Bayesian Forecasting of Interest Rates: Do Priors Matter?

Fernando C. Barbi*
São Paulo School of Economics (FGV)

March 17, 2014

Abstract

Central Banks that commit to an Inflation Target monetary regime are bound to respond to inflation expectation spikes and product hiatus widening in a clear and transparent way by abiding to a Taylor rule. There are various reports of central banks being more responsive to inflationary than to deflationary shocks rendering the monetary policy response to be indeed non-linear. Besides that there is no guarantee that coefficients remain stable during time. Central Banks may switch to a dual target regime to consider deviations from inflation and the output gap. An estimation of the a Taylor rule may therefore have to consider a non-linear model with time varying parameters. This paper uses Bayesian forecasting methods to predict short-term interest rates. We take two different approaches: from a theoretic perspective we focus on an augmented version of the Taylor rule and include the Real Exchange Rate, the Credit-to-GDP and the Net Public Debt-to-GDP ratios. We also take an "atheoretic" approach based on the Expectations Theory of the Term Structure to model short-term interest. The selection of priors is particularly relevant for predictive accuracy yet, ideally, forecasting models should require as little *a priori* expert insight as possible. We present recent developments in prior selection, in particular we propose the use of hierarchical hyper-g priors for better forecasting in a framework that can be easily extended to other key macroeconomic indicators.

Keywords: forecasting, Bayesian model averaging, hyper-g prior, monetary policy

JEL: C53, E27, E43, E47

*The author is grateful to Pedro Valls, Paulo Picchetti, Alkimar Moura, Emerson Marçal and Neil Ericsson for their comments. Email: fcbarbi@gmail.com

1 Introduction

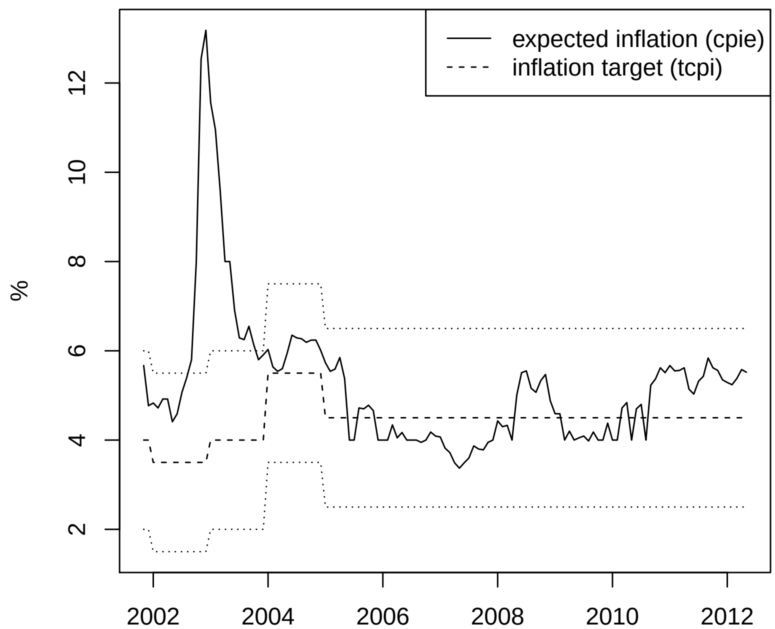
Since Taylor (1993) uncovered the Federal Reserve monetary policy response to inflation and product gaps, the Taylor rule is a central tool in the conduction of monetary policy as it conveys the necessary transparency to the monetary authority reaction to price shocks. Petersen (2007) sums up the essence of the rule: "the Taylor rule corresponds to a guide-post to good monetary policy: a mechanism that constrains monetary policy to be systematic, consistent, and rule-like. Monetary policy that is systematic, consistent, and rule-like characterizes a transparent and credible monetary policy, and therefore alleviates the time-inconsistency problems associated with discretionary monetary policy." Consistent with the rational expectation behavior of agents Clarida et al. (2000) suggest changing the original rule to a forward-looking reaction function.

This paper introduces a new way to estimate the Central Bank decision on monetary policy rate (MPR), selic, by using a Bayesian Model Averaging (BMA) framework based on hyper-g priors. Our first contribution is to test the forecasting accuracy of policy rates using a methodology that handles possible changes in the structure of the rule. The choice for BMA also considered previous interest rates forecasting attempts, notably using multivariate methods as did Lima et al. (2006) whose results suggest that VAR/VEC models perform poorly in terms of forecasting accuracy of interest rates. The motivation for this paper is to empirically test the hyper-g priors proposed by Liang et al. (2008). Our work extends Chua et al. (2011), who got enticing results forecasting interest rates for Australia. We used flexible hyper-g priors and controlled for the "supermodel effect" to reach very accurate forecasts.

Our second contribution is to discuss the best forecasting approach given that there are two clear distinct viewpoints: the "theoretic" approach proposes a causal relation between inflation expectations and the level of interest rates while the "atheoretic" approach is guided by expectations over the policy rate path shaping the yield curve. The forecasting capability of the monetary policy instrument reflects the quality of Central Bank communication: in an ideal monetary policy context agents should be able to correctly anticipate rate movements keeping uncertainty to a minimum to reduce[1] sovereign risk premium, that would otherwise

[1]Other sources of sovereign risk, notably fiscal policy, are not a concern in this work.

Figure 1: Expected Inflation (cpie) x Target Inflation (tcpi)

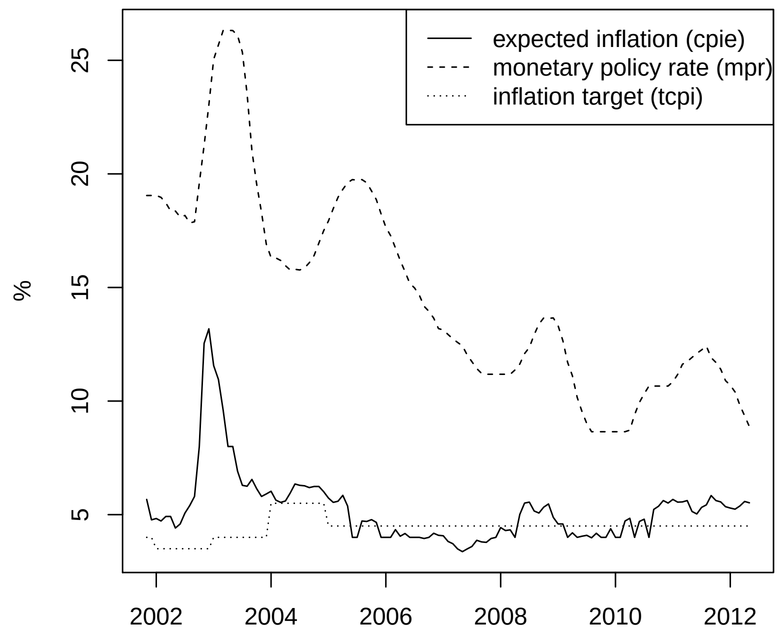


be charged by Brazilian bonds holders.

The very essence of the commitment to a target inflation comes from the Central Bank credibility. According to Minella et al. (2003) "private agents should believe that the central bank will act consistently within the inflation-targeting framework. Gaining credibility, however, takes time. In the context of large shocks, even with a strong response by the monetary authority, expectations will tend to deviate from the targets. In this case, communication with the market public to explain the reasons of the non-fulfillment of the targets becomes crucial. Furthermore, it is important that expectations converge to the target over a certain time horizon." Albeit transparency is paramount to credibility, the Brazilian Central Bank (BCB) does not publish its reaction function (Taylor rule) leaving the market to infer its dynamic structure. This study offers a novel estimation methodology of the reaction function in an effort to improve the market understanding of the evolution of monetary policy.

Figure 1 presents the expected inflation 12 months ahead (cpie) and inflation target (tcpi)to show that inflation was tamed in 2005/2006. Before that some adjustments were made to

Figure 2: Expected inflation (cpie) x Monetary Policy Rate (mpr)

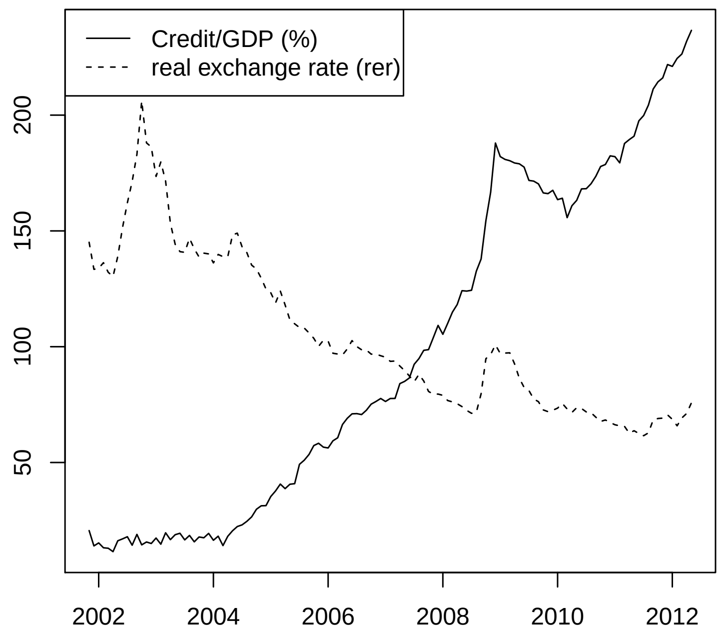


the target and its fluctuation band that later was set to ±2%. Only after 2006 the BCB gained the necessary credibility over inflation control. Figure 2 shows the evolution of inflation expectations (cpie) and the movements in the monetary policy rate.

In Garcia et al. (2011) the authors test the response of an advanced economy and an emerging economy to demand, cost-push and premium risk shocks if the Central Bank adopts a Taylor augmented (or hybrid) rule that incorporates the real exchange rate. Their simulations indicate that the origin of the demand shock matters: if domestic, the deterioration of trade balances induces an undesirable depreciation of the currency but if this positive demand shock originated abroad, a strong balance may induce a currency appreciation. The authors also conclude that emerging markets might gain by adopting hybrid Taylor rules specially when facing risk premium shocks.

Emerging economies bear some common aspects. Access to credit markets is restricted so consumption is more related to current income when compared to developed economies in which agents can smooth consumption through credit. In Brazil credit availability has

Figure 3: Credit/GDP x Real Exchange Rate (rer)



increased significantly as can be seen in figure 3 where the credit/GDP ratio expansion is presented to show how credit has risen fivefold in 2012 over its 2002 level. We can also see that after 2003 the exchange rate tended to appreciate in real terms. Emerging markets present some restrictions to capital mobility that renders foreign exchange interventions effective. Although these macroeconomic variables are relevant, it is not clear if and how the BCB incorporates the exchange rate and the expanded credit base in its monetary policy decisions.

What is generally accepted, though, is that policymakers are unwilling to allow sharp exchange rate movements whenever the pass-through effect is significant and movements in exchange rate might impact domestic inflation. This is particularly relevant in the context of depreciations in which there is a raising risk for domestic inflation that may materialize when the Marshall-Lerner condition[2] is valid. On the other hand, depreciation may expose the economy to balance sheet risks if a significant portion of domestic debt is denominated in

[2]The Marshall-Lerner condition states that exchange rates affect output when the sum of import and export elasticities is above 1 (Obstfeld and Rogoff (1996)).

foreign currencies, thus leading to the *Fear of Floating* syndrome (Calvo and Reinhart (2002)).

The alternative movement towards exchange rate appreciations bears risk for the competitiveness of manufactured goods in the international market. Brazil has undergone a relative reduction in the participation of the industrial sector on output, suggesting a de-industrialization process (Oreiro and Feijó (2010)), caused not only by the long-run appreciation of the Real but also from a fierce competition from China[3].

The mere volatility of exchange rate may also be undesirable as it introduces uncertainty in long-term contracts. Central Banks might be willing to intervene in foreign exchange markets not to set levels but to reduce volatility. Garcia et al. (2011) verify that placing some weight on the exchange rate in an open economy Taylor rule substantially reduces the volatility of output and inflation when facing demand shocks.

It can be expected that since inflation was tamed in 2006 the BCB may have changed its preferences about inflation hiatus and output gap by reweighting the coefficients in the reaction function, the Taylor rule that will be presented in the next section. In the inflation targeting period where inflation expectation data is available, from 2001 to 2012, the BCB has been ruled by three presidents whose mandates coincide with the Presidential mandates: Fernando Henrique Cardoso (from 1999 to 2003), Luiz Inácio Lula da Silva (from 2003 to 2010) and Dilma Rousseff (from 2011 to 2014).

Minella and Souza-Sobrinho (2009) showed that the exchange rate transmission channel is important in Brazil. It might be argued that in this context the BCB might have adopted a Taylor augmented rule to respond to increases in the expected inflation, in spite of the fact that this option was never explicitly communicated to the market[4]. If the appreciation of the Real is relevant for domestic inflation, interest rates might be calibrated to target specific exchange rate levels rather than to set specific inflation targets or output gaps.

The linearity of the Taylor rule might suggest that the monetary authority reacts symmet-

[3]Some other possible causes would be structural shortcomings that increase production costs such as high taxation, infrastructure bottlenecks and restricted exporters access to cheap financing.

[4]Central banks that follow the inflation targeting regime may enhance their credibility by improving communications with the market. Blinder et al. (2008) define central bank communication "as the provision of information by the central bank to the public regarding such matters as the objectives of monetary policy, the monetary policy strategy, the economic outlook, and the outlook for future policy decisions. Nowadays, it is widely accepted that the ability of a central bank to affect the economy depends critically on its ability to influence market expectations about the future path of overnight interest rates, and not merely on their current level."

rically to either positive or negative deviations of inflation from its target. Petersen (2007) uses a Logistic Smooth Transition Regressions (LSTR) to verify that the Fed follows a non-linear Taylor rule even not considering either the smoothing or the forward-looking aspects of the rule. Castro (2008) corrects these omissions and extends the investigation to verify that besides the Fed both the ECB (European Central Bank) and the BoE (Bank of England) adopted non-linear Taylor rules. This later work also uses LSTR technology to model the transitions between monetary policy regimes modeled as endogenous breaks. The author justifies the choice by "allowing for endogenous regime switches – contrary to the Markov-switching models – it also provides economic intuition for the nonlinear policy behavior of the central bank and it is able to explain why and when the central bank changes its policy rule."

We will proceed after this introduction to present two theories for short-term interest rate in section 2, Bayesian inference and forecasting methodologies are introduced in section 3, data and results are presented in section 4 and we propose some conclusions in section 5.

2 Theories for short-term interest rate

2.1 The Taylor rule

Essential to any central bank toolkit the Taylor rule was empirically devised by John Taylor in early 1990. This generated a new strand in monetary policy literature in a time when money supply was losing the prominent place it held since the end of 1960[5]. At a time when most believed that a new instrument of monetary policy should be chosen, the new rule brought much needed light. Agents would be able to predict the future path of the short-term basic rate. The rule simplicity may be an important reason for its choice as the *de facto* instrument of monetary policy and a key macroeconomic indicator for asset prices throughout the economy.

[5] Friedman (1968) stated his option for monetary policy instrument: "The first requirement is that the monetary authority should guide itself by magnitudes that it can control, not by ones that it cannot control. If, as the authority has often done, it takes interest rates or the current unemployment percentage as the immediate criterion of policy, it will be like a space vehicle that has taken a fix on the wrong star. No matter how sensitive and sophisticated its guiding apparatus, the space vehicle will go astray. (...) I believe that a monetary total is the best currently available immediate guide or criterion for monetary policy - and I believe that it matters much less which particular total is chosen than that one be chosen."

Originally the Taylor rule was stated as

$$i_t = \bar{i} + 1.5(\pi_t - \bar{\pi}) + 0.5(y_t - \bar{y}) + \nu_t$$

where i_t is the nominal interest rate, $\bar{i}$ is the natural interest rate, π_t is the inflation rate, $\bar{\pi}$ is the inflation target, y_t is output and $\bar{y}$ is the natural product. The difference between inflation and its target is the "inflation hiatus", $(\pi_t - \bar{\pi})$, and the difference between product and potential output is the "product gap", $(y_t - \bar{y})$.

Yet this apparent simplicity can be quite elusive as this rule proved to be rather difficult to estimate. An important issue to estimate a Taylor rule is that the natural interest rate and the natural product are both non-observed variables. The natural interest rate according to Woodford (2003a) is "the equilibrium rate of return in the case of fully flexible prices" and the natural product is the product that would exist should there not be any frictions such as monopolistic competition or price and wage rigidities. Let's briefly review four issues concerning the estimation of this rule.

The first problem with the Taylor rule specification is that the key parameters of natural interest rate and the natural product are not known by the policymaker at the time when a decision must be made. Orphanides (2001) discusses the relevance of using historical, unrevised data available to the policymaker: "Retrospectively, the "appropriate" policy setting for a particular quarter may appear different with subsequent renditions of the data necessary to evaluate the rule for that quarter. Through a distorted glass, the interpretation of historical episodes may change."

Certain parameters are either not observed or not precisely measurable when a decision is made. The usual way to estimate these non-observed components in the econometric literature is to filter the output time series with a Hodrick-Prescott filter to generate a smoothed series that would represent the long run equilibrium. Another form of eliciting both the natural interest rate and the ouput gap is to impose a structural model as did Laubach and Williams (2003). The key insight of this model is that both unobserved components share a common factor, g_t, assumed to have a unit root:

$$\bar{i}_t = \gamma_t g_t + z_t \quad \text{and} \quad \bar{y}_t = \bar{y}_{t-1} + g_{t-1} + \epsilon_t$$

The stochastic process z_t may be a stationary auto-regressive process (generally of low order) but it may also have a unit root. The authors postulate that natural output is I(2) and explain that "as Stock and Watson (1998) point out, when the disturbance to the growth rate component has small variance, such a test statistic has a high false-rejection rate."

A second issue in estimating the Taylor rule concerns the forward-looking aspect of the reaction function as discussed by Clarida et al. (2000). There is a delay from 6 to 12 months to transmit unanticipated shocks to the MPR to inflation Walsh (2010). This requires that the Central Bank focuses in the future inflation, generally 12 months ahead. There is also an "expectation channel" that captures the effects of monetary policy shocks through changes in inflation expectations. The Taylor rule can be written as:

$$i_t = \bar{i}_t + \beta_{\pi^e}(\pi^e_{t+k|t} - \bar{\pi}_t) + \beta_\pi(\pi_t - \bar{\pi}_t) + \beta_y(y_t - \bar{y}_t) + \nu_t$$

The expectations provided by the FOCUS bulletin, a survey made weekly next to market participants, is generally accepted as the best indicator of future inflation and frequently cited in the Monetary Policy Committee (COPOM) minutes. Note that in the above equation the current inflation term is preserved: its significance indicates some degree of backward-looking by the Central Bank.

A third issue concerns the inertia of interest rates. The inertia in the interest rate rule is explained by Woodford (2003b) as a way to influence the market about the future path of interest rates: "an effective response by the Fed to inflationary pressures, say, requires that the private sector be able to believe that the entire future path of short rates has changed. A policy that maintains interest rates at a higher level for a period of time once they are raised – or even following initial small interest-rate changes by further changes in the same direction, in the absence of a change in conditions that makes this unnecessary – is one that, if understood by the private sector, will allow a moderate adjustment of current short rates to have a significant effect on long rates."

Some authors distinguish between the Taylor rule and the central bank response by stating that the response rule incorporates inertia in interest rates while the Taylor rule is just the sum of the original three components. We use these expressions interchangeably, for us the Taylor rule incorporates inertia as in:

$$i_t = \rho i_{t-1} + (1-\rho)\left[\bar{i}_t + \beta_{\pi^e}(\pi^e_{t+k|t} - \bar{\pi}_t) + \beta_y(y_t - \bar{y}_t)\right] + \nu_t$$

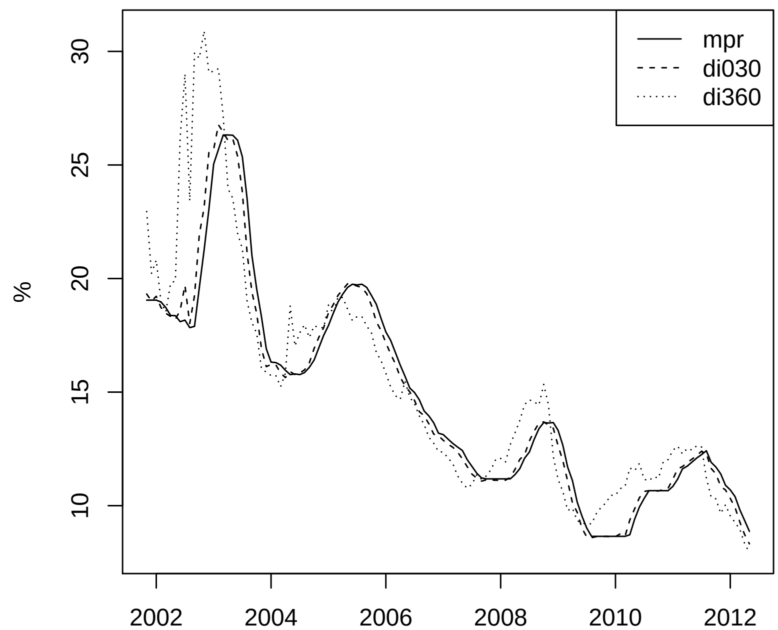
Even if we agree with Lansing (2002) that the interest rate smoothing is an illusion caused by the use of final data, as opposed to real-time data used by the central bank, there is still some degree of auto-correlation in the interest rate series, albeit smaller than the traditionally estimated at 0.8. Rudebush (2002) opposes the concept of inertia on quarterly interest rates by studying the behavior of interest rates at the short-term end of the yield curve. He believes that the illusion of inertia is the result of persistent shocks that cause the central bank to deviate from the policy rule. Yet, he does not deny the existence of smoothing on a monthly basis. More recently Coibion and Gorodnichenko (2012) offer another explanation to the persistence of interest rates: "the higher order auto-regressive process for interest smoothing may have been capturing the central bank's response to the private sector's information set. This suggests a novel potential explanation for deviations of actual interest rates from standard Taylor rule prescriptions." To be practical this study considers an auto-regressive process for the MPR.

A fourth issue to be addressed is the stability of the parameters of the reaction function. A linear model with time varying parameters is:

$$i_t = \rho_t i_{t-1} + \beta_{i,t}\bar{i}_t + \beta_{\pi^e,t}(\pi^e_{t+k|t} - \bar{\pi}_t) + \beta_{y,t}(y_t - \bar{y}_t) + \nu_t$$

Figure 3 suggests that the policy rate increases as the inflation expectation rises towards the inflation target upper bound and that it might be reduced when inflation is under control but output is negatively affected.

Figure 4: Interest rates: MPR, DI030, DI360



2.2 The Expectations Theory of the Term Structure

The Expectations Theory of the Term Structure (Walsh (2010, Chap. 10)) states that monetary policy affects long-term interest rates by directly influencing short-term rates and by altering market expectations of future short-term rates

$$(1 + i_{n,t})^n = \prod_{h=0}^{n-1}(1 + i_{t+h})$$

where $i_{n,t}$ is the nominal yield to maturity at time t on an n period discount bond and i_t the short-term rate. Taking the logs and recalling that $log(1 + x) \approx x$

$$i_{n,t} = \sum_{h=1}^{n-1} i_{t+h}$$

$$(1 + i_{n,t})^n = (1 + i_t)(1 + i_{n-1,t+1})^{n-1}$$

In figure 4 we can compare the short rate (mpr), the 30 days (di030) and one year (di360) maturities. They exhibit a common behavior in time but with closer horizons, the curves get smoother.

The long-term nominal interest rate depends on current short-rate and market expectations of future short-rates

$$I_t^n = E_t\left[\frac{1}{n}\sum_{h=1}^{n-1} i_{t+h}\right] + E_t\phi_t^n$$

where I_t^n is the long-term nominal interest rate at horizon n and ϕ_t^n is average risk premium.

In the rest of this paper we address this approach as "atheoretic" as it does not involve a causal relation among interest rates and explanatory variables. This is a naming convention and does not imply that the referred theory is not valid. Actually, the Expectations Theory implies a long-term relationship between short-term and long-term rates that was tested for cointegration with an error correction model (Johansen (1995)) as in

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \delta X_{t-i} + \epsilon_t$$

where $X_t = (i_t, i_t^{30}, i_t^{60}, i_t^{90}, i_t^{120}, i_t^{180}, i_t^{360})$ and i_t^{360} is the one-year interbank swap contract. Tests in table 1 suggest a cointegration rank of $r = 5$ with 2 or 3 lags.

3 Bayesian Inference and Forecasting

Bayesian inference allows for pre-existing knowledge about the observed phenomenon to be factored in the model, embedded in a prior distribution of the parameters to be estimated. Generally the prior brings expert insight that will be confronted with observed data. From this observation, a posterior distribution is generated that may confirm the prior insight or contradict it. Either way the outcome is directly influenced by the choice of priors. The need for a prior distribution over unknown parameters may be seen as a limitation of the method. This requirement is usually met with uncompromising, uninformative priors that forsake the opportunity to bring additional previous knowledge to the inference process.

The advantages of adopting Bayesian methods are, on the other hand, significant. First, Bayesian inference supports both linear and non-linear models in the same estimation framework and therefore avoids the need for detecting and treating structural breaks in the data. Second, Bayesian inference is very efficient as it transforms an optimization problem, the cornerstone of the maximum likelihood approach, into an integral calculation that is less computationally intense. Closed form posteriors (like the ones generated from conjugate Normal-Gamma priors) are popular ways to increase computational efficiency even more. Yet, with the average computer power available today, any distribution can be quickly sampled using MCMC (Markov Chain Monte Carlo) algorithms, as we do.

Given the potential upside from the adoption of Bayesian inference a thriving literature tries to elicit an objective process for prior selection. In this paper we present recent prior proposals and test them by forecasting the short-term monetary policy rate in Brazil. To get results that are comparable to those in existing literature we assume a linear relation between $\mathbf{y}$ $(N \times 1)$ and the covariates $\mathbf{X}$ $(N \times K)$ for the model M_s as in

$$\mathbf{y} = \mathbf{X}\beta_s + \epsilon_s \qquad \text{where} \qquad \epsilon_s \sim N(0, \sigma_s^2 \mathbf{I}) \tag{1}$$

When comparing two candidate models, the *posterior odds* ratio combines the marginal posterior probabilities of two competing models M_i and M_j for a direct comparison between them given the observed data y. Posterior odds are written as

$$\frac{p(M_i|y,X)}{p(M_j|y,X)} = \frac{p(M_i|X)}{p(M_j|X)} \frac{p(y|M_i,X)}{p(y|M_j,X)}$$

because by the Bayes theorem: $p(M_i|y,X)p(y|X) = p(y|M_i,X)p(M_i|X)$.

The posterior odds ratio can be decomposed into the product of the *prior odds ratio* and the marginal likelihood ratios, also known as the *Bayes Factor* (Kass and Raftery (1995)) defined as

$$B(M_i, M_j) \equiv \frac{p(y|M_i,X)}{p(y|M_j,X)}$$

A Bayes Factor $B(M_i, M_j)$ higher than one indicates the superiority of model M_i over M_j.

Prior selection affects Bayes Factors. The use of an *improper prior*[6] is not permitted in model selection since posterior model probabilities and Bayes Factors may be indeterminate

[6] i.e. a distribution density probability function that does not integrate to 1, an "infinitely spread-out prior".

as improper priors are determined only up to an arbitrary multiplicative constant (Berger and Pericchi (2001)).

Congdon (2003, Chap. 2) tackles prior selection from an identification perspective: "[T]he major issues in identifying mixture models using parametric densities $f(y|\theta_j)$ are the general question of identifiability in the face of possibly flat likelihoods (Bohning (2000)), and the specification of appropriate priors that are objective, but also effective in estimation. Thus, Wasserman (2000) cites the hindrance in mixture modelling arising from the fact that improper priors yield improper posteriors. More generally, vague priors even if proper, may lead to poorly identified posterior solutions, especially for small samples. Various approaches to prior specification in mixture modelling have been proposed and often mildly informative proper priors based on subject matter knowledge may be employed."

Before discussing model averaging we briefly present the process of model selection to show how models can be sorted conditionally on the observed data. In a model averaging context this ranking translates into a weighting function.

3.1 The Choice of Priors

Zellner (1986) proposes a family of gaussian priors on the coefficients β_s of equation 1 known as the *g-priors*. These priors are considered as fixed because they do not adapt to observed data and only reflect the researcher's perspective and confidence on parameter distributions:

$$p(\sigma_s) = 1/\sigma_s$$
$$\pi(\beta_s|\sigma_s, M_s) \sim N(0, \frac{g}{\sigma_s}(X'X)^{-1})$$

As Liang et al. (2008) we assume a null-based approach[7] to calculate the Bayes factor as

$$BF(M_s, M_N) = (1+g)^{(n-k-1)/2}(1+g(1-R_s^2))^{-(n-1)/2}$$

where R_s is the determination coefficient for model M_s and k is model's size (number of regressors).

[7]In this case, the base model is the null model $M_N : y_t = \epsilon_t$

When setting priors, there are generally two criteria for prior selection: the Bartlett's paradox and the Information paradox (Liang et al. (2008)). The Bartlett's paradox states that uninformative priors tend to favor the parsimonious null model.

The Information paradox states that Bayes Factor may not perform well when faced with information overload. As model's M_s coefficient of determination, R_s^2, goes to 1, the Bayes Factor, when compared to the null model, should go to infinity. Actually for some priors (like the Zellner g-prior) as posterior probability increases the Bayes Factor is bounded and therefore looses its selection capability.

Feldkircher and Zeugner (2009) argue that, beyond Bayes Factors, marginal likelihood characteristics should affect priors choice. They argue that the elicitation of g should be guided by consistency concerns and the use of g as a penalty term to enforce parameter parsimony. Consistency assures that the choice of g is such that posterior model probabilities asymptotically uncover 'the true model' M_T, i.e. $p(M_T|Y) \to 1$ as $N \to \infty$.

Considering the implications of the choice of g to parsimony, Liang et al. (2008) point out that under Uniform prior, the choice of g determines model selection: large values for g tend to concentrate the prior on parsimonious models with a few large coefficients and conversely small g tend to concentrate the prior on saturated models with small coefficients.

The hyperparameter g conveys how much certainty the researcher has about that coefficients are zero. A small g means little prior coefficient variance and implies confidence that the coefficients are indeed zero. In contrast, a large g suggests uncertainty and may not be robust to noise innovations and risks over-fitting by using more terms than necessary to reproduce the data generating process.

Fernandez et al. (2001) lists other automatic prior generation criteria. Some popular priors are (N is the sample size and K the model size):

1. Unit Information Prior (UIP): $g = N$.

2. Risk Inflation Criterion Prior (RIC): $g = K^2$.

3. Benchmark RIC (BRIC): $g = max(N, K^2)$.

4. Empirical Bayes Local (EBL): $g_s = \arg\max_g p(y|M_s, X, g)$.

Liang et al. (2008) show that the UIP, RIC, and BRIC priors do not resolve the Information paradox for fixed N and K since the choices of g are fixed values not depending on the information in the data. Only the EBL approach has the desirable behavior (see Appendix).

3.2 Hyper-g priors

Liang et al. (2008) introduce a family of priors $\pi(g)$ on $g(g > 0)$ as in

$$\pi(g) = \frac{a-2}{2}(1+g)^{-a/2}$$

The authors propose the *hyper-g* priors, a new class of flexible-g hierarchical priors based on the hypergeometric distribution, with a closed form posterior distribution of g given by

$$p(g|X, M_s) = \frac{p_s + a - 2}{2F(\frac{n-1}{2}, 1, \frac{p_s+a}{2}, R_s^2)}(1+g)^{(n-1-p_s-a)/2}\left[1 + (1-R_s^2)g\right]^{-(n-1)/2}$$

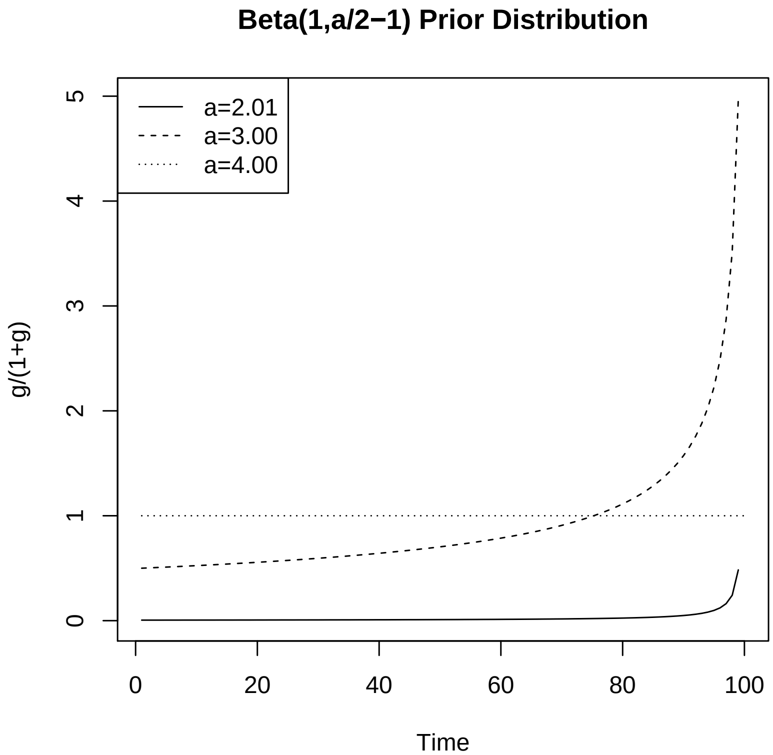
where n is the sample size, p_s is the number of covariates in model M_s and $F()$ is the gaussian hypergeometric distribution as given by

$$F(a, b, c, x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(b-c)}\int_0^1 y^{b-1}(1-y)^{c-b-1}(1-xy)^{-a}dy$$

The authors show that this prior solves the Information paradox.

The hyper-g priors fare better when compared to fixed-g priors, specially when facing noisy data because hyper-g priors dilute the posterior mass among models while fixed-g priors incorrectly favors one model. This is a symptom of the "supermodel effect". As documented by Feldkircher and Zeugner (2009) the supermodel effect refers to the concentration of posterior mass on a few models as indicated by the choice of g. The hyper-g prior is less exposed to this effect than a fixed-g prior because it adjusts the distribution of posterior mass according to the information provided by the data. The authors empirically showed that: "[S]mall degrees of noise trigger a concentration of posterior mass under the hyper-g prior and the empirical Bayes approach. A surge in noise is reflected in a wider spread of posterior mass among models under flexible priors, whereas fixed priors still concentrate on a small number of models."

Figure 5: Hyper-g Priors for different value of a



Hyper-g priors are shrinkage priors. Shrinkage priors provide a continuous alternative to point mass mixture priors. Komaki (2006) shows that the use of shrinkage priors for Bayesian predictive distributions asymptotically dominates improper priors (such as the Jeffreys' prior). Liang et al. (2008) show that the hyper-g prior translates into a Beta[8]prior on the shrinkage factor[9] $\gamma = \frac{g}{1+g}$ in a hierarchical setup that further simplifies the prior definition:

$$\frac{g}{1+g} \sim Beta\left(1, \frac{a}{2} - 1\right) \qquad \text{therefore} \qquad E(\frac{g}{1+g}) = \frac{2}{a}$$

The researcher's choice now concerns $a \in (2, 4]$ where $\pi(g)$ is always proper. The choice $a = 4$ makes the prior uniform, while as a moves to 2 the prior mass of the shrinkage factor concentrates close to 1 as can be seen in figure 5.

[8]If $X \sim Beta(a, b), f(x) = \Gamma(a + b)/(\Gamma(a)\Gamma(b))x^{a-1}(1 - x)^{b-1}$ and $E(X) = a/(a + b)$

[9]The "shrinkage factor" γ affects the distribution of the coefficients since $E(\beta_i|\gamma) = (1 - \gamma)\beta_i + \gamma E(\beta_i)$.

3.3 Bayesian Model Averaging

From the previous Taylor rule discussion it is clear that there is a high level of uncertainty about model composition and dynamics. The Central Banker preferences may evolve from a pure inflation target to a dual target (inflation hiatus and output gap) and extend the original rule to include other factors (RER and Credit-to-GDP) that reflect different transmission channels of Monetary Policy. There is a Bayesian tool that is flexible enough to allow for both parameter and model uncertainties.

Bayesian Model Averaging[10] (BMA) is an efficient tool to handle the model uncertainty problem. As parameter uncertainty is intrinsically handled by Bayesian inference and different models are weighted according to their corresponding Bayes Factors the researcher does not have to choose a single model and can work with a set of models, with different structures, weighed according to the observed data. This supersedes the structural model choice as all models are always considered[11]. Ley and Steel (2007) show the influence of prior choice to BMA results.

BMA has been increasingly considered since Raftery et al. (1997) and Fernandez et al. (2001) showed the improved forecasting performance of this class of models. BMA factors in model uncertainty by basing inference on a weighted average of all possible models. In the Bayesian framework weighting comes naturally from posterior model probabilities (PMP). The PMP $p(M_s|y, X, g)$ for model M_s conditional on data (y, X) is proportional to the marginal likelihood $p(y|M_s, X)$ times a model prior $p(M_s)$ as in:

$$p(M_s|y, X) \propto p(y|M_s, X)p(M_s|X)$$

If K_s is the number of covariates in model M_s there are 2^{K_s} possible sampling models, depending on whether we include or exclude each of the regressors. The posterior model probabilities is given by

$$p(M_s|y, X, g) = \frac{p(y|M_s, X, g)p(M_s|X, g)}{p(y|M_s, X, g)} = \frac{p(y|M_s, X, g)p(M_s|X, g)}{\sum_{j=1}^{2^K} p(y|M_j, X, g)p(M_j)}$$

[10]Hoeting et al. (1999) offer a detailed tutorial and software tools.

[11]Weighting with a zero-weight prior permanently excludes a model from the model set.

As shown by Feldkircher and Zeugner (2009) the Bayes Factor for hyper-g priors allows for the comparison of any two models M_s and M_j by assessing their relative weights:

$$B(M_s, M_j) \equiv \frac{p(y|M_s, X, g)}{p(y|M_j, X, g)} = (1+g)^{\frac{K_j - K_s}{2}} \left(\frac{1 - \frac{g}{1+g} R_s^2}{1 - \frac{g}{1+g} R_j^2} \right)^{-\frac{N-1}{2}}$$

Model averaging is the marginal posterior distribution of the parameter set $\theta = (\beta, \alpha, \sigma)$ obtained by mixing posterior model probabilities:

$$p(\theta|y, X) = \sum_{j=1}^{2^K} p(\theta|y, X, M_j) p(M_j|y, X)$$

A common criticism of the BMA methodology is that it is always possible to find a better fitting model than the weighted average model. One may face this situation under two perspectives. The pragmatic researcher may treat BMA results as a benchmark to be beaten while the cautious researcher may prefer to use an encompassing method. As the dynamics of the underlying process may change the optimal model at each point in time BMA seems to be the ideal tool for inference, yet the true quality of a model may stem from its predictive power, this is why we take the proposed models a step further and test their predictive performance.

3.4 Bayesian Forecasting

Geweke and Whiteman (2006) stress the importance of the posterior predictive probability distribution for Bayesian forecasting: "A principal attraction of the Bayesian structure is its internal logical consistency, a useful and sometimes distinguishing property in applied economic forecasting. But the external consistency of the structure is also critical to successful forecasting: a set of bad models, no matter how consistently applied, will produce bad forecasts. (...) One of the most useful tools in the evaluation of external consistency is the posterior predictive distribution".

We recall that $X = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$. If we define the h-step forecasts by

$X_h = (x_{N+1}, \ldots, x_{N+h})$ the predictive posterior probability at $N + h$ is

$$p(M_s|X_h, y, X) = \frac{p(y_h|X_h, y, X)p(X_h|y, X)}{\sum_{j=1}^{2^K} p(y_h|X_h, y, X, M_j)p(M_j|X_h, y, X)}$$

where the predictive likelihood for one-step ahead is

$$p(y_h|X_h, y, X) = \int p(y_h|\theta_j, X_h, y, X)p(\theta_j|X_h, y, X)\, d\theta_j$$

The predictive likelihood can be obtained by simulation from the posterior $N - N_b$ draws (N_b is the size of the burnin sample) of the parameters θ_s^j as in

$$p(y_h|X_h, y, X, M_s) = \frac{1}{N - N_b} \sum_{j=N_b+1}^{N} p(y_h|\theta_s^j, X_h, y, X, M_s)$$

To gauge quality of the out-of-sample forecasting we use the Log Predictive Probability (LPS). LPS is an approximation to the expected loss with a logarithmic rule related to the well-known Kullback-Leibler criterion (see Appendix). It is defined for a h-step ahead predictive likelihood as

$$LPS(X_h, y, X) \equiv -\frac{1}{h} \sum_{i=1}^{h} \ln p(y_{N+i}|X_h, y, X)$$

The smaller the value of LPS the better is the model in terms of forecasting. We use a BMA version of LPS to take into account the predictive likelihood of all the models. When analyzing forecast results, we provide both out-of-sample MAPE and LPS measures. The MAPE shows how the set of models predicted with the existing dataset while LPS indicates the model predictive capabilities and may be considered a better measure of quality as it is not limited by the quality of the dataset used neither in estimation nor forecasting.

4 Data and Results

Table 2 brings data sources while table 3 presents basic descriptive statistics for the full sample, from November 2001 up to May 2012. This sample choice is determined by data availability. In order to conduct out-of-sample forecasting we split the sample in two parts: a training period from November 2001 up to September 2009 and a forecasting period from October 2009 up to May 2012. The out-of-sample dynamic forecast is performed with a rolling window over forecasting period.

(Place Table 2 about here)

(Place Table 3 about here)

To measure model predictive accuracy four statistics were used: the p-value for the well-known Diebold and Mariano (1995) predictive accuracy test, the in-sample and out-of-sample MAPE[12] (Mean Average Percentage Error) measures and the Log Predictive Score (LPS). All sampling was conducted by a MCMC birth/death algorithm (see Appendix) with $N = 100.000$ draws and $N_b = 50.000$ burnin sample size.

(Place Table 4 about here)

Table 4 is ordered according to posterior inclusion probabilities (PIP). PIP is the sum of the posterior model probabilities (PMP) for all models wherein the covariate was included. The next columns present the estimated coefficient posterior mean and standard deviation. All models are estimated with UIP (Unit Information Priors) as it is the underlying prior for the BIC selection criterion (see Appendix). Model M01 is a simple AR(1) used as a reference. Model M02 is standard Taylor. M03 is an augmented Taylor. Model M04 is a mix between extended Taylor regressors and yield curve components. DI030 stands for the average interbank rate for the 30 day maturity contract and so on until DI360 that covers maturities of one year.

The Diebold Mariano test (see Appendix) is conducted comparing the residuals for each model from an optimum series of zeros (perfect match). The p-values indicate that the null hypothesis (that model residuals are zero) is rejected for all models except M01, what could be expected since this model presents a very low in-sample MAPE of 0.05% that is much lower than the second best of 0.23%. Anyway, this test does not look adequate to sort models in this BMA framework as all models generate the same p-value close to zero.

Of all models, M01 has the best in-sample, with the lowest MAPE, but the lowest out-of-

[12] $MAPE = \sum_{t=1}^{T} |\epsilon_t|/T$ where ϵ_t is the forecast error given by $\epsilon_t = y_t - y_{t|t-1}$ (see Appendix).

sample MAPE is from M04. Model M04 also has the lowest LPS, suggesting a better predictive capacity. The most relevant covariates are the one period lagged MPR, MPR(-1), and the DI030 contract. Both present not only high PIPs but are also statistically significant (we assume their distributions to follow gaussian normal distributions). The other most frequent covariates, DI090 and DI060, should appear in only 25% and 24% of the models (out of $2^{16} = 65536$ possible) but are not statistically significant. Neither are the real exchange rate (RER) or its lead (1 period look-ahead rate), RER(+1). As M04 has the best indicator for predictive performance, next we will test this model with different fixed and flexible priors. Worth mentioning is the fact that even when most covariates are not statistically significant, when they are present the model's predictive capacity increases because their presence reduces the error variance (bias-variance trade-off).

(Place Table 5 about here)

Table 5 compares different fixed priors: RIC, BRIC and EBL. The qualitative choice of covariates is not changed, MPR(-1) and DI030 are omnipresent with roughly the same coefficients: 0.42 for MPR(-1) and 0.79 for DI030. The difference might induce to believe that DI030 is more relevant to predict the MPR but as the standard-error for DI030 is larger than the one for MPR(-1) it is not clear that the coefficients so distant apart so both covariates are equally relevant.

This result can be interpreted as good news: the BCB appears to be conducting communications with the market in an efficient manner because agents are able to predict the monetary policy instrument level and price it in the yield curve. If the MAPE measures are basically the same for all three models, the relevant difference is again in the Log Predictive Score (LPS). The EBL prior generates a significantly lower LPS of -0.800 when compare to the -0.194 from RIC and BRIC. BRIC's LPS is actually a little smaller than RIC's but this is not reported due to the rounding display settings. So far EBL would be the best prior to use. We now estimate with hyper-g priors and expect an even better predictive capacity.

(Place Table 6 about here)

Table 6 presents model M04 with hyper-g priors with $a = 2.01$, $a = 3$ and $a = 4$. The in-sample and out-of-sample indicators have not changed, suggesting they are not adequate measures of the differences. LPS is basically the same for all three, between -0.58 and -.0.59 suggesting that they offer good forecasting capacity but not as much as the models estimated with EBL priors. A noteworthy result is that the relative importance of the covariates changes. In previous models, DI030 and the lagged MPR had basically the same relevance, now the same qualitative result applies: MPR(-1) has an average coefficient of 0.51 while DI030 has a coefficient circa 0.47 with a somewhat higher standard error. This clearly indicates that both covariates offer the same degree of predictive capacity. DI030 is present on average in 93% of the models, while before it would be present in 100% of the models. The importance of this covariate dropped in terms of PIP from 100% to 93% now, but as the standard error also dropped we understand this to be an important covariate for prediction.

(Place Table 7 about here)

As a final result we test all the priors in a pure yield curve model in table 7. The results were qualitative compatible to the ones with the mixed, Taylor + Yield, M04 model. The EBL prior leads to the lowest LPS but this time the results are coherent: a lower LPS translates into a lower MAPE for both in and out-of-samples. EBL priors generate the lowest in-sample MAPE (0.16%) of all models studied while its out-of-sample (0.39%) is also amongst the lowest ones. This performance is acceptable in a production environment.

We could verify that priors matter: the flexible g priors offer better predictive performance that the fixed-g priors, as expected. The use of flexible priors with different hyper-parameters also allowed for lower in and out-of-sample MAPEs but Log Predictive Score is worse when compared to the EBL prior model.

5 Conclusions

Albeit priors allow for expert insights, forecasting models should require as little previous knowledge as possible. By verifying that prior choice impacts the predictive accuracy this paper offers guidance for a better forecasting framework that might reduce the need for previous knowledge about macroeconomic variable paths.

The empirical results confirm the relevance of the prior choice for improved predictive performance. The key to prior choice seems to be the flexibility of the prior to adapt to different noise levels in data. While we expected the hyper-g priors to dominate both in-sample and out-of-sample predictions, we verified that Local Empirical Bayes (EBL) priors, in which the prior distribution is also estimated from the data, provide superior predictive performance. This result reinforces EBL priors as good alternatives as they are proper, consistent priors and do not suffer from the Information paradox limitations. In terms of performance, hyper-g priors may offer some advantage given their closed form posterior distributions. The collected data on time performance during our experiments does not allow for a definite conclusion on this.

The discussion over the best empiric model to estimate an Taylor rule and the high quality predictions of the short-term interest rates in an emerging market riddled with high volatility, suggest that the lack of a sound theoretic model is not an impediment to construct good forecasting models for macroeconomic indicators. Future research could extend this exploration to forecast other macroeconomic relevant indicators notably economic activity, price and wage inflations, and exchange rate.

Finally we may also conclude in favor of the Brazilian Central Bank communication efforts since the interbank market predicts with great accuracy the level of the monetary policy instrument, selic. Finally, another interesting conclusion is that the best predictive model comes from a combination of a forward-looker Taylor rule based model, with an atheoretic yield curve model. Although the yield curve offers enough information for a good predictive performance our results indicate that the best path is to mix covariates from both models and choose a hyper-g prior.

References

Berger, J. O. and Pericchi, L. (2001) "Objective Bayesian methods for model selection: Introduction and comparison". In *Model Selection*, vol. 38 of *IMS Lecture Notes - Monograph Series*, (ed. P. Lahiri), pp. 135-193. Institute of Mathematical Statistics.

Blinder, A., Ehrmann, M., Fratzsche, M., Haan, J. D., and De Haan, J. 2008. "Central Bank Communication and Monetary Policy a Survey of Theory and Evidence," Journal of Economic Literature, American Economic Association, vol. 46(4), pages 910-45, December.

Bohning, D. 2000 "Computer Assisted Analysis of Mixtures and Applications". Monographs on Statistics and Applied Probability, 81. London: Chapman & Hall.

Calvo, G.A., Reinhart, C.M. 2002. "Fear Of Floating," The Quarterly Journal of Economics, MIT Press, vol. 117(2), pages 379-408, May.

Castro, V. 2008. "Are Central Banks Following a Linear or Nonlinear (Augmented) Taylor Rule?" Warwick Economic Research Papers No.872

Chua, C.L., Suardi, S. and Tsiaplias, S. 2011. "Predicting Short-Term Interest Rates: Does Bayesian Model Averaging Provide Forecast Improvement?," Melbourne Institute Working Paper Series wp2011n01, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Clarida, R., Gal, J., Gertler, M. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." Quarterly Journal of Economics, Feb.

Congdon, P. 2003. "Applied Bayesian Modelling". John Wiley & Sons, Ltd.

Coibion, O., Gorodnichenko, Y. 2012. "Why Are Target Interest Rate Changes So Persistent?" American Economic Journal: Macroeconomics, 4(4): 12662.

Diebold, F.X. and Mariano, R.S. 1995. "Comparing predictive accuracy," Journal of Business and Economic Statistics, 13, 253-263.

Feldkircher, M., Zeugner, S. 2009. “Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging,” IMF Working Papers 09/202, International Monetary Fund.

Fernandez, C.E. Ley and M. Steel (2001) “Benchmark priors for Bayesian model averaging”. Journal of Econometrics 100(2), 381427.

Friedman, M. (1968) “The Role of Monetary Policy”. The American Economic Review. Vol.58 No.1 (p.1-17). March 1968.

Garcia, C., Restrepo, J., Roger, S. 2011. “How much should inflation targeters care about the exchange rate?” Journal of International Money and Finance. Volume 30, Issue 7, November 2011.

George, E.I. and Foster, D.P. 2000. “Calibration and Empirical Bayes Variable Selection,” Biometrika. 2000, 87, 731-747.

Geweke, J., Whiteman, C., 2006. “Bayesian Forecasting,” Handbook of Econometrics and Forecasting, Elsevier, vol.I.

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. 1999. “Bayesian Model Averaging: A Tutorial”. Statistical Science, 14, No. 4:382-417.

Johansen, S. 1995. “Likelihood-Based Inference in Cointegrated Vector Autoregressive Models,” Oxford University Press.

Kass, R., Raftery, A. 1995. “Bayes Factor,” Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 773-795.

Komaki, F. (2006). “Shrinkage priors for Bayesian prediction,” the Annals of Statistics, vol. 34, 808-819.

Lansing, K. 2002. “Real-Time Estimation of Trend Output and the Illusion of Interest Rate Smoothing” (January 29, 2002). FRBSF Economic Review 2002.

Laubach, T., Williams, J. 2003. “Measuring the Natural Rate of Interest”. The Review of Economics and Statistics, MIT Press, vol. 85(4), pages 1063-1070, November.

Lehmann, E.L. and Casella, G. 1998. “Theory of Point Estimation”. Second edition. Springer-Verlag New York, Inc.

Ley, E. and Steel, M. 2007. “On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions”. World Bank Working Paper 4238, June 2007.

Liang, F., Paulo, R., Molina, G., Clyde, M.A., and Berger, J.O. 2008. “Mixtures of g Priors for Bayesian Variable Selection”. Journal of the American Statistical Association 103, 410-423.

Lima, E.J.A., Luduvice, F. and Tabak, B.M. 2006. “Forecasting Interest Rates: an application for Brazil,” Working Papers Series 120, Central Bank of Brazil, Research Department.

Minella, A., Freitas,P.S., Goldfajn,I. Muinhos, M.K. 2003. “Inflation Targeting in Brazil: Constructing Credibility under Exchange Rate Volatility.” BCB Working Paper Series n.77.

Minella, A., Souza-Sobrinho,N.F. 2009. “Monetary Channels in Brazil through the Lens of a Semi-Structural Model,” Working Papers Series 181, Central Bank of Brazil, Research Department.

Obstfeld,M. and Rogoff,K.S. 1996. “Foundations of International Macroeconomics,” MIT Press Books, The MIT Press, edition 1, volume 1, number 0262150476.

Orphanides,A. 2001. “Monetary Policy Rules Based on Real-Time Data.” American Economic Review, American Economic Association, vol. 91(4), pages 964-985, September.

Oreiro,J.L., Feijó, C.A. 2010. “Desindustrialização: conceituação, causas, efeitos e o caso brasileiro”, Rev. Econ. Polit. vol.30 no.2 São Paulo Apr./June 2010

Petersen, K. 2007. “Does the Federal Reserve Follow a Non-Linear Taylor Rule?” Working papers 2007-37, University of Connecticut, Department of Economics.

Pfaff, B. 2008. “Analysis of Integrated and Cointegrated Time Series with R”. Second Edition. Springer, New York. ISBN 0-387-27960-1

Raftery, A.E., Madigan, D. and Hoeting, J.A. 1997. "Bayesian model averaging for regression models." Journal of the American Statistical Association, 92, 179-191.

Raftery, A.E. 1998. "Bayes Factors and BIC: Comment on *A Critique of the Bayesian Information Criterion for Model Selection* by David L. Weakliem" Socilogical Methods & Research, Vol. 27 No. 3, February 1999 411-427.

Ravn, M.O. and Uhlig, H., 2002. "On adjusting the Hodrick-Prescott filter for the frequency of observations," The Review of Economics and Statistics, MIT Press, vol. 84(2), pages 371-375.

Rudebusch, G., 2002. "Term structure evidence on interest rate smoothing and monetary policy inertia," Journal of Monetary Economics, Volume 49, Issue 6, September 2002, Pages 1161-1187

Stephens, M. 2010. "Bayesian analysis of mixture models with an unknown number of componentsan alternative to reversible jump methods". The Annals of Statistics 28 (2000), no. 1, 40–74. doi:10.1214/aos/1016120364. http://projecteuclid.org/euclid.aos/1016120364.

Stock, J.H. and Watson, M.W. 2003. "Has the Business Cycle Changed and Why?," NBER Chapters, in: NBER Macroeconomics Annual 2002, Volume 17, pages 159-230.

Taylor, J. 1993. "Discretion versus rules in practice". Carnegie Rochester Conference Series in Public Policy 39.

Walsh, C. 2010. "Monetary Theory and Policy, Third Edition," MIT Press Books, The MIT Press, 3 ed.

Wasserman, L. 2000 "Asymptotic inference for mixture models using data-dependent priors". J. Roy. Stat. Soc., Ser. B, Stat. Methodol. 62(1), 159-180.

Woodford, M. 2003a. "Interest and Prices: Foundations of a Theory of Monetary Policy". Princeton, NJ: Princeton University Press, 2003.

Woodford, M., 2003b. "Optimal Interest-Rate Smoothing". Review of Economic Studies, Wiley Blackwell, vol. 70(4), pages 861-886, October.

Zellner, A. 1986. "On assessing prior distributions and Bayesian regression analysis with g-prior distributions"'. In Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, (eds. P. K. Goel and A. Zellner), pp. 233-243. North-Holland/Elsevier.

Technical Appendix

Bayes factors consistency

Liang et al. (2008) showed that EBL and hyper-g priors have consistent Bayes factors when the true model is not the null one. The null-based Bayes factors for these priors are:

$$BF_{EBL}(M_s, M_N) = \frac{(1+g)^{\frac{(n-p-1)}{2}}}{(1+(1-R_s^2))g^{\frac{(n-1)}{2}}} \leq (1-R_s^2)^{\frac{-(n-1)}{2}}$$

$$BF_{h-g}(M_s, M_N) = \int L(g)\pi(g)\,dg = \int \left(1 - R_s^2 \frac{g}{1+g}\right)^{-\frac{(n-1)}{2}} \frac{\pi(g)}{(1+g)^{p/2}}\,dg$$

Assuming that the true model is not the null one both Bayes factors are consistent, ie. they tend to zero as the sample size increases

$$\underset{n\to\infty}{plim}\, BF(M_s, M_N) = 0$$

Unit Information Prior (UIP)

Under Kass and Raftery (1995) recommendation, we chose UIP as our reference prior: "if one were willing to use this prior as a reference prior, the Schawrz criterion would be a reasonably good approximation of the log to the Bayes factor."

The Unit Information Prior is a multivariate normal prior with mean at the maximum likelihood estimate and variance equal to the expected information matrix for one observation. Roughly speaking, this prior corresponds to assigning the same amount of information to the conditional prior of β as is contained in one observation.

As discussed in Raftery (1998) Unit Information Priors are the implicit priors used in BIC (bayesian information criteria) indicators. "[T]he unit information prior is well spread-out relative to the likelihood, and is relatively flat within the part of parameter space where the likelihood is substantial, without being much larger outside it. (...) In this situation, we can say that the likelihood dominates the prior. The unit information prior usually leads us to be in this (often desirable) situation. (...) Most of the criticisms of the unit information prior on which BIC is based imply that it is too spread out".

Risk Inflation Criterion Prior (RIC)

George and Foster (2000) named their prior proposal as Risk Inflation Criterion (RIC) because it asymptotically minimizes the maximum predictive risk inflation. Let $X = (x_1, \ldots, x_p)$ and the set of models $\gamma = 1, \ldots, 2^p$. The dimensions are X_γ is $(n \times q)$, β_γ is $(p \times 1)$. Let's assume the model

$$Y = X\beta + \epsilon \quad \text{where} \quad \epsilon \sim N_n(0, \sigma^2 I)$$

The sum of squares errors SSE_γ is the regression sum of squares for the γth model.

$$SSE_\gamma = Y'Y - \hat{\beta}'_\gamma X'_\gamma X_\gamma \hat{\beta}_\gamma$$

The penalised sum of squares criterion entails picking the γth model that minimizes $SSE_\gamma/\hat{\sigma}^2 + Fq_\gamma$ where F can be interpreted as a 'dimensionality penalty'. For $F = \log(n)$ we have BIC and for $F = 2\log(p)$ we have RIC.

Empirical Bayes Priors

Empirical Bayes was developed by George and Foster (2000). "To avoid some of the difficulties of a fully Bayes approach, we propose an empirical Bayes approximation that uses the data to estimate c and w. Although such an approximation ignores the uncertainty of the estimates by treating them as known, as opposed to a fully Bayes marginalisation over c and w, it at least avoids using arbitrary choices of c and w which may be at odds with the data." (c and w are hyper-parameters)

Empirical Bayes estimators are flexible and robust. According to Lehmann and Casella (1998) they have "proven to be an effective technique of constructing estimators that perform well under both Bayesian and frequentist criteria. One reason for this is that empirical Bayes estimators tend to be more robust against misspecification of the prior distribution."

In a hierarchical setup the desired model is conditioned on a θ parameter that is distributed according to a γ hyper-parameter that must be estimated based on the sample data X_i:

$$X_i|\theta \sim f(x|\theta)$$

$$\theta|\gamma \sim \pi(\theta|\gamma)$$

The marginal likelihood of $\mathbf{X}$ has density:

$$m(\mathbf{x}|\gamma) = \int \prod f(x_i|\theta)\pi(\theta|\gamma)\, d\theta$$

We estimate $\hat{\gamma}(\mathbf{x})$ by MLE of γ and determine the estimator $\hat{\theta}$ for θ by minimizing the empirical posterior loss:

$$\hat{\theta} = \underset{\theta}{argmin} \int L(\theta, \delta(\mathbf{x}))\pi(\theta|\mathbf{x}, \hat{\gamma}(\mathbf{x}))\, d\theta$$

The Empirical Bayesian estimator $\hat{\theta}$ is a Bayes estimator when the posterior density $\pi(\theta|\mathbf{x}, \hat{\gamma}(\mathbf{x}))$ is proper.

MCMC birth/death algorithm

Stephens (2010) propose the birth/death MCMC (BDMCMC), an alternative method to the Reversible Jump MCMC (RJMCMC), of constructing an ergodic Markov chain with stationary distribution, when the number of components of mixture model is unknown.

The BDMCMC algorithm view the parameters of the model as a marked (ie.weighed to sum 1) point process, with each point representing a component of the mixture. The scheme allows the number of components to vary by allowing new components to be "born" and existing components to "die". These births and deaths occur in continuous time, and the relative rates at which they occur determine the stationary distribution of the process. BDMCMC is a continuous-time version of RJMCMC, with a limit on the permitted types of moves to simplify implementation.

Algorithm 1.Simulation of a process with appropriate stationary distribution.

Assume $\pi = (\pi_1, \ldots, \pi_k)$ are the mixture proportions and $\phi = (\phi_1, \ldots, \phi_k)$ are the component specific parameters. There are k components. Let the likelihood be $L(k, \pi, \phi, \eta) = p(x^n|k, \pi, \phi, \eta)$, $y = (\pi_1, \phi_1), \ldots, (\pi_k, \phi_k)$ and $y \setminus (\pi_i, \phi_i) = y$ less the term (π_i, ϕ_i). To simulate a process with appropriate stationary distribution we iterate:

1. Let the birth rate $\beta(y) = \lambda_b$ (a constant).
2. Calculate the death rate $\delta(y_i)$ for each component, the death rate for component j being given by $\delta_j(y) = \lambda_b \frac{L(y \setminus (\pi_j, \phi_j))}{L(y)} \frac{p(k-1|\omega,\eta)}{kp(k|\omega,\eta)}$ where $j = (1, \ldots, k)$
3. Calculate the total death rate $\delta(y) = \sum_j \delta_j(y)$
4. Simulate time to next jump from an exponential distribution with mean $1/(\beta(y) + \delta(y))$
5. Simulate the type of jump: birth or death with respective probabilities, where $p(birth) = \beta(y)/(\beta(y) + \delta(y))$ and $p(death) = \delta(y)/(\beta(y) + \delta(y))$
6. Adjust y to reflect the birth or death. Birth: Simulate the point (π, ϕ) at which a birth takes place from the density $Beta(y, \pi, \phi) = k(1-\pi)^{k-1}\tilde{p}(\phi|\omega, \eta)$ by simulating π and ϕ independently from densities $k(1-\pi)^{k-1}$ and $\tilde{p}(\phi|\omega, \eta)$. Death: Select a component (π_i, ϕ_i) to die with probability $\delta_j(y)/\delta_(y)$. Note: $\phi_1, \ldots, \phi_k$ is independent and identically distributed from $\tilde{p}(\phi|\omega, \eta)$.
7. Return to step 2

Algorithm 2.Simulation of a Markov chain with appropriate stationary distribution.

This algorithm uses Gibbs sampling to simulate a value for $\Theta^{(t+1)} = \theta^{(t+1)}$ given the state $\Theta^{(t)} = \theta^{(t)}$ as time t:

1. Sample $(k^{(t)'}, \pi^{(t)'}, \phi^{(t)'})$ by running the birth-death process for a fixed time t_0 starting from $(k^{(t)}, \pi^{(t)}, \phi^{(t)})$ and fixing (ω, η) to be $(\omega^{(t)}, \eta^{(t)})$. Set $k^{(t+1)} = k^{(t)'}$.
2. Sample $(z^n)^{(t+1)}$ from $p(z^n|k^{(t+1)}, \pi^{(t)'}, \phi^{(t)'}, \eta^{(t)}, \omega^{(t)}, x^n)$.
3. Sample $\eta^{(t+1)}, \omega^{(t+1)}$ from $p(\eta, \omega|k^{(t+1)}, \pi^{(t)'}, \omega^{(t)'}, x^n, z^n)$.
4. Sample $\pi^{(t+1)}, \phi^{(t+1)}$ from $p(\pi, \phi|k^{(t+1)}, \eta^{(t+1)}, \omega^{(t+1)}, x^n, z^n)$.

Diebold Mariano Test for Predictive Accuracy

Assume $y_{t+h|t}$ are the true unknown values of the h-ahead forecast for series y_t. There are two competing forecast models, each generates residuals given by $\epsilon^i_{t+h|t} = y_{t+h|t} - y^i_{t+h|t}$ for $i = 1, 2$.

The Diebold-Mariano test is based on the squared error loss difference $d_t = (\epsilon^1_{t+h|t})^2 - (\epsilon^2_{t+h|t})^2$ and the test statistic is $S = \frac{d_t}{\sqrt{a\hat{var}(d_t)}}$.

The null of equal predictive accuracy is $H_o : S \sim N(0, 1)$ implying that $E(d_t) = 0$.

In our implementation, the model residuals were compared to the perfect predictor that generates a series of zeros. By rejecting the null hypothesis, the expected model residual is not zero and the model does not offer a good out-of-sample predictor.

Measures of Forecast Accuracy

Let ϵ_t be the forecast error given by $\epsilon_{t+1} = y_{t+1} - y_{t+1|t}$ where $y_{t+1|t}$ is the forecast of y_{t+1} made at time t. As a way of measuring the quadratic loss criterion, the Root Mean Square Error (RMSE) is a frequent choice.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{t+1} - y_{t+1|t})^2}$$

Unfortunately, this measure has some flaws: performance may be significantly affected by outliers, and it is inherently scale dependent, in that its magnitude depends not only on forecast accuracy, but also on the level of the underlying series. For example, a forecast 10% in excess of an actual value of $1,000,000$ result in a substantially greater RMSE than one 10% percent above an actual of $1,000$.

The Mean Absolute Percentage Error (MAPE) favored by practitioners, which expresses error as a fraction of the associated actual value, avoids the scale dependency of the RMSE.

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{t+1} - y_{t+1|t}}{y_{t+1}}\right|$$

MAPE also has disadvantages. Differently than RMSE, summary MAPE measures may be skewed by small actuals; indeed, the MAPE is infinite for an actual value of 0. MAPE also exhibits a counter-intuitive asymmetry: a forecast of 5 units on an actual of 10 produces an absolute percentage error of 50 percent, while a forecast of 10 units on an actual of 5 gives an APE of 100 percent.

LPS and Kullback-Leibler

Fernandez et al. (2001) relates Log Predictive Probability (LPS) to the Kullback-Leibler (KL) divergence. LPS for a h-step ahead predictive likelihood is

$$LPS(X_h, y, X) \equiv -\frac{1}{h}\sum_{i=1}^{h} \ln p(y_{N+i}|X_h, y, X)$$

The Kullback-Leibler divergence between the actual sampling density and the out-of-sample predictive density is given by

$$KL(p(y_h|X_h), p(y_h|X_h, y, X)) =$$
$$\int_{\mathbb{R}} \{\ln p(y_h|X_h)\} p(y_h|X_h)\, dy_h - \int_{\Re} \{\ln p(y_h|X_h, y, X)\} p(y_h|X_h)\, dy_h$$

where the first term is the negative entropy of the sampling density and the second can be seen as a theoretical counterpart of LPS for a given value of y_h. The entropy has a constant value regardless of the choice of M_s so that LPS is the variable associated the KL measure.

Table 1: Cointegration tests of interest rates

lag order=2	eigen test				trace test			
	test	10%	5%	1%	test	10%	5%	1%
$r \leq 6$	7.58	10.49	12.25	16.26	7.58	10.49	12.25	16.26
$r \leq 5$	20.84	16.85	18.96	23.65	28.42	22.76	25.32	30.45
$r \leq 4$	53.63	23.11	25.54	30.34	82.05	39.06	42.44	48.45
$r \leq 3$	68.83	29.12	31.46	36.65	150.88	59.14	62.99	70.05
$r \leq 2$	103.08	34.75	37.52	42.36	253.96	83.2	87.31	96.58
$r \leq 1$	126.28	40.91	43.97	49.51	380.23	110.42	114.9	124.75
$r \leq 0$	162.64	46.32	49.42	54.71	542.88	141.01	146.76	158.49
lag order=3	eigen test				trace test			
	test	10%	5%	1%	test	10%	5%	1%
$r \leq 6$	6.12	10.49	12.25	16.26	6.12	10.49	12.25	16.26
$r \leq 5$	18.33	16.85	18.96	23.65	24.46	22.76	25.32	30.45
$r \leq 4$	41.44	23.11	25.54	30.34	65.90	39.06	42.44	48.45
$r \leq 3$	47.72	29.12	31.46	36.65	113.61	59.14	62.99	70.05
$r \leq 2$	53.87	34.75	37.52	42.36	167.49	83.2	87.31	96.58
$r \leq 1$	73.47	40.91	43.97	49.51	240.95	110.42	114.9	124.75
$r \leq 0$	120.2	46.32	49.42	54.71	361.16	141.01	146.76	158.49

Tested with linear trend in cointegration with Pfaff (2008) routines.

Table 2: Data description and sources

Data	Description	Source
mpr	Monetary Policy annual rate (selic)	BCB (series 1178)
cpi	Current Consumer Price Inflation (ipca)	BCB (series 433)
cpie	Expected Inflation median 12 months ahead	BCB [1]
tcpi	Target Consumer Price Inflation	BCB [2]
eai	Economic Activity Index (seasonally adj.)	Serasa-Experian
credit/gdp	Tot credit non-earmarked funds/GDP	BCB (12127,12128)
debt/gdp	Net Public Debt/GDP (%)	BCB (series 4513)
rer	Real effective exchange rate index	BCB (11753)
comm	Commodity Index - Brazil - Index	BCB (series 20048)
cci	Consumer confidence index	BCB (series 4393)
ffr	Federal Funds Rate	FRED
di030	Swap reference rate 30-day (end period)	BCB (series 7816)
di060	Swap reference rate 60-day (end period)	BCB (series 7817)
di090	Swap reference rate 90-day (end period)	BCB (series 7818)
di120	Swap reference rate 120-day (end period)	BCB (series 7819)
di180	Swap reference rate 180-day (end period)	BCB (series 7820)
di360	Swap reference rate 360-day (end period)	BCB (series 7821)

Annualized data (constructed as a rolling window of the last 12 months)
[1] daily data was averaged to a monthly basis
[2] http://www.bcb.gov.br/Pec/metas/TabelaMetaseResultados.pdf

Table 3: Descriptive Statistics (full sample)

	Obs	Min	Median	Mean	Max	SE	Skew	E/K
mpr	126	8.65	13.52	14.73	26.32	4.57	0.73	-0.16
mprhp	126	9.80	14.06	14.71	21.40	3.86	0.28	-1.41
cpi	126	2.92	5.85	6.38	16.04	2.92	1.81	3.02
cpie	126	3.37	4.97	5.23	13.18	1.62	2.63	8.82
tcpi	126	3.50	4.50	4.45	5.50	0.46	0.20	1.26
hiatus	126	-1.13	0.53	0.78	9.68	1.78	2.78	9.42
eai	126	111.90	137.00	137.22	163.00	16.50	0.12	-1.35
gap	126	-6.75	0.38	0.00	5.58	2.43	-0.43	0.37
credit/gdp	126	11.50	77.66	98.50	236.61	71.76	0.29	-1.41
debt/gdp	126	35.04	46.67	46.50	62.86	6.77	0.34	-0.79
rer	126	61.56	96.73	103.64	206.11	34.09	0.81	-0.22
comm	126	57.30	101.53	102.27	138.00	15.19	-0.52	1.54
cci	126	84.40	133.74	132.94	170.18	21.31	-0.41	-0.69
ffr	126	0.25	1.25	1.95	5.25	1.77	0.78	-0.82
di030	126	8.32	13.35	14.66	26.77	4.64	0.76	-0.12
di060	126	8.18	13.33	14.69	27.17	4.71	0.78	-0.11
di090	126	8.08	13.31	14.73	27.49	4.77	0.80	-0.07
di120	126	8.01	13.39	14.78	27.75	4.83	0.83	-0.03
di180	126	7.94	13.46	14.86	28.20	4.93	0.91	0.15
di360	126	7.93	13.85	15.18	30.90	5.27	1.20	1.02

hiatus=cpie-tcpi, gap=eai-HPfilter(eai,λ), mprhp=HPfilter(mpr,λ)
SE = standard error, Skew = skewness, E/Kurt = Excess Kurtosis ($K - 3$)
Ravn and Uhlig (2002) recommend $\lambda = 129600$ for monthly data

Table 4: BMA Models with Unit Information Priors

Model M01 (UIP)				Model M02 (UIP)				Model M03 (UIP)				Model M04 (UIP)			
	PIP	Mean	S.E.		PIP	Mean	S.E.		PIP	Mean	S.E.		PIP	Mean	S.E.
mpr(-1)	100%	0.986	0.014	mpr(-1)	100%	0.948	0.019	mpr(-1)	100%	0.928	0.031	mpr(-1)	100%	0.436	0.061
				hiatus	100%	0.201	0.037	hiatus	100%	0.226	0.059	di030	95%	0.722	0.287
				gap	99%	0.096	0.028	gap	87%	0.078	0.041	di090	25%	-0.038	0.255
				mprhp	12%	0.003	0.014	credit/gdp	26%	-0.001	0.003	di060	24%	-0.042	0.305
								ffr	23%	0.013	0.031	di120	23%	-0.049	0.172
								debt/gdp	18%	-0.008	0.026	di180	23%	-0.024	0.087
								rer	16%	0.001	0.004	di360	18%	-0.008	0.030
								mprhp	12%	-0.001	0.026	rer	11%	0.000	0.001
								rer(+1)	12%	0.000	0.003	debt/gdp	10%	-0.001	0.006
								comm	12%	0.000	0.002	rer(+1)	10%	0.000	0.001
												mprhp	9%	-0.001	0.011
												comm	9%	0.000	0.001
												credit/gdp	9%	0.000	0.000
												gap	9%	0.000	0.007
												ffr	8%	0.000	0.007
												hiatus	8%	0.000	0.010
in-sample MAPE			0.05%	in-sample MAPE			0.24%	in-sample MAPE			0.23%	in-sample MAPE			0.29%
Diebold-Mariano			0.789	Diebold-Mariano			0.000	Diebold-Mariano			0.000	Diebold-Mariano			0.000
out-of-sample MAPE			0.56%	out-of-sample MAPE			0.75%	out-of-sample MAPE			0.97%	out-of-sample MAPE			0.54%
Log Predictive Score			0.770	Log Predictive Score			0.708	Log Predictive Score			0.681	Log Predictive Score			0.218

Table 5: BMA with fixed g priors

Model M04 (RIC)				Model M04 (BRIC)				Model M04 (EBL)			
	PIP	Mean	S.E.		PIP	Mean	S.E.		PIP	Mean	S.E.
mpr(-1)	100%	0.426	0.042	mpr(-1)	100%	0.425	0.042	mpr(-1)	100%	0.421	0.018
di030	100%	0.794	0.203	di030	100%	0.797	0.202	di030	100%	0.795	0.119
di120	26%	-0.043	0.104	di090	26%	-0.060	0.149	di120	47%	-0.080	0.087
di060	24%	-0.096	0.223	di120	25%	-0.042	0.100	di090	24%	-0.064	0.117
di180	23%	-0.022	0.061	di060	24%	-0.091	0.226	di180	19%	-0.020	0.043
di090	22%	-0.050	0.147	di180	23%	-0.022	0.059	di060	11%	-0.049	0.145
di360	18%	-0.008	0.025	di360	16%	-0.007	0.024	di360	3%	-0.001	0.008
mprhp	8%	-0.001	0.007	mprhp	8%	-0.001	0.007	mprhp	2%	0.000	0.002
debt/gdp	7%	0.000	0.004	debt/gdp	8%	0.000	0.004	rer	2%	0.000	0.000
rer	7%	0.000	0.001	rer(+1)	7%	0.000	0.001	debt/gdp	1%	0.000	0.001
rer(+1)	7%	0.000	0.001	credit/gdp	6%	0.000	0.000	comm	1%	0.000	0.000
ffr	7%	0.000	0.005	rer	6%	0.000	0.001	gap	1%	0.000	0.001
hiatuse	7%	0.000	0.007	ffr	6%	0.000	0.004	credit/gdp	1%	0.000	0.000
gap	7%	0.000	0.004	gap	6%	0.000	0.004	rer(+1)	1%	0.000	0.000
comm	6%	0.000	0.001	hiatuse	6%	0.000	0.006	ffr	0%	0.000	0.001
credit/gdp	6%	0.000	0.000	comm	6%	0.000	0.001	hiatuse	0%	0.000	0.001
in-sample MAPE			0.29%	in-sample MAPE			0.29%	in-sample MAPE			0.29%
Diebold-Mariano			0.000	Diebold-Mariano			0.000	Diebold-Mariano			0.000
out-of-sample MAPE			0.54%	out-of-sample MAPE			0.54%	out-of-sample MAPE			0.54%
Log Predictive Score			-0.194	Log Predictive Score			-0.194	Log Predictive Score			-0.800

Table 6: BMA with hyper-g priors

Model M04 (hyper h=2.01)				Model M04 (hyper h=3.0)				Model M04 (hyper h=4.0)			
	PIP	Mean	S.E.		PIP	Mean	S.E.		PIP	Mean	S.E.
mpr(-1)	100%	0.516	0.038	mpr(-1)	100%	0.512	0.039	mpr(-1)	100%	0.512	0.040
di030	93%	0.475	0.168	di030	94%	0.487	0.167	di030	92%	0.478	0.184
di060	8%	0.024	0.173	di060	7%	0.019	0.169	di060	8%	0.027	0.192
di090	3%	0.003	0.144	di090	4%	0.001	0.130	di090	5%	0.010	0.181
di120	3%	-0.010	0.095	di120	3%	-0.009	0.071	di180	4%	-0.009	0.054
di180	2%	-0.005	0.038	di180	3%	-0.006	0.039	di120	3%	-0.013	0.096
di360	2%	-0.002	0.014	di360	3%	-0.002	0.015	di360	3%	-0.003	0.017
mprhp	1%	0.000	0.004	rer(+1)	1%	0.000	0.000	rer(+1)	1%	0.000	0.001
rer	1%	0.000	0.000	debt/gdp	1%	0.000	0.002	rer	1%	0.000	0.000
debt/gdp	1%	0.000	0.002	rer	1%	0.000	0.000	debt/gdp	1%	0.000	0.002
rer(+1)	1%	0.000	0.000	mprhp	1%	0.000	0.003	mprhp	1%	0.000	0.004
credit/gdp	1%	0.000	0.000	hiatuse	1%	0.000	0.001	credit/gdp	1%	0.000	0.000
gap	0%	0.000	0.001	gap	1%	0.000	0.001	hiatuse	1%	0.000	0.001
hiatuse	0%	0.000	0.001	credit/gdp	1%	0.000	0.000	ffr	1%	0.000	0.001
comm	0%	0.000	0.000	ffr	0%	0.000	0.001	gap	1%	0.000	0.001
ffr	0%	0.000	0.001	comm	0%	0.000	0.000	comm	0%	0.000	0.000
in-sample MAPE			0.29%	in-sample MAPE			0.29%	in-sample MAPE			0.29%
Diebold-Mariano			0.000	Diebold-Mariano			0.000	Diebold-Mariano			0.000
out-of-sample MAPE			0.54%	out-of-sample MAPE			0.54%	out-of-sample MAPE			0.54%
Log Predictive Score			-0.586	Log Predictive Score			-0.587	Log Predictive Score			-0.592

Table 7: BMA with mixture priors

Model M05 (UIP)				Model M03 (RIC)				Model M03 (EBL)			
di030	100%	2.372	0.294	di030	100%	2.232	0.483	di030	100%	2.454	0.150
di090	93%	-2.071	1.025	di090	68%	-1.228	1.294	di090	100%	-2.458	0.559
di120	46%	0.546	0.728	di060	35%	-0.350	0.953	di120	65%	0.820	0.640
di180	38%	0.180	0.285	di120	32%	0.184	0.780	di180	34%	0.170	0.249
di060	17%	-0.046	0.583	di180	28%	0.114	0.328	di060	3%	0.015	0.179
di360	16%	0.014	0.057	di360	21%	0.020	0.084	di360	2%	0.002	0.020
in-sample MAPE			0.17%	in-sample MAPE			0.19%	in-sample MAPE			0.16%
Diebold-Mariano			0.001	Diebold-Mariano			0.000	Diebold-Mariano			0.001
out-sample MAPE			0.46%	out-sample MAPE			0.60%	out-sample MAPE			0.39%
Log Predictive Score			0.462	Log Predictive Score			0.800	Log Predictive Score			0.050
Model M04 - hyper h=2.01				Model M04 - hyper h=3.0				Model M04 - hyper h=4.0			
di030	100%	2.231	0.368	di030	100%	2.221	0.365	di030	100%	2.238	0.364
di090	54%	-0.659	0.786	di090	59%	-0.753	0.839	di090	58%	-0.762	0.865
di060	43%	-0.668	0.788	di060	39%	-0.594	0.776	di060	40%	-0.611	0.784
di120	10%	0.061	0.352	di120	12%	0.078	0.392	di120	13%	0.082	0.409
di180	6%	0.028	0.127	di180	9%	0.039	0.150	di180	10%	0.044	0.158
di360	4%	0.005	0.028	di360	5%	0.006	0.032	di360	5%	0.006	0.033
in-sample MAPE			0.17%	in-sample MAPE			0.17%	in-sample MAPE			0.17%
Diebold-Mariano			0.001	Diebold-Mariano			0.001	Diebold-Mariano			0.001
out-sample MAPE			0.39%	out-sample MAPE			0.39%	out-sample MAPE			0.39%
Log Predictive Score			0.113	Log Predictive Score			0.111	Log Predictive Score			0.109