Interest Rate Rules in Practice - the Taylor Rule or a Tailor-Made Rule?

Adam Check*

November 9, 2015

Abstract

This paper investigates the nature of the Federal Open Market Committee's (FOMC's) interest rate rule, with a focus on which variables have been relevant to the FOMC over the past 40 years. I consider a large number of potential variables, including alternate measures of inflation, aggregate real activity, and sectoral variables. Based on inclusion probabilities derived from Bayesian Model Averaging (BMA) over a sample from 1970-2007, I find that the FOMC responds to changes in unemployment rather than to changes in GDP growth. Additionally, I find that the FOMC reacts not only to inflation and aggregate output, but also to measures of sectoral activity, such as changes in commodity prices. Finally, I find that using BMA improves out-of-sample forecasting performance over baseline Taylor-type interest rate rules.

JEL Codes: C11, C52, E52, E58

*I can be contacted at ajc@uoregon.edu

1 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 produce the most accurate description of policy. Furthermore, there are a host of additional 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 used 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. 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 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 1, 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 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 while Clarida et al. (2000) assume the FOMC is forward looking and responds to forecasts.

Table 1: 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	Real-Time 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	M2, PCom
Davig and Doh (2008)	GDP Deflator	GDP Gap	Contemp.	-
Coibion and Gorodnichenko (2011)	GDP Deflator	GDP Gap & Growth	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 actually pays attention 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

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

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 coefficient value when it is included, with the weight on zero being very high. This shrinkage toward zero typically

[2]See Koop (2003).

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.

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 implies 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 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.

2 Data and Interest Rate Rule Specification

In my analysis, I consider a total of 13 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, 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.

Table 2: 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 data, 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 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.

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

[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.

Table 3: 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%

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 weights 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]

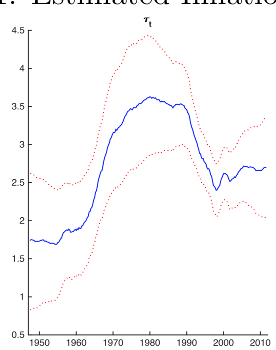
For a measure of the inflation target, I use Matlab code that accompanies the paper by

[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.

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 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 1, and we can see that the 5% upper bound of the target is not binding.

Figure 1: 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:

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

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

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

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]

3 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 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.

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

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

Distribution. Because there are 14 potential regressors, there are $2^{13} = 8,192$ 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 was much more volatile. Precise prior statistics are provided in Table 4 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.

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

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

Table 4: 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

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 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 5, 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

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

probability, while the change in the unemployment rate receives an inclusion probability of 95.8%.

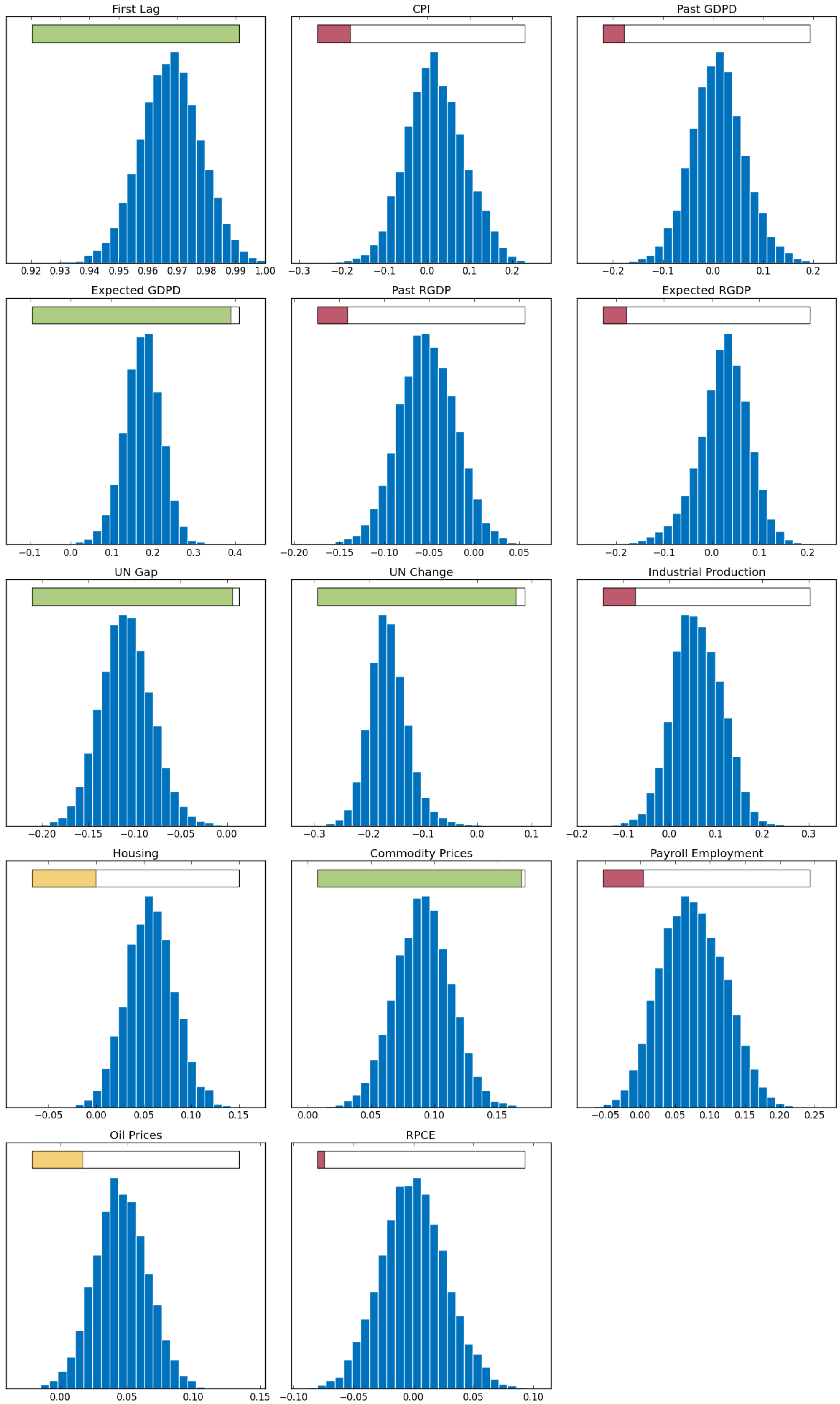
Table 5: Inclusion Probabilities from BMA

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%

Histograms for the conditional posterior distribution of each coefficient are presented in Figure 2. 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 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

Figure 2: Coefficient Histograms



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 prices, oil prices, and RPCE. In table 6, 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 6: 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 so 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 including 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 7. 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 7: 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 8. 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.

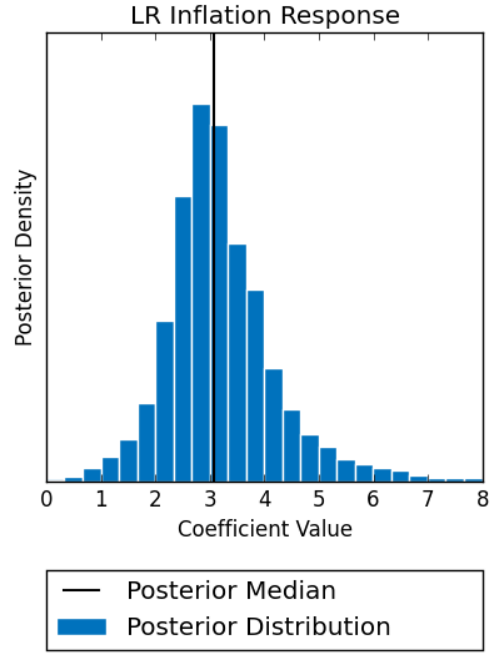
Table 8: Probability of "Generalized Taylor Rule"

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 DSGE model, the inflation response coefficient needs to be greater than or equal to 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. Like the histograms presented earlier, Figure 3 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

Figure 3:



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.

4 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

[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)

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 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,

[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.

and is expressed 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.

Table 9: 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

First I compare the three forecasting metrics for rolling sample vs. recursive estimation, and show the results in Table 9. 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.

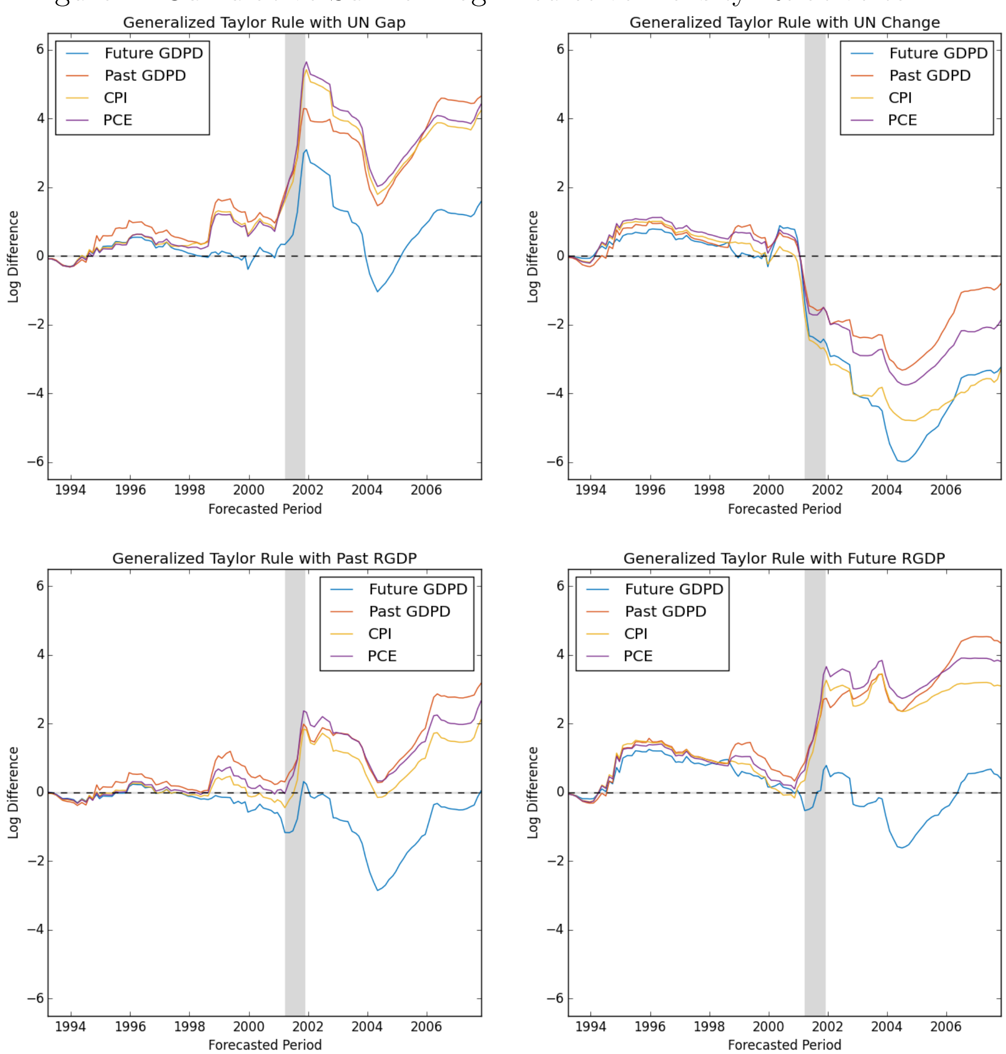
Table 10: Forecasting Performance of Taylor Rules, Relative to BMA

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

Next, I present a table of forecasting metrics for a variety of Taylor rules, relative to the statistics of BMA, and show the results in Table 10. Here, I have normalized the MAFE and RMSFE by dividing these statistics for each Taylor rule by the values listed in table 9 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 4, we see that much of the difference in the SLPD's 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, especially because of the four measures of real output that I include in this study, the change in the unemployment rate is the least commonly used.

Figure 4: Cumulative Sum of Log Predictive Density Relative to BMA



5 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 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 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

[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.

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. 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.

References

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

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.

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.

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.

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. Technical report.

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.

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

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.

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

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.

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.

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.

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*.