Del Negro, Marco; Hasegawa, Raiden B.; Schorfheide, Frank

Working Paper

Dynamic prediction pools: An investigation of financial frictions and forecasting performance

Staff Report, No. 695

Provided in Cooperation with:
Federal Reserve Bank of New York

Suggested Citation: Del Negro, Marco; Hasegawa, Raiden B.; Schorfheide, Frank (2014) : Dynamic prediction pools: An investigation of financial frictions and forecasting performance, Staff Report, No. 695, Federal Reserve Bank of New York, New York, NY

This Version is available at:
https://hdl.handle.net/10419/120816

Federal Reserve Bank of New York
Staff Reports

Dynamic Prediction Pools:

An Investigation of Financial Frictions and Forecasting Performance

Marco Del Negro
Raiden B. Hasegawa
Frank Schorfheide

Staff Report No. 695
October 2014

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

Marco Del Negro, Raiden B. Hasegawa, and Frank Schorfheide

Federal Reserve Bank of New York Staff Reports, no. 695
October 2014
JEL classification: C53, E31, E32, E37

Abstract

We provide 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 dynamic stochastic general equilibrium (DSGE) models, with and without financial frictions, for output growth and inflation in the period 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 doesn't 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 Bayesian model averaging, optimal (static) pools, and equal weights. 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.

Key words: Bayesian estimation, DSGE models, financial frictions, forecasting, Great Recession, linear prediction pools

---

Del Negro: Federal Reserve Bank of New York (e-mail: marco.delnegro@ny.frb.org). Hasegawa: Wharton School of the University of Pennsylvania (e-mail: raiden@wharton.upenn.edu). Schorfheide: University of Pennsylvania (e-mail: schorf@ssc.upenn.edu). The authors are grateful for helpful comments and suggestions from Gianni Amisano, Francis X. Diebold, Tom Engsted, Bartosz Maćkowiak, Francesco Ravazzolo, Shaun Vahey, and seminar participants at Princeton University, the University of Pennsylvania, the Bank of England, the EFAB@Bayes250 Conference at Duke University, the University of Venice, New York University, the 2014 Euro Area Business Cycle Network (EABCN) Conference in London, the 2014 Society for Nonlinear Dynamics and Econometrics (SNDE) Conference, Université de Montréal, the 2014 Macro-Finance Conference at Aarhus, the 2014 European Central Bank Workshop on Forecasting, the 2014 Computing in Economics and Finance (CEF) Conference in Oslo, and the 2014 National Bureau of Economic Research Summer Institute. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 1061725. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

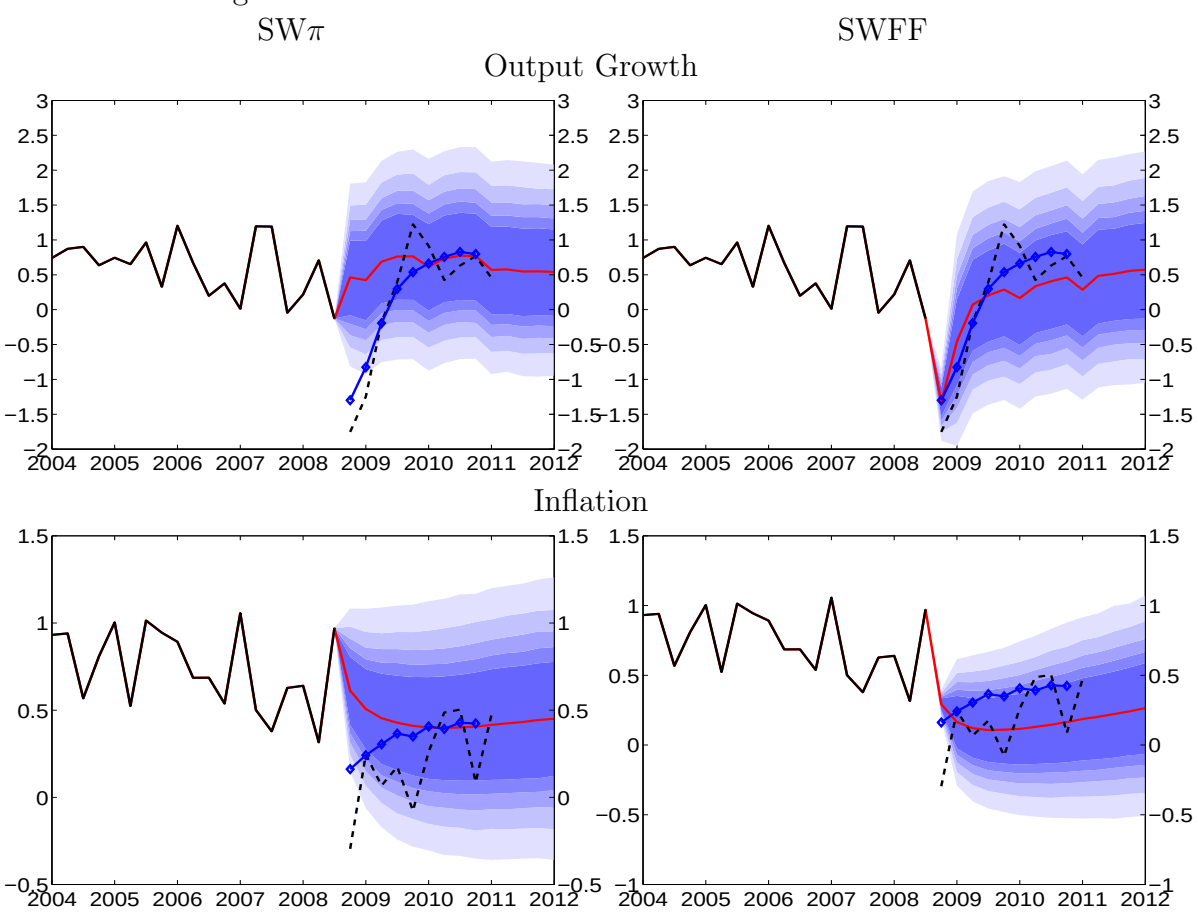
1 Introduction

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. The SW model augments the neoclassical stochastic growth model by price and wage rigidities as well as various adjustment mechanisms, but it has no credit market frictions. The only financial time series that is used in the estimation of the SW model is the short-term nominal interest rate, which serves as the monetary policy instrument in the model. 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) and others, Del Negro and Schorfheide (2013) and Del Negro et al. (Forthcoming) show that an enlarged SW model with financial frictions (henceforth SWFF) as in Bernanke et al. (1999) 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 obtained in the aftermath of the Lehman crisis using 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?

We document 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, where 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

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.

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, and may provide a partial justification for the reliance on DSGE models without financial frictions prior to the Great Recession. The time-variation in the relative forecasting performance of these 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 develop a new method 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 the 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. A particle filter is used to track the evolution of λ_t conditional on the observed data. The law of motion of λ_t is indexed by some hyperparameters, which can be integrated out under their posterior distribution at each forecast origin. Our empirical analysis focuses on the sequence of model weights as well as the forecast performance of the dynamic prediction pool.

Our paper builds on the growing literature on density forecast combination, and specifically on the work by Hall and Mitchell (2007) and Geweke and Amisano (2011).[1] These authors 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. The pools are optimal in the sense that the weights are chosen to maximize the pool's historic forecast performance. If one thinks of models as stocks, and of predictive densities as returns, the optimal-pool approach can be seen as choosing the weights so to optimize the portfolio's historical 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. While Waggoner and Zha (2012) also emphasize the importance of time-variation in the combination weights, 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 (depending on the hyperparameter settings) 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

[1] Our paper is also loosely related to the large body of work on the combination of point forecasts dating back to a seminal paper by Bates and Granger (1969), and more specifically to the strand of that literature using time-variation in the weights, e.g., Terui and van Dijk (2002) and Guidolin and Timmermann (2009).

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 it is impractical in the DSGE model applications that we have in mind. In a typical setting at a policy institution, we think that it is unrealistic (the candidate models may be maintained by different divisions or modeling groups within the institution) 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.

We obtain a non-linear state space model and conduct inference about the weights using a particle filter. As such, our approach bears many similarities to that proposed by Billio et al. (2013), who discuss time-varying model convolutions. Where we differ from Billio et al. (2013) is in the specification of the law of motion of the weights. Our setting is designed to encompass Hall and Mitchell (2007)'s and Geweke and Amisano (2011)'s static pools: as the hyperparameter characterizing the persistence in the law of motion of λ_t our procedure specializes to theirs (except that we determine the combination weight as the mean instead of the mode of the posterior distribution). Thus, we let the data determine the degree of time-variation in the weights, via the posterior distribution on the persistence parameter, as opposed to imposing it a priori as in Billio et al. (2013). In addition, our approach allows for a prior distribution of the weights that is uniform over $[0, 1]$, thereby making inference about the pool's weights arguably more informative about the models' relative forecasting performance.

Throughout this paper we focus on linear prediction pools, albeit with time-varying weights. There also exists a literature on nonlinear pools. For instance, Gneiting and Ranjan (2013) take beta transformations of cumulative densities from univariate prediction pools to ensure that the predictive density of the pool is well calibrated. Fawcett et al. (2013)

[2]This is certainly the case for medium-scale DSGE models, the application considered here, which in some cases are not even re-estimated every quarter. For models that are not estimated using full information methods, e.g. the FRB/US model used at the Board of Governors, the requirement that the model is re-estimated in light of the information coming from the pool's weights is simply unfeasible.

let the combination weights depend on the variable that is being forecast and use sieves to approximate the variable-dependent weights. Exploring these generalizations is beyond the scope of our paper.

Our empirical application focuses on 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 and Bayesian model averaging (BMA). We find that after prolonged periods in which one model performs better than the other, all three 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 the dynamic pool's hyperparameters is consistent with these findings. It shows that the data favor the dynamic relative to the static pool, but also that discrepancies in the forecasting performance tend to be persistent. We also allow for the unconditional mean of the distribution to be different from equal weights, which is the case if one model were to be on average better than the other, but the data offer little evidence in favor of this hypothesis. Finally, the data favor a specification in which the weights' distribution may shift rapidly away from the unconditional mean.

In terms of the substantive questions motivating this paper, we find that the posterior distribution of the 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. In particular, we show that by the time the Lehman 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 this model before the crisis, they had no reason for not relying on it at all. 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.

Our findings beg the question of whether they are the consequence of using 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., see Brunnermeier and Sannikov (Forthcoming), Dewachter and Wouters (2012), or the estimated DSGE model of Bocola (2013)). Work by Alessandri and Mumtaz (2014) on forecasting with regime switching models provides complementary evidence on the presence of non-linearities.

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. Throughout the paper we focus on the combination of two DSGE model and leave the extension to more than two model for future research. 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.

2 A Stylized Principle-Agent Framework

Estimation of large-scale DSGE models can be tedious and 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 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 principle-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.[3] 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 off 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

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

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 $\left\{p(\tilde{y}_\tau|\mathcal{I}_{\tau-1}^m, \mathcal{M}_m)\right\}_{\tau=1}^{t+1}$,

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

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.

In this paper, we propose to compute the weights $\hat{\lambda}_t$ using a method that we refer to as dynamic prediction pool. Before doing so, we review the computation of $\hat{\lambda}_t$ using Bayesian model averaging (BMA) and a static prediction pool. For now, we will consider sequences of one-step-ahead forecasts and postpone the discussion of multi-step forecasts until Section4.3.

3 Bayesian Model Averaging and Static Pools

The idea of BMA (Section 3.1) is to assign discrete prior probabilities to each model under consideration and then to update these probabilities based on Bayes Theorem in light of the data. Predictive distributions for future observations are then generated as a weighted average of the predictive distributions of the individual models. 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). Static prediction pools (Section 3.2) can be viewed as an extension of the classic Bates and Granger (1969) approach of combining point forecasts to the combination of density forecasts. The basic idea of the Bates-Granger approach is to construct a linear combination of point forecasts and to use past data to determine the combination weight that minimizes the mean-squared forecast error of the combined forecast assuming that the optimal weight is constant over time. (Linear) static pools are linear combinations of predictive densities with weights estimated by maximizing the predictive performance of the pool on past observations.

To simplify the exposition, 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 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. ADD 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. Thus, the domain of λ is restricted to take one of the following two values:

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

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

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

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

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

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

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

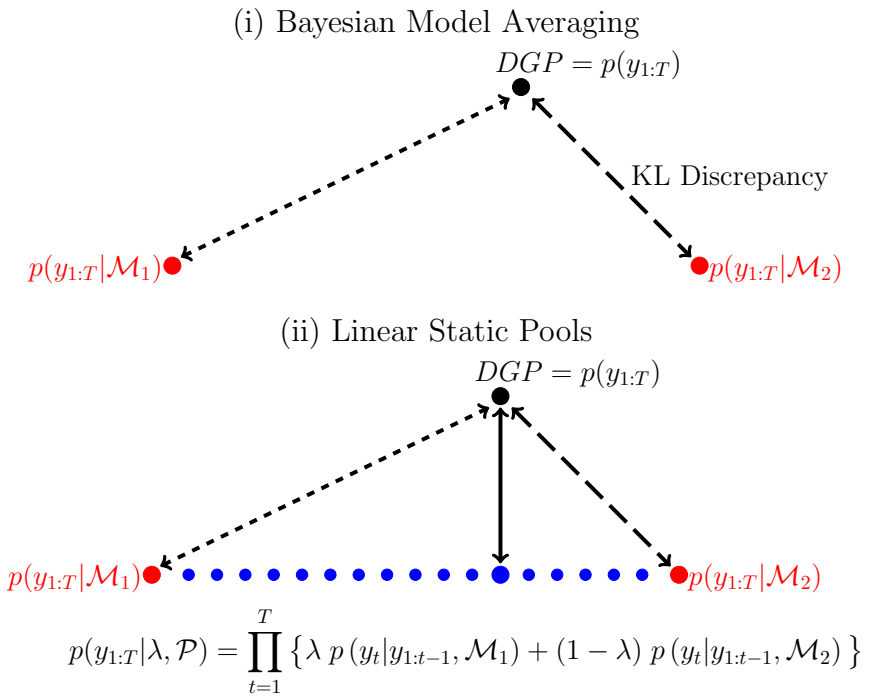
After taking expectations of the right-hand-side of (3) conditional on $y_{1:T}$ to integrate out the unknown λ we obtain

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

which takes the form of (2).

ADDed by FS

Figure 2: Combining Models if the Model Space is Incomplete



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. He 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 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 have 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

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

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.

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

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

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

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

and leads to

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

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).[4]

Instead of weighting the model-based predictive density by the posterior mean of λ the approach proposed in Hall and Mitchell (2007) and Geweke and Amisano (2011) amounts to defining $\hat{\lambda}_t$ as the mode of $p(\lambda|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$ under a uniform prior distribution $p(\lambda|\mathcal{P})$, which is also the argmax of the likelihood function (8). We refer to the resulting pool of densities as maximum (likelihood) static pool (MSP). Accordingly, we use $\hat{\lambda}_t^{MSP}$ to denote the estimate of λ and $p_{MSP}(y_{t+1}|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$ to denote the resulting linear prediction pool. Geweke and Amisano (2011) show that the sequence of pooling weights has the property that it converges to the value of λ that minimizes the KL distance between the DGP $p(Y_{1:T})$ and the family of distributions $p(y_{1:T}|\lambda, \mathcal{P})$. Unless, the DGP is identical to either $\mathcal{M}_1$ or $\mathcal{M}_2$, the sequences $\hat{\lambda}_T^{BSP}$ and $\hat{\lambda}_T^{MSP}$ do not converge to either one or zero as $T \longrightarrow \infty$.

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 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 the z_t^1 as well as a measure of interest rate spreads. The assumption stated in (12)

[4]Of course, in any period t (11) could still be interpreted as observation being generated by $p(y_{t+1}|y_{1:t}, \mathcal{M}_1)$ with probability $\hat{\lambda}_t^{BSP}$ and by $p(y_{t+1}|y_{1:t}, \mathcal{M}_2)$ with probability $1 - \hat{\lambda}_t^{BSP}$ where these probabilities follow a complicated nonlinear law of motion.

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 $p(z_t|y_t, z_{1:t-1}, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$. 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

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

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

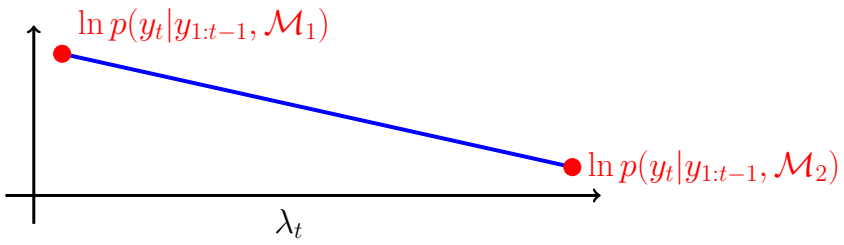
$$\hat{\lambda}_t^{BMA} = \frac{\hat{\lambda}_{t-1}^{BMA} p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1)}{\hat{\lambda}_{t-1}^{BMA} p(y_t|\mathcal{I}_{t-1}^1, \mathcal{M}_1) + (1-\lambda_{t-1}^{BMA})p(y_t|\mathcal{I}_{t-1}^2, M_2)}. \qquad (14)$$

Note, however, that the policymaker never explicitly uses the additional information $z_{1:t}$, because the density $p(z_t|y_t, z_{1:t-1}, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$ 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 $\mathcal{I}_t^{\mathcal{P}}$ 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, which extends the domain of λ to the unit interval. The information 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, $t = 1, 2, \ldots$.

Figure 3: Dynamic Pools: Likelihood Function



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

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

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

Expecting the optimal combination weights to exhibit some degree of persistence, we follow the literature on Bayesian non-parametric function estimation (see, e.g., the survey by Griffin et al. (2011)) and impose a stochastic-process prior on $\lambda_{1:T}$ that implies a "smooth" evolution over time. This prior 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

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 $\lambda_t$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 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 described in detail in Appendix C to approximate the sequence of

densities $p(\lambda_t|\theta, \mathcal{I}_t^{\mathcal{P}}, P)$.[5] 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

Our application focuses on multi-step forecasts of output growth and inflation. Policy makers are generally more interested in forecasts of average output growth or inflation over the next h period rather than the growth rates between period $t+h-1$ and period $t+h$. Accordingly, we define

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

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

[5] Recent surveys of particle-filtering methods for nonlinear state-space models in econometrics are provided by Giordani et al. (2011) and Creal (2012).

$\{p(\tilde{y}_\tau|\mathcal{I}^m_{\tau-h},\mathcal{M}_m)\}^{t+h}_{\tau=1}$, 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 $\{p(\tilde{y}_\tau|\mathcal{I}^m_{\tau-h},\mathcal{M}_m)\}^{t}_{\tau=1}$ at the observed value $\tilde{y}_\tau = y_t$. This leads to a sequence of pseudo-posterior distributions $p^{(h)}(\lambda_t|\theta,\mathcal{I}^{\mathcal{P}}_t,\mathcal{P})$, which are updated according to Bayes Theorem:

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

where

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

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

$$\hat{\lambda}^{DP}_{t+h|t}(\theta) = \int^1_0 \lambda_{t+h}\left[\int^1_0 p(\lambda_{t+h}|\lambda_t)p^{(h)}(\lambda_t|\theta,\mathcal{I}_{t-h},\mathcal{P})d\lambda_t\right]d\lambda_{t+h}. \quad (25)$$

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

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

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 $\hat{\lambda}^{DP}_{t+h|t}(\theta)$ using the pseudo-posterior distribution $p^{(h)}(\theta|\mathcal{I}_t,\mathcal{P})$, which is obtained from

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

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)) that combines a bootstrap particle filter with a random-walk Metropolis-Hastings algorithm and is described in Appendix C. We drop the (θ) argument to denote the marginal posterior mean of λ that is obtained by integrating out θ and write

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

Moreover, removing θ from the conditioning set, we let

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

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

4.4 Further Discussion

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. In fact, Waggoner and Zha (2012) start directly from such an interpretation. To simplify the exposition, we focus on the case $h=1$, $\mu=0$, and $\sigma=1$. If $s_t=1$ then the y_t is given by $p(y_t|\mathcal{I}_{t-1}^1,\mathcal{M}_1)$ and if $s_t=2$, then $p(y_t|\mathcal{I}_{t-1}^2,\mathcal{M}_2)$. The probability of being in state $s_t=1$ is determined by the hidden process x_t, and conditional on x_t is equal to $\lambda_t=\Phi(x_t)$. Under the assumption that $\mu=0$, and $\sigma=1$ the unconditional probability of being in state $s_t=1$ is

$$\mathbb{P}(s_t=1) = \int \Phi(x_t)\varphi(x_t)dx_t = 1/2, \tag{30}$$

where $\varphi(\cdot)$ is the probability density function (pdf) of a $N(0,1)$ random variable. Conditional on x_{t-1} the regime probability is

$$\mathbb{P}(s_t=1|x_{t-1}) = \int \Phi(\rho x_{t-1}+\sqrt{1-\rho^2}+\epsilon_t)\varphi(\epsilon_t)d\epsilon_t. \tag{31}$$

We can also calculate the probability of staying in regime 1. Using Bayes theorem

$$p(x_{t-1}|s_{t-1}=1) = \frac{\Phi(x_{t-1})\varphi(x_{t-1})}{\int \Phi(x_{t-1})\varphi(x_{t-1})} = 2\Phi(x_{t-1})\varphi(x_{t-1}).$$

Therefore,

$$\mathbb{P}(s_t=1|s_{t-1}=1) = \int \left[\int \Phi(\rho x_{t-1}+\sqrt{1-\rho^2}+\epsilon_t)\varphi(\epsilon_t)d\epsilon_t\right] 2\Phi(x_{t-1})\varphi(x_{t-1})dx_{t-1}. \tag{32}$$

However, unlike in the Hamilton (1989)-style Markov regime-switching framework considered by Waggoner and Zha (2012), our setup does not have a first-order Markov structure in terms of s_t, that is, $\mathbb{P}(s_t=1|s_{t-1},s_{t-2},\ldots) \neq \mathbb{P}(s_t=1|s_{t-1}=1)$. In our setting, 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.[6]

[6] 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

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.

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. (Forthcoming). 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 various 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 formal specification of the two models is presented in Appendix A

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

model (unlike our model) is isomorphic to Hamilton's regime switching model if x_t evolves exogenously.

Table 1: Timing of Information Sets for 2009

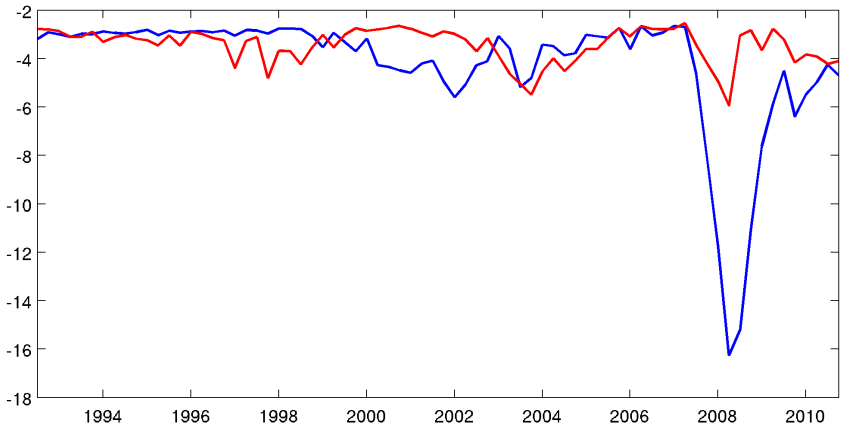
Forecast	End of Est.	Forecast			
Origin	Sample t	$h=1$	$h=2$	$h=3$	$h=4$
Jan 1, 09	08:Q3	08:Q4	09:Q1	09:Q2	09:Q3
Apr 1, 09	08:Q4	09:Q1	09:Q2	09:Q3	09:Q4
Jul 1, 09	09:Q1	09:Q2	09:Q3	09:Q4	10:Q1
Oct 1, 09	09:Q2	09:Q3	09:Q4	10:Q1	10:Q2

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. Precise data definitions are provided in Appendix B.

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 have subsequently been revised by the statistical agencies. Due to the real-time analysis, the definition of the information sets $\mathcal{I}_t^m$ requires some care. We consider four forecasts per year, corresponding to the information set available to the modelers on January 1st, April 1st, July 1st, and October 1st of each year. An example of the timing convention for 2009 is given in Table 1. For instance, on January 1st 2009, the modelers have NIPA data available until 2008:Q3. At this point in time, a preliminary estimate for 2008:Q4 has not yet been published.

When plotting results over time we use the convention that t equals the quarter corresponding to the latest NIPA data. Thus, t would correspond to 2008:Q3 for forecasts made on January 1st 2009. In turn, the horizon $h = 1$ corresponds to a "nowcast" of the fourth quarter of 2008. While NIPA data for our forecast origins are only available with a one-quarter delay, financial data are available in real time. In our empirical analysis we will distinguish between unconditional and semi-conditional forecasts. The unconditional forecasts are based on the information set $\mathcal{I}_t^m$ 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 $\mathcal{I}_{t+}^m$, which includes the $t+1$ federal funds rate and spreads in addition to $\mathcal{I}_t^m$. For instance, on January 1st 2009, the semi-conditional forecasts use financial variables

Figure 4: Log Scores Comparison: SWFF vs SWπ



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

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 Appendix A. 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 $p(\bar{y}_{t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m)$ (or $p(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^m, \mathcal{M}_m)$) from the modelers. The computation of these densities is described in Appendix C.

Figure 4 depicts the log scores $p(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^m, \mathcal{M}_m)$ (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. 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.[7]

6 Results from the Dynamic Prediction Pool

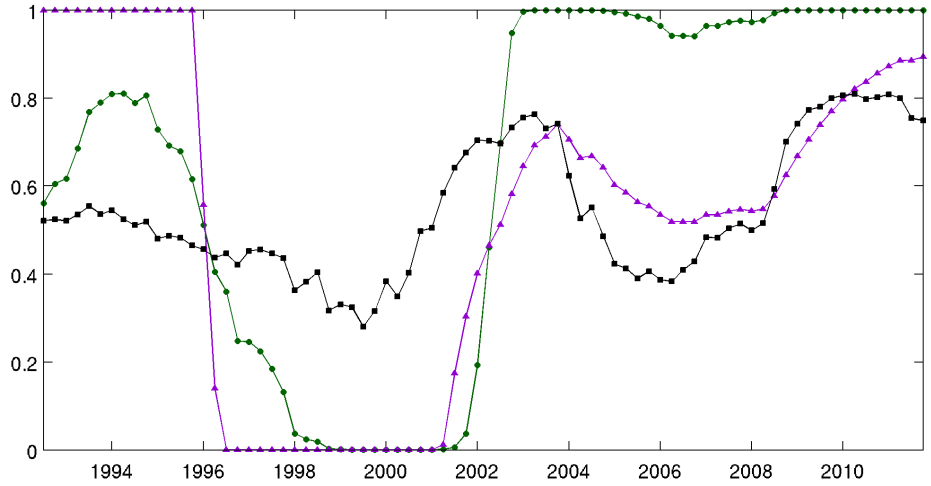
We now apply the dynamic pools methodology discussed in Section 6 to the forecasts of four-quarter-ahead average output growth and inflation obtained from the SWπ model and the SWFF model. We ask the following questions: How do the weights assigned to the two DSGE models by the Bayesian dynamic pool (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)? 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 .

6.1 Pooling Weights

We documented in Section 5 that there is no dominant model in the race between SWFF and SWπ, and that there are medium-frequency swings in the relative forecasting performance of the two models. Arguably, this is a common situation in model-based forecasting. How do the various model combination techniques described in Sections 3 and 4 deal with these features? To what extent does the weight given to a model approach one after a fairly long period in which this model has been dominant? And how quickly does this weight change as the relative performance reverses?

[7]Similarly, Kolasa and Rubaszek (2013) 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 ($\hat{\lambda}^{DP}_{t+h|t}$ – black), (maximum likelihood) static pool ($\hat{\lambda}^{MSP}_{t}$ – purple), and BMA ($\hat{\lambda}^{BMA}_{t}$ – green). Prior 1 is used for the DP: $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$.

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 ($\hat{\lambda}^{DP}_{t+h|t}$ – black) as well as the static pool ($\hat{\lambda}^{MSP}_{t}$ – purple) with weights estimated by maximum likelihood, and BMA ($\hat{\lambda}^{BMA}_{t}$ – 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 flips around 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

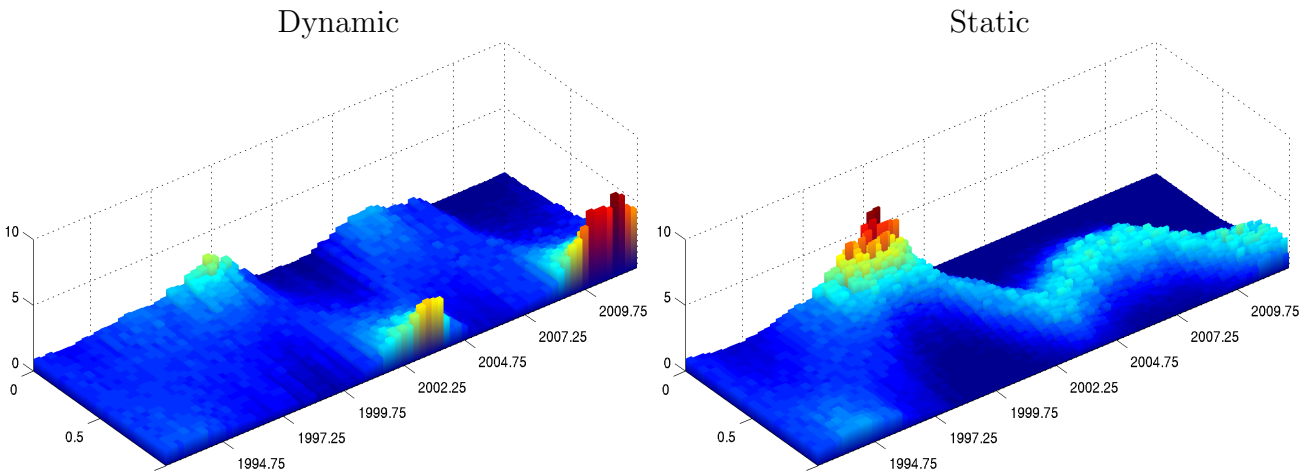
weight on the SWFF model from around 0.7 to 0.4. Due to the poor forecast performance of the SWπ model during the Great Recession, the weight of the SWFF model increases to 0.8 by 2010 and stayed there until the end of our sample period in 2011:Q2.

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 model rather than the mean of the posterior distribution of λ to form the combination weights. After 2002 movements in the static pool weights $\hat{\lambda}_t^{MSP}$ mirror those in $\hat{\lambda}_{t+h|t}^{DP}$ with two differences. First, $\hat{\lambda}_{t+h|t}^{DP}$ moves more rapidly than $\hat{\lambda}_t^{MSP}$. For example, starting in 2008 $\hat{\lambda}_{t+h|t}^{DP}$ drifts upwards toward the SWFF model faster than the SP weight. Second, movements in $\hat{\lambda}_{t+h|t}^{DP}$ tend to exhibit less inertia than those in $\hat{\lambda}_t^{MSP}$. As the difference in forecasting performance narrows by the end of 2009, for example, $\hat{\lambda}_t^{MSP}$ continues to rise while $\hat{\lambda}_{t+h|t}^{DP}$ 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. As a robustness check, we also compute the evolution of the BMA weights for the unconditional forecasts that do not use any current quarter information about the financial variables. Under this scenario, $\hat{\lambda}_t^{BMA}$ is close to zero before the financial crisis, in spite of the fact that the log scores of the two models are not very different than those shown in Figure 4. This finding illustrates the well known lack of robustness of BMA weights: minor changes in model or data specification can lead to very different outcomes in terms of marginal likelihood comparisons. Figures A-2 and A-3 in Appendix D provide further details.

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.

Figure 6: Posterior distribution of λ Over Time: Static vs. Dynamic Pools



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

There is no separate panel for BMA, because, in a nutshell, BMA takes the posterior mass of $p^{(h)}_{BSP}(\lambda|\mathcal{I}^{\mathcal{P}}_t,\mathcal{P})$ and allocates it to the endpoints $\lambda=0$ and $\lambda=1$. Under both posterior distributions the probability mass is fairly uniformly allocated to the unit interval prior to 1997. During this period the main difference in the weights $\hat{\lambda}^{DP}_{t+h|t}$ and $\hat{\lambda}^{MSP}_t$ stems from the fact that the former is computed as posterior mean, whereas the latter is computed as posterior mode. The mode of $p^{(h)}_{BSP}(\lambda|\mathcal{I}^{\mathcal{P}}_t,\mathcal{P})$ exhibits a large swing from $\lambda=1$ to $\lambda=0$ in 1996. After 2000, two features of the SP distribution emerge. First, the mass is rather tightly concentrated around the mode. For instance, at the end of the sample $p^{(h)}_{BSP}(\lambda|\mathcal{I}^{\mathcal{P}}_t,\mathcal{P})$ puts virtually no mass on $\{\lambda\leq.5\}$. REMOVE Second, the distribution $p^{(h)}_{BSP}(\lambda|\mathcal{I}^{\mathcal{P}}_t,\mathcal{P})$ moves sluggishly over time: under static pools, new information has only a small effect on the overall distribution.

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)

REMOVE by MDN: This is consistent with BMA essentially assigning posterior weight one to the SWFF model after 2002. [I DON'T FOLLOW]

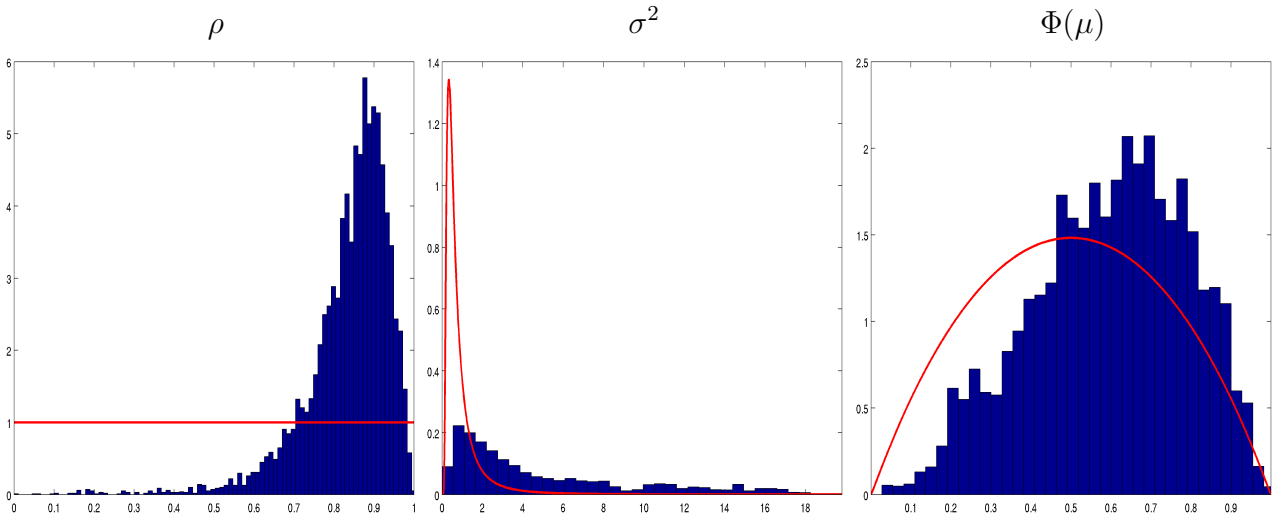
depending on which model has been the best forecaster in recent periods. Loosely speaking, we can think of the distribution as a seesaw, whose slope depends on the recent gap in log scores between the two models. A narrowing of the gap leads to a flattening of the seesaw, shifting the mean toward equal 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 Hyperparameter 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 as shown in Figure A-4 in Appendix D. The Appendix also documents the effect of shifting prior mass

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).

toward one by considering Beta distributions for ρ with (mean, standard deviation) of (0.8, 0.1) and (0.9, 0.2), respectively.

The center panel of Figure 7 focuses on the hyperparameter σ^2, which affects the shape of the marginal distribution of λ_t. As discussed in Section 4.1, $\sigma=1$ implies a uniform prior for the weight λ_t (regardless of ρ, as long as $\mu=0$). Values of σ less than one imply a unimodal (inverse-U shaped) prior for λ_t that peaks at equal weights (or $\Phi^{-1}(\mu)$ if $\mu\neq 0$). Values of σ greater than one, on the other hand, generate a prior that is U-shaped. The U-shaped prior leads to a convex posterior density (as opposed to the linear posterior obtained by combining a uniform prior distribution with the linear likelihood function depicted in Figure 3). The convexity of the posterior implies that the mean is shifted toward one of the endpoints ($\lambda_t=0$ or $\lambda_t=1$). In the context of the seesaw metaphor, $\sigma<1$ is equivalent to putting weight on the fulcrum, thereby limiting oscillations. Vice versa, $\sigma>1$ favors sharper changes in the posterior mean of λ_t.

The center panel of Figure 7 is based on the hyperparameter prior distribution[8]

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

Compared to the inverse Gamma prior $\mathcal{IG}(2,1)$ distribution the posterior mass is shifted toward the right and σ^2 is greater than one with very high probability. Thus, the data favor a parameterization in which the posterior mean is more sensitive to the arrival of new information.

Finally, 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, which is also based on Prior 2 in (34), shows the prior and posterior distributions of $\Phi(\mu)$ – that is, μ mapped into the space of the model weight λ. The prior is $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$: translated into the unit interval, this prior is centered at equal weights on both DSGE models and assigns a probability of approximately 68% to the interval [0.25, 0.75] Interestingly, in spite of the disparity in log scores between the two models during the Great Recession, 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.

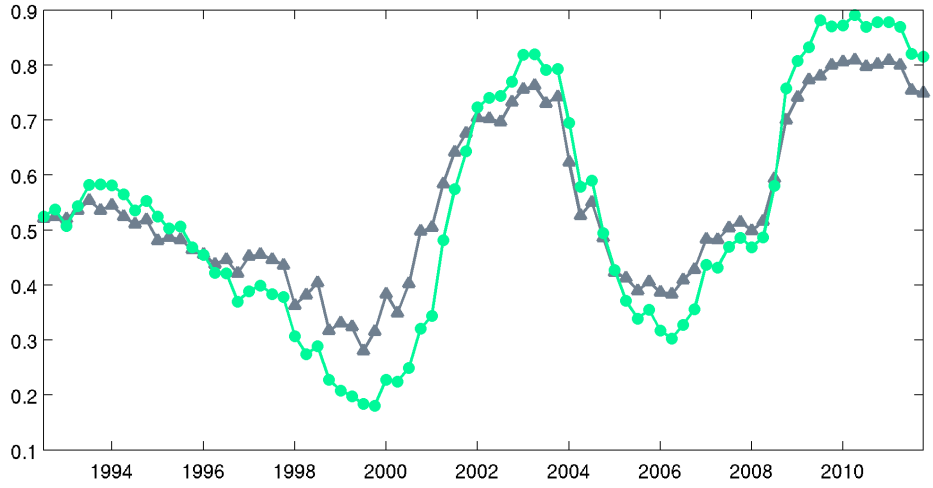
Figure 8 shows the effect of estimating the hyperparameters μ and σ on the evolution of the posterior mean $\hat{\lambda}^{DP}_{t+h|t}$. The figure compares the weight on the SWFF under Prior 1 in (33) and Prior 2 in (34). The swings of $\hat{\lambda}^{DP}_{t+h|t}$ are more pronounced under Prior 2 than 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

[8]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.

Figure 8: $\hat{\lambda}_{t+h|t}^{DP}$: Fixed vs. Estimated μ and σ



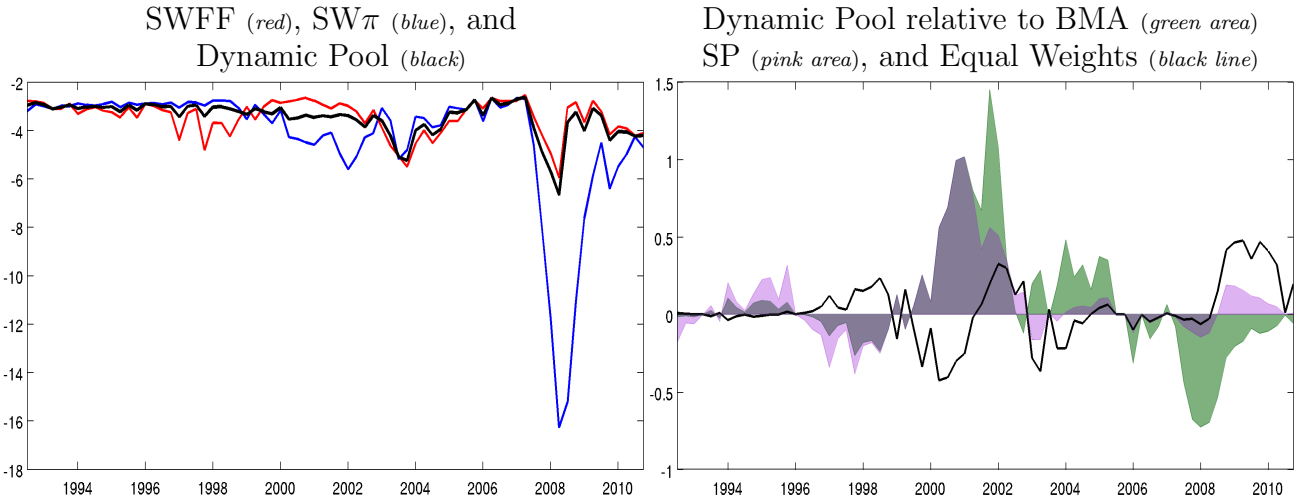
Notes: The figure shows the weight $\hat{\lambda}_{t+h|t}^{DP}$ for t =1992:Q1-2011:Q2 computed using the Prior 1 $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$ (black), and Prior 2 $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$,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

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 log predictive score of the dynamic pool, $\ln p_{DP}^{(h)}(\bar{y}_{t+h,h}|\mathcal{I}_{t^+}^{\mathcal{P}}, \mathcal{P})$, lies in between the score of the two DSGE models. Importantly, 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.

How does the dynamic pool compare to other forecast combination methods? 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. A comparison of the left and right panels is very instructive to understand the forecasting performance

Figure 9: Log Scores Comparison Over Time



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

of the various procedures. 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 forecast performance differentials arise. Following the long period in which SWπ is the dominant model both BMA and MSP are caught off guard by the change in regime. DP is caught off guard as well, as evidenced by the fact that it is forecasting worse than equal weights, but not as much and most importantly for not as long, given that it reacts quickly by increasing the weight on the financial friction model (recall Figure 5). In fact, DP performs better than equal weights in the last part of the dot-com bust period.

This episode offers two lessons for model combination. First, static approaches are not robust to regime changes, which is perhaps a key reason for why they have been found to perform worse than equal weights (see Amisano and Geweke (2013)). Second, the equal weights approach can be outperformed as long as there are persistent periods where one model is dominant. During these periods the dynamic pool adjusts to put more weight on the dominant model, and therefore performs better than equal weights. This is precisely what happens also at the end of the Great Recession period, when DP outperforms equal weights. During this period BMA gains an advantage from the fact that its weight on the

Table 2: Cumulative Log Scores / Differentials

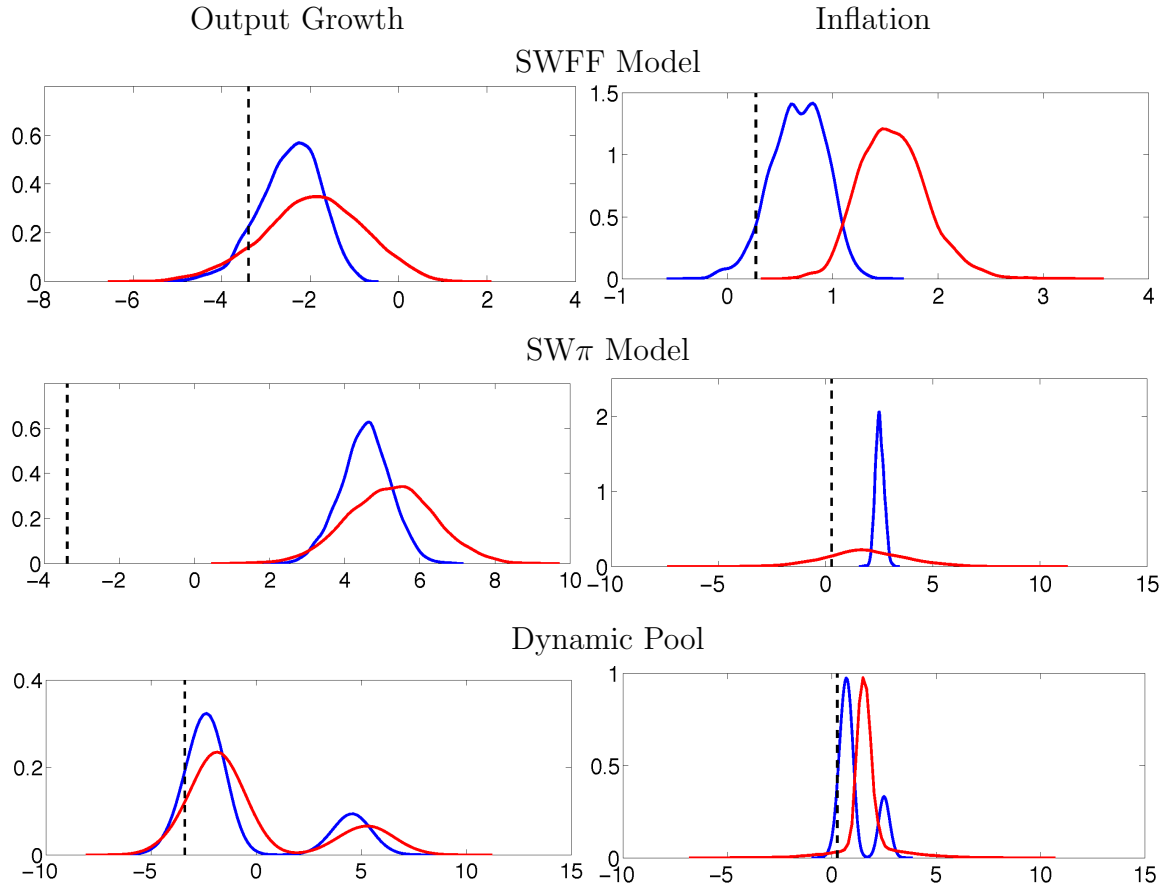
DP Prior	Log Score DP (1)	Differentials EW (2)	BMA (3)	MSP (4)
Prior 1: $\rho \sim U(0,1)$, $\mu = 0$, $\sigma^2 = 1$	-256.91	1.34	4.07	4.95
Prior 2: $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$	-256.43	1.82	4.55	5.43
Prior 3: $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu = 0$, $\sigma^2 \sim \mathcal{IG}(2,1)$	-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 specifications 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 1992:Q1-2011:Q2.

SWFF model is close to one (but as we discussed earlier, this result is not robust to a switch from semi-conditional to unconditional forecasts). The log score difference between the dynamic and the static pool during the recession is relatively modest.

Table 2 shows the cumulative log scores for three specifications of the dynamic pool, as well as the difference between the dynamic pools cumulative log scores and that of equal weights, BMA, and the (maximum likelihood) static pool, 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 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.

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.

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.

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

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

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:

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

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 in 2012 the policymakers did make the policy instrument explicitly contingent on the state of the labor market.[9] 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

[9] 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.

Figure 10.[10] 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 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. (Forthcoming), and therefore the predictive distribution of inflation is not nearly as dispersed.

[10] 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.

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 the rather different predictive densities obtained from the component models. 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. However, 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. 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.

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 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 an 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.

References

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

Amisano, G. and J. Geweke (2013): "Prediction Using Several Macroeconomic Models," *ECB Working Paper*, 1537.

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. (2013): "The Pass-Through of Sovereign Risk," *Manuscript, FRB of Minneapolis and Northwestern University.*

Brunnermeier, M. K. and Y. Sannikov (Forthcoming): "A Macroeconomic Model with a Financial Sector," *American Economic Review.*

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 (Forthcoming): "Risk Shocks," *American Economic Review*.

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 (Forthcoming): "Inflation in the Great Recession and New Keynesian Models," *American Economic Journal: Macroeconomics*.

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 (2013): "Generalised Density Forecast Combinations," *Manuscript, Bank of England*.

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.

Griffin, J., F. Quintana, and M. Steel (2011): "Flexible and Nonparametric Modeling," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk, Oxford University Press.

Guidolin, M. and A. Timmermann (2009): "Forecasts of US short-term interest rates: A flexible forecast combination approach," *Journal of Econemetrics*, 150, 297–311.

Hall, S. G. and J. Mitchell (2007): "Combining Density Forecasts," *International Journal of Forecasting*, 23, 1–13.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econemetrica*, 57, 357–384.

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 (2013): "Forecasting with DSGE Models with Financial Frictions," *Manuscript, National Bank of Poland*.

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.

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.

Terui, N. and H. K. van Dijk (2002): "Combined forecasts from linear and nonlinear time series models," *International Journal of Forecasting*, 18, 421–438.

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.

Appendix for *Time-varying Prediction Pools*

Marco Del Negro, Raiden B. Hasegawa, and Frank Schorfheide

A Detailed Description of DSGE Models

A.1 Model 1: The Smets-Wouters Model with Time-Varying Inflation Target (SWπ)

Model Specification. We use a slightly modified version of the Smets and Wouters (2007) model. Following Del Negro and Schorfheide (2013), we detrend the non-stationary model variables by a stochastic rather than a deterministic trend. This approach makes it possible to express almost all equilibrium conditions in a way that encompasses both the trend-stationary total factor productivity process in Smets and Wouters (2007), as well as the case where technology follows a unit root process. Let $\tilde{z}_t$ be the linearly detrended log productivity process which follows the autoregressive law of motion

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \sigma_z \varepsilon_{z,t}. \tag{A-1}$$

We detrend all non stationary variables by $Z_t = e^{\gamma t + \frac{1}{1-\alpha}\tilde{z}_t}$, where γ is the steady state growth rate of the economy. The growth rate of Z_t in deviations from γ, denoted by z_t, follows the process:

$$z_t = \ln(Z_t/Z_{t-1}) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z \epsilon_{z,t}. \tag{A-2}$$

All variables in the following equations are expressed in log deviations from their non-stochastic steady state. Steady state values are denoted by $*$-subscripts and steady state formulas are provided in the technical appendix of Del Negro and Schorfheide (2013). The consumption Euler equation is given by:

$$\begin{aligned} c_t = &-\frac{(1-he^{-\gamma})}{\sigma_c(1+he^{-\gamma})}\left(R_t - \mathbb{E}_t[\pi_{t+1}] + b_t\right) + \frac{he^{-\gamma}}{(1+he^{-\gamma})}\left(c_{t-1} - z_t\right) \\ &+ \frac{1}{(1+he^{-\gamma})}\mathbb{E}_t\left[c_{t+1} + z_{t+1}\right] + \frac{(\sigma_c - 1)}{\sigma_c(1+he^{-\gamma})}\frac{w_* L_*}{c_*}\left(L_t - \mathbb{E}_t[L_{t+1}]\right), \end{aligned} \tag{A-3}$$

where c_t is consumption, L_t is labor supply, R_t is the nominal interest rate, and π_t is inflation. The exogenous process b_t drives a wedge between the intertemporal ratio of the marginal utility of consumption and the riskless real return $R_t - \mathbb{E}_t[\pi_{t+1}]$, and follows an AR(1) process with parameters ρ_b and σ_b. The parameters σ_c and h capture the degree of relative risk aversion and the degree of habit persistence in the utility function, respectively. The following condition expresses the relationship between the value of capital in terms of consumption q_t^k and the level of investment i_t measured in terms of consumption goods:

$$q_t^k = S'' e^{2\gamma}(1+\bar{\beta})\Big(i_t - \frac{1}{1+\bar{\beta}}\left(i_{t-1} - z_t\right) - \frac{\bar{\beta}}{1+\bar{\beta}}\mathbb{E}_t\left[i_{t+1} + z_{t+1}\right] - \mu_t\Big), \tag{A-4}$$

which is affected by both investment adjustment cost (S'' is the second derivative of the adjustment cost function) and by μ_t, an exogenous process called the "marginal efficiency of investment" that affects the rate of transformation between consumption and installed capital. The exogenous process μ_t follows an AR(1) process with parameters ρ_μ and σ_μ. The parameter $\bar{\beta} = \beta e^{(1-\sigma_c)\gamma}$ depends on the intertemporal discount rate in the utility function of the households β, the degree of relative risk aversion σ_c, and the steady-state growth rate γ.

The capital stock, $\bar{k}_t$, evolves as

$$\bar{k}_t = \left(1 - \frac{i_*}{\bar{k}_*}\right)\left(\bar{k}_{t-1} - z_t\right) + \frac{i_*}{\bar{k}_*} i_t + \frac{i_*}{\bar{k}_*} S'' e^{2\gamma}(1+\bar{\beta})\mu_t, \tag{A-5}$$

where $i_*/\bar{k}_*$ is the steady state ratio of investment to capital. The arbitrage condition between the return to capital and the riskless rate is:

$$\frac{r_*^k}{r_*^k + (1-\delta)}\mathbb{E}_t[r_{t+1}^k] + \frac{1-\delta}{r_*^k + (1-\delta)}\mathbb{E}_t[q_{t+1}^k] - q_t^k = R_t + b_t - \mathbb{E}_t[\pi_{t+1}], \tag{A-6}$$

where r_t^k is the rental rate of capital, r_*^k its steady state value, and δ the depreciation rate. Given that capital is subject to variable capacity utilization u_t, the relationship between $\bar{k}_t$ and the amount of capital effectively rented out to firms k_t is

$$k_t \;=\; u_t - z_t + \bar{k}_{t-1}. \tag{A-7}$$

The optimality condition determining the rate of utilization is given by

$$\frac{1-\psi}{\psi} r_t^k = u_t, \tag{A-8}$$

where ψ captures the utilization costs in terms of foregone consumption. Real marginal costs for firms are given by

$$mc_t \;=\; w_t + \alpha L_t - \alpha k_t, \tag{A-9}$$

where w_t is the real wage and α is the income share of capital (after paying markups and fixed costs) in the production function. From the optimality conditions of goods producers it follows that all firms have the same capital-labor ratio:

$$k_t \;=\; w_t - r_t^k + L_t. \tag{A-10}$$

The production function is:

$$y_t \;=\; \Phi_p \left(\alpha k_t + (1-\alpha) L_t\right) + \mathcal{I}\{\rho_z < 1\}(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t, \tag{A-11}$$

if the log productivity is trend stationary. The last term $(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t$ drops out if technology has a stochastic trend, because in this case one has to assume that the fixed costs are proportional to the trend. Similarly, the resource constraint is:

$$y_t = g_t + \frac{c_*}{y_*}c_t + \frac{i_*}{y_*}i_t + \frac{r_*^k k_*}{y_*}u_t - \mathcal{I}\{\rho_z < 1\}\frac{1}{1-\alpha}\tilde{z}_t, \tag{A-12}$$

where again the term $-\frac{1}{1-\alpha}\tilde{z}_t$ disappears if technology follows a unit root process. Government spending g_t is assumed to follow the exogenous process:

$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t} + \eta_{gz}\sigma_z \varepsilon_{z,t}.$$

Finally, the price and wage Phillips curves are, respectively:

$$\pi_t = \frac{(1-\zeta_p\bar{\beta})(1-\zeta_p)}{(1+\iota_p\bar{\beta})\zeta_p((\Phi_p-1)\epsilon_p+1)}mc_t + \frac{\iota_p}{1+\iota_p\bar{\beta}}\pi_{t-1} + \frac{\bar{\beta}}{1+\iota_p\bar{\beta}}\mathbb{E}_t[\pi_{t+1}] + \lambda_{f,t}, \tag{A-13}$$

and

$$w_t = \frac{(1-\zeta_w\bar{\beta})(1-\zeta_w)}{(1+\bar{\beta})\zeta_w((\lambda_w-1)\epsilon_w+1)}\left(w_t^h - w_t\right) - \frac{1+\iota_w\bar{\beta}}{1+\bar{\beta}}\pi_t + \frac{1}{1+\bar{\beta}}\left(w_{t-1} - z_t - \iota_w\pi_{t-1}\right) + \frac{\bar{\beta}}{1+\bar{\beta}}\mathbb{E}_t\left[w_{t+1} + z_{t+1} + \pi_{t+1}\right] + \lambda_{w,t}, \tag{A-14}$$

where ζ_p, ι_p, and ϵ_p are the Calvo parameter, the degree of indexation, and the curvature parameter in the Kimball aggregator for prices, and ζ_w, ι_w, and ϵ_w are the corresponding

parameters for wages. w_t^h measures the household's marginal rate of substitution between consumption and labor, and is given by:

$$w_t^h = \frac{1}{1-he^{-\gamma}}\left(c_t - he^{-\gamma}c_{t-1} + he^{-\gamma}z_t\right) + \nu_l L_t, \tag{A-15}$$

where ν_l characterizes the curvature of the disutility of labor (and would equal the inverse of the Frisch elasticity in absence of wage rigidities). The mark-ups $\lambda_{f,t}$ and $\lambda_{w,t}$ follow exogenous ARMA(1,1) processes

$$\lambda_{f,t} = \rho_{\lambda_f}\lambda_{f,t-1} + \sigma_{\lambda_f}\varepsilon_{\lambda_f,t} + \eta_{\lambda_f}\sigma_{\lambda_f}\varepsilon_{\lambda_f,t-1}, \text{ and}$$

$$\lambda_{w,t} = \rho_{\lambda_w}\lambda_{w,t-1} + \sigma_{\lambda_w}\varepsilon_{\lambda_w,t} + \eta_{\lambda_w}\sigma_{\lambda_w}\varepsilon_{\lambda_w,t-1},$$

respectively. Finally, the monetary authority follows a generalized feedback rule:

$$\begin{aligned} R_t &= \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1\pi_t + \psi_2(y_t - y_t^f)\right) \\ &\quad + \psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m, \end{aligned} \tag{A-16}$$

where the flexible price/wage output y_t^f is obtained from solving the version of the model without nominal rigidities (that is, Equations (A-3) through (A-12) and (A-15)), and the residual r_t^m follows an AR(1) process with parameters ρ_{r^m} and σ_{r^m}.

In order to capture the rise and fall of inflation and interest rates in the estimation sample, we replace the constant target inflation rate by a time-varying target inflation. The interest-rate feedback rule of the central bank (A-16) is modified as follows

$$\begin{aligned} R_t &= \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1(\pi_t - \pi_t^*) + \psi_2(y_t - y_t^f)\right) \\ &\quad + \psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m. \end{aligned} \tag{A-17}$$

The time-varying inflation target evolves according to:

$$\pi_t^* = \rho_{\pi^*}\pi_{t-1}^* + \sigma_{\pi^*}\epsilon_{\pi^*,t}, \tag{A-18}$$

where $0 < \rho_{\pi^*} < 1$ and $\epsilon_{\pi^*,t}$ is an iid shock. We model π_t^* as following a stationary process, although our prior for ρ_{π^*} will force this process to be highly persistent. A detailed justification of this modification of the policy rule is provided in Del Negro and Schorfheide (2013).

Model Solution and State-Space Representation. We use the method in Sims (2002) to solve the log-linear approximation of the DSGE model. We collect all the DSGE model parameters in the vector θ, stack the structural shocks in the vector ϵ_t, and derive a state-space representation for our vector of observables y_t. The state-space representation is comprised of the transition equation:

$$s_t = \mathcal{T}(\theta)s_{t-1} + \mathcal{R}(\theta)\epsilon_t, \tag{A-19}$$

which summarizes the evolution of the states s_t, and the measurement equation:

$$y_t = \mathcal{Z}(\theta)s_t + \mathcal{D}(\theta), \tag{A-20}$$

which maps the states onto the vector of observables y_t, where $\mathcal{D}(\theta)$ represents the vector of steady state values for these observables.

The measurement equations for real output, consumption, investment, and real wage growth, hours, inflation, and interest rates are given by:

$$\begin{aligned}
\textit{Output growth} &= \gamma + 100\left(y_t - y_{t-1} + z_t\right) \\
\textit{Consumption growth} &= \gamma + 100\left(c_t - c_{t-1} + z_t\right) \\
\textit{Investment growth} &= \gamma + 100\left(i_t - i_{t-1} + z_t\right) \\
\textit{Real Wage growth} &= \gamma + 100\left(w_t - w_{t-1} + z_t\right), \\
\textit{Hours} &= \bar{l} + 100 l_t \\
\textit{Inflation} &= \pi_* + 100\pi_t \\
\textit{FFR} &= R_* + 100 R_t
\end{aligned} \tag{A-21}$$

where all variables are measured in percent, where π_* and R_* measure the steady state level of net inflation and short term nominal interest rates, respectively and where $\bar{l}$ captures the mean of hours (this variable is measured as an index). To incorporate information about low-frequency movements of inflation the set of measurement equations (A-21) is augmented by

$$\begin{aligned}
\pi_t^{O,40} &= \pi_* + 100\mathbb{E}_t\left[\frac{1}{40}\sum_{k=1}^{40}\pi_{t+k}\right] \\
&= \pi_* + \frac{100}{40}\mathcal{Z}(\theta)_{(\pi,.)}(I - \mathcal{T}(\theta))^{-1}\left(I - [\mathcal{T}(\theta)]^{40}\right)\mathcal{T}(\theta)s_t,
\end{aligned} \tag{A-22}$$

where $\pi_t^{O,40}$ represents observed long run inflation expectations obtained from surveys (in percent per quarter), and the right-hand-side of (A-22) corresponds to expectations obtained

from the DSGE model (in deviation from the mean π_*). The second line shows how to compute these expectations using the transition equation (A-19) and the measurement equation for inflation. $\mathcal{Z}(\theta)_{(\pi,.)}$ is the row of $\mathcal{Z}(\theta)$ in (A-20) that corresponds to inflation. The SWπ model is estimated using the observables in expressions (A-21) and (A-22).

A.2 Model 2: Smets-Wouters Model with Financial Frictions (SWFF)

Model Specification. We now add financial frictions to the SW model building on the work of Bernanke et al. (1999), Christiano et al. (2003), De Graeve (2008), and Christiano et al. (Forthcoming). In this extension, 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 may vary as a function of the entrepreneurs' leverage and their riskiness. Adding these frictions to the SW model amounts to replacing equation (A-6) with the following conditions:

$$E_t\left[\tilde{R}^k_{t+1} - R_t\right] = b_t + \zeta_{sp,b}\left(q^k_t + \bar{k}_t - n_t\right) + \tilde{\sigma}_{\omega,t} \tag{A-23}$$

and

$$\tilde{R}^k_t - \pi_t = \frac{r^k_*}{r^k_* + (1-\delta)} r^k_t + \frac{(1-\delta)}{r^k_* + (1-\delta)} q^k_t - q^k_{t-1}, \tag{A-24}$$

where $\tilde{R}^k_t$ is the gross nominal return on capital for entrepreneurs, n_t is entrepreneurial equity, and $\tilde{\sigma}_{\omega,t}$ captures mean-preserving changes in the cross-sectional dispersion of ability across entrepreneurs (see Christiano et al. (Forthcoming)) and follows an AR(1) process with parameters ρ_{σ_ω} and σ_{σ_ω}. The second condition defines the return on capital, while the first one determines the spread between the expected return on capital and the riskless rate.[11] The following condition describes the evolution of entrepreneurial net worth:

$$\begin{aligned} n_t \;=\; & \zeta_{n,\tilde{R}^k}\left(\tilde{R}^k_t - \pi_t\right) - \zeta_{n,R}\left(R_{t-1} - \pi_t\right) + \zeta_{n,qK}\left(q^k_{t-1} + \bar{k}_{t-1}\right) + \zeta_{n,n} n_{t-1} \\ & - \frac{\zeta_{n,\sigma_\omega}}{\zeta_{sp,\sigma_\omega}} \tilde{\sigma}_{\omega,t-1}. \end{aligned} \tag{A-25}$$

[11] Note that if $\zeta_{sp,b} = 0$ and the financial friction shocks $\tilde{\sigma}_{\omega,t}$ are zero, (A-23) and (A-24) coincide with (A-6).

State-Space Representation. The SWFF model uses in addition spreads as observables. The corresponding measurement equation is

$$\text{Spread} = SP_* + 100 I\!E_t \left[\tilde{R}^k_{t+1} - R_t \right], \tag{A-26}$$

where the parameter SP_* measures the steady state spread.

A.3 Prior Distribution

The prior distributions for the SWπ and the SWFF model are summarized in Table A-1. The joint prior distribution is obtained as the product of the marginals listed in the table. This prior is then truncated to ensure that for each parameter in the support of the prior the linearized DSGE model has a unique stable rational expectations equilibrium.

Table A-1: Priors

	Density	Mean	St. Dev.		Density	Mean	St. Dev.
Panel I: SWπ							
Policy Parameters							
ψ_1	Normal	1.50	0.25	ρ_R	Beta	0.75	0.10
ψ_2	Normal	0.12	0.05	ρ_{r^m}	Beta	0.50	0.20
ψ_3	Normal	0.12	0.05	σ_{r^m}	InvG	0.10	2.00
Nominal Rigidities Parameters							
ζ_p	Beta	0.50	0.10	ζ_w	Beta	0.50	0.10
Other "Endogenous Propagation and Steady State" Parameters							
α	Normal	0.30	0.05	π^*	Gamma	0.75	0.40
Φ	Normal	1.25	0.12	γ	Normal	0.40	0.10
h	Beta	0.70	0.10	S''	Normal	4.00	1.50
ν_l	Normal	2.00	0.75	σ_c	Normal	1.50	0.37
ι_p	Beta	0.50	0.15	ι_w	Beta	0.50	0.15
r_*	Gamma	0.25	0.10	ψ	Beta	0.50	0.15
(Note $\beta = (1/(1 + r_*/100))$							
ρs, σs, and ηs							
ρ_z	Beta	0.50	0.20	σ_z	InvG	0.10	2.00
ρ_b	Beta	0.50	0.20	σ_b	InvG	0.10	2.00
ρ_{λ_f}	Beta	0.50	0.20	σ_{λ_f}	InvG	0.10	2.00
ρ_{λ_w}	Beta	0.50	0.20	σ_{λ_w}	InvG	0.10	2.00
ρ_μ	Beta	0.50	0.20	σ_μ	InvG	0.10	2.00
ρ_g	Beta	0.50	0.20	σ_g	InvG	0.10	2.00
η_{λ_f}	Beta	0.50	0.20	η_{λ_w}	Beta	0.50	0.20
η_{gz}	Beta	0.50	0.20				
ρ_{π^*}	Beta	0.50	0.20	σ_{π^*}	InvG	0.03	6.00
Panel II: SWFF							
SP_*	Gamma	2.00	0.10	$\zeta_{sp,b}$	Beta	0.05	0.005
ρ_{σ_w}	Beta	0.75	0.15	σ_{σ_w}	InvG	0.05	4.00

Notes: Smets and Wouters (2007) original prior is a $Gamma(.62, .10)$. The following parameters are fixed in Smets and Wouters (2007): $\delta = 0.025$, $g_* = 0.18$, $\lambda_w = 1.50$, $\varepsilon_w = 10$, and $\varepsilon_p = 10$. In addition, for the model with financial frictions we fix the entrepreneurs' steady state default probability $\bar{F}_* = 0.03$ and their survival rate $\gamma_* = 0.99$. The columns "Mean" and "St. Dev." list the means and the standard deviations for Beta, Gamma, and Normal distributions, and the values s and ν for the Inverse Gamma (InvG) distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region. The prior for $\bar{l}$ is $\mathcal{N}(-45, 5^2)$.

B Data

Real GDP (GDPC), the GDP price deflator (GDPDEF), nominal personal consumption expenditures (PCEC), and nominal fixed private investment (FPI) are constructed at a quarterly frequency by the Bureau of Economic Analysis (BEA), and are included in the National Income and Product Accounts (NIPA). Average weekly hours of production and nonsupervisory employees for total private industries (AWHNONAG), civilian employment (CE16OV), and civilian noninstitutional population (LNSINDEX) are produced by the Bureau of Labor Statistics (BLS) at the monthly frequency. The first of these series is obtained from the Establishment Survey, and the remaining from the Household Survey. Both surveys are released in the BLS Employment Situation Summary (ESS). Since our models are estimated on quarterly data, we take averages of the monthly data. Compensation per hour for the nonfarm business sector (COMPNFB) is obtained from the Labor Productvity and Costs (LPC) release, and produced by the BLS at the quarterly frequency. All data are transformed following Smets and Wouters (2007). Let Δ denote the temporal difference operator. Then:

$$
\begin{aligned}
\textit{Output growth} &= 100 * \Delta LN((GDPC)/LNSINDEX) \\
\textit{Consumption growth} &= 100 * \Delta LN((PCEC/GDPDEF)/LNSINDEX) \\
\textit{Investment growth} &= 100 * \Delta LN((FPI/GDPDEF)/LNSINDEX) \\
\textit{Real Wage growth} &= 100 * \Delta LN(COMPNFB/GDPDEF) \\
\textit{Hours} &= 100 * LN((AWHNONAG * CE16OV/100)/LNSINDEX) \\
\textit{Inflation} &= 100 * \Delta LN(GDPDEF).
\end{aligned}
$$

The federal funds rate is obtained from the Federal Reserve Board's H.15 release at the business day frequency. We take quarterly averages of the annualized daily data and divide by four. In the estimation of the DSGE model with financial frictions we measure *Spread* as the annualized Moody's Seasoned Baa Corporate Bond Yield spread over the 10-Year Treasury Note Yield at Constant Maturity. Both series are available from the Federal Reserve Board's H.15 release. Like the federal funds rate, the spread data is also averaged over each quarter and measured at the quarterly frequency. This leads to:

$$
\begin{aligned}
FFR &= (1/4) * FEDERAL\ FUNDS\ RATE \\
Spread &= (1/4) * (BaaCorporate\ -\ 10yearTreasury)
\end{aligned}
$$

The long-run inflation expectations are obtained from the Blue Chip Economic Indicators survey and the Survey of Professional Forecasters (SPF) available from the FRB Philadelphia's Real-Time Data Research Center. Long-run inflation expectations (average CPI inflation over the next 10 years) are available from 1991:Q4 onwards. Prior to 1991:Q4, we use the 10-year expectations data from the Blue Chip survey to construct a long time series that begins in 1979:Q4. Since the Blue Chip survey reports long-run inflation expectations only twice a year, we treat these expectations in the remaining quarters as missing observations and adjust the measurement equation of the Kalman filter accordingly. Long-run inflation expectations $\pi_t^{O,40}$ are therefore measured as

$$\pi_t^{O,40} = (\text{10-YEAR AVERAGE CPI INFLATION FORECAST} - 0.50)/4.$$

where 0.50 is the average difference between CPI and GDP annualized inflation from the beginning of the sample to 1992. We divide by 4 because the data are expressed in quarterly terms.

Many macroeconomic time series get revised multiple times by the statistical agencies that publish the series. In many cases the revisions reflect additional information that has been collected by the agencies, in other instances revisions are caused by changes in definitions. For instance, the BEA publishes three releases of quarterly GDP in the first three month following the quarter. Thus, in order to be able to compare DSGE model forecasts to real-time forecasts made by private-sector professional forecasters or the Federal Reserve Board, it is important to construct vintages of real time historical data. We follow the work by Edge and Gürkaynak (2010) and construct data vintages that are aligned with the publication dates of the Blue Chip survey. A detailed description of how this data set is constructed is provided in Del Negro and Schorfheide (2013).

C Computational Details

C.1 DSGE Models

The parameter estimation for the two DSGE models is described in detail in Del Negro and Schorfheide (2013). Thus, this Appendix focuses on the computation of h-step predictive

densities $p(y_{t:t+h}|\mathcal{I}_{t-1}^m, \mathcal{M}_m)$. Starting point is the state-space representation of the DSGE model. The transition equation

$$s_t = \mathcal{T}(\theta)s_{t-1} + \mathcal{R}(\theta)\epsilon_t,\ \epsilon_t \sim N(0, \mathcal{Q}) \tag{A-27}$$

summarizes the evolution of the states s_t. The measurement equation:

$$y_t = \mathcal{Z}(\theta)s_t + \mathcal{D}(\theta), \tag{A-28}$$

maps the states onto the vector of observables y_t, where $\mathcal{D}(\theta)$ represents the vector of steady states for these observables. To simplify the notation we omit model superscripts/subscripts and we drop $\mathcal{M}_m$ from the conditioning set. We assume that the forecasts are based on the $\mathcal{I}_{t-1}$ information set. Let θ denote the vector of DSGE model parameters. For each draw θ^i, $i = 1, \ldots, N$, from the posterior distribution $p(\theta|\mathcal{I}_{t-1})$, execute the following steps:

1. Evaluate
   $\mathcal{T}(\theta), \mathcal{R}(\theta), \mathcal{Z}(\theta), \mathcal{D}(\theta)$.

2. Run the Kalman filter to obtain $s_{t-1|t-1}$ and $P_{t-1|t-1}$.

3. Compute $\hat{s}_{t|t-1} = s_{t|\mathcal{I}_{t-1}}$ and $\hat{P}_{t|t-1} = P_{t|\mathcal{I}_{t-1}}$ as

   (a) Unconditional forecasts: $\hat{s}_{t|t-1} = \mathcal{T}s_{t-1|t-1}$, $\hat{P}_{t|t-1} = \mathcal{T}P_{t-1|t-1}\mathcal{T}' + \mathcal{R}\mathcal{Q}\mathcal{R}'$.

   (b) Semiconditional forecasts (using time t spreads, and FFR): after computing $\hat{s}_{t|t-1}$ and $\hat{P}_{t|t-1}$ using the "unconditional" formulas, run time t updating step of Kalman filter using a measurement equation that only uses time t values of these two observables.

4. Compute recursively for $j = 1, .., h$ the objects $\hat{s}_{t+j|t-1} = \mathcal{T}s_{t+j-1|t-1}$, $\hat{P}_{t+j|t-1} = \mathcal{T}P_{t+j-1|t-1}\mathcal{T}' + \mathcal{R}\mathcal{Q}\mathcal{R}'$ and construct the matrices

$$\hat{s}_{t:t+k|t-1} = \begin{bmatrix} \hat{s}_{t|t-1} \\ \vdots \\ \hat{s}_{t+k|t-1} \end{bmatrix}$$

   and

$$\hat{P}_{t:t+k|t-1} = \begin{bmatrix} \hat{P}_{t|t-1} & \hat{P}_{t|t-1}\mathcal{T}' & \ldots & \hat{P}_{t|t-1}\mathcal{T}^{k\,\prime} \\ \mathcal{T}\hat{P}_{t|t-1} & \hat{P}_{t+1|t-1} & \ldots & \hat{P}_{t+1|t-1}\mathcal{T}^{k-1\,\prime} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{T}^k\hat{P}_{t|t-1} & \mathcal{T}^{k-1}\hat{P}_{t+1|t-1} & \ldots & \hat{P}_{t+k|t-1} \end{bmatrix}.$$

This leads to: $s_{t:t+h}|(\theta, \mathcal{I}_{t-1}) \sim N(\hat{s}_{t:t+h|t-1}, \hat{P}_{t:t+h|t-1})$.

5. The distribution of $y_{t:t+h} = \tilde{\mathcal{D}} + \tilde{\mathcal{Z}} s_{t:t+h}$ is

$$y_{t:t+h}|(\theta, \mathcal{I}_{t-1}) \sim N(\tilde{\mathcal{D}} + \tilde{\mathcal{Z}}\hat{s}_{t:t+h|t-1}, \tilde{\mathcal{Z}}\hat{P}_{t:t+h|t-1}\tilde{\mathcal{Z}}'),$$

where $\tilde{\mathcal{Z}} = I_{h+1} \otimes \mathcal{Z}$ and $\tilde{\mathcal{D}} = 1_{h+1} \otimes \mathcal{D}$ (note $I_1 = 1_1 = 1$)

6. Compute

$$p(y^o_{t:t+h}|\theta, \mathcal{I}_{t-1}) = p_N(y^o_{t:t+h}; \tilde{\mathcal{D}} + \tilde{\mathcal{Z}}\hat{s}_{t:t+h|t-1}, \tilde{\mathcal{Z}}\hat{P}_{t:t+h|t-1}\tilde{\mathcal{Z}}'), \tag{A-29}$$

where $y^o_{t:t+h}$ are the actual observations and $p_N(x; \mu, \Sigma)$ is the probability density function of a $N(\mu, \Sigma)$.

7. For linear functions $Fy_{t:t+h}$ (e.g., four quarter averages, etc.) where F is a matrix of fixed coefficients the predictive density becomes

$$p(Fy^o_{t:t+h}|\theta, \mathcal{I}_{t-1}) = p_N(Fy^o_{t:t+h}; F\tilde{\mathcal{D}} + F\tilde{\mathcal{Z}}\hat{s}_{t:t+h|t-1}, F\tilde{\mathcal{Z}}\hat{P}_{t:t+h|t-1}\tilde{\mathcal{Z}}'F'). \tag{A-30}$$

In the application we choose the matrix F such that $Fy_{t:t+h} = \bar{y}_{t+h,h} = \frac{1}{h}\sum_{j=1}^{h} y_{t+j}$ and let

$$p(\bar{y}^o_{t+h,h}|\mathcal{I}_{t-1}) = \frac{1}{N}\sum_{i=1}^{N} p(\bar{y}^o_{t+h,h}|\theta^i, \mathcal{I}_{t-1}). \tag{A-31}$$

C.2 Dynamic Prediction Pool

In each period t, the principal has to conduct inference about λ_t to generate $\hat{\lambda}^{DP}_{t+h|t}(\theta)$, where $\theta = (\rho, \mu, \sigma)'$ is the vector of hyperparameters. Some of the results that we are reporting in the main part of the paper are conditional on a particular value of θ, while others are obtained by integrating out θ under the relevant pseudo posterior distribution.

We use a bootstrap particle filter to update the sequence of pseudo posteriors $p^{(h)}(\lambda_t|\theta, \mathcal{I}^{\mathcal{P}}_t, \mathcal{P})$. Let $s_t = [x_t, \lambda_t]'$ and assume that the period $t-1$ particles $\{s^j_{t-1}, W^j_{t-1}\}_{j=1}^N$ approximate the moments of $p^{(h)}(\lambda_{t-1}|\theta, \mathcal{I}^{\mathcal{P}}_{t-1}, \mathcal{P})$:

$$\frac{1}{N}\sum_{j=1}^{N} f(s^j_{t-1})W^j_{t-1} \approx \int f(s_{t-1})p^{(h)}(s_{t-1}|\theta, \mathcal{I}^{\mathcal{P}}_{t-1}, \mathcal{P})ds_{t-1}. \tag{A-32}$$

By $\approx$ we mean that under suitable regularity conditions (see, for instance, Chopin (2004)) the Monte Carlo average satisfies a strong law of large numbers and a central limit theorem. An initial set of particles can be generated by *iid* sampling from $x_0 \sim N(\mu, \sigma^2)$, letting $s_0^j = [x_0^j, \Phi(x_0^j)]$, and setting $W_0^j = 1$. The bootstrap particle filter involves the following recursion:

1. Propagate particles forward:

$$\tilde{x}_t^j = (1-\rho)\mu + \rho x_{t-1}^j + \sqrt{1-\rho^2}\sigma\varepsilon_t^j,\ \varepsilon_t^j \sim N(0,1). \tag{A-33}$$

2. Compute $\tilde{\lambda}_t^j = \Phi(\tilde{x}_t^j)$ and let $\tilde{s}_t^j = [\tilde{x}_t^j, \tilde{\lambda}_t^j]'$.

3. Compute the incremental weights

$$\tilde{w}_t^j = p^{(h)}(\bar{y}_{t,h}|\lambda_t^j, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}) = \tilde{\lambda}_t^j p(\bar{y}_{t,h}|\mathcal{I}_{t-h}^1, \mathcal{M}_1) + (1-\tilde{\lambda}_t^j)p(\bar{y}_{t,h}|\mathcal{I}_{t-h}^2, \mathcal{M}_2). \tag{A-34}$$

   The predictive density $p^{(h)}(\bar{y}_{t,h}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})$ can be approximated by

$$\hat{p}^{(h)}(\bar{y}_{t,h}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}) = \frac{1}{N}\sum_{j=1}^{N} \tilde{w}_t^j W_{t-1}^j. \tag{A-35}$$

4. Update the weights according to

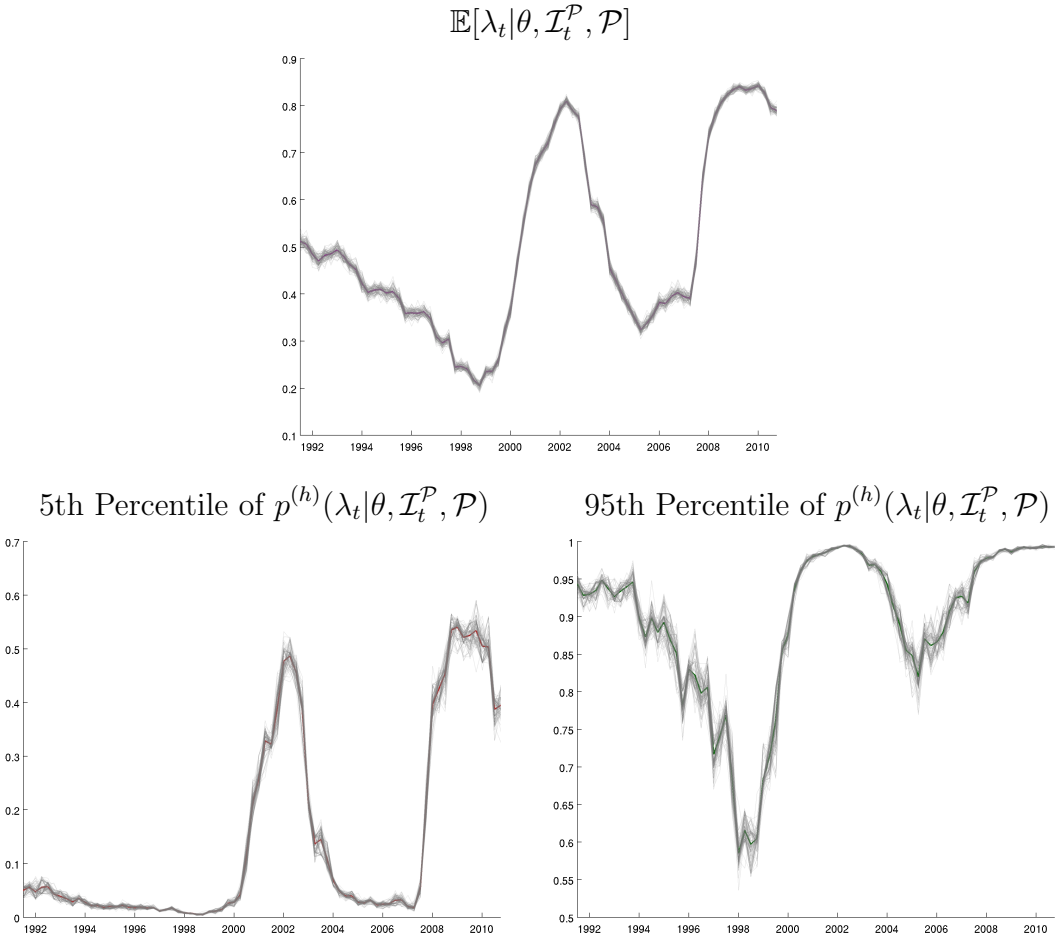
$$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{N}\sum_{j=1}^{N} W_{t-1}^j}. \tag{A-36}$$

5. Resample (using multinomial resampling) the particles if the distribution of particle weights becomes very uneven. Let $ESS = N^2 / \sum_{j=1}^{N} \tilde{w}_t^j W_{t-1}^j$. (a) If $ESS < (2/3)N$ resample the particles and let s_t^j denote the value of the resampled particle j and set its weight $W_t^j = 1$. (b) If $ESS \geq (2/3)N$ let $s_t^j = \tilde{s}_t^j$ and $W_t^j = \tilde{W}_t^j$.

6. The particle system $\{s_t^j, W_t^j\}_{j=1}^N$ approximates

$$\frac{1}{N}\sum_{j=1}^{N} f(s_{t-1}^j) W_{t-1}^j \xrightarrow{p} \int f(s_{t-1}) p^{(h)}(s_{t-1}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}) ds_{t-1} \tag{A-37}$$

Figure A-1: Accuracy of Particle Filter Approximation: $N = 1000$, $N_{rep} = 100$



Notes: Figure depicts results from N_rep runs of the particle filter. θ is given by $\rho = 0.9$, $\mu = 0$, and $\sigma = 1$.

In Figure A-1 we graphically examine the Monte Carlo variance of our estimate of $p^{(h)}(\lambda_t|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$. The figure is based on $N = 1,000$ particles and $N_{rep} = 100$ independent runs of the particle filter. We set $\rho = 0.9$, $\mu = 0$, and $\sigma = 1$. The accuracy deteriorates somewhat as ρ approaches one, because the innovation in state-transition equation decreases and do does the degree of particle mutation. For the limit case $\rho = 1$ filtering becomes unnecessary because $\lambda_t = \lambda$.

The predictive densities can be combined to form the pseudo-likelihood function

$$p^{(h)}(\bar{y}_{1:t,h}|\theta, \mathcal{P}) = \prod_{t=1}^{T} p^{(h)}(\bar{y}_{t,h}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}). \tag{A-38}$$

The pseudo-likelihood has the particle filter approximation

$$\hat{p}^{(h)}(\bar{y}_{1:t,h}|\theta,\mathcal{P}) = \prod_{t=1}^{T} \hat{p}^{(h)}(\bar{y}_{t,h}|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P}), \tag{A-39}$$

where $\hat{p}^{(h)}(\bar{y}_{t,h}|\theta,\mathcal{I}_{t-1}^{\mathcal{P}},\mathcal{P})$ was defined in (A-35). We use the pseudo-likelihood function to conduct inference with respect to θ:

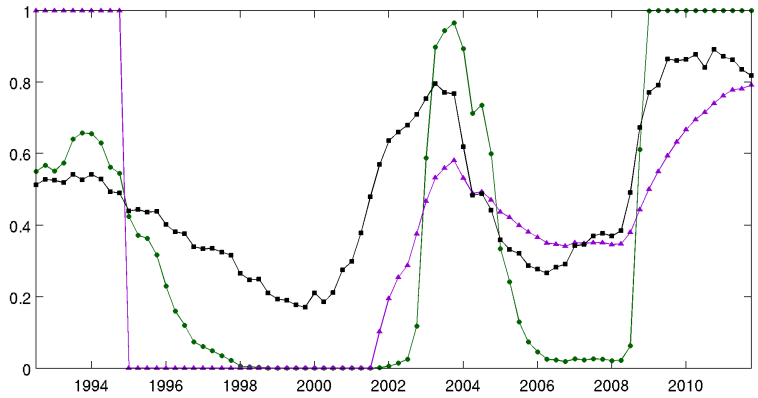
$$p^{(h)}(\theta|\mathcal{I}_t^{\mathcal{P}},\mathcal{P}) \propto \hat{p}^{(h)}(\bar{y}_{1:t,h}|\theta,\mathcal{P})p(\theta). \tag{A-40}$$

In order to generate draws from pseudo-posterior $p^{(h)}(\theta|\mathcal{I}_t^{\mathcal{P}},\mathcal{P})$, we embedd the particle-filter approximation of the pseudo-likelihood function in an otherwise standard random-walk Metropolis-Hastings algorithm. A theoretical justification for this procedure is provided in Andrieu et al. (2010). The random-walk Metropolis-Hastings (RWMH) algorithm is identical to Algorithm 1 in Del Negro and Schorfheide (2013). Due to the low dimensionality of the hyperparameter vector θ and the high degree of accuracy of the particle filter approximation, the posterior sampler is very efficient. All results reported in the main text are based on 5,000 particles and 10,000 draws from the RWMH algorithm.

D Additional Tables and Figures

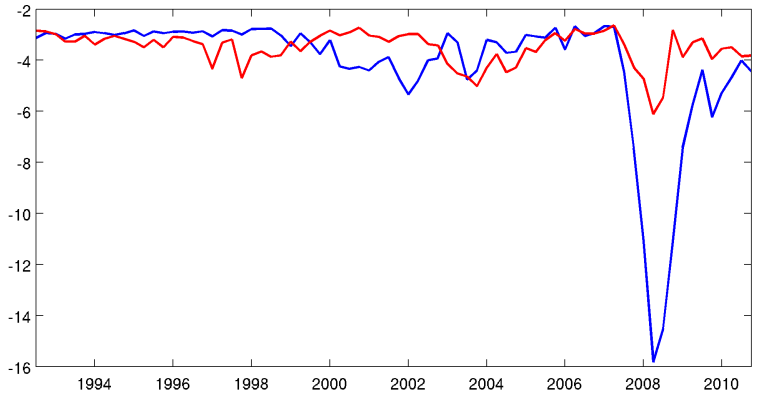
Additional empirical results are presented in Figures A-2 to A-5.

Figure A-2: Weights in Real Time: BMA, Static, and Dynamic Pools – $\mathcal{I}_t^m$ Excludes Time $t+1$ Information from Financial Variables



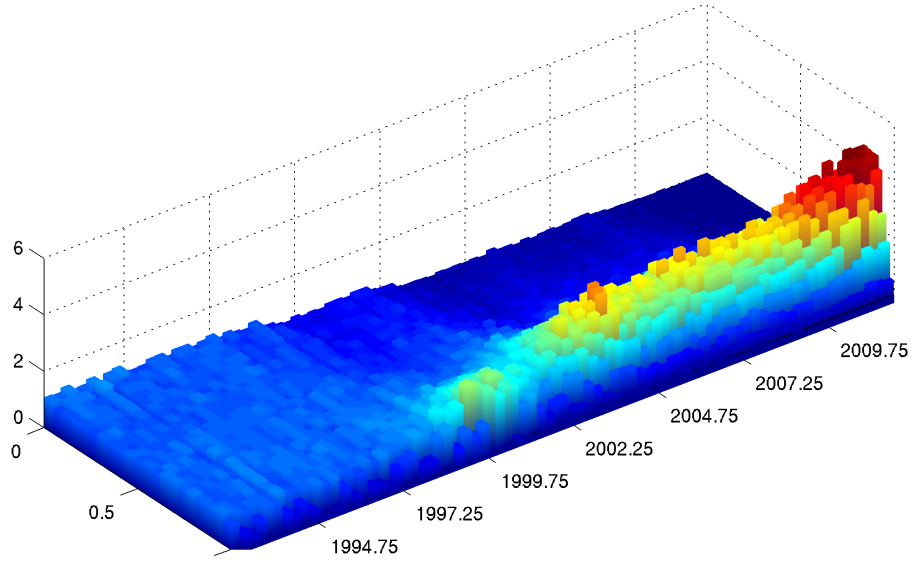
Notes: The figure shows the weight on the SWFF model in forecast pools, computed using real time information only, over the period 1992:Q1-2011:Q2 for three different pooling techniques: BMA ($\hat{\lambda}_t^{BMA}$ – green), (maximum likelihood) static pool ($\hat{\lambda}_t^{MSP}$ – purple), and dynamic pools ($\hat{\lambda}_{t+h|t}^{DP}$ – black).

Figure A-3: Log Scores Comparison: SWFF vs. SWπ Without Time $t+1$ Information from Financial Variables



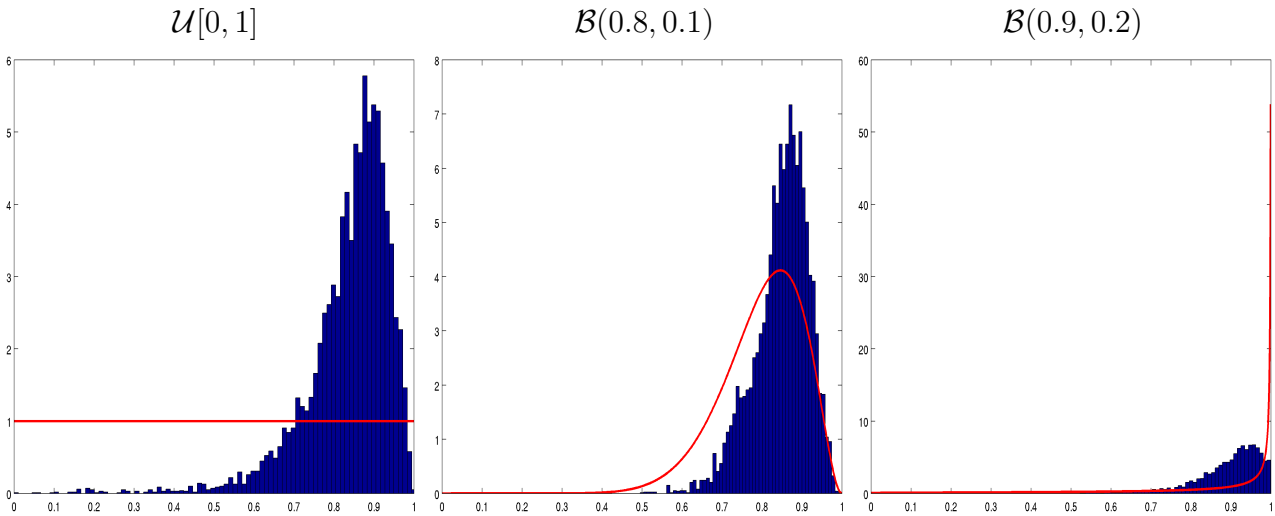
Notes: The figure shows the log scores $p(\bar{y}_{t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m)$ for SWFF (red), and SWπ (blue) over the period 1992:Q1-2011:Q2.

Figure A-4: Posterior $p^{(h)}(\rho|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$ Over Time



Notes: The figure shows the posterior $p^{(h)}(\rho|\mathcal{I}_t^{\mathcal{P}}, \mathcal{P})$ for t =1992:Q1-2011:Q2 based on the hyperparameter Prior 1: $\rho \sim \mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$.

Figure A-5: End-of-Sample Posterior of the Hyperparameter ρ Under Different Priors



Notes: The three panels show the posterior $p(\rho|\mathcal{I}_T^{\mathcal{P}}, \mathcal{P})$ (histogram) under three priors (red line): $\mathcal{U}[0,1]$, $\mu = 0$, $\sigma = 1$ (left); $\rho \sim \mathcal{B}(0.8, 0.1)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$ (center); and $\mathcal{B}(0.9, 0.2)$, $\mu \sim \mathcal{N}(0, \Phi^{-1}(0.75))$, $\sigma^2 \sim \mathcal{IG}(2,1)$ (right).