models is lower than the variance only models, this exercise is still of interest. There seems to be a view among economists that monetary policy has changed since the 1970s, see Clarida et al. (2000) and the large related literature. The following exercise aims to shed light on whether the decision of rs vs tvp has important implications for characterizing changes in monetary policy.

Panel (a) in figure 5 shows the long run response to inflation from the tvp b model. The response to inflation rises gradually from the mid 1960s to mid 1970s, when there is a small dip. The peak of the response is in the early years of the Volcker regime and since the mid 1980s the response has been relatively stable. Interestingly, the Taylor principle is not satisfied only in the first four quarters of the sample (1965:Q1 - 1965:Q4).\(^7\) The results from the rs b3 model paint a slightly different picture. Panels (b), (c) and (d) show the smoothed probabilities of the three regimes. The color and line-style of the probabilities in these panels corresponds to that in panel (a) where the horizontal line shows the long-run responses under each of the three regimes. The average of the long-run response to inflation under the two models is similar but the dynamics are not. The regime probabilities show that the first regime (black line in (a) with probabilities in (b)) is in place for majority of the sample. The long run-response to inflation in this regime is around 3, while the response is significantly lower in the other two regimes. The regime probabilities show that there are multiple switches between the first “hawkish” regime and the other two “dovish” regimes. Importantly, this suggests that there were multiple instances in the 1970s where the Taylor principle was not satisfied. This is in contrast with the tvp b model where this occurs only briefly in the beginning of the sample.

Another way to highlight this difference is by looking at the impulse response of the fed funds rate to an inflation shock. Figure 6 shows the response of the fed funds rate to a negative 1 unit inflation shock for two time periods: 1975:Q1 and 1982:Q3. The response

\(^7\)Note that the Taylor principle requires that interest rates increase more than one for one in response to an increase inflation.
under the tvp b model is very similar for the two dates. On the other hand, the rs b3 model implies that the response was a lot more muted and short-lived in 1982:Q3 relative to 1975:Q1.\footnote{Of course, these two dates were chosen specifically to make this point and the differences are smaller on average.} This difference in the estimated dynamics of the policy coefficients has important implications for the identification of monetary policy shocks. Figure 7 plots the monetary policy shocks from the tvp b and rs b3 models. Unsurprisingly, there are big differences between the two series during the 1970s and early 1980s, where the tvp b specification finds slow gradual movements in the response to inflation while the rs b3 specification finds frequent jumps between the three regimes. One may expect that if the number of regimes in the rs b model are allowed to increase that the dynamic pattern may look similar to the tvp b case. But in practice, for this three variable model it is difficult to achieve identification for a specification with more than 3 or 4 regimes.\footnote{Sims et al. (2008) are able to achieve identification of more than 3 regimes in a 3 variable model but they use a restrictive form of time variation.} Thus we conclude it is likely that researchers using the rs or tvp model with changes allowed only in the monetary policy rule would not arrive at identical conclusions about changes in the behavior of monetary policy.

### 4.3 Forecast Performance

An important use of structural VAR models is in forecasting. Thus it is natural to compare the forecast performance of the regime switching vs. time varying parameter framework. To that end, we use a recursive estimation scheme to perform a pseudo out-of-sample forecasting exercise, similar to Clark and Ravazzolo (2014) and D’Agostino et al. (2013). The simulation exercise begins in 1975:Q3. For the first run we estimate the model using data from 1965:Q3 to 1975:Q3. Using these estimates we forecast upto 8 quarters ahead. Then we add 1 quarter of data and re-estimate the model using data from 1965:Q3 to 1975:Q4, and so on. The forecast performance is evaluated using the root mean square forecast error (rmsfe). The numbers are reported in table 3. We will focus on the results for the fed funds rate but results for unemployment and inflation are provided as well. First comparing across the rs and tvp specifications, we notice that the rmsfe (for all horizons) are lower for rs models with changing variances compared to the corresponding tvp specifications. This suggests that in addition to providing a better in sample fit, the rs models with changing variances also perform well out of sample. Within the rs specifications, the rs v3 model provides the best forecasts and allowing for time variation in the monetary policy parameters
increases the rmsfe. This is consistent with the picture emerging from the marginal likelihood calculations. Now turning to comparison within tvp models, we notice that the forecast performance aligns with the marginal likelihood comparison as well. But here the values of the rmsfe for the three different specifications are “closer” to each other. One way to highlight this is to notice that the rmsfe for the tvp b model is actually lower than that for the rs b2 model, whereas the marginal likelihood suggested a better fit for the rs b2 model. Thus it appears that the “penalty” imposed by the Inverse-Wishart prior specification on the marginal likelihood does not affect the forecasting performance of the model as much. Overall, the forecasting exercise gives results that are consistent with the marginal likelihood and provides a useful insight for the tvp b model.

5 Conclusion

Regime-switching and time-varying parameter models are both popular in applied macroeconomic research. This paper is the first to compare the performance of these models to evaluate monetary policy in the US. In a structural VAR setting we find that both frameworks agree on a specification where the variance of the shocks change but not the parameters of the monetary policy rule. This is established by using marginal likelihood calculations and a forecast performance evaluation exercise. Thus researchers using either the rs or tvp framework would arrive at similar answers to monetary policy questions as long as they use the best-fit model. On the other hand, if researchers insist on using a specification where only the monetary policy parameters change, they would find potentially conflicting results about the changes in monetary policy.

References


Koop, G. and S. Potter (2010). A flexible approach to parametric inference in nonlinear and

Koop, G. and S. M. Potter (2007). Estimation and forecasting in models with multiple

Liu, Z., D. F. Waggoner, and T. Zha (2011). Sources of macroeconomic fluctuations: A


Sargent, T., N. Williams, and T. Zha (2006). Shocks and government beliefs: The rise and


Sims, C. A. and T. Zha (2006). Were there regime switches in u.s. monetary policy? The
American Economic Review 96(1), 54–81.
<table>
<thead>
<tr>
<th>Model</th>
<th>Marginal Likelihood</th>
<th>Log Likelihood</th>
<th>Log Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-484.85</td>
<td>-458.00</td>
<td>-33.48</td>
</tr>
<tr>
<td>Regime Switching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs v2</td>
<td>-410.46</td>
<td>-366.46</td>
<td>-55.50</td>
</tr>
<tr>
<td>rs v3</td>
<td>-393.42</td>
<td>-354.72</td>
<td>-57.38</td>
</tr>
<tr>
<td>rs b2</td>
<td>-435.10</td>
<td>-404.53</td>
<td>-46.57</td>
</tr>
<tr>
<td>rs b3</td>
<td>-404.15</td>
<td>-378.04</td>
<td>-53.74</td>
</tr>
<tr>
<td>rs v2b2</td>
<td>-406.68</td>
<td>-368.44</td>
<td>-62.01</td>
</tr>
<tr>
<td>Time Varying Parameter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tvp v</td>
<td>-396.76</td>
<td>-302.29</td>
<td>-58.41</td>
</tr>
<tr>
<td>tvp b</td>
<td>-821.35</td>
<td>-448.12</td>
<td>-375.84</td>
</tr>
<tr>
<td>tvp vb</td>
<td>-837.36</td>
<td>-284.64</td>
<td>-402.99</td>
</tr>
</tbody>
</table>

Table 1: The table shows the marginal likelihood estimates using the Harmonic Mean Estimator. The log-likelihood and log prior are evaluated at the posterior mean of the parameter estimates.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{u\pi}$</th>
<th>$\Sigma$</th>
<th>$\alpha_{21}$</th>
<th>$S$</th>
<th>$Q$</th>
<th>$\beta_{0,i}$</th>
<th>$A_{0,i}$</th>
<th>$W$</th>
<th>$\log(\sigma_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>tvp b</td>
<td>1.6</td>
<td>-17.84</td>
<td>-5.52</td>
<td>-30.76</td>
<td>-331.91</td>
<td>5.42</td>
<td>3.18</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>tvp vb</td>
<td>1.32</td>
<td>N/A</td>
<td>-5.52</td>
<td>-31.76</td>
<td>-338.05</td>
<td>5.42</td>
<td>3.18</td>
<td>-34.81</td>
<td>-2.76</td>
</tr>
</tbody>
</table>

Table 2: The table shows the contribution of the various parameter blocks to the value of the log prior evaluated at the posterior mean of the parameter estimates.
<table>
<thead>
<tr>
<th>Horizon</th>
<th>rs v3</th>
<th>rs b2</th>
<th>rs v2b2</th>
<th>tvp v</th>
<th>tvp b</th>
<th>tvp vb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal Funds Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.94</td>
<td>1.06</td>
<td>0.93</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>1.45</td>
<td>1.68</td>
<td>1.44</td>
<td>1.50</td>
<td>1.59</td>
<td>1.63</td>
</tr>
<tr>
<td>4</td>
<td>2.09</td>
<td>2.38</td>
<td>2.08</td>
<td>2.16</td>
<td>2.25</td>
<td>2.43</td>
</tr>
<tr>
<td>8</td>
<td>3.17</td>
<td>3.50</td>
<td>3.29</td>
<td>3.26</td>
<td>3.31</td>
<td>3.78</td>
</tr>
<tr>
<td>Inflation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.03</td>
<td>1.03</td>
<td>1.02</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>2</td>
<td>1.20</td>
<td>1.25</td>
<td>1.25</td>
<td>1.23</td>
<td>1.24</td>
<td>1.23</td>
</tr>
<tr>
<td>4</td>
<td>1.56</td>
<td>1.65</td>
<td>1.66</td>
<td>1.57</td>
<td>1.61</td>
<td>1.58</td>
</tr>
<tr>
<td>8</td>
<td>2.24</td>
<td>2.49</td>
<td>2.55</td>
<td>2.33</td>
<td>2.45</td>
<td>2.48</td>
</tr>
<tr>
<td>Unemployment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.26</td>
<td>0.27</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>2</td>
<td>0.52</td>
<td>0.54</td>
<td>0.51</td>
<td>0.53</td>
<td>0.54</td>
<td>0.53</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>1.00</td>
<td>0.96</td>
<td>0.98</td>
<td>1.01</td>
<td>0.99</td>
</tr>
<tr>
<td>8</td>
<td>1.39</td>
<td>1.39</td>
<td>1.42</td>
<td>1.43</td>
<td>1.45</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Table 3: This table shows the root mean squared forecast error calculated using the posterior mean for the different specifications.
Figure 1: Log Likelihood evaluated at posterior mean
The figure shows the smoothed probability of the variance regimes at the posterior mean from the rs v3 model.
Figure 3: Posterior Mean of Standard Deviations

The figure shows the posterior mean of the time-varying standard deviations in the tvp v model.
Figure 4: Monetary Policy Shocks

The figure shows the monetary policy shocks from the two best fit models. The dashed blue line is the tvp v model while the solid black line is the rs v3 model.
Panel (a): The green line shows the long-run response to inflation in the tvp b model. The black, blue and red lines show the long-run responses to inflation in the rs b3 model in ascending order.
The figure shows the impulse response of the fed funds rate to a one-unit inflation shock. The dashed blue line is the tvp b model while the solid black line is the rs b3 model.
The figure shows the monetary policy shocks from two models with change in policy coefficients. The dashed blue line is the tvp b model while the solid black line is the rs b3 model.