Dynamic Prediction Pools:
An Investigation of Financial Frictions and Forecasting Performance

Marco Del Negro

Federal Reserve Bank

of New York

Raiden B. Hasegawa

University of Pennsylvania

Frank Schorfheide*

University of Pennsylvania

CEPR and NBER

This Version: November 19, 2015

Abstract

We apply a novel methodology for estimating time-varying weights in linear prediction pools, which we call Dynamic Pools, and use it to investigate the relative forecasting performance of DSGE models with and without financial frictions for output growth and inflation from 1992 to 2011. We find strong evidence of time variation in the pool's weights, reflecting the fact that the DSGE model with financial frictions produces superior forecasts in periods of financial distress but does not perform as well in tranquil periods. The dynamic pool's weights react in a timely fashion to changes in the environment, leading to real-time forecast improvements relative to other methods of density forecast combination, such as equal-weights combination, Bayesian model averaging, optimal static pools, and dynamic model averaging. We show how a policymaker dealing with model uncertainty could have used a dynamic pool to perform a counterfactual exercise (responding to the gap in labor market conditions) in the immediate aftermath of the Lehman crisis.

JEL CLASSIFICATION: C53, E31, E32, E37.
KEY WORDS: Bayesian estimation, DSGE Models, Financial Frictions, Forecasting, Great Recession, Linear Prediction Pools.

1   Introduction

Many macroeconomists paid scant attention to financial frictions models before the recent Great Recession. As a consequence, most of the dynamic stochastic general equilibrium (DSGE) models used by monetary policy-making institutions at the onset of the recession in 2007 were variants of the Smets and Wouters (2007) (henceforth, SW) model, which have no credit market frictions. Yet financial frictions mechanisms, e.g. Kiyotaki and Moore (1997) and Bernanke et al. (1999), that could be built into estimable DSGE models were available long before the Great Recession. Christiano et al. (2003) was the first paper to incorporate a credit market friction into a large-scale DSGE model suitable for prediction and policy analysis.

Building on the work of Christiano et al. (2003), Christiano et al. (2014), and others, Del Negro and Schorfheide (2013) show that an enlarged SW model with financial frictions along the lines of Bernanke et al. (1999) (henceforth SWFF) would have done a much better job forecasting the dynamics of real GDP growth and inflation in the aftermath of the Lehman Brothers collapse in the fall of 2008. The left panels of Figure 1, which is taken from Del Negro and Schorfheide (2013), show real-time forecasts of real GDP growth and GDP deflator inflation from our version of the SW model without financial frictions (henceforth SWπ). The figure highlights that this model was blindsided by the subsequent drop in output and inflation in the last quarter of 2008. The right panels show forecasts obtained using the financial frictions model SWFF, which is designed to account for real-time information coming from financial market spreads. The figure suggests that DSGE models with credit market imperfections might have provided policymakers with a reasonable outlook for the economy in the aftermath of the crisis. To our knowledge, however, these models were not used. Why not?

This paper documents that the forecasting performance of the SWFF model is better than that of the SWπ model during financially turbulent times, but it is worse during tranquil times. Our measure of forecasting performance throughout the paper is the log predictive density score for a joint forecast of average output growth and inflation over a period of four quarters. This evidence is consistent with Stock and Watson (2003), who find, using reduced form models, that asset prices are not particularly useful on average in forecasting output and inflation. The time-variation in the relative forecasting performance of these

Figure 1: DSGE Model Forecasts of the Great Recession
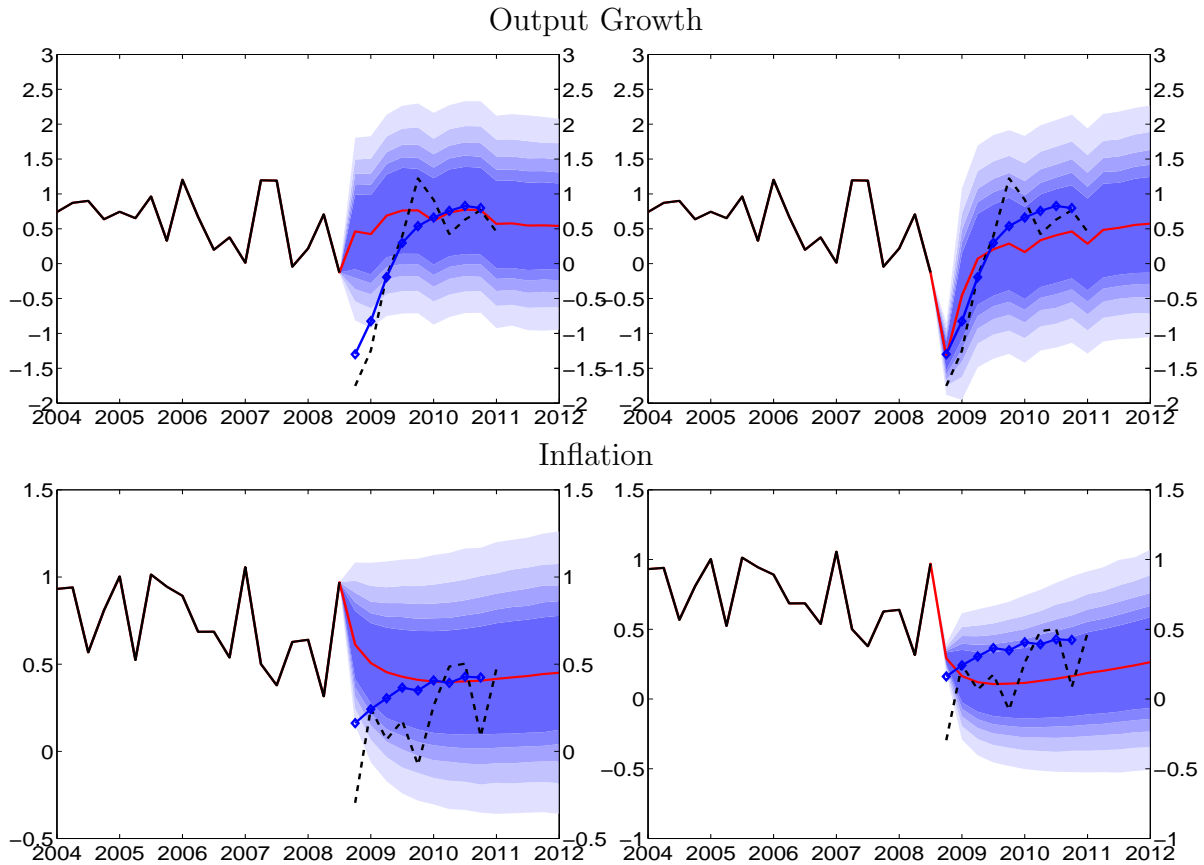
Notes: This figure is taken from Del Negro and Schorfheide (2013). The panels show for each model real time data on real GDP growth (upper panel) and inflation (GDP deflator, lower panel) (black line); multi-step (fixed origin) posterior mean forecasts (red line); credible bands from the predictive distributions (shaded blue areas; these are the 50, 60, 70, 80, and 90 percent bands, in decreasing shade); Blue Chip forecasts (blue diamonds); and the actual realizations according to the May 2011 vintage (black dashed line). All the data are in quarter-on-quarter percent.

two models, which is a feature more broadly encountered by forecasters and policymakers dealing with the issue of model uncertainty, raises an important question: How should the models be combined for real time predictions and policy analysis?

To address this question, we extend existing methods of combining predictive densities from recursively estimated econometric models using time-varying weights. This dynamic linear prediction pool relies on a sequence of weights that follow an exogenous process, respecting the constraint that the model weights have to lie on a simplex. Throughout the paper we focus on the combination of two models (SWπ and SWFF), which means that the

time-varying weights correspond to a process λ_t on the unit interval. Our setup takes the form of a state-space model, in which a nonlinear state-transition equation determines the law of motion of λ_t and the linear combination of the predictive densities obtained from the two DSGE models provides the measurement equation.

Most closely related to our work is the literature on density forecast combination, specifically Hall and Mitchell (2007), Geweke and Amisano (2011), Waggoner and Zha (2012), and Billio et al. (2013). Hall and Mitchell (2007) and Geweke and Amisano (2011) use the sequence of predictive densities, which measure the likelihood of ex-post outcomes from the perspective of a model's ex-ante forecast distribution, to construct optimal linear pools.[1] These pools are optimal in the sense that the weights are chosen to maximize the pool's historic forecast performance. Because this optimization is performed under the assumption that "in population" the optimal combination weight does not vary over time, we refer to the resulting prediction pool as static.

Waggoner and Zha (2012) extend the Geweke and Amisano (2011) approach to a setting in which the combination weights follow a Markov switching process. Our approach differs from theirs in several dimensions: First, we are using a smooth autoregressive process for the evolution of the pool's weights instead of using a Markov-switching process, allowing for potentially slow rather than drastic changes in the combination weights. Second, Waggoner and Zha (2012) emphasize the joint estimation of model parameters and combination weights, which implies that observations in periods in which a model receives little weight, should be heavily discounted in the estimation of that model's parameters. While this joint estimation is conceptually interesting and potentially desirable, we think that in a typical setting at a policy institution, it is unrealistic that DSGE models are re-estimated when being pooled.[2] Third, and most important, the Waggoner and Zha (2012) approach requires that all models under consideration share the same set of observables, a requirement that is not met when the key difference across models is the set of observables, as is the case in many interesting applications including the one considered here. To justify our inference procedure for the model parameters we propose a principal-agent framework in which the policymaker has to aggregate predictive densities that she receives from two DSGE modelers.

---

[1]Our paper focuses on linear pools. Nonlinear pools are discussed, for instance, in Gneiting and Ranjan (2013) and Fawcett et al. (2014).

[2]Candidate models may be maintained by different divisions; may not be re-estimated every quarter; or may not be estimated using full information methods, e.g., the Board of Governor's FRB/US model.

Billio et al. (2013) propose a general approach to combine predictive densities using time-varying weights that nests static linear pools, the Markov-switching weight specification of Waggoner and Zha (2012), and the dynamic linear pool considered in our paper. While the framework of Billio et al. (2013) is very general, their simulation experiments and empirical applications are necessarily restricted to very specific laws of motion for the model weights. The law of motion used in our paper differs from those applied in their computational work. We control the prior persistence, volatility, and the mean of the weight process through hyperparameters that are estimated in real time. *A priori* uniform model weights are nested as a special case. As Billio et al. (2013), we also use sequential Monte Carlo techniques to solve a nonlinear filtering problem. More specifically, we use a particle MCMC method, see Andrieu et al. (2010), that combines a particle filter for the hidden model combination weights with a Metropolis-Hastings algorithm that is used for posterior inference on the hyperparameters.

Our empirical application considers forecasts of four-quarter-ahead average output growth and inflation obtained from combinations of the SWπ and the SWFF model. We compare the weight evolution implied by our dynamic pools to that implied by the static pool approach, Bayesian model averaging (BMA), and the dynamic model averaging (DMA) approach proposed by Raftery et al. (2010). We find that after prolonged periods in which one model performs better than the other, all four approaches – quite naturally – put more weight on the best-performing model. BMA often goes to the extreme of assigning weights of 0 and 1 to the competing models, while both the static and the dynamic pool tend to maintain the benefits of model diversification. BMA and the static pool are characterized by a slow reaction to reversals in the relative forecasting performance, however. Because they are caught flat-footed when the environment changes, these procedure perform worse in terms of real-time accuracy than naive combination procedures such as equal weights. Conversely, dynamic pool weights can change rapidly, implying limited losses relative to the equal weights approach when reversals occurs. Because these reversals tend to be persistent, the dynamic pool eventually outperforms equal weights in our application.

The posterior distribution of model weights tends to be tilted toward the model without financial friction during tranquil times, but shifts rapidly in the other direction once financially turbulent times start.[3] In particular, we show that by the time the Lehman

---

[3]This finding seems to be consistent with evidence from reduced-form regime-switching models reported

crisis struck, the real-time distribution of λ_t was putting considerable mass on the SWFF model. We conclude that while macroeconomists had some reason for not relying exclusively on the SWFF model before the crisis, there was no justification for ignoring it.[4] We also perform a counterfactual policy exercise using the real-time weight distribution available in the aftermath of the Lehman crisis. We show that a policymaker would have had reasons to switch from the historical policy Taylor-type rule to a policy that strongly responds to measures of gaps in the labor market, had she used dynamic pools to address the issue of model uncertainty.

The remainder of this paper is organized as follows. Section 2 discusses the pricipal-agent setting which we use to describe the model combination environment. Section 3 provides a unified framework for characerizing static model combination procedures such as Bayesian model averaging and static pools, and sets the stage for the introduction of dynamic pools, which are described in Section 4. Section 5 describes the models, the data on which they are estimated, and the evolution in their forecasting performance for output and inflation over time. Section 6 presents the results and Section 7 concludes. An online appendix contains detailed description of the DSGE models, a description of the data set, computational details, and additional empirical results.

2   A Stylized Principle-Agent Framework

Estimation of large-scale DSGE models can be computationally costly. In central banks, this task is typically delegated to (groups of) expert modelers. Each modeling group builds, estimates, and maintains their model and generates model output that can be used in the policy-making process, e.g., an account of current and historical events, forecasts, and policy scenarios. Individuals involved in making policy decisions typically face the problem of aggregating or pooling the output from different models. In this paper, we focus on the pooling of density forecasts from two DSGE models. Our general approach is neither restricted to

---

in Alessandri and Mumtaz (2014).

[4]Our findings may in part be due to the use of linearized DSGE models. A non-linear model with financial friction may look (and forecast) very much like one without friction in tranquil times, but have very different dynamics when the financial constraints become binding, e.g., Brunnermeier and Sannikov (2014), Dewachter and Wouters (2012), and Bocola (2016). However, none of these models are detailed enough to be usable in a real-time forecasting application.

DSGE models applications nor to the combination of only two models and we comment on these generalizations in the conclusion.

To characterize the density forecast combination problem, we consider a stylized principal-agent framework. In this setting each agent is an econometric modelers who estimates a DSGE model and uses it to generate density forecasts. We assume that there are two modelers, $\mathcal{M}_1$ and $\mathcal{M}_2$, who each maintain their own DSGE model.[5] The principal is a policymaker who aggregates the density forecast of the modelers. The modelers are rewarded based on the accuracy of their prediction. Both modelers generate predictive densities of the form $p(y_t|\mathcal{I}_{t-1}^m, \mathcal{M}_m)$ for a vector y_t, $t = 1, \ldots, T$ of variables of interest. Let $y_{1:t}$ denote the sequence $\{y_1, .., y_t\}$ and note that for each of the two models the information set $\mathcal{I}_{t-1}^m$ may, in general, be larger than $y_{1:t-1}$. For instance, the (model-specific) information sets may include additional variables $z_{1:t-1}^m$ such that $\mathcal{I}_{t-1}^m = \{y_{1:t-1}, z_{1:t-1}^m\}$ and

$$p(y_t|\mathcal{I}_{t-1}^m, \mathcal{M}_m) = p(y_t|y_{1:t-1}, z_{1:t-1}^m, \mathcal{M}_m). \tag{1}$$

In the application in Section 6, y_t includes output growth and inflation, z_t includes consumption, investment, hours per capita, real wage growth, the federal funds rate, long-run inflation expectations, for both models, and, in addition, spreads for one of the two models.

We assume that the agents are rewarded based on the log predictive score $\ln p(y_t|\mathcal{I}_{t-1}^m, \mathcal{M}_m)$. The log predictive score is a proper scoring rule that induces truth telling, i.e., the agents maximize their expected payoff by reporting the actual predictive density associated with their model. Moreover, Agent 1's payoff is independent of the accuracy of the forecast of Agent 2 (and vice versa) which eliminates strategic interactions. This assumption reflects the observation that modeling groups in policy-making institutions typically do not take into account that the output of their model may be combined with the output of other models. They simply strive to provide the best characterization of the economy obtainable within their particular modeling framework.

In every period t, the principal receives the predictive densities $p(y_{t+1}|\mathcal{I}_t^m, \mathcal{M}_m)$ and combines them using a linear prediction pool

$$p(y_{t+1}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \hat{\lambda}_t p(y_{t+1}|\mathcal{I}_t^1, \mathcal{M}_1) + (1 - \hat{\lambda}_t)p(y_{t+1}|\mathcal{I}_t^2, \mathcal{M}_2). \tag{2}$$

---

[5]Because each modeler maintains only one model we use the notation $\mathcal{M}_m$ for both modelers and models.

The remainder of this paper focuses on the construction of the sequence of weights $\hat{\lambda}_t$. We assume that the principal's information set $\mathcal{I}_t^{\mathcal{P}}$ contains the following objects: (i) the sequence of actual observations $y_{1:t}$; and (ii) the sequence of predictive densities $\left\{p(\tilde{y}_\tau | \mathcal{I}_{\tau-1}^m, \mathcal{M}_m)\right\}_{\tau=1}^{t+1}$, where $\tilde{y}_\tau$ denotes the generic argument of the predictive density function. Based on this information, the principal can evaluate the first t predictive densities $\left\{p(\tilde{y}_\tau | \mathcal{I}_{\tau-1}^m, \mathcal{M}_m)\right\}_{\tau=1}^{t}$ at the realized value $\tilde{y}_\tau = y_\tau$ and generate a density forecast for the future (and thus far unobserved) y_{t+1} by pooling the modelers' density forecasts.

3 Bayesian Model Averaging and Static Pools

To simplify the exposition of BMA (Section 3.1) and static pools (Section 3.2), we will start from the assumption that neither modeler includes additional variables z_t^m in his model and the information sets of the modelers are simply $\mathcal{I}_t^m = y_{1:t}$, $m = 1, 2$. A generalization to larger information sets $\mathcal{I}_t^m$ that might differ across modelers is provided in Section 3.3. Starting point for both BMA and the static pool is the assumption that the policymaker combines the predictive densities provided by the modelers using the following mixture:

$$p(y_{t+1} | \lambda, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \lambda p(y_{t+1} | y_{1:t}, \mathcal{M}_1) + (1 - \lambda) p(y_{t+1} | y_{1:t}, \mathcal{M}_2), \tag{3}$$

The main difference between BMA and static pools lies in the assumption about the domain of λ. Under both approaches, the policymaker has to infer λ based on past observations.

3.1 Bayesian Model Averaging

In the context of our principal-agent setup, the premise of BMA is that the policymaker believes that one of the two DSGE models is correctly specified but she does not know which one.[6] Thus, the domain of λ is restricted to take one of the following two values:

$$\lambda \in \{1, 0\}.$$

---

[6]BMA was advocated by Leamer (1978) as a way to account for model uncertainty. Hoeting et al. (1999) provide a review of BMA techniques. Forecasting applications in the econometrics literature include, among others, Min and Zellner (1993) and Wright (2008).

From the policymaker's perspective $\mathcal{M}_1$ ($\mathcal{M}_2$) is correct if $\lambda = 1$ ($\lambda = 0$). In this setup the joint density of a sequence of observations $y_{1:T}$, which is called marginal likelihood in Bayesian analysis, is given by

$$p(y_{1:T}|\lambda, \mathcal{P}) = \begin{cases} p(y_{1:T}|\mathcal{M}_1) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \mathcal{M}_1) & \text{if } \lambda = 1, \\ p(y_{1:T}|\mathcal{M}_2) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \mathcal{M}_2) & \text{if } \lambda = 0. \end{cases} \quad (4)$$

The policymaker can conduct inference on λ by sequentially updating the probability that $\lambda = 1$. Let λ_0^{BMA} denote the prior probability of $\lambda = 1$ and

$$\hat{\lambda}_t^{BMA} = \mathbb{P}(\{\lambda = 1\}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) \quad (5)$$

the posterior probability conditional on the policymaker's information set. A straightforward application of Bayes Theorem leads to the recursive updating formula

$$\hat{\lambda}_t^{BMA} = \frac{\hat{\lambda}_{t-1}^{BMA} p(y_t|y_{1:t-1}, \mathcal{M}_1)}{\hat{\lambda}_{t-1}^{BMA} p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1 - \lambda_{t-1}^{BMA}) p(y_t|y_{1:t-1}, \mathcal{M}_2)}. \quad (6)$$

After taking expectations of the right-hand-side of (3) conditional on $y_{1:T}$ to integrate out the unknown λ we obtain

$$p_{BMA}(y_{t+1}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \hat{\lambda}_t^{BMA} p(y_{t+1}|y_{1:t}, \mathcal{M}_1) + (1 - \hat{\lambda}_t^{BMA}) p(y_{t+1}|y_{1:t}, \mathcal{M}_2), \quad (7)$$

which takes the form of (2).

3.2 Static Pools

Strictly speaking, a policymaker engaging in Bayesian model averaging operates under the belief that the model space is complete in the sense that one of the two models is correct. She uses the data to infer which of the two models is the correct one. Alternatively, and probably more realistically, the policymaker may be concerned that the model space is incomplete and neither model is correctly specified. This case is depicted in a stylized manner in the top panel of Figure 2. The figure shows the two DSGE models, represented by $p(y_{1:T}|\mathcal{M}_m)$ and a generic data generating process (DGP), represented by the density $p(y_{1:T})$. The length of the arrows depicts the Kullback-Leibler (KL) discrepancy between the DGP and two

Figure 2: Combining Models if the Model Space is Incomplete

(i) Bayesian Model Averaging

$DGP = p(y_{1:T})$

KL Discrepancy

$p(y_{1:T}|\mathcal{M}_1)$

$p(y_{1:T}|\mathcal{M}_2)$

(ii) Linear Static Pools

$DGP = p(y_{1:T})$

$p(y_{1:T}|\mathcal{M}_1)$

$p(y_{1:T}|\mathcal{M}_2)$

$$p(y_{1:T}|\lambda, \mathcal{P}) = \prod_{t=1}^{T} \left\{ \lambda \, p\left(y_t|y_{1:t-1}, \mathcal{M}_1\right) + (1 - \lambda) \, p\left(y_t|y_{1:t-1}, \mathcal{M}_2\right) \right\}$$
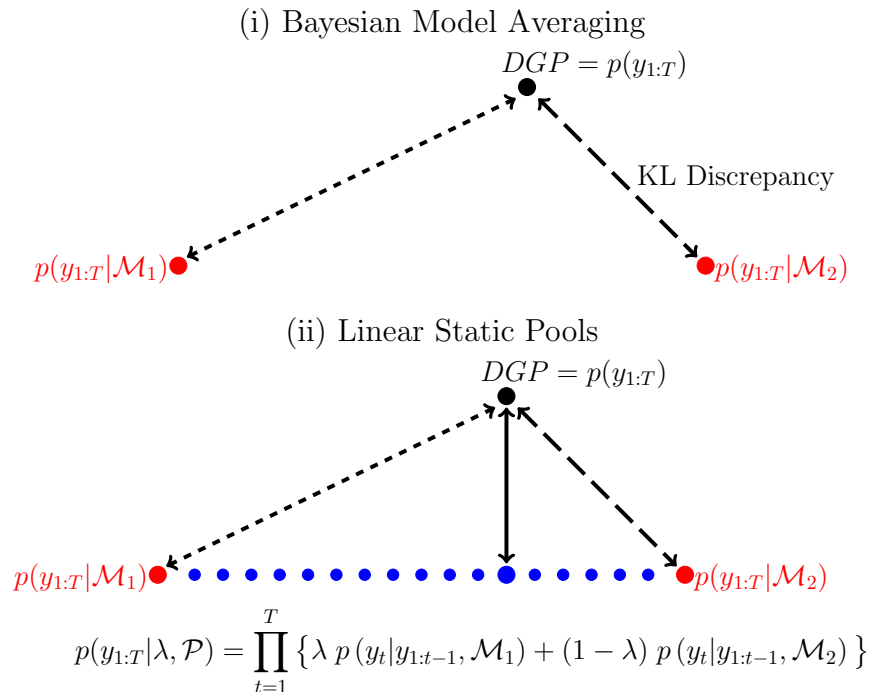
models. The posterior model probability $\hat{\lambda}_T^{BMA}$ recursively defined in (6) has the property that under a stable DGP it converges to one almost surely if $\mathcal{M}_1$ is closer (in terms of the KL discrepancy) to the DGP than $\mathcal{M}_2$. Vice versa, if $\mathcal{M}_2$ is closer to the DGP, $\hat{\lambda}_T^{BMA}$ converges to zero as $T \longrightarrow \infty$. This basic result has been proved in the literature under various conditions, e.g., Dawid (1984) or Geweke and Amisano (2011), and it implies that BMA assigns all the weight to a single model.

The optimal static prediction pool approach advocated by Hall and Mitchell (2007) and Geweke and Amisano (2011) extends the domain of λ from the set $\{0, 1\}$ to the unit interval $\lambda \in [0, 1]$, which allows for a convex combination (rather than a selection) of the one-step-ahead predictive densities $p(y_t|y_{1:t-1}, \mathcal{M}_1)$ and $p(y_t|y_{1:t-1}, \mathcal{M}_2)$ in (3). This approach generalizes (4) to

$$p(y_{1:T}|\lambda, \mathcal{P}) = \prod_{t=1}^{T} \{\lambda p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1 - \lambda) p(y_t|y_{1:t-1}, \mathcal{M}_2)\}, \quad \lambda \in [0, 1]. \quad (8)$$

A stylized representation of this extension is depicted in the bottom panel of Figure 2. The dots connecting the end points $p(y_{1:T}|\mathcal{M}_1)$ and $p(y_{1:T}|\mathcal{M}_2)$ represent the models associated

with the likelihood function given in (8). Hall and Mitchell (2007) interpret this approach as an extension of the Bates and Granger (1969) point forecast combination method to the case of density forecasts, whereas Geweke and Amisano (2011) view it as a way of implementing Bayesian inference on an incomplete model space. By creating a linear combination of the two models with non-degenerate weights, it is possible to reduce the KL discrepancy to the DGP. Of course, as indicated in Figure 2 there is no guarantee that a linear pool actually covers the DGP.

To conduct Bayesian inference about the unknown weight λ the policymaker can start from a prior density $p(\lambda|\mathcal{P})$ and recursively update the believes about λ based on Bayes Theorem as follows. Suppose $p(\lambda|\mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$ denotes the posterior density conditional on $\mathcal{I}_{t-1}^{\mathcal{P}}$, then

$$p(\lambda|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \frac{\left[\lambda p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1-\lambda)p(y_t|y_{1:t-1}, \mathcal{M}_2)\right]p(\lambda|\mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})}{\int_0^1 \left[\lambda p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1-\lambda)p(y_t|y_{1:t-1}, \mathcal{M}_2)\right]p(\lambda|\mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})d\lambda}. \quad (9)$$

This formula generalizes the recursive computation of the posterior probability of $\{\lambda = 1\}$ in (6). The numerator of (9) equals the one-step-ahead predictive density $p(y_t|\mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$. As in the case of BMA, the policymaker can form a one-step-ahead prediction by integrating out λ from (2), which amounts to setting

$$\hat{\lambda}_t^{BSP} = \mathbb{E}[\lambda|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}] = \int_0^1 \lambda p(\lambda|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})d\lambda, \quad (10)$$

and leads to

$$p_{BSP}(y_{t+1}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \hat{\lambda}_t^{BSP}p(y_{t+1}|y_{1:t}, \mathcal{M}_1) + (1 - \hat{\lambda}_t^{BSP})p(y_{t+1}|y_{1:t}, \mathcal{M}_2). \quad (11)$$

We use the abbreviation BSP for Bayesian static pool. The key difference between BMA and the static pool is the updating rule for $\hat{\lambda}_t$. In the former case it is given by (6), whereas in the latter case it is given by (9) and (11).[7]

3.3 Modelers' Use of Larger Information Sets

The policymaker may only care about a subset of variables, e.g., output growth and inflation as in our application, that are used by the modelers to estimate their DSGE models. In other

---

[7] Hall and Mitchell (2007) and Geweke and Amisano (2011) define $\hat{\lambda}_t$ as the argmax of the likelihood function (8). We refer to their pool as maximum (likelihood) static pool (MSP), using the notation $\hat{\lambda}_t^{MSP}$ and $p_{MSP}(y_{t+1}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$.

words, the modelers may use additional variables z_t^m (in addition to y_t) in their econometric analysis and these additional variables may differ across modelers. In this case, the formula in (6) could be justified as follows. In slight abuse of notation, define the vector $z_t = z_t^1 \cup z_t^2$ as the union of the variables contained in z_t^1 and z_t^2 and let $\mathcal{I}_t = \{y_{1:t}, z_{1:t}\}$. Moreover, assume that the DSGE models are specified such that additional information through variables in z_t that are not included in the original model specification, do not change the predictive density in the sense that

$$p(y_t|\mathcal{I}_{t-1}^m, \mathcal{M}_m) = p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_m). \tag{12}$$

In our empirical application y_t is composed of output growth and inflation, z_t^1 includes interest rates, hours worked, consumption growth, investment, and real wage growth. z_t^2 comprises z_t^1 as well as a measure of interest rate spreads. The assumption stated in (12) implies that modeler $\mathcal{M}_1$ would not change his forecast in view of the interest rate spread series, because she is using a model that is not designed to incorporate this information.

Suppose now that the policymaker's beliefs about the evolution of z_t take the form of a density $p(z_t|y_t, z_{1:t-1}, \mathcal{I}_{t-1}^P, \mathcal{P})$. That is, let us say that the policymaker uses "her own" model to obtain projections for the z_t variables because she believes that neither $\mathcal{M}_1$ or $\mathcal{M}_2$ is suitable for forecasting them (e.g., spreads). Then we can write

$$
\begin{aligned}
&p(y_t, z_t|z_{1:t-1}, \lambda, \mathcal{I}_{t-1}^P, \mathcal{P}) \tag{13} \\
&= \left[\lambda p(y_t|y_{1:t-1}, z_{1:t-1}, \mathcal{M}_1) + (1-\lambda)p(y_t|y_{1:t-1}, z_{1:t-1}, \mathcal{M}_2)\right] \times p(z_t|y_t, z_{1:t-1}, \mathcal{I}_{t-1}^P, \mathcal{P}).
\end{aligned}
$$

Under BMA, if we let $\hat{\lambda}_t^{BMA}$ be the probability of $\lambda = 1$ conditional on the extended information set $(z_{1:t}, \mathcal{I}_t^P)$, then we can generalize (6) to

$$\hat{\lambda}_t^{BMA} = \frac{\hat{\lambda}_{t-1}^{BMA} p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1)}{\hat{\lambda}_{t-1}^{BMA} p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1) + (1-\lambda_{t-1}^{BMA})p(y_t|\mathcal{I}_{t-1}^2, M_2)}. \tag{14}$$

Note, however, that the policymaker never explicitly uses the additional information $z_{1:t}$, because the density $p(z_t|y_t, z_{1:t-1}, \mathcal{I}_{t-1}^P, \mathcal{P})$ does not depend on λ and cancels out in the calculation of $\hat{\lambda}_t^{BMA}$. Thus, as long as the policymaker's beliefs about the evolution of the z_t's are independent of whether model $\mathcal{M}_1$ or $\mathcal{M}_2$ is correct, she could update posterior probabilities for $\mathcal{M}_1$ and $\mathcal{M}_2$ based on the predictive densities that she receives from the models and the information set $\mathcal{I}_t^P$ which excludes $z_{1:t}$. There is no need to evaluate a conditional density for z_t. A similar argument applies to the static pool. The information

Figure 3: Dynamic Pools: Likelihood Function
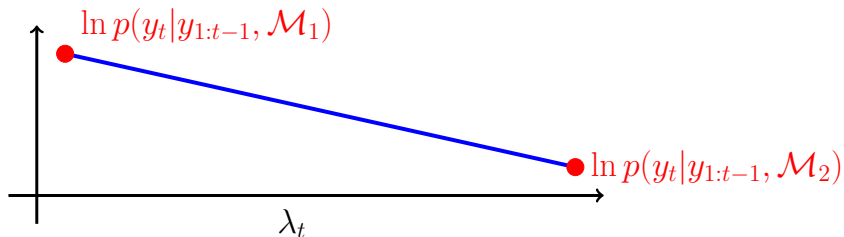
set $\mathcal{I}_{t-1}^{\mathcal{P}}$ in (9) can be replaced by $(z_{1:t}, \mathcal{I}_t)$ and the density $p(z_t|\cdot)$ appears in both the numerator and denominator of (9) and therefore cancels and does not affect $p(\lambda|z_{1:t}, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = p(\lambda|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$. In other words, if the policymaker has her own model to forecast z_t, then she will use only information from y_t to discriminate between $\mathcal{M}_1$ and $\mathcal{M}_2$.

4 Dynamic Pools

In view of structural changes in the macro economy over the past six decades it is plausible to let the pooling weight λ evolve over time. Thus, in the remainder of the paper, we focus on a model combination procedure in which the fixed λ is replaced by a sequence λ_t, $t = 1, 2, \ldots$. To simplify the exposition we assume again that $\mathcal{I}_t = \mathcal{I}_t^m = y_{1:t}$, that is, neither modeler uses additional variables z_t^m to estimate the DSGE model. Following the same arguments as in Section 3.3, everything goes through when the information set is expanded to include z_t^m. The policymaker's likelihood function takes the form

$$p(y_{1:T}|\lambda_{1:T}, \mathcal{P}) = \prod_{t=1}^{T} \{\lambda_t p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1 - \lambda_t) p(y_t|y_{1:t-1}, \mathcal{M}_2)\}, \quad \lambda_t \in [0, 1]. \quad (15)$$

The time t contribution to the log likelihood function is depicted in Figure 3. Note that a straight maximization with respect to $\lambda_{1:T}$ yields the uninteresting corner solutions

$$\hat{\lambda}_t = \begin{cases} 1 & \text{if } p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1) > p(y_t|\mathcal{I}_{t-1}^2, \mathcal{M}_2) \\ 0 & \text{if } p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1) < p(y_t|\mathcal{I}_{t-1}^2, \mathcal{M}_2) \end{cases}, \quad t = 1, \ldots, T$$

and is not useful for predictive purposes because the forecast of y_{t+1} requires knowledge of λ_{t+1}.

4.1 A Stochastic Process Prior for the Model Weights

In order to allow for some degree of persistence in the optimal combination weights, we assume that $\lambda_{1:T}$ evolves according to a stochastic process. Our prior for the process $\lambda_{1:T}$ is indexed by three hyperparameters: ρ controls the persistence of λ_t, μ its long-run mean, and σ the shape of the density associated with the marginal distribution of λ_t. Let

$$
\begin{aligned}
x_t &= (1-\rho)\mu + \rho x_{t-1} + \sqrt{1-\rho^2}\sigma\varepsilon_t, \quad \varepsilon_t \sim iid\ N(0,1), \quad x_0 \sim N(\mu, \sigma^2), \\
\lambda_t &= \Phi(x_t),
\end{aligned}
\tag{16}
$$

where $\Phi(.)$ is the cumulative distribution function (cdf) of a $N(0,1)$ random variable.[8] The law of motion for λ_t nests several interesting special cases. Suppose that $\mu = 0$ and $\sigma = 1$. In this case the marginal distribution of x_t is $N(0,1)$ and the marginal distribution of λ_t is $U[0,1]$ for each t. The hyperparameter ρ controls the persistence of the x_t and λ_t processes. The closer ρ is equal to one, the more slowly the weights on the two DSGE models change. In the limit, if $\rho = 1$, the dynamic pool reduces to the static pool: $\lambda_t = \lambda$ for each t and $\lambda \sim U[0,1]$. Choosing a positive (negative) value of μ shifts the mean of the unconditional distribution of the model weight λ_t from $1/2$ to one (zero). Finally, if $\sigma < 1$, then the variance of x_t is less than one and the probability integral transform based on the $N(0,1)$ cdf generates a density for λ_t with an inverted-U shape, increasing the probability that both models receive equal weight. Vice versa, values $\sigma > 1$ generate a U-shaped density for λ_t, increasing the probability that λ is either close to one or to zero. In sum, the process (16) generates a sequence of priors $p(\lambda_t|\lambda_{t-1}, \theta, \mathcal{P})$, where θ stacks the hyperparameters (ρ, μ, σ).

4.2 Posterior Inference

To conduct inference about the sequence of weights λ_t the policymaker has to combine the likelihood function (15) with the prior (16). The dynamic pool can be viewed as a nonlinear state-space model in which λ_t is the hidden state, (16) describes the state transition, and the convex combination of time t predictive densities in (3) with λ replaced by λ_t is the measurement equation. Let $p(\lambda_{t-1}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$ denote the posterior distribution of λ_{t-1} given

---

[8] In order to extend the approach to more than two models one could proceed as in Billio et al. (2013), replacing x_t by a vector process and mapping it onto the simplex using, say, a logistic transformation or treating the elements of an x_t vector as concentration parameters of a Dirichlet distribution. However, these approaches generally do not lead to a uniform marginal distribution of λ_t.

the policymaker's time $t-1$ information set $\mathcal{I}_{t-1}^{\mathcal{P}}$ (which includes $y_{1:t-1}$). The time t forecasting step of a nonlinear filter for the state-space model generates the the one-step-ahead predictive density for λ_t:

$$p(\lambda_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}) = \int p(\lambda_t|\theta, \lambda_{t-1}, \mathcal{P})p(\lambda_{t-1}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})d\lambda_{t-1}. \tag{17}$$

The application of Bayes Theorem leads to the policymaker's posterior distribution of λ_t and corresponds to the updating step of a nonlinear filter:

$$p(\lambda_t|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \frac{[\lambda_t p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1-\lambda_t)p(y_t|y_{1:t-1}, \mathcal{M}_2)]p(\lambda_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})}{p(y_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})}. \tag{18}$$

Because of the nonlinearity of the state-space representation of the dynamic pool, we use a bootstrap particle filter (see online appendix for details) to approximate the sequence of densities $p(\lambda_t|\theta, \mathcal{I}_t^{\mathcal{P}}, P)$. Based on the posterior distribution of λ_t, we can define

$$\hat{\lambda}_{t+1|t}^{DP}(\theta) = \mathbb{E}[\lambda_{t+1}|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}] = \int_0^1 \lambda_{t+1}p(\lambda_{t+1}|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P})d\lambda_{t+1}, \tag{19}$$

which leads to

$$p_{DP}(y_{t+1}|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \hat{\lambda}_{t+1|t}^{DP}(\theta)p(y_{t+1}|y_{1:t}, \mathcal{M}_1) + (1 - \hat{\lambda}_{t+1|t}^{DP})p(y_{t+1}|y_{1:t}, \mathcal{M}_2). \tag{20}$$

The dependence on the hyperparameter θ can be eliminated by integrating out θ using the posterior distribution $p(\theta|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$, which is discussed in more detail below.

4.3 Multi-Step Forecasting

Policy makers are generally more interested in forecasts of average growth rates over the next h periods than the growth rate between period $t+h-1$ and period $t+h$. Accordingly, we define the object of interest as

$$\bar{y}_{t+h,h} = \frac{1}{h}\sum_{s=1}^h y_{t+s} \tag{21}$$

and assume that the policymaker receives the densities $p(\bar{y}_{t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m)$ from the modelers. To the extent that the policymaker is concerned about the misspecification of the h-step predictive densities that she receives from the modelers, it is reasonable to adopt a loss-function-based approach and estimate the sequence of combination weights λ_t separately

for each horizon h. As is common in the literature on predictive regressions and multi-step estimation, e.g., Schorfheide (2005) and the references cited therein, we ignore the overlap between $\bar{y}_{t,h}$, $\bar{y}_{t-1,h}$, etc. and assume that the policymaker conducts inference based on a pseudo-likelihood function of the form

$$p^{(h)}(\bar{y}_{1:T,h}|\lambda_{1:T},\mathcal{P}) = \prod_{t=1}^{T} \left\{ \lambda_t p(\bar{y}_{t,h}|\mathcal{I}_{t-h}^1,\mathcal{M}_1) + (1-\lambda_t)p(\bar{y}_{t,h}|\mathcal{I}_{t-h}^2,\mathcal{M}_2) \right\} \tag{22}$$

using the same prior distribution as in (16).

In period t the information set $\mathcal{I}_t^{\mathcal{P}}$ of the policymaker consists of the following objects: (i) the sequence of actual observations $\bar{y}_{1:t,h}$; and (ii) the sequence of predictive densities $\{p(\tilde{y}_\tau|\mathcal{I}_{\tau-h}^m,\mathcal{M}_m)\}_{\tau=1}^{t+h}$, where $\tilde{y}_\tau$ denotes the generic argument of the probability density function. Based on this information, the policymaker can evaluate the first t densities $\{p(\tilde{y}_\tau|\mathcal{I}_{\tau-h}^m,\mathcal{M}_m)\}_{\tau=1}^{t}$ at the observed value $\tilde{y}_\tau = y_t$. This leads to a sequence of pseudo-posterior distributions $p^{(h)}(\lambda_t|\theta,\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$, which are updated according to Bayes Theorem:

$$p^{(h)}(\lambda_t|\theta,\mathcal{I}_t^{\mathcal{P}},\mathcal{P}) = \frac{[\lambda_t p(\bar{y}_{t,h}|\mathcal{I}_{1:t-h}^1,\mathcal{M}_1) + (1-\lambda_t)p(\bar{y}_{t,h}|\mathcal{I}_{1:t-h}^2,\mathcal{M}_2)]p^{(h)}(\lambda_t|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P})}{p^{(h)}(\bar{y}_{t,h}|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P})}, \tag{23}$$

where

$$p(\lambda_t|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P}) = \int p(\lambda_t|\theta,\lambda_{t-1},\mathcal{P})p(\lambda_{t-1}|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P})d\lambda_{t-1}. \tag{24}$$

Equations (23) and (24) generalize (18) and (17), respectively, to multi-step forecasting. Iterating the law of motion of (x_t,λ_t) in (16) forward for h periods, we can construct

$$\hat{\lambda}_{t+h|t}^{DP}(\theta) = \int_0^1 \lambda_{t+h} \left[\int_0^1 p(\lambda_{t+h}|\lambda_t)p^{(h)}(\lambda_t|\theta,\mathcal{I}_{t-h},\mathcal{P})d\lambda_t \right] d\lambda_{t+h}. \tag{25}$$

Conditional on θ the h-step ahead predictive density of the dynamic pool is then given by

$$p_{DP}^{(h)}(\bar{y}_{t+h}|\theta,\mathcal{I}_t^{\mathcal{P}},\mathcal{P}) = \hat{\lambda}_{t+h|t}^{DP}(\theta)p(\bar{y}_{t+h}|\mathcal{I}_t^1,\mathcal{M}_1) + \left(1-\hat{\lambda}_{t+h|t}^{DP}(\theta)\right)p(\bar{y}_{t+h}|\mathcal{I}_t^2,\mathcal{M}_2). \tag{26}$$

We construct h-step ahead predictive densities for BMA and the static pool in a similar manner, using the corresponding pseudo-likelihood function.

The hyperparameter θ can be integrated out from the combination weight $\hat{\lambda}_{t+h|t}^{DP}(\theta)$ using the pseudo-posterior distribution $p^{(h)}(\theta|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$, which is obtained from

$$p^{(h)}(\theta|\mathcal{I}_t^{\mathcal{P}},\mathcal{P}) \propto \left(\prod_{t=1}^{T} p^{(h)}(\bar{y}_{t,h}|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P}) \right) p(\theta), \tag{27}$$

where $p(\theta)$ is a prior distribution for the hyperparameter vector θ and the likelihood increments $p^{(h)}(\bar{y}_{t,h}|\theta, \mathcal{I}^{\mathcal{P}}_{t-1}, \mathcal{P})$ appear as normalization constants in the denominator of (23) and are generated as a byproduct of the nonlinear filter that iterates over the forecasting step (24) and the updating step (23). In order to generate draws from the pseudo-posterior we use a particle-MCMC technique (see, Andrieu et al. (2010) as well as the online appendix) that combines a bootstrap particle filter with a random-walk Metropolis-Hastings algorithm. We drop the (θ) argument to denote the marginal posterior mean of λ that is obtained by integrating out θ and write

$$\hat{\lambda}^{DP}_{t+h|t} = \int \hat{\lambda}^{DP}_{t+h|t}(\theta)p^{(h)}(\theta|\mathcal{I}_t, \mathcal{P})d\theta. \tag{28}$$

Moreover, removing θ from the conditioning set, we let

$$p^{(h)}_{DP}(\bar{y}_{t+h}|\mathcal{I}^{\mathcal{P}}_t, \mathcal{P}) = \hat{\lambda}^{DP}_{t+h|t}p(\bar{y}_{t+h}|\mathcal{I}^1_t, \mathcal{M}_1) + \left(1 - \hat{\lambda}^{DP}_{t+h|t}\right)p(\bar{y}_{t+h}|\mathcal{I}^2_t, \mathcal{M}_2). \tag{29}$$

The empirical application in Section 6 focuses on the marginal posterior of the weights $p^{(h)}(\lambda_t|\mathcal{I}^{\mathcal{P}}_t, \mathcal{P})$, the h-step ahead weight $\hat{\lambda}^{DP}_{t+h|t}$ and the log scores $\ln p^{(h)}_{DP}(\bar{y}_{t+h}|\mathcal{I}_t, \mathcal{P})$.

4.4 Regime Switching and Dynamic Model Averaging

We motivated the prediction pool as a linear combination of two predictive densities. Alternatively, the prediction pool could be reinterpreted as a two-state regime switching model. Let s_t be the regime indicator. If $s_t = 1$ then the predictive y_t is given by $p(y_t|\mathcal{I}^1_{t-1}, \mathcal{M}_1)$ and if $s_t = 2$, then $p(y_t|\mathcal{I}^2_{t-1}, \mathcal{M}_2)$. Under our proposed dynamic pool, the probability of being in state $s_t = 1$ is determined by the hidden process x_t. However, unlike Waggoner and Zha (2012)'s dynamic pool, our setup does not have a first-order Markov structure in terms of s_t. It can be verified (see NBER Working Paper 20575 version of this paper) that higher-order lags of s_t provide additional information about the hidden process x_t and the probability of transitioning to $s_t = 1$ depends on the entire regime history.[9]

Raftery et al. (2010) proposed a model combination method, called dynamic model averaging (DMA), which can be interpreted as a computational short-cut to the implementation

---

[9]Chang et al. (2014) proposed a regime switching model with an autoregressive factor in which the hidden state s_t is directly tied to the latent factor x_t through a threshold rule of the form $s_t = \mathcal{I}\{x_t \geq \tau\}$. Their model (unlike our model) is isomorphic to Hamilton's regime switching model if x_t evolves exogenously.

of the Waggoner and Zha (2012) approach.[10] Rather than trying to estimate the Markov transition probabilities for s_t and forecasting and updating the s_t process using a filter, the forecast under the DMA approach is obtained by downweighing the current (say, $t-1$) state probabilities:

$$\mathbb{P}\{s_t = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\} = \frac{\left[\mathbb{P}\{s_{t-1} = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\}\right]^{\alpha}}{\sum_{m=1,2} \left[\mathbb{P}\{s_{t-1} = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\}\right]^{\alpha}}. \tag{30}$$

The tuning parameter α can be interpreted as forgetting factor. If $\alpha = 1$, then the probability that $s_t = m$ is equal to the probability that $s_{t-1} = m$. If $\alpha = 0$, then the probability that $s_t = m$ is simply $1/2$. Using (30), the updating equation can be expressed as

$$\mathbb{P}\{s_t = m | \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}\} = \frac{p(y_t | \mathcal{I}_{1:t-1}^{m}, \mathcal{M}_m)\left[\mathbb{P}\{s_{t-1} = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\}\right]^{\alpha}}{\sum_{m=1,2} p(y_t | \mathcal{I}_{1:t-1}^{m}, \mathcal{M}_m)\left[\mathbb{P}\{s_{t-1} = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\}\right]^{\alpha}}. \tag{31}$$

Note that the updating formula implies that

$$\mathbb{P}\{s_t = m | \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}\} \propto \prod_{\tau=1}^{t-1} \left[p(y_{t-\tau} | \mathcal{I}_{1:t-\tau-1}^{k}, \mathcal{M}_m)\right]^{\alpha^{\tau}}. \tag{32}$$

In the empirical analysis we will compare our proposed dynamic pools approach to DMA. To adapt DMA to multi-step forecasting we replace $p(y_t | \mathcal{I}_{1:t-1}^{m}, \mathcal{M}_m)$ by $p(\bar{y}_{t+h} | \mathcal{I}_t^{m}, \mathcal{M}_m)$.

5   Two DSGE Models: Specification, Estimation, and Forecasting

Our empirical analysis is based on two DSGE models. The first DSGE model, $\mathcal{M}_1$, is a modified version of the Smets and Wouters (2007) model. The SW model is based on earlier work by Christiano et al. (2005) and Smets and Wouters (2003). It is a medium-scale DSGE model, which augments the standard neoclassical stochastic growth model with nominal price and wage rigidities as well as habit formation in consumption and investment adjustment costs. We modify the SW model by introducing a time-varying inflation target and including long-run inflation expectations into the set of observables that is used to estimate the model. A justification for this modification is provided in Del Negro and Schorfheide (2013). We call the resulting specification SWπ model.

---

[10]Koop and Korobilis (2012) used DMA to average over inflation density forecasts.

The second DSGE model, $\mathcal{M}_2$, is obtained by adding financial frictions to the SWπ model and builds on work by Bernanke et al. (1999), Christiano et al. (2003), De Graeve (2008), and Christiano et al. (2014). In this DSGE model, banks collect deposits from households and lend to entrepreneurs who use these funds as well as their own wealth to acquire physical capital, which is rented to intermediate goods producers. Entrepreneurs are subject to idiosyncratic disturbances that affect their ability to manage capital. Their revenue may thus be too low to pay back the bank loans. Banks protect themselves against default risk by pooling all loans and charging a spread over the deposit rate. This spread varies exogenously due to changes in the riskiness of entrepreneurs' projects and endogenously as a function of the entrepreneurs' leverage. We refer to the second DSGE model as SWFF model. All ingredients of the SWFF model were publicly available prior to 2008. As such, the model does not include some of the features that may have been found to be relevant following the crisis.

The SWπ model is estimated based on quarterly data on U.S. output growth, consumption growth, investment growth, real wage growth, hours worked, inflation, the federal funds rate, and ten-year ahead inflation expectations. The history of these series generates the information set $\mathcal{I}_t^m$. The estimation of the SWFF model is based on nine variables: the same eight time series used for the SWπ model and an additional time series for spreads. In the empirical analysis we will assume that the policymaker is interested in output growth and inflation forecasts, which means that the vector y_t includes those two series. The remaining series are part of the vectors z_t^1 and z_t^2, respectively. The formal specifications of the two models and precise data definitions are provided in the online appendix.

Our empirical analysis is based on the real-time data set constructed in Del Negro and Schorfheide (2013). This data set reconstructs the actual information sets available to modelers on January 1st, April 1st, July 1st, and October 1st of each year, accounting for the fact that the macroeconomic time series are released with a delay and are subsequently revised by the statistical agencies. For instance, on January 1st 2009, the modelers have access to NIPA data until 2008:Q3. We adopt the convention that $h = 1$ corresponds to the nowcast of 2008:Q4, $h = 2$ refers to 2009:Q1, and so forth. When plotting results over time we use the convention that t equals the quarter corresponding to the latest NIPA data. Thus, t corresponds to 2008:Q3 for forecasts made on January 1st 2009.

While NIPA data for the forecast origins are only available with a one-quarter delay,

Figure 4: Log Scores Comparison: SWFF vs SWπ
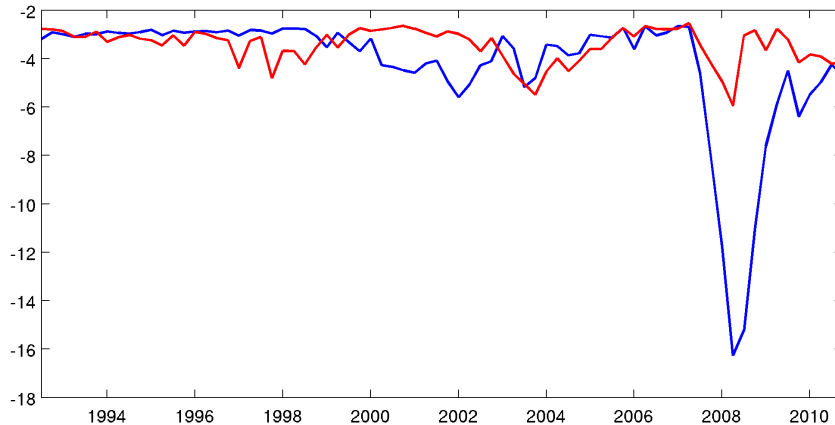
Notes: The figure shows the log scores $p(\bar{y}_{t+h,h}|\mathcal{I}^m_{t+}, \mathcal{M}_m)$ for SWFF (red), and SWπ (blue) over the period 1992:Q1-2011:Q2.

financial data are available in real time. The empirical analysis distinguishes between unconditional and semi-conditional forecasts. The unconditional forecasts are based on the information set $\mathcal{I}^m_t$ that use financial data only up to the same quarter (t) for which NIPA data are available. The semi-conditional forecasts are based on the information set $\mathcal{I}^m_{t+}$, which includes the $t+1$ federal funds rate and spreads in addition to $\mathcal{I}^m_t$. For instance, on January 1st 2009, the semi-conditional forecasts use interest rates and spreads from 2008:Q4. All of the subsequent results are based on the semi-conditional forecasts, if not otherwise explicitly indicated, because that is the natural information set for a forecaster.

Both DSGE models are recursively estimated using Bayesian techniques. A summary of the prior distribution is provided in Table A-1 of the online appendix. A detailed discussion of the prior can be found in Del Negro and Schorfheide (2013). Each estimation sample starts in 1964:Q1. Our forecast origins (and hence the endpoints of the estimation samples) range from 1992:Q1 to 2011:Q2. In order to pool the DSGE model forecasts, the policymaker needs to receive the predictive densities $p(\bar{y}_{t+h,h}|\mathcal{I}^m_t, \mathcal{M}_m)$ (or $p(\bar{y}_{t+h,h}|\mathcal{I}^m_{t+}, \mathcal{M}_m)$) from the modelers.

Figure 4 depicts the log scores $p(\bar{y}_{t+h,h}|\mathcal{I}^m_{t+}, \mathcal{M}_m)$ (semi-conditional forecasts) for the predictions of four-quarter-ahead ($h = 4$) average output growth and inflation for the SWπ (blue) and the SWFF (red) model. The forecast origins range from 1992:Q1 to 2011:Q2.[11]

---

[11] Figure A-9 in the online appendix shows the log predictive densities using one-quarter ahead forecasts.

These scores are subsequently used as the inputs for the policymaker's density combination. Two features of Figure 4 are apparent. First, the financial frictions model grossly outperform the SWπ model during the recent financial crisis, which is not surprising in light of Figure 1. Second, the relative forecasting performance of the two models varies over time, with the SWFF outperforming the SWπ model during periods of financial turmoil, such as the early-2000s dot-com bust and the Great Recession, and the opposite occurring during more tranquil periods. Del Negro and Schorfheide (2013) report similar findings using 12-periods moving averages of four-quarter-ahead rolling RMSEs for both output growth and inflation.[12]

# 6 Results from the Dynamic Prediction Pool

We now apply the dynamic pools methodology discussed in Section 6 to the forecasts of four-quarter-ahead average output growth and inflation obtained from the SWπ model and the SWFF model. We ask the following questions: How do the weights assigned to the two DSGE models by the Bayesian dynamic pool (DP) procedure evolve over time and how do the DP weights compare to those obtained from BMA or the static pool (MSP) (Section 6.1)? In particular, do the DP weights change rapidly enough when estimated in real time to offer useful guidance to policymakers and/or forecasters? How do the hyperparameters of the DP procedure affect the speed of reaction to changes in the relative forecasting performance of the component models (Section 6.2)? How do the various pooling procedures perform in terms of real-time forecasting accuracy? Section 6.3 presents results for joint forecasts of output and inflation (additional results for the two series individually can be found in the online appendix). Finally, how can a policymaker use the DP procedure to perform a counterfactual policy analysis (Section 6.4)? While we use the semi-conditional forecast to compute the forecast accuracy statistics as well as in the estimation of the combination weights, we sometimes drop the '+' superscript in t^+ to simplify the notation.

The pattern is similar to that of Figure 4 except that the log scores are much noisier. In any case, nowcasting is arguably not the best use of DSGE models, hence in our application we focus on projections for four-quarter ahead averages.

[12]Similarly, Kolasa and Rubaszek (2015) show that in normal times DSGE models without financial frictions perform better than models with frictions in normal times. However, they also find that models with frictions in the housing market perform better in forecasting during the Great Recession than models with standard financial frictions like the one considered here – a result that deserves further study.

Figure 5: Weights in Real Time: Dynamic Pool, BMA, and Static Pool
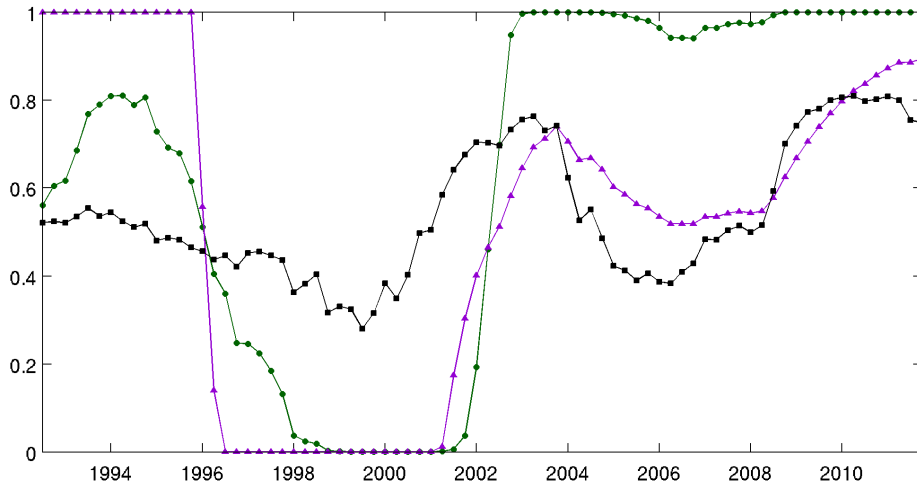
Notes: The figure shows the weight on the SWFF model in forecast pools, computed using real time information only, over the period 1992:Q1-2011:Q2 for three different pooling techniques: dynamic pool ($\hat{\lambda}_{t|t}^{DP}$ – black), (maximum likelihood) static pool ($\hat{\lambda}_t^{MSP}$ – purple), and BMA ($\hat{\lambda}_t^{BMA}$ – green). Prior 1 is used for the DP: $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$.

6.1 Pooling Weights

Figure 5 shows the weight on the SWFF model in forecast pools over the period 1992:Q1-2011:Q2 obtained from the proposed dynamic pooling technique ($\hat{\lambda}_{t|t}^{DP}$ – black) as well as the static pool ($\hat{\lambda}_t^{MSP}$ – purple) with weights estimated by maximum likelihood, and BMA ($\hat{\lambda}_t^{BMA}$ – green). All of the weights and the hyperparameter estimates for the DP are computed in real time, based on information that would have been available to the policymaker at the time of the combination of the model forecasts. The DP weights reported in the figure are based on the following prior distribution for the hyperparameter vector θ:

$$\text{Hyperparameter Prior 1: } \rho \sim \mathcal{U}[0,1], \quad \mu = 0, \quad \sigma = 1. \tag{33}$$

Under this prior distribution, the marginal distribution of λ_t is also $\mathcal{U}[0,1]$.

By construction, the DP weight of the SWFF model starts close to 0.5 at the beginning of the sample. The weight falls through the 1990s because the SWπ model forecasts better than SWFF, but the drop is not very rapid because the log score differential is not substantial. As soon as the relative forecasting performance of the SWπ and the SWFF model switches in 1999 (see Figure 4) the DP weight starts rising rapidly. The weight on the SWFF model peaks in 2003 after a five year period in which the financial friction model outperformed the

SWπ model. From 2004 to 2006 the relative ranking of the forecasts is reversed and the weight on the SWFF model from around 0.7 to 0.4. Due to the poor forecast performance of the SWπ model during the Great Recession, the weight of the SWFF model increases to 0.8 by 2010 and stayed there until the end of our sample period in 2011:Q2.

The MSP weights of the static pool evolve markedly different from the DP weights in the first half of the sample and exhibit a bang-bang behavior for the first ten years. In part this is the consequence of using the mode rather than the mean of the posterior distribution of λ to form the combination weights, as we further discuss later. After 2002 movements in the static pool weights $\hat{\lambda}_t^{MSP}$ mirror those in $\hat{\lambda}_{t|t}^{DP}$ with two differences. First, $\hat{\lambda}_{t|t}^{DP}$ moves more rapidly than $\hat{\lambda}_t^{MSP}$. For example, starting in 2008 $\hat{\lambda}_{t|t}^{DP}$ drifts upwards toward the SWFF model faster than the SP weight. Second, movements in $\hat{\lambda}_{t|t}^{DP}$ tend to exhibit less inertia than those in $\hat{\lambda}_t^{MSP}$. For example, as the difference in forecasting performance narrows by the end of 2009, $\hat{\lambda}_t^{MSP}$ continues to rise while $\hat{\lambda}_{t|t}^{DP}$ has stopped increasing. In general the MSP weights react more sluggishly to new information because the assumption of constant weights implies that the historical forecast performance is not discounted as strongly as in the dynamic combination approach.

The BMA weight behaves as expected from the discussion in Section 3.1: as soon as enough information accumulates in the tranquil 1990s that the SWπ model fares better than its competitor, the SWFF weight approaches zero and remains there for almost four years. Around 2002, after almost two years in which SWπ was outperformed by SWFF, the weight rapidly shifts to the opposite extreme, reaching one by the end of 2002. Except for a small drop in 2006, the BMA weight of the SWFF stays close to one until the end of the sample.[13]

In order to understand the difference between the behavior of the combination procedures, it is instructive to look at the entire posterior distribution of the weights. The two panels of Figure 6 show the posterior distributions $p_{DP}^{(h)}(\lambda_t|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ and $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$, respectively. There is no separate panel for BMA, because, in a nutshell, BMA takes the posterior mass of $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ and allocates it to the endpoints $\lambda=0$ and $\lambda=1$. Under both posterior distributions the probability mass is fairly uniformly allocated across the unit interval prior

---

[13]Results for forecasts that do not condition on the $t+1$ financial information are reported in Figures A-6 and A-7 of the online appendix. For the unconditional forecasts $\hat{\lambda}_t^{BMA}$ is close to from 2006 to 2008 whereas the DP and MSP weights are similar to the semi-conditional case. This illustrates the well-known lack of robustness of BMA weights: minor changes in model specification or observables can lead to very different outcomes in terms of marginal likelihood comparisons.

Figure 6: Posterior distribution of λ Over Time
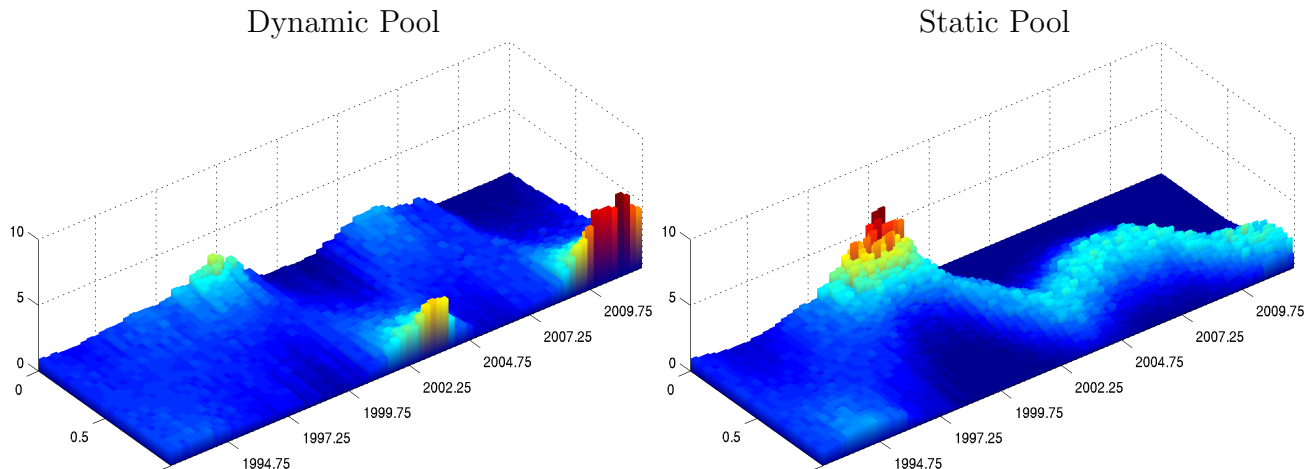
Notes: The two panels show the posterior distributions $p_{DP}^{(h)}(\lambda_t|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ (dynamic) and $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$, respectively, for t =1992:Q1-2011:Q2. Prior 1 is used for the DP specification: $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$.

to 1997. During this period the main difference in the weights $\hat{\lambda}_{t|t}^{DP}$ and $\hat{\lambda}_t^{MSP}$ stems from the fact that the former is computed as posterior mean, whereas the latter is computed as posterior mode. The mode of $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ swings from $\lambda = 1$ to $\lambda = 0$ in 1996, and remains at 0 until the end of 2001. These swings in the mode of the distribution explain the bang-bang pattern of $\hat{\lambda}_t^{MSP}$ highlighted in Figure 5. After 1997, two features of the posterior distribution associated with the static pool emerge. First, the mass is rather tightly concentrated around the mode. For instance, at the end of our sample $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ puts virtually no mass on $\{\lambda \leq .5\}$. Second, the distribution $p_{BSP}^{(h)}(\lambda|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$ moves sluggishly over time. For the static pool new information has only a small effect on the overall posterior of λ.

Starting in 1998, the posterior distribution of λ_t (dynamic pool) looks very different from the posterior of λ (static pool). While the latter is concentrated around the mean, the former is characterized by a "sloshing" of the mass from one side to the other mirroring the alternations in the relative forecasting performance shown in Figure 4. In all periods but those of transition, the marginal distribution of λ_t is similar to that depicted in Figure 3, in that it tilts in one direction or another (the mode distribution is either one or zero) depending on which model has been the best forecaster in recent periods. Loosely speaking, we can think of the distribution as a seesaw, whose slope depends on the recent gap in log scores between the two models. A narrowing of the gap leads to a flattening of the seesaw,

shifting the mean toward equal weights, and a reversal of the gap causes the seesaw to tilt in the opposite direction. The frequency and speed of the oscillations in the seesaw, and its fulcrum, depend on the hyperparameters ρ, μ, and σ introduced in Section 4.1, whose posterior distribution we are going to discuss next.

6.2 How Dynamic Is the Dynamic Pool? A Look at the Hyper-parameter Estimates

The law of motion for λ_t is based on three hyperparameters: ρ, σ, and μ. We will now discuss the role of these hyperparameters and their posterior estimates, starting with the most important one: ρ. To continue with the seesaw analogy, ρ affects the frequency of the seesaw oscillations. If ρ is close to zero, the seesaw's slope is only determined by the disparity in log scores in the *current* period. As ρ increases, the changes in relative performance of the two models have to become persistent in order to alter the seesaw's inclination. In other words, ρ determines the "forgetting factor" of dynamic pools. As ρ approaches one, there is no discounting of past information and dynamic pool turns into a (Bayesian) static pools.

The first set of results is based on the prior in (33). The left panel of Figure 7 shows the end-of-sample ($t = T$) posterior $p^{(h)}(\rho|\mathcal{I}_T^{\mathcal{P}}, \mathcal{P})$ (histogram) together with the $\mathcal{U}[0, 1]$ prior for ρ (red line). As one might have expected, the posterior distribution concentrates around relative high values of ρ, indicating that the distribution of the weights does not react much to temporary changes in log scores. The mode of the posterior is between 0.8 and 0.9, with most of the mass being in the $[0.75, 0.95]$ interval. Importantly, the mass drops rapidly as ρ approaches one, suggesting that the data do not favor a static pool over the proposed dynamic pool. The posterior distribution $p^{(h)}(\rho|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$ evolves over time. Until 1998 the posterior of ρ is fairly flat, but subsequently most of the probability mass shifts toward the interval $[0.75, 0.95]$. After 2004 the general shape of the posterior stays very similar but the concentration of probability mass continues until the end of the sample.[14]

Posterior distributions for σ^2 and μ are depicted in the center panel and the right panel

---

[14]The evolution of the posterior is depicted in Figure A-2 of the online appendix. The appendix also documents the effect of shifting prior mass toward one by considering Beta distributions for ρ with (mean, standard deviation) of (0.8, 0.1) and (0.9, 0.2), respectively.

Figure 7: Prior and End-of-Sample Posterior of the Hyperparameters
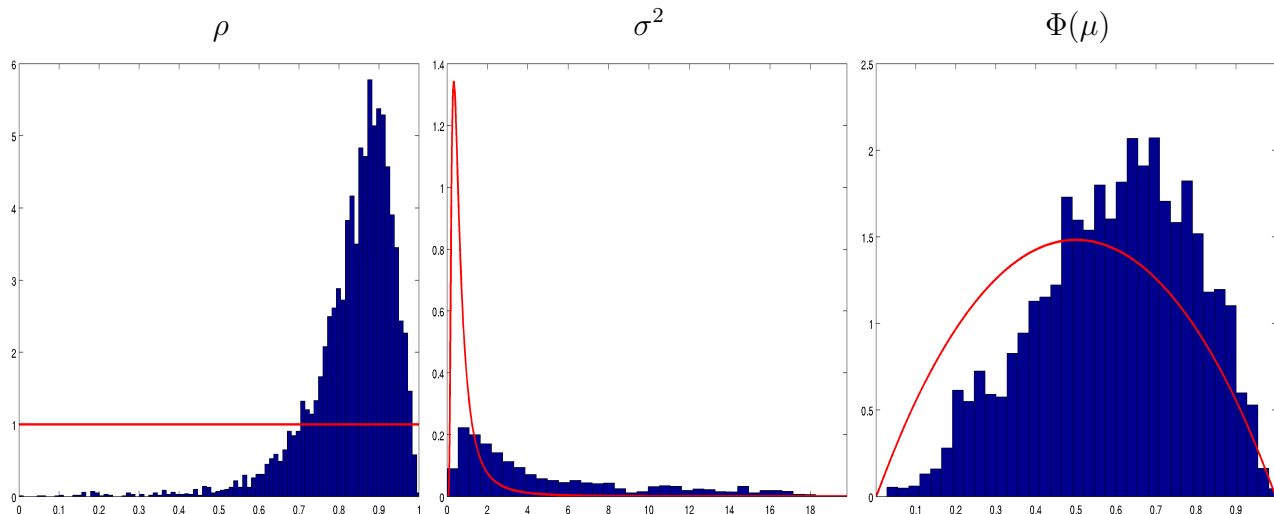
Notes: The left, center, and right panels show the posteriors (histogram) $p^{(h)}(\rho|\mathcal{I}_T^{\mathcal{P}},\mathcal{P})$, $p^{(h)}(\sigma^2|\mathcal{I}_T^{\mathcal{P}},\mathcal{P})$, and $p^{(h)}(\Phi(\mu)|\mathcal{I}_T^{\mathcal{P}},\mathcal{P})$, respectively, together with the prior (red line). The underlying priors are: Prior 1 is $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$ (left panel); Prior 2 is $\rho \sim \mathcal{B}(0.8,0.1)$, $\mu \sim \mathcal{N}(0,\Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$ (center and right panel).

of Figure 7. These results are based on the following prior distribution:[15]

$$\text{Hyperparameter Prior 2: } \rho \sim \mathcal{B}(0.8,0.1), \quad \mu \sim \mathcal{N}(0,\Phi^{-1}(0.75)), \quad \sigma^2 \sim \mathcal{IG}(2,1). \quad (34)$$

Relative to the inverse Gamma prior $\mathcal{IG}(2,1)$ distribution for σ^2 the posterior mass is shifted toward the right and σ^2 is greater than one with very high probability. Recall that for values of $\sigma^2 > 1$ the prior distribution of λ_t is U-shaped, which tends to shift the posterior mean of λ_t toward one of the endpoints ($\lambda_t = 0$ or $\lambda_t = 1$). Thus, the estimation results indicate that the data favor a parameterization in which the posterior mean is more sensitive to the arrival of new information.

The hyperparameter μ determines the location of the seesaw fulcrum, which corresponds to equal weights on both models if $\mu = 0$. The right panel of Figure 7 shows the prior and posterior distributions of $\Phi(\mu)$ – that is, μ mapped into the space of the model weight λ.

---

[15]The arguments for the Beta distribution $\mathcal{B}(\cdot,\cdot)$ refer to the mean and standard deviation. The arguments for the inverse Gamma distribution $\mathcal{IG}(a,b)$ refer to the "natural" shape (a) and scale (b) parameter. The prior for μ is chosen so that its mean, when translated into the $[0,1]$ interval, is 0.5 while the one standard deviation range is approximately the interval $[0.25, 0.75]$.

Figure 8: $\hat{\lambda}^{DP}_{t|t}$: Fixed vs. Estimated μ and σ
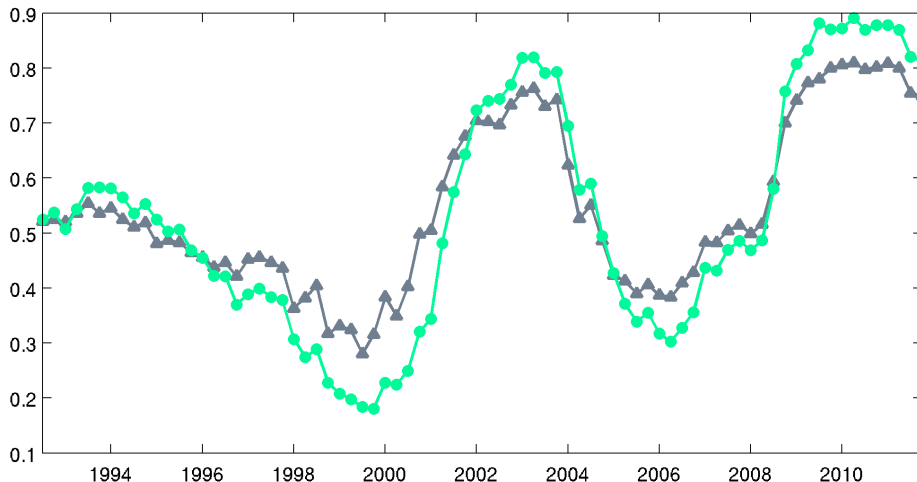
Notes: The figure shows the weight $\hat{\lambda}^{DP}_{t|t}$ for $t =$1992:Q1-2011:Q2 computed using the Prior 1 $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$ (black), and Prior 2 $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$, (light green).

The implied prior density of $\Phi(\mu)$ is centered at equal weights on both DSGE models and assigns a probability of approximately 68% to the interval [0.25, 0.75]. Because the relative forecasting performance of the DSGE models alternates over our sample period the data are not very informative as to whether one model is better than the other on average. The end-of-sample $t = T$ posterior of $\Phi(\mu)$ is slightly shifted toward one relative to the prior, but nonetheless assigns substantial probability to weights less than 0.5.

Figure 8 shows the effect of estimating the hyperparameters μ and σ on the evolution of the posterior mean $\hat{\lambda}^{DP}_{t|t}$. The figure compares the weight on the SWFF under Prior 1 in (33) and Prior 2 in (34). The swings of $\hat{\lambda}^{DP}_{t|t}$ are more pronounced under Prior 2 than under Prior 1. As discussed above, under Prior 2 the posterior distribution of σ assigns most of its mass to values greater than one, which amplifies the movements in the evolution of the model weights.

6.3   The Dynamic Pool's Forecasting Performance

After examining the evolution of model weights under dynamic pooling, static pooling, and BMA, we now turn to the real time forecast performance. Does the dynamic pool of DSGE models generate more accurate density forecasts than the static pool or the posterior weighted
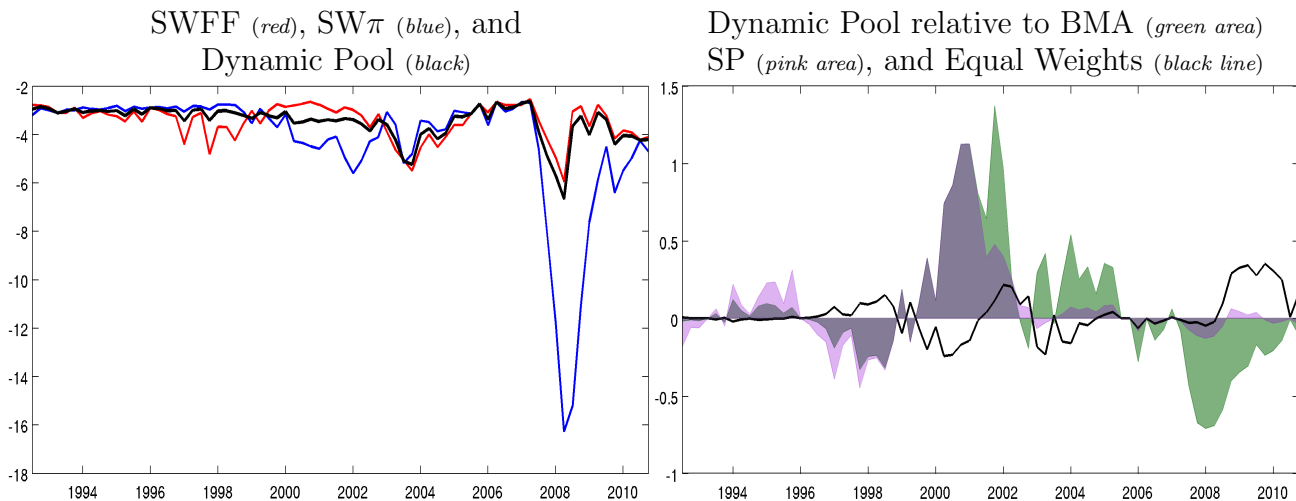
model average? The answer is a qualified yes in our application: the dynamic pool fares significantly better than the static pool (with maximum likelihood weights) and BMA. The dynamic pool is also more accurate than a model average with time-invariant equal weights, but the accuracy difference is smaller. Throughout, we measure forecast accuracy in terms of log predictive score differentials.

The left panel of Figure 9 compares the log score of the dynamic pool (black line) over time to that of its two components: the SWFF model (red) and SWπ model (blue). Recall that the forecasts are generated based on an information set that includes current interest rates and spreads (denoted by $\mathcal{I}_{t+}^{\mathcal{P}}$). The predictive density associated with the dynamic pool, $p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t+}^{\mathcal{P}}, \mathcal{P})$ is a linear combination of the predictive densities of the two DSGE models, $p^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t+}^m, \mathcal{M}_m)$, and, for every $\bar{y}_{t+h,h}$, has to lie between the two model-specific density values. Because the logarithm is a monotone transformation, the log predictive score of the pool has to be bounded from above and below by the log predictive scores of the individual models. Of course, *ex ante* it is not known to the policymaker which of the two DSGE models will attain the higher log score. Thus, it is important to note that in most time periods the dynamic pool's log score is in close proximity to that of the best performing DSGE model. In particular, during the Great Recession, when SWπ performs poorly in terms of forecast accuracy, $\ln p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t+}^{\mathcal{P}}, \mathcal{P})$ closely tracks the log predictive score of the DSGE model with financial frictions.

The right panel of Figure 9 shows the log predictive score differences between the dynamic pool and the following alternatives: BMA (green area), maximum likelihood static pools (purple area), and equal weights (black line). Positive differentials favor the dynamic pool. In the early part of the sample there are no major differences in forecasting performance. However, around the time of the dot-com bust in early 2000, large performance differentials arise. Following an eight-year period period in which SWπ has been the dominant model, both BMA and MSP are caught off guard by the change in the relative forecast performance of the two DSGE models. The DP also struggles to adjust, as evidenced by the fact that it is forecasting worse than the equal-weights combination scheme, but not as much and most importantly not for as long. The DP reacts quickly by increasing the weight on the financial friction model and performs better than equal weights between 2001 and 2003.

From 2005 to 2008 the DP weight on SWFF is close to 50%, which is why the differential with respect to equal-weight combination is essentially zero. Starting in 2009 the

Figure 9: Log Scores Comparison Over Time

SWFF (*red*), SWπ (*blue*), and   Dynamic Pool relative to BMA (*green area*)
Dynamic Pool (*black*)   SP (*pink area*), and Equal Weights (*black line*)

Notes: The left panel shows the log scores $\ln p(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^m, \mathcal{M}_m)$ for SWFF (red), and SWπ (blue), and the log score for the dynamic pools (black) $\ln p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^{\mathcal{P}}, \mathcal{P})$ over the period 1992:Q1-2011:Q2. The right panel shows log score differences between dynamic pool $\ln p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^{\mathcal{P}}, \mathcal{P})$ and the following alternatives: BMA ($\ln p_{BMA}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^{\mathcal{P}}, \mathcal{P})$, green area), maximum likelihood static pool ($\ln p_{MSP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^{\mathcal{P}}, \mathcal{P})$, purple area), and equal weights (black line). We use Prior 2 for the DP: $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2, 1)$.

DP recognizes the strong performance of the SWFF model and is able to out-perform the equally-weighted pool. At the same time BMA gains an advantage from the fact that its weight on the SWFF model has been close to one since 2004, which of course came at the cost of relatively poor performance between 2004 and 2006. The log score differential between DP and MSP during and after the Great Recession is relatively modest. Our results offer the following lesson: dynamic pools can perform well in an environment with persistent shifts in the relative forecasting performance of the underlying models. Static approaches are generally slow to adjust and the use of equal combination weights is suboptimal if one of the models in the pool is dominant for an extended period of time.

Table 1 shows the cumulative log scores for three DP specifications, as well as the difference between the cumulative log scores of the dynamic pools and that of equal weights, BMA, and the static pool (MSP), respectively. These results can be summarized as follows. First, there is not a major difference among the three dynamic pool specifications. Estimating σ^2 leads to a roughly one log point improvement in forecasting performance (recall that the weights react faster to a change in the environment because the estimated σ is greater than one). On the other hand, the parameter μ – that is, allowing for the fact that one of

Table 1: Cumulative Log Scores / Differentials

DP Prior	Log Score	Differentials		
	DP	EW	BMA	MSP
	(1)	(2)	(3)	(4)
Prior 1: $\rho \sim U(0,1)$, $\mu = 0$, $\sigma^2 = 1$	-256.91	1.34	4.07	4.95
Prior 2: $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$	-256.43	1.82	4.55	5.43
Prior 3: $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu = 0$, $\sigma^2 \sim \mathcal{IG}(2,1)$	-255.97	2.28	5.01	5.89

Notes: The table shows in column (1) the cumulative log score $\sum_{t=1}^{T} \ln p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t+}^{\mathcal{P}}, \mathcal{P})$ for various speci-fications of the dynamic pool. Columns (2) through (4) show for each specification the difference between the DP cumulative log scores and that of equal weights, BMA, and MSP, respectively. The cumulative log scores are computed over the period 1993:Q1-2011:Q2.

the model can be better on average than the other – does not seem to improve the forecast performance. The dynamic pools outperform both BMA and the static pool by a substantial margin: the differences are larger than four log points. The gain relative to the combination based on equal weights, however, is modest. The log score differentials range from 1.3 to 2.3 log points.

We also compared the forecasting performance of the dynamic pool to that of the DMA approach described in Section 4.4.[16] We consider three different choices for the tuning parameter α: 0.9, 0.95, and 0.99. The closer α is equal to one, the more persistent the beliefs about the hidden state. In our application the DMA weights for $\alpha = 0.99$ are essentially identical to the BMA weights. Prior to 2004 the DMA weights are very similar for our three choices of α. Starting in 2005 some differences emerge. The smaller α, the more strongly the weight on the SWFF model falls between 2005 and 2006. However, by 2009, regardless of α the weight on the financial friction model has increased to one again. The DP generally attains higher cumulative log scores than the DMA pool. Depending on the prior distribution used for the DP, the score differentials relative to DMA ($\alpha = 0.99$) range from 2.7 to 3.8 and are slightly smaller than the differentials relative to BMA. As α is reduced from 0.99 to 0.90, the differentials drop to the range of -0.1 to 0.8. Recall that $\alpha = 0$ delivers equal weighting which generates log score differentials ranging from 1.3 to

---

[16]Detailed results are presented in Figure A-5 and Table A-2 of the online appendix.

2.3.

We started from the premise that the policymaker believes that neither $\mathcal{M}_1$ nor $\mathcal{M}_2$ provide an accurate description of how macroeconomic time series evolve. In turn, the policymaker created a dynamic pool of DSGE models. If the combination weights are chosen optimally and we abstract from uncertainty about the optimal weights, we should observe that the pool forecasts on average as least as well as the best DSGE model. Thus, in addition to computing the sum of log scores over the entire out-of-sample forecast period reported in Table 1, we also computed recursive sums of log predictive scores (summing from period 1 to t for $t = 1, \ldots, T$).[17] Until 1996 the DP and the two DSGE models essentially attain the same cumulative out-of-sample log predictive score. From 1995 to 2001 the DP is better than the weaker of the two DSGE models. At this point, the weight estimate is still to noisy to reap the full benefits from the pooling. If the predictive scores are summed until any period between 2001 and 2008, then the DP in fact beats both DSGE models, consistent with the notion an incomplete model space.

During the Great Recession the SWπ model performs very poorly which leads to very large log score differentials over a few quarters. Because the DP does not assign 100% of the weight to the SWFF model immediately, the end-of-sample sum of log predictive scores for the DP is slightly worse than for the SWFF model (-254.77) yet still much better than for the SWπ model (-314.35). This highlights that tracking the optimal dynamic combination weights can be challenging but it is preferable to benchmarks such as the use of equal weights.

## 6.4   Policy Experiments Under Model Uncertainty

Forecasting is not the only reason to study model combination. Models in general, and DSGE models in particular, can be used for counterfactual policy analysis. To the extent that projected outcomes differ across models, the question arises on how to best combine these projections. Our dynamic pool provides a natural framework for the combination of models in a counterfactual analysis. In this section we provide an illustrative example of how the dynamic pool could have been used to study the effect of switching to a different monetary policy rule during the financial crisis.

---

[17]Detailed results are presented in Figure A-4 of the online appendix.

Figure 10: Predictive Distribution of Four-Quarter-Ahead Average Output Growth and Inflation under Baseline and Counterfactual Rule using Real-Time 2008:Q3 Information
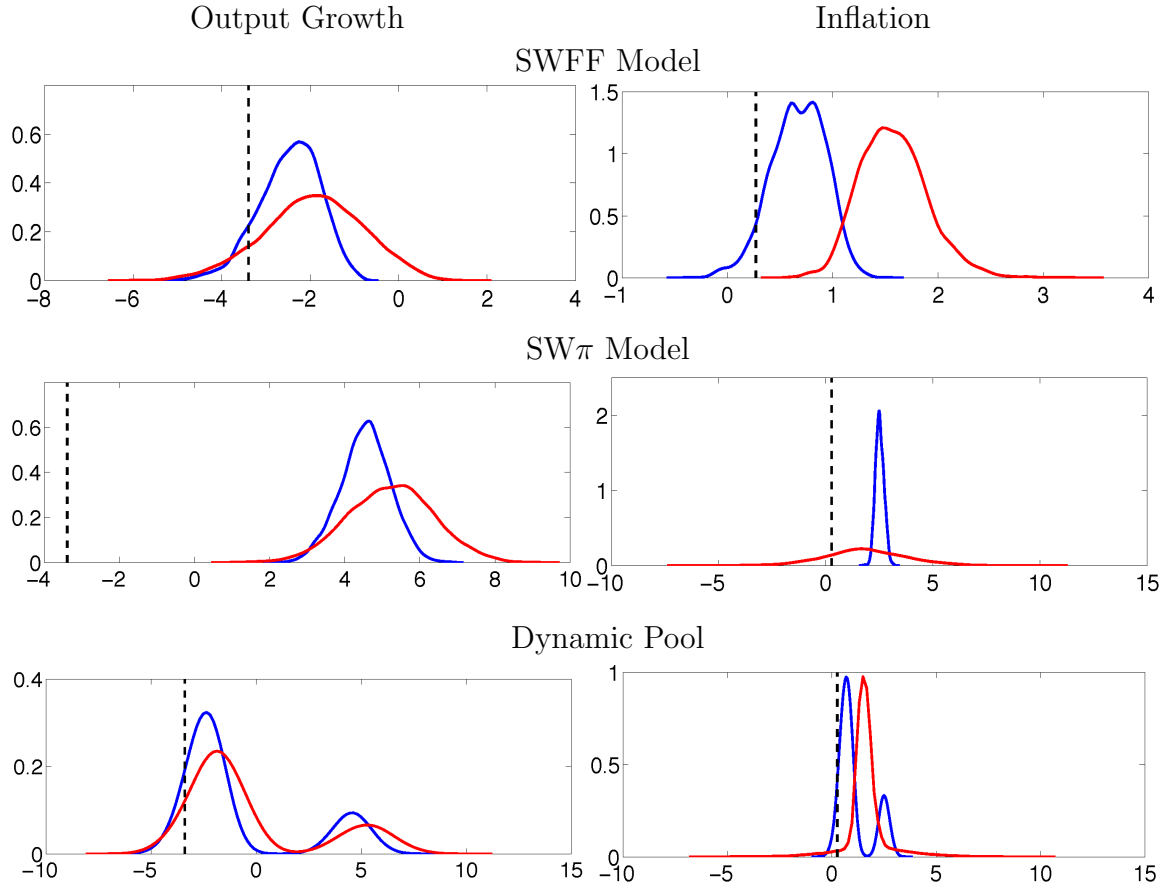
Notes: Predictive densities under the historical (blue) and counterfactual (red) rule, respectively, for four-quarter output growth (left panel) and inflation (right panel). Predictive densities are reported for the SWFF model, the SWπ model, and the dynamic pool (DP, Prior 1). Forecast origin is 2008:Q3$^+$. Actual outcomes are depicted with vertical dashed lines.

In both the SWFF model and the SWπ model monetary policy is represented by an interest rate feedback rule of the form

$$
\begin{aligned}
R_t &= \rho_R R_{t-1} + (1 - \rho_R)\left(\psi_1(\pi_t - \pi_t^*) + \psi_2(y_t - y_t^f)\right) \\
&\quad + \psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m.
\end{aligned}
\tag{35}
$$

Here R_t is the federal funds rate, π_t is inflation, π_t^* is an exogenously-varying target inflation rate, y_t is output, y_t^f is potential output (defined as the level of output that would prevail in the absence of nominal rigidities and inefficient mark-up shocks), and r_t^m is an exogenous monetary policy shock that follows an AR(1) process. Based on their respective information

sets $\mathcal{I}_t^m$, the modelers have generated a posterior distribution for all DSGE model parameters including the policy rule coefficients.

Our counterfactual experiment studies the effect on output and inflation of replacing the estimated policy rule (35) by the following rule:

$$R_t = \rho_R R_{t-1} + (1 - \rho_R)\left(\psi_1(\pi_t - \pi_t^*) + 0.2L_t\right) + r_t^m. \tag{36}$$

Here L_t is hours worked in deviation from steady state. Under this alternative rule, the central bank responds directly to labor market conditions instead of indirectly through aggregate output. In the counterfactual analysis, we use the posterior estimates for the coefficients ρ_R and ψ_1 and fix the response to hours worked at 0.2. The rationale for considering (36) is that policymakers in 2012 did make their instrument explicitly contingent on the state of the labor market.[18] We therefore ask what would have happened to output and inflation had the policymaker immediately after the Lehman crisis considered targeting labor market conditions. Specifically, we use January 1, 2009 (i.e, the information set includes NIPA data until 2008:Q3 as well as financial data from 2008:Q4) as the forecast origin for this counterfactual.

Actual and counterfactual predictions for output growth and inflation are depicted in Figure 10.[19] Each panel shows two predictive densities: one of them is obtained under the estimated policy rule (blue) and one obtained under the counterfactual policy rule (red). A comparison of the first and second row of the figure indicates that the SWFF model and the SWπ model generate vastly different output growth and inflation forecasts based on the 2008:Q3 macroeconomic data. The actual values of output and inflation during the forecast period are depicted with vertical dashed lines. Under the SWFF model the actual outcomes, depicted with vertical dashed lines, are not too far from the mode of the predictive distribution, as shown by the blue distributions in the top row of Figure 10. Under

---

[18]According to the December 2012 FOMC statement: "If the outlook for the labor market does not improve substantially, the Committee will continue its purchases of Treasury and agency mortgage-backed securities, and employ its other policy tools as appropriate, until such improvement is achieved in a context of price stability." The policy instruments considered in this quote was large scale asset purchases, whose effects we do not incorporate in the models considered here. For this reason we consider the short term rate as the policy instrument for this experiment.

[19]We do not show the projections for the federal funds rate, but we note that for both models the interest rate projection under both policies does not violate the zero lower bound constraint on nominal interest rates.

the baseline monetary policy, the financial friction DSGE model only slightly overpredicts four-quarter average output growth and inflation after the Lehman crisis. The SWπ model, on the other hand, generates a forecast that severely overpredicts the actuals. For instance, while the posterior predictive distribution of output growth peaks at about 4.5% actual output growth was less than -3% over the forecast period.

The predicted effect of the counterfactual policy on output growth is qualitatively similar across the two models: output growth increases and the predictive distribution becomes a bit more diffuse. The shift of the predictive distribution is more pronounced under the financial frictions model than under SWπ. The effect of the change in policy rule on inflation, however, is quite different. Under the SWFF model inflation rises by about 100 basis points, but the dispersion of the predictive density remains roughly constant. The SWπ model, on the other hand, predicts that inflation might fall relative to the baseline policy rule, but the outcome is much more uncertain, i.e., the predictive density is much more spread out. In the SWπ model, both price and wage markup shocks are important determinants of inflation in the (see Smets and Wouters (2007) and King and Watson (2012)). Policies that respond very strongly to the level of the gap in economic activity fare very poorly in terms of controlling inflation when faced with markup shocks (see Chung et al. (2014)), as these shocks move activity and inflation in opposite directions. The counterfactual policy (36) is one such policy, which explains the diffuse predictive distribution of inflation. Conversely, in the SWFF model markup shocks do not play as large a role as in the SWπ model, as discussed in Del Negro et al. (2015), and therefore the predictive distribution of inflation is not nearly as dispersed.

How can a policymaker aggregate such disparate predictions arising from multiple models? The bottom row of Figure 10 plots predictive distributions for our proposed dynamic prediction pool. By the end of 2008, the weights on the two DSGE models in the dynamic pool are approximately 50% each. At the end of 2008, the dynamic pool generates a bimodal predictive distribution for output growth and inflation, which is a reflection of the rather different predictive densities obtained from the component models. The realized value of output growth and inflation falls close to the mode associated with the contribution of the SWFF predictive density. Because the predictive density associated with the DP is slightly lower than the predictive density of the SWFF model, the overall log-predictive score of the DP is slightly smaller than that of the SWFF model, which is what was depicted in Figure 9.

Because both DSGE model predict a stimulative effect on output, under the counterfactual policy the predictive distribution for output growth shifts to the right and the right tail of the distribution lengthens. More interestingly, the predictive distribution for inflation turns from being bimodal under the estimated policy rule to being unimodal (centered around 1.5%) and fat-tailed under the counterfactual policy. Given a loss function that is specified in terms of output growth and inflation, the policymaker could use the predictive distributions from the dynamic pool to choose between the two policy rules.

7   Conclusion

This paper provides a methodology for estimating time-varying weights for linear prediction pools. In our application we combine predictive densities from two DSGE models, with and without financial frictions. We introduce an informational friction to justify not estimating the combination weights and DSGE model weights jointly. Given the computational difficulties with generating draws from posterior distributions of large-scale DSGE models as well as the institutional arrangements within central banks, we think that this framework is attractive. The same method could be used to combine other classes of time series models and it could be extended to the combination of more than two models.

In our empirical analysis we find that the model weights in the dynamic pool vary substantially over time. In times without financial distress the SWπ model forecasts output growth and inflation more accurately than the DSGE model with financial frictions. This ranking changes, however, in more turbulent times such as the dot-com bust in early 2000 and the Great Recession in 2007-09. The model weights of the dynamic pool adjust accordingly and, in each time period, the resulting model mixture forecasts (almost) as well as the best (among the two) DSGE model. These results suggest that our dynamic pool may be viewed as a gross approximation of a more elaborate nonlinear DSGE model that inherits the dynamics of the SWπ model during financial tranquility and the dynamics of the financial frictions model during turbulent financial times. Our results provide further motivation for developing an estimable version of such a non-linear model.

References

ALESSANDRI, P. AND H. MUMTAZ (2014): "Financial Conditions and Density Forecasts for US Output and Inflation," *Manuscript, Queen Mary University*.

ANDRIEU, C., A. DOUCET, AND R. HOLENSTEIN (2010): "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B*, 72, 269–342.

BATES, J. AND C. W. GRANGER (1969): "The Combination of Forecasts," *Operations Research*, 20, 451–468.

BERNANKE, B., M. GERTLER, AND S. GILCHRIST (1999): "The Financial Accelerator in a Quantitative Business Cycle Framework," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, North Holland, Amsterdam, vol. 1C.

BILLIO, M., R. CASARIN, F. RAVAZZOLO, AND H. K. VAN DIJK (2013): "Time-varying Combinations of Predictive Densities using Nonlinear Filtering," *Journal of Econometrics*, 177, 213–232.

BOCOLA, L. (2016): "The Pass-Through of Sovereign Risk," *Journal of Political Economy*, Forthcoming.

BRUNNERMEIER, M. K. AND Y. SANNIKOV (2014): "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104, 379–421.

CHANG, Y., Y. CHOI, AND J. Y. PARK (2014): "Regime Switching Model with Endogenous Autoregressive Latent Factor," *Manuscript, Indiana University*.

CHOPIN, N. (2004): "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Annals of Statistics*, 32, 2385–2411.

CHRISTIANO, L., R. MOTTO, AND M. ROSTAGNO (2003): "The Great Depression and the Friedman-Schwartz Hypothesis," *Journal of Money, Credit and Banking*, 35, 1119–1197.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

CHRISTIANO, L. J., R. MOTTO, AND M. ROSTAGNO (2014): "Risk Shocks," *American Economic Review*, 104, 27–65.

CHUNG, H., E. P. HERBST, AND M. KILEY (2014): "Effective Monetary Policy Strategies in New-Keynesian Models: A Re-Examination," in *NBER Macroeconomics Annual 2014*, ed. by J. Parker and M. Woodford, University of Chicago Press, vol. 29.

CREAL, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31, 245–296.

DAWID, A. (1984): "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Series A*, 147, 278–292.

DE GRAEVE, F. (2008): "The External Finance Premium and the Macroeconomy: US Post-WWII Evidence," *Journal of Economic Dynamics and Control*, 32, 3415 – 3440.

DEL NEGRO, M., M. P. GIANNONI, AND F. SCHORFHEIDE (2015): "Inflation in the Great Recession and New Keynesian Models," *American Economic Journal: Macroeconomics*, 7, 168–196.

DEL NEGRO, M. AND F. SCHORFHEIDE (2013): "DSGE Model-Based Forecasting," in *Handbook of Economic Forecasting, Volume 2*, ed. by G. Elliott and A. Timmermann, Elsevier.

DEWACHTER, H. AND R. WOUTERS (2012): "Endogenous Risk in a DSGE Model with Capital-Constrained Financial intermediaries," *National Bank of Belgium Working Paper*, 235.

EDGE, R. AND R. GÜRKAYNAK (2010): "How Useful Are Estimated DSGE Model Forecasts for Central Bankers," *Brookings Papers of Economic Activity*, 41, 209–259.

FAWCETT, N., G. KAPETANIOS, J. MITCHELL, AND S. PRICE (2014): "Generalised Density Forecast Combinations," *Bank of England Working Paper*, 492.

GEWEKE, J. AND G. AMISANO (2011): "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

GIORDANI, P., M. K. PITT, AND R. KOHN (2011): "Bayesian Inference for Time Series State Space Models," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk, Oxford University Press.

GNEITING, T. AND R. RANJAN (2013): "Combining Predictive Distributions," *Electronic Journal of Statistics*, 7, 1747–1782.

HALL, S. G. AND J. MITCHELL (2007): "Combining Density Forecasts," *International Journal of Forecasting*, 23, 1–13.

HERBST, E. AND F. SCHORFHEIDE (2016): *Bayesian Estimation of DSGE Models*, Princeton University Press.

HOETING, J. A., D. MADIGAN, A. E. RAFTERY, AND C. T. VOLINSKY (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–417.

KING, R. G. AND M. W. WATSON (2012): "Inflation and Unit Labor Cost," *Journal of Money, Credit and Banking*, 44, 111–149.

KIYOTAKI, N. AND J. MOORE (1997): "Credit Cycles," *Journal of Political Economy*, 105, 211–248.

KOLASA, M. AND M. RUBASZEK (2015): "Forecasting with DSGE Models with Financial Frictions," *International Journal of Forecasting*, 31, 1–19.

KOOP, G. AND D. KOROBILIS (2012): "Forecasting Inflation Using Dynamic Model Averaging," *International Economic Review*, 53, 867–886.

LEAMER, E. E. (1978): *Specification Searches*, Wiley, New York.

MIN, C.-K. AND A. ZELLNER (1993): "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics*, 56, 89–118.

RAFTERY, A. E., M. KARNY, AND P. ETTLER (2010): "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52, 52–66.

SCHORFHEIDE, F. (2005): "VAR Forecasting Under Misspecification," *Journal of Econometrics*, 128, 99–136.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Smets, F. and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123 – 1175.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586 – 606.

Stock, J. H. and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829.

Waggoner, D. and T. Zha (2012): "Confronting model misspecification in macroeconomics," *Journal of Econometrics*, 171, 167184.

Wright, J. (2008): "Bayesian Model Averaging and Exchange Rate Forecasting," *Journal of Econometrics*, 146, 329–341.