Federal Reserve Bank of New York
Staff Reports

Online Estimation of DSGE Models

Michael Cai
Marco Del Negro
Edward Herbst
Ethan Matlin
Reca Sarfati
Frank Schorfheide

Staff Report No. 893
August 2019

Online Estimation of DSGE Models

Michael Cai, Marco Del Negro, Edward Herbst, Ethan Matlin, Reca Sarfati,
and Frank Schorfheide

Federal Reserve Bank of New York Staff Reports, no. 893
August 2019
JEL classification: C11, C32, C53, E32, E37, E52

Abstract

This paper illustrates the usefulness of sequential Monte Carlo (SMC) methods in approximating DSGE model posterior distributions. We show how the tempering schedule can be chosen adaptively, explore the benefits of an SMC variant we call *generalized tempering* for "online" estimation, and provide examples of multimodal posteriors that are well captured by SMC methods. We then use the online estimation of the DSGE model to compute pseudo-out-of-sample density forecasts of DSGE models with and without financial frictions and document the benefits of conditioning DSGE model forecasts on nowcasts of macroeconomic variables and interest rate expectations. We also study whether the predictive ability of DSGE models changes when we use priors that are substantially looser than those commonly adopted in the literature.

Key words: adaptive algorithms, Bayesian inference, density forecasts, online estimation, sequential Monte Carlo methods

---

Del Negro, Matlin, Sarfati: Federal Reserve Bank of New York (emails: marco.delnegro@ny.frb.org, ethan.matlin@ny.frb.org, rebecca.sarfati@ny.frb.org). Cai: Northwestern University (email: michaelcai@u.northwestern.edu). Herbst: Board of Governors of the Federal Reserve System (email: edward.p.herbst@frb.gov). Schorfheide: University of Pennsylvania (email: schorf@ssc.upenn.edu). This paper was prepared for the Econometrics Journal special session at the 2018 Royal Economics Society Conference. The authors thank participants at that session, as well as those at various other conferences and seminars, for helpful comments. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) PSC Regular Memory (Bridges) through allocation TG-SES190003. It also used the Federal Reserve Bank of Dallas's BigTex computing resources, for which the authors are very grateful. Schorfheide acknowledges financial support from the National Science Foundation under Grant SES 1851634.

1 Introduction

The goal of this paper is to provide a framework for performing "online" estimation of medium- and large-scale DSGE models using sequential Monte Carlo (SMC) techniques that can be parallelized. We borrow the term online estimation from the statistics and machine-learning literature to describe the task of re-estimating a model frequently as new data become available. While our framework is tailored toward central bank applications with its focus on forecasting and the use of real-time data, it should also be relevant to academic researchers who are interested in estimating complex models in a stepwise fashion. That is, either sequentially increasing the sample information used in the estimation or mutating preliminary estimates from a relatively simple (linear dynamics, heterogeneous agents models with a coarse approximation) model, which can be computed quickly, into estimates of a more complex model (nonlinear dynamics, heterogeneous agents models with a finer approximation), which would take a long time to compute from scratch.

SMC methods have been traditionally used to solve nonlinear filtering problems, an example being the bootstrap particle filter of Gordon et al. (1993). Subsequently, Chopin (2002) showed how to adapt particle filtering techniques to conduct posterior inference for a static parameter vector. The first paper that applied SMC techniques to posterior inference for the parameters of a (small-scale) DSGE model was Creal (2007). Subsequent work by Herbst and Schorfheide (2014, 2015) fine-tuned the algorithm so that it could be used for the estimation of medium- and large-scale models.

In order to frame the paper's contributions, a brief summary of how SMC works is in order. SMC algorithms approximate a target posterior distribution by creating intermediate approximations for a sequence of bridge distributions. Under a version of SMC called likelihood tempering, these bridge distributions are posteriors constructed based on a likelihood sequence that is generated by raising the full-sample likelihood function to the power of ϕ_n, where ϕ_n increases from zero to one. At each stage, the current bridge distribution is represented by a swarm of particles. Each particle is associated with a value and a weight. Weighted averages of the particle values converge to expectations under the stage-n distribution. The transition from stage $n-1$ to n involves changing the particle weights and values so that the swarm adapts to the new distribution.

This paper makes several contributions. First, we replace the fixed tempering schedule for the DSGE model likelihood function by an adaptive tempering schedule. While adaptive

tempering schedules have been used in the statistics literature before, e.g., Jasra et al. (2011), their use for the estimation of DSGE model parameters is new. Adaptive tempering is particularly attractive for online estimation because in some periods the new observation(s) may be quite informative and shift the posterior distribution substantially, whereas in other periods the posterior distribution remains essentially unchanged. In the former case, it is desirable to use a larger number of intermediate stages to reach the new posterior to maintain accuracy. In the latter case one would like to keep the number of intermediate stages small to reduce the runtime. Our adaptive schedules are controlled by a single tuning parameter and we assess how this tuning parameter affects the accuracy-runtime trade-off for the algorithm.

Second, we modify the SMC algorithm so that the initial particles are drawn from a previously computed posterior distribution instead of the prior distribution. This initial posterior can result from estimating the model on a shorter sample or a simpler version of the same model (e.g., linear versus non-linear, as discussed above) on the full sample. In the former case, our approach can be viewed as a form of generalized data tempering. Under standard data tempering, the likelihood sequence is generated by adding progressively more observations. Our approach is more general in that it allows users to add information from fractions of observations and accommodates data revisions, which are pervasive in macro applications.

Third, we contribute to the literature that assesses the real-time pseudo-out-of-sample forecast performance of DSGE models. Here *real-time* means that for a forecast using a sample ending at time t, the data vintage used to estimate the model is one that would have been available to the econometrician at the time. *Pseudo-out-of-sample* means that the forecasts were however produced ex-post.[1] We use the proposed SMC techniques to recursively estimate the Smets and Wouters (2007) model and a version of this model with financial frictions. Our forecast evaluation exercises extend previous results in Del Negro and Schorfheide (2013) and Cai et al. (forthcoming). Relative to the former study, we now have access to a longer sample that includes the recovery from the Great Recession, and we focus on log predictive scores rather than probability integral transforms. Relative to the latter paper, we are focusing on density rather than point forecasts. Moreover, both of the earlier papers were based on the widely-used Random Walk Metropolis Hastings (RWMH) algorithm, whereas the current paper utilizes the SMC algorithm described above.

We find that the DSGE model with financial frictions generates more accurate forecasts

[1] Cai et al. (forthcoming) provide a genuine real-time forecast evaluation that uses the NY Fed DSGE model's forecasts.

than the one without during and after the Great Recession, by and large confirming the results of Del Negro and Schorfheide (2013) and Cai et al. (forthcoming). We also consider the effects of dramatically loosening the prior informativeness on forecasting performance. Despite the emergence of multiple modes in the posterior distribution, our SMC-based results show that the large increase in the prior standard deviation has surprisingly small effects on forecasting accuracy, thereby debunking the notion that priors in DSGE models are chosen to improve the model's predictive ability.

The remainder of this paper is organized as follows. In Section 2, we outline the basic structure of an SMC algorithm designed for posterior inference on a static parameter vector θ. We review different tempering approaches and present an algorithm for the adaptive choice of the tempering schedule. Section 3 provides an overview of the DSGE models that are estimated in this paper. In Section 4, we study various dimensions of the performance of SMC algorithms: we assess the accuracy and runtime tradeoffs of adaptive tempering schedules, we document the benefits of generalized data tempering for online estimation, and we demonstrate the ability of SMC algorithms to capture multimodal posteriors. Section 5 contains various pseudo-out-of-sample forecasting assessments for models that are estimated by SMC. Finally, Section 6 concludes. An Online Appendix provides further details on model specifications, prior distributions, computational aspects. It also contains additional empirical results.

2 Adaptive SMC Algorithms for Posterior Inference

SMC techniques to generate draws from posterior distributions of a static parameter θ are emerging as an attractive alternative to MCMC methods. SMC algorithms can be easily parallelized and, properly tuned, may produce more accurate approximations of posterior distributions than MCMC algorithms. Chopin (2002) showed how to adapt particle filtering techniques to conduct posterior inference for a static parameter vector. Textbook treatments of SMC algorithms are provided, for instance, by Liu (2001) and Cappé et al. (2005). This section reviews the standard SMC algorithm (Section 2.1), contrasts our generalized tempering approach with existing alternatives (Section 2.2), and finally describes our adaptive tempering algorithm (Section 2.3).

The first paper that applied SMC techniques to posterior inference in a small-scale DSGE models was Creal (2007). Herbst and Schorfheide (2014) develop the algorithm further,

provide some convergence results for an adaptive version of the algorithm building on the theoretical analysis of Chopin (2004), and show that a properly tailored SMC algorithm delivers more reliable posterior inference for large-scale DSGE models with a multimodal posterior than the widely used RWMH algorithm. Creal (2012) provides a recent survey of SMC applications in econometrics. Durham and Geweke (2014) show how to parallelize a flexible and self-tuning SMC algorithm for the estimation of time series models on graphical processing units (GPU). The remainder of this section draws heavily from the more detailed exposition in Herbst and Schorfheide (2014, 2015).

2.1 SMC Algorithms for Posterior Inference

SMC combines features of classic importance sampling and modern MCMC techniques. The starting point is the creation of a sequence of intermediate or bridge distributions $\{\pi_n(\theta)\}_{n=0}^{N_\phi}$ that converge to the target posterior distribution, i.e., $\pi_{N_\phi}(\theta) = \pi(\theta)$. At any stage the (intermediate) posterior distribution $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^i, W_n^i\}_{i=1}^N$ in the sense that the Monte Carlo average

$$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} W_n^i h(\theta^i) \xrightarrow{a.s.} \mathbb{E}_{\pi_n}[h(\theta_n)]. \tag{1}$$

as $N \longrightarrow \infty$, for each $n = 0, \dots N_\phi$. The bridge distributions are posterior distributions constructed from stage-n likelihood functions:

$$\pi_n(\theta) = \frac{p_n(Y|\theta)p(\theta)}{\int p_n(Y|\theta)p(\theta)d\theta} \tag{2}$$

with the convention that $p_0(Y|\theta) = 1$, i.e., the intial particles are drawn from the prior, and $p_{N_\phi}(Y|\theta) = p(Y|\theta)$. The actual form of the likelihood sequences depend on the tempering approach and will be discussed in Section 2.2 below. We adopt the convention that the weights W_n^i are normalized to average to one.

The SMC algorithm proceeds iteratively from $n = 0$ to $n = N_\phi$. Starting from stage $n-1$ particles $\{\theta_{n-1}^i, W_{n-1}^i\}_{i=1}^N$ each stage n of the algorithm targets the posterior π_n and consists of three steps: *correction*, that is, reweighting the stage $n-1$ particles to reflect the density in iteration n; *selection*, that is, eliminating a highly uneven distribution of particle weights (degeneracy) by resampling the particles; and *mutation*, that is, propagating the particles forward using a Markov transition kernel to adapt the particle values to the stage n bridge density.

Algorithm 1 (Generic SMC Algorithm)**.**

1. **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^i \overset{iid}{\sim} p(\theta)$ and $W_1^i = 1$, $i = 1, \ldots, N$.

2. **Recursion.** For $n = 1, \ldots, N_\phi$,

   (a) **Correction.** Reweight the particles from stage $n-1$ by defining the incremental weights

   $$\tilde{w}_n^i = \frac{p_n(Y|\theta_{n-1}^i)}{p_{n-1}(Y|\theta_{n-1}^i)} \tag{3}$$

   and the normalized weights

   $$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \ldots, N. \tag{4}$$

   (b) **Selection (Optional).** Resample the swarm of particles $\{\theta_{n-1}^i, \tilde{W}_n^i\}_{i=1}^N$ and denote resampled particles by $\{\hat{\theta}^i, W_n^i\}_{i=1}^N$, where $W_n^i = 1$ for all i.

   (c) **Mutation.** Propagate the particles $\{\hat{\theta}_i, W_n^i\}$ via N_{MH} steps of an MH algorithm with transition density $\theta_n^i \sim K_n(\theta_n|\hat{\theta}_n^i; \zeta_n)$ and stationary distribution $\pi_n(\theta)$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

   $$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^N h(\theta_n^i) W_n^i. \tag{5}$$

3. For $n = N_\phi$ ($\phi_{N_\phi} = 1$) the final importance sampling approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by:

   $$\bar{h}_{N_\phi,N} = \sum_{i=1}^N h(\theta_{N_\phi}^i) W_{N_\phi}^i. \tag{6}$$

The algorithm can be initialized with *iid* draws from the prior density $p(\theta)$, provided the prior density is proper, with the initial weights W_0^i set equal to one. The correction step is a classic importance sampling step, in which the particle weights are updated to reflect the stage n distribution $\pi_n(\theta)$. The selection step is optional. On the one hand, resampling adds noise to the Monte Carlo approximation, which is undesirable. On the other hand, it equalizes the particle weights, which increases the accuracy of subsequent importance sampling approximations. The decision of whether or not to resample is typically based on

a threshold rule for the variance of the particle weights. We can define an effective particle sample size as:

$$\widehat{ESS}_n = N \Big/ \left(\frac{1}{N} \sum_{i=1}^{N} (\tilde{W}_n^i)^2 \right) \tag{7}$$

and resample whenever $\widehat{ESS}_n$ falls below a threshold $\underline{N}$. An overview of specific resampling schemes is provided, for instance, in the books by Liu (2001) or Cappé et al. (2005) (and references cited therein). We are using systematic resampling in the applications below.

The mutation step changes the particle values. In the absence of the mutation step, the particle values would be restricted to the set of values drawn in the initial stage from the prior distribution. This would clearly be inefficient, because the prior distribution is typically a poor proposal distribution for the posterior in an importance sampling algorithm. As the algorithm cycles through the N_ϕ stages, the particle values successively adapt to the shape of the posterior distribution. This is the key difference between SMC and classic importance sampling. The transition kernel $K_n(\theta_n|\hat{\theta}_n;\zeta_n)$ has the following invariance property:

$$\pi_n(\theta_n) = \int K_n(\theta_n|\hat{\theta}_n;\zeta_n)\pi_n(\hat{\theta}_n)d\hat{\theta}_n. \tag{8}$$

Thus, if $\hat{\theta}_n^i$ is a draw from π_n, then so is θ_n^i. The mutation step can be implemented by using one or more steps of a Metropolis-Hastings (MH) algorithm. The probability of mutating the particles can be increased by blocking the elements of the parameter vector θ or by iterating the MH algorithm over multiple steps. The vector ζ_n summarizes the tuning parameters of the MH algorithm.

The SMC algorithm produces as a by-product an approximation of the marginal likelihood. Note that

$$\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_n^i \tilde{W}_{n-1}^i \approx \int \frac{p_n(Y|\theta)}{p_{n-1}(Y|\theta)} \left[\frac{p_{n-1}(Y|\theta)p(\theta)}{\int p_{n-1}(Y|\theta)p(\theta)d\theta} \right] d\theta = \frac{\int p_n(Y|\theta)p(\theta)d\theta}{\int p_{n-1}(Y|\theta)p(\theta)d\theta}. \tag{9}$$

Thus, it can be shown that the approximation

$$\hat{p}(Y) = \prod_{n=1}^{N_\phi} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{w}_n^i W_{n-1}^i \right) \tag{10}$$

converges almost surely to $p(Y)$ as the number of particles $N \longrightarrow \infty$.

2.2 Likelihood, Data, and Generalized Tempering

The stage-n likelihood functions are generated in different ways. Under likelihood tempering, one takes power transformations of the entire likelihood function:

$$p_n(Y|\theta) = [p(Y|\theta)]^{\phi_n}, \quad \phi_n \uparrow 1, \tag{11}$$

The advantage of likelihood tempering is that one can make, through the choice of ϕ_n, consecutive posteriors arbitrarily "close" one another. Under data tempering, sets of observations are gradually added to the likelihood function, that is,

$$p_n(Y|\theta) = p(y_{1:\lfloor \phi_n T \rfloor}), \quad \phi_n \uparrow 1, \tag{12}$$

where $\lfloor x \rfloor$ is the largest integer that is less or equal to x. Data tempering is particularly attractive in time series applications. But because individual observations are not divisible, the data tempering approach is less flexible.

Our approach generalizes both likelihood and data tempering as follows. Imagine one has draws from the posterior

$$\tilde{\pi}(\theta) \propto \tilde{p}(\tilde{Y}|\theta)p(\theta), \tag{13}$$

where the posterior $\tilde{\pi}(\theta)$ differs from the posterior $\pi(\theta)$ because either the sample (Y versus $\tilde{Y}$), or the model ($p(Y|\theta)$ versus $\tilde{p}(\tilde{Y}|\theta)$), or both, are different.[2] We define the stage-n likelihood function as:

$$p_n(Y|\theta) = [p(Y|\theta)]^{\phi_n}[\tilde{p}(\tilde{Y}|\theta)]^{1-\phi_n}, \quad \phi_n \uparrow 1. \tag{14}$$

First, if one sets $\tilde{p}(\cdot) = 1$, then (14) is identical to likelihood tempering. Second, if one sets $\tilde{p}(\cdot) = p(\cdot)$, $Y = y_{1:\lfloor \phi_m T \rfloor}$, and $\tilde{Y} = y_{1:\lfloor \phi_{m-1} T \rfloor}$, where $\phi_m \uparrow 1$ is the same sequence as in (12), then our approach generalizes data tempering by allowing for a gradual transition between $y_{1:\lfloor \phi_{m-1} T \rfloor}$ and $y_{1:\lfloor \phi_m T \rfloor}$. This may be important if the additional sample $y_{\lfloor \phi_{m-1} T \rfloor + 1:\lfloor \phi_m T \rfloor}$ substantially affects the likelihood (e.g., $y_{\lfloor \phi_{m-1} T \rfloor + 1:\lfloor \phi_m T \rfloor}$ includes the Great Recession).

Third, by allowing Y to differ from $\tilde{Y}$ we can accommodate data revisions between time $\lfloor \phi_{m-1} T \rfloor$ and $\lfloor \phi_m T \rfloor$. For online estimation conducted in central banks, one can use the most recent estimation to jump-start a new estimation on revised data, without starting from scratch. Finally, by allowing $p(\cdot)$ and $\tilde{p}(\cdot)$ to differ, one can transition between the

[2] It is straightforward to generalize our approach to also encompass differences in the prior.

posterior distribution of two models that share the same parameters. We will evaluate the accuracy of the generalized tempering approach in Section 4 and use it in the real-time forecast evaluation of Section 5.

2.3 Adaptive Algorithms

The implementation of the SMC algorithm requires the choice of several tuning constants. First and foremost, the user has to choose the number of particles N. As shown in Chopin (2004), Monte Carlo averages computed from the output of the SMC algorithm satisfy a CLT as the number of particles increases to infinity. This means that the variance of the Monte Carlo approximation decreases at the rate $1/N$. Second, the user has to determine the tempering schedule ϕ_n and the number of bridge distributions N_ϕ. Third, the threshold level $\underline{N}$ for $\widehat{ESS}_n$ needs to be set to determine whether the resampling step should be executed in iteration n. Finally, the implementation of the mutation step requires the choice of the number of MH steps, N_{MH}, the number of blocks into which the parameter vector θ is partitioned, N_{blocks}, and the parameters ζ_n that control the Markov transition kernel $K_n(\theta_n|\hat{\theta}_n^i;\zeta_n)$.

Our implementation of the algorithm starts from a choice of N, $\underline{N}$, N_{MH}, and N_{blocks}. The remaining features of the Algorithm are determined adaptively. As in Herbst and Schorfheide (2015), we use a random-walk Metropolis Hastings (RWMH) algorithm to implement the mutation step. The proposal density takes the form $N(\hat{\theta}^i, c_n^2, \tilde{\Sigma}_n)$. The scaling constant c and the covariance matrix $\tilde{\Sigma}_n$ can be easily chosen adaptively; see Herbst and Schorfheide (2015, Algorithm 10). Based on the MH rejection frequency, c can be adjusted to achieve a target rejection rate of approximately 25-40%. For $\tilde{\Sigma}_n$ one can use an approximation of the posterior covariance matrix computed at the end of the stage n correction step.

In the current paper, we will focus on the adaptive choice of the tempering schedule, building on work by Jasra et al. (2011), Del Moral et al. (2012), Schäfer and Chopin (2013), Geweke and Frischknecht (2014), and Zhou et al. (2015). The key idea is to choose ϕ_n to target a desired level $\widehat{ESS}_n^*$. Roughly, the closer the desired $\widehat{ESS}_n^*$ to the previous $\widehat{ESS}_{n-1}$, the smaller the increment $\phi_n - \phi_{n-1}$ and the information increase in the likelihood function. In order to formally describe the choice of ϕ_n define the functions:

$$w^i(\phi) = [p(Y|\theta_{n-1}^i)]^{\phi-\phi_{n-1}}, \quad W^i(\phi) = \frac{w^i(\phi)W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N w^i(\phi)W_{n-1}^i}, \quad \widehat{ESS}(\phi) = N/\left(\frac{1}{N}\sum_{i=1}^N(\tilde{W}_n^i(\phi))^2\right)$$

We will choose ϕ to target a desired level of ESS:

$$f(\phi) = \widehat{ESS}(\phi) - \alpha\widehat{ESS}_{n-1} = 0, \tag{15}$$

where α is a tuning constant that captures the tolerated deterioration in ESS. The algorithm can be summarized as follows:

Algorithm 2 (Adaptive Tempering Schedule)**.**

1. If $f(\phi) \geq 0$, then set $\phi_n = 1$.
2. If $f(\phi) < 0$, let $\phi_n \in (\phi_{n-1}, 1)$ be the smallest value of ϕ_n such that $f(\phi_n) = 0$.

For the tempering schedule to be "well-formed" it must be monotonically increasing. This is guaranteed by construction of the adaptive schedule for the following reason. Suppose that $\phi_{n-1} < 1$. First, note that $f(\phi_{n-1}) = (1-\alpha)\widehat{ESS}_{n-1} > 0$. Second, if $f(1) \geq 0$, we set $\phi_n = 1 > \phi_{n-1}$ and the algorithm terminates. Alternatively, if $f(1) < 0$, then by continuity of $f(\phi)$ and the compactness of the interval $[\phi_{n-1}, 1]$, there exists at least one root of $f(\phi) = 0$. We define ϕ_n to be the smallest one.[3]

3 DSGE Models

In the subsequent applications we consider three DSGE models. The precise specifications of these models, including their linearized equilibrium conditions, measurement equations, and prior distributions, are provided in Section B of the Online Appendix. The first model is a small-scale New Keynesian DSGE model that has been widely studied in the literature (see Woodford (2003) or Gali (2008) for textbook treatments). The particular specification used in this paper is based on (An and Schorfheide, 2007; henceforth, AS). The model economy consists of final goods producing firms, intermediate goods producing firms, households, a central bank, and a fiscal authority. Labor is the only factor of production. Intermediate goods producers act as monopolistic competitors and face downward sloping demand curves for their products. They face quadratic costs for adjusting their nominal prices, which generates price rigidity and real effects of unanticipated changes in monetary policy in the model. The model solution can be reduced to three key equations: a consumption Euler

[3] A refined version of Algorithm 2 that addresses some numerical challenges in finding the root of $f(\phi)$ is provided in Section A.2 of the Online Appendix.

equation, a New Keynesian Phillips curve, and a monetary policy rule. Fluctuations are driven by three types of exogenous shocks and the model is estimated based on output growth, inflation, and federal funds rate data. We will subsequently refer to this model as the AS model.

The second model is the Smets and Wouters (2007) model, henceforth SW, which is based on earlier work by Christiano et al. (2005) and Smets and Wouters (2003), and is the prototypical medium-scale New Keynesian model. In the SW model, capital is a factor of intermediate goods production and in addition to price stickiness, the model also features nominal wage stickiness. In order to generate a richer autocorrelation structure, the model includes investment adjustment costs, habit formation in consumption, and partial dynamic indexation of prices and wages to lagged values. The SW model is estimated using the following seven macroeconomic time series: output growth, consumption growth, investment growth, real wage growth, hours worked, inflation, and the federal funds rate.

Whenever we want to produce forecasts that are conditional on (survey) expectations of future interest rates (which is important whenever we need to capture the effective lower bound on nominal interest rates from 2009-2015, or forward guidance), we add these expectations to the vector of observables and anticipated monetary policy shocks in the interest rate feedback rules, following Del Negro and Schorfheide (2013).

In Section 5, we also consider an extension of the SW model based on Del Negro and Schorfheide (2013) that includes a time-varying target inflation rate to capture low frequency movements of inflation. To anchor the estimates of the target inflation rate, we include long-run inflation expectations into the set of observables. We refer to this model as SWπ.

The third DSGE model, SWFF, is obtained by adding financial frictions to the SWπ model and builds on work by Bernanke et al. (1999b), Christiano et al. (2003), De Graeve (2008), and Christiano et al. (2014).[4] In this DSGE model, banks collect deposits from households and lend to entrepreneurs who use these funds as well as their own wealth to acquire physical capital, which is rented to intermediate goods producers. Entrepreneurs are subject to idiosyncratic disturbances that affect their ability to manage capital. Their revenue may thus be too low to pay back the bank loans. Banks protect themselves against default risk by pooling all loans and charging a spread over the deposit rate. This spread varies exogenously due to changes in the riskiness of entrepreneurs' projects and endogenously

[4]SWFF was also introduced in Del Negro and Schorfheide (2013). Its projections for the Great Recession are studied in Del Negro et al. (2015).

Table 1: Configuration of SMC Algorithm for Different Models

	AS	SW
Number of particles	$N = 3,000$	$N = 12,000$
Fixed tempering schedule	$N_\phi = 200$, $\lambda = 2$	$N_\phi = 500$, $\lambda = 2.1$
Mutation	$N_{MH} = 1$, $N_{blocks} = 1$	$N_{MH} = 1$, $N_{blocks} = 3$
Selection/Resampling	$\underline{N} = N/2$	$\underline{N} = N/2$

as a function of the entrepreneurs' leverage. The estimation of the SWFF model includes an additional time series for spreads.

4 SMC Estimation at Work

We now illustrate various dimensions of the performance of the SMC algorithm. Section 4.1 documents the shape of the adaptive tempering schedule as well as speed-versus-accuracy trade-offs when tuning the adaptive tempering. Section 4.2 uses generalized tempering for the online estimation of DSGE models and document its runtime advantages. Finally, in Section 4.3 we show that the SMC algorithm is able to reveal multimodal features of DSGE model posteriors.

4.1 Adaptive Likelihood Tempering

We described in Section 2.3 how the tempering schedule for the SMC algorithm can be generated adaptively. The tuning parameter α controls the desired level of reduction in ESS in (15). The closer α is to one, the smaller the desired ESS reduction, and therefore the smaller the change in the tempering parameter and the larger the number of tempering steps. We will explore the shape of the adaptive tempering schedule generated by Algorithm 2, the runtime of SMC Algorithm 1, and the accuracy of the resulting Monte Carlo approximation as a function of α. Rather than reporting results for individual DSGE model parameters, we consider the standard deviation of the log marginal data density (MDD) defined in (10), computed across multiple runs of the SMC algorithm, as a measure of accuracy (and precision) of the Monte Carlo approximation.

Table 2: AS Model: Fixed and Adaptive Tempering Schedules

	Fixed	α =0.90	α =0.95	α =0.97	α =0.98
Mean log(MDD)	-1032.55	-1034.13	-1032.46	-1032.04	-1031.92
StdD log(MDD)	0.73	1.54	0.61	0.29	0.21
Schedule Length	200.00	111.92	218.77	350.04	505.79
Resamples	14.61	15.34	15.02	15.00	14.00
Runtime [Min]	2.00	0.83	1.60	2.56	3.70

Notes: Results are based on $N_{run} = 200$ runs of the SMC algorithm. We report averages across runs for the runtime, schedule length, and number of resampling steps.

We consider the AS and SW models, estimated based on data from 1966:Q4 to 2016:Q3.[5] In addition to the adaptive tempering schedule, we also consider a fixed tempering schedule of the form

$$\phi_n = \left(\frac{n}{N_\phi}\right)^\lambda. \tag{16}$$

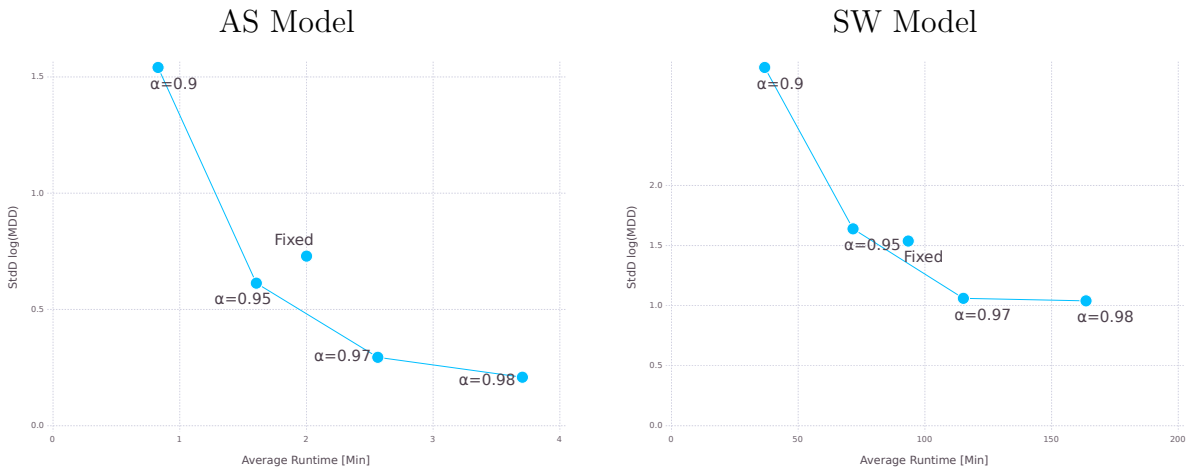
This schedule has been used in the SMC applications in Herbst and Schorfheide (2014) and Herbst and Schorfheide (2015). The user-specified tuning parameters for the SMC algorithm are summarized in Table 1.

Results for the AS model based on adaptive likelihood tempering, see (11), are summarized in Table 2. We also report results for the fixed tempering schedule (16) in the second column of the table. The remaining columns show results for the adaptive schedule with different choices of the tolerated ESS reduction α. The adaptive schedule length is increasing from approximately 112 stages for $\alpha = 0.90$ to 506 stages for $\alpha = 0.98$. As mentioned above, the closer α is to one, the smaller is the increase in ϕ_n. This leads to a large number of stages which, in turn, increases the precision of the Monte Carlo approximation. The standard deviation of $\log \hat{p}(Y)$ is 1.54 for $\alpha = 0.9$ and 0.21 for $\alpha = 0.98$.

The runtime of the algorithm increases approximately linearly in the number of stages. The number of times that the selection step is executed is approximately constant as a function of α. However, because N_ϕ is increasing, the fraction of stages at which the particles are resampled decreases from 14% to 3%. The mean of the log MDD is increasing as the precision increases. This is the result of Jensen's inequality. MDD approximations obtained

[5] For the estimation of both models, we use a pre-sample from 1965:Q4 to 1966:Q3.

Figure 1: Trade-Off Between Runtime and Accuracy



Notes: AS results are based on $N_{run} = 200$ and SW results are based on $N_{run} = 100$ runs of the SMC algorithm.

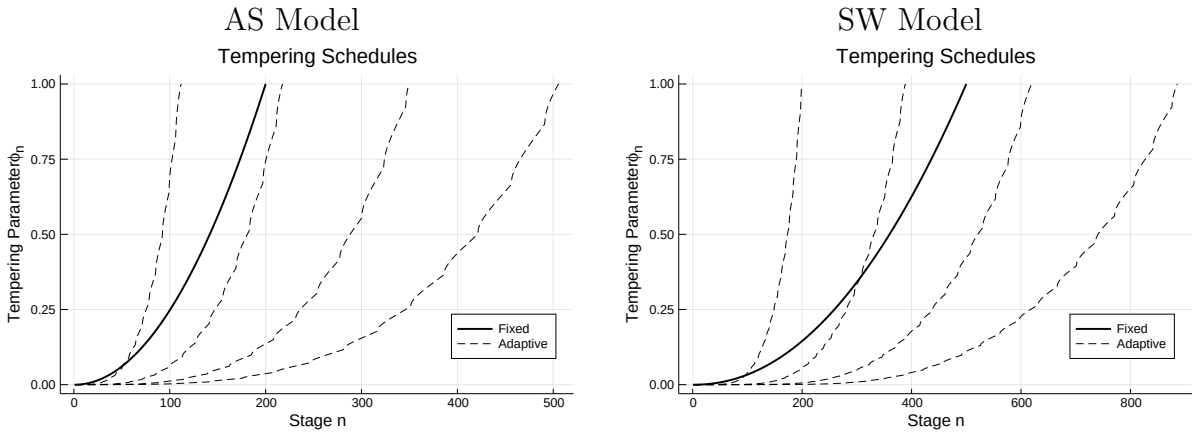
from SMC algorithms tend to be unbiased, which means that log MDD approximations exhibit a downward bias.

The left panel of Figure 1 depicts the time-accuracy curve for the AS model. This curve is convex, indicating that increasing the schedule length from 112 to 219 stages generated a drastic increase in accuracy, whereas implicitly increasing N_ϕ from 350 to 506 had a relatively small effect. Connecting the $\alpha = 0.90$ and $\alpha = 0.95$ dots, it appears that the fixed schedule is slightly inefficient in terms of trading off accuracy on runtime.

In addition to the AS model, we also evaluate the posterior of the SW model using the SMC algorithm with adaptive likelihood tempering. Because the dimension of the parameter space of the SW model is much larger than that of the AS model, we are using more particles and multiple blocks in the mutation step when approximating its posterior (see Table 1). The right panel of Figure 1 depicts the time-accuracy curve for the SW model. For a given α, the log MDD approximation for SW is less accurate than for AS, which is consistent with the SW model having more parameters that need to be integrated out and a less regular likelihood surface. Similar to the results for the AS model, the time-accuracy curve is also convex for the SW model, and the point corresponding to the fixed-schedule lies slightly within the curve.

In Figure 2 we plot the fixed and adaptive tempering schedules for both models. All of the adaptive schedules are convex. Very little information, less than under the fixed

Figure 2: Tempering Schedules



Notes: The figure depicts (pointwise) median ϕ_n values across $N_{run} = 200$ for AS and $N_{run} = 100$ for SW. The solid lines represent the fixed schedule, parameterized according to Table 1. The dashed lines represent a range of adaptive schedules: $\alpha = 0.90, 0.95, 0.97, 0.98$.

schedule, is added to the likelihood function in the early stages, whereas a large amount of information is added during the later stages. This is consistent with the findings in Herbst and Schorfheide (2014, 2015) who examined the performance of the SMC approximations under the fixed schedule (16) for various values of λ.

4.2 Generalized Tempering for Online Estimation

To provide a timely assessment of economic conditions and to produce accurate forecasts and policy projections, econometric modelers at central banks have to re-estimate their DSGE models regularly, e.g., once a quarter or once a year. A key impediment to online estimation of DSGE models with the RWMH algorithm is that any amendment to the previously-used dataset requires a full re-estimation of the DSGE model, which can be quite time-consuming and often requires supervision.[6] SMC algorithms that are based on data tempering, on the other hand, allow for efficient online estimation of DSGE models. This online estimation entails combining the adaptive tempering schedule in Algorithm 2 with the generalized tempering in (14). As mentioned above, our algorithm is also amenable to data revisions.

[6]This is particularly true when the proposal covariance matrix is constructed from the Hessian of the log likelihood function evaluated at the global mode, which can be very difficult to find.

Scenario 1. In the following illustration, we partition the sample into two subsamples: $t = 1, \ldots, T_1$ and $t = T_1 + 1, \ldots, T$, and allow for data revisions by the statistical agencies between periods $T_1 + 1$ and T. We assume that the second part of the sample becomes available after the model has been estimated on the first part of the sample using the data vintage available at the time, $\tilde{y}_{1:T_1}$. Thus, in period T we already have a swarm of particles $\{\theta^i_{T_1}, W^i_{T_1}\}_{i=1}^N$ that approximates the posterior

$$p(\theta|\tilde{y}_{1:T_1}) \propto p(\tilde{y}_{1:T_1}|\theta)p(\theta).$$

Following (14) with $Y = y_{1:T}$ and $\tilde{Y} = \tilde{y}_{1:T_1}$, we define the stage (n) posterior as

$$\pi_n(\theta) = \frac{p(y_{1:T}|\theta)^{\phi_n} p(\tilde{y}_{1:T_1}|\theta)^{1-\phi_n} p(\theta)}{\int p(y_{1:T}|\theta)^{\phi_n} p(\tilde{y}_{1:T_1}|\theta)^{1-\phi_n} p(\theta) d\theta}.$$

We distinguish notationally between y and $\tilde{y}$ because some observations in the $t = 1, \ldots, T_1$ sample may have been revised. The incremental weights are given by

$$\tilde{w}^i_n(\theta) = p(y_{1:T}|\theta)^{\phi_n - \phi_{n-1}} p(\tilde{y}_{1:T_1}|\theta)^{\phi_{n-1} - \phi_n}$$

and it can be verified that

$$\frac{1}{N}\sum_{i=1}^N \tilde{w}^i_n W^i_{n-1} \approx \frac{\int p(y_{1:T}|\theta)^{\phi_n} p(\tilde{y}_{1:T_1}|\theta)^{1-\phi_n} p(\theta) d\theta}{\int p(y_{1:T}|\theta)^{\phi_{n-1}} p(\tilde{y}_{1:T_1}|\theta)^{1-\phi_{n-1}} p(\theta) d\theta}. \tag{17}$$

Now define the conditional marginal data density (CMDD)

$$\text{CMDD}_{2|1} = \prod_{n=1}^{N_\phi} \left(\frac{1}{N}\sum_{i=1}^N \tilde{w}^i_{(n)} W^i_{(n-1)} \right) \tag{18}$$

with the understanding that $W^i_{(0)} = W_{T_1}$. Because the product of the terms in (17) simplify, and because $\phi_{N_\phi} = 1$ and $\phi_1 = 0$, we obtain:

$$\text{CMDD}_{2|1} \approx \frac{\int p(y_{1:T}|\theta)p(\theta)d\theta}{\int p(\tilde{y}_{1:T_1}|\theta)p(\theta)d\theta} = \frac{p(y_{1:T})}{p(\tilde{y}_{1:T_1})}. \tag{19}$$

Note that in the special case of no data revisions ($\tilde{y}_{1:T_1} = y_{1:T_1}$) the expression simplifies to $\text{CMDD}_{2|1} \approx p(y_{T_1+1:T}|y_{1:T_1})$. We consider this case in our simulations below.

We assume that the DSGE model has been estimated using likelihood tempering based on the sample $y_{1:T_1}$, where $t = 1$ corresponds to 1966:Q4 and $t = T_1$ corresponds to 2007:Q1. The second sample, $y_{T_1+1:T}$, starts in 2007:Q2 and ends in 2016:Q3.[7] We now consider

[7] We use a recent data vintage and abstract from data revisions in this exercise.

Table 3: AS Model: Generalized Tempering

	α =0.90	α =0.95	α =0.97	α =0.98
Mean log(MDD)	-1034.12	-1032.48	-1032.10	-1031.95
StdD log(MDD)	1.31	0.62	0.35	0.23
Schedule Length	24.18	47.20	75.56	106.45
Runtime [Min]	0.25	0.46	0.73	1.02

Notes: Results are based on $N_{run} = 200$ runs of the SMC algorithm, starting from particles that represent $p(\theta|Y_{1:T_1})$. We report averages across runs for the runtime, schedule length, and number of resampling steps.

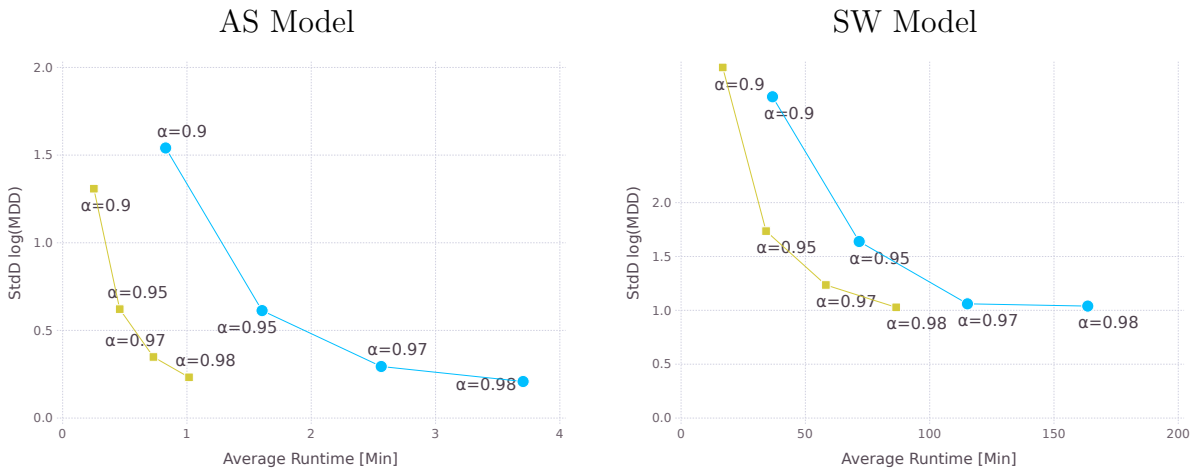
two ways of estimating the log MDD $\log p(y_{1:T})$. Under full-sample estimation, we ignore the existing estimate based on $y_{1:T_1}$ and use likelihood tempering based on the full-sample likelihood $p(y_{1:T}|\theta)$. Under generalized tempering, we start from the existing posterior based on $y_{1:T_1}$ and use generalized data tempering to compute $CMDD_{2|1}$ in (18). We then calculate $\log p(y_{1:T_1}) + \log CMDD_{2|1}$.

Note that our choice of the sample split arguably stacks the cards against the generalized tempering approach. This is because the second period is quite different from the first, as it includes the Great Recession, the effective lower bound constraint on the nominal interest rate, and unconventional monetary policy interventions such as large-scale asset purchases and forward guidance.

We begin with the following numerical illustration for the AS model. For each of the $N_{run} = 200$ SMC runs, we first generate the estimate of $\log p(y_{1:T_1}|\theta)$ by likelihood tempering and then continue with generalized tempering to obtain $\log CMDD_{2|1}$. The two numbers are added to obtain an approximation of $\log p(y_{1:T})$ that can be compared to the results from the full-sample estimation reported in Table 2. The results from the generalized tempering approach are reported in Table 3. Comparing the entries in both tables, note that the mean and standard deviations (across runs) of the log MDDs are essentially the same. The main difference is that generalized tempering reduces the schedule length and the runtime by a factor of roughly 2/3 because it starts from the posterior distribution $p(\theta|y_{1:T_1})$.

In Figure 3 we provide scatter plots of average runtime versus the standard deviation of the log MDD for the AS model and the SW model. The blue circles correspond to full sample estimation and are identical to the blue circles in Figure 1. The yellow squares correspond to generalized tempering. As in Table 3, the runtime does not reflect the time it

Figure 3: Trade-Off Between Runtime and Accuracy



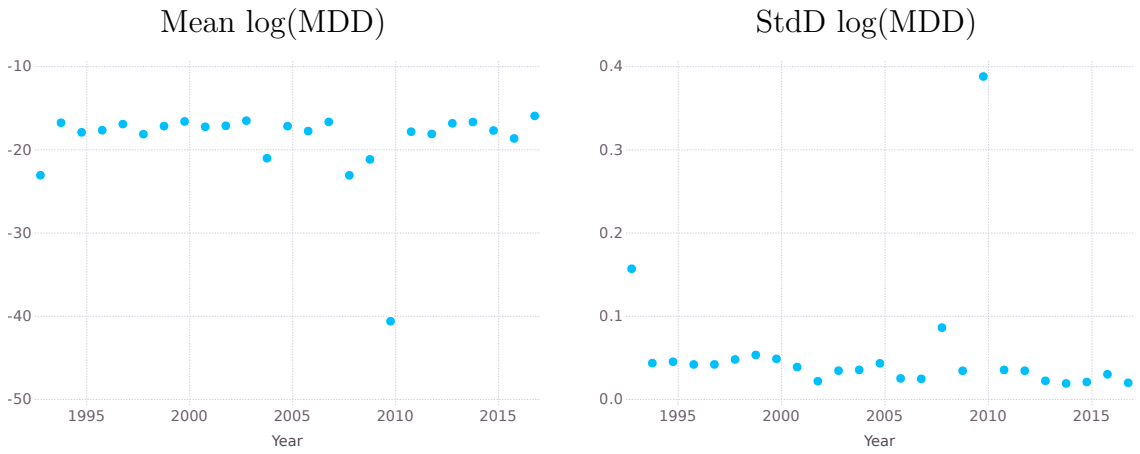
Notes: AS results are based on $N_{run} = 200$ and SW results are based on $N_{run} = 100$ runs of the SMC algorithm. Yellow squares correspond to generalized tempering and blue circles correspond to full sample estimation.

took to compute $p(\theta|y_{1:T_1})$ because the premise of the analysis is that this posterior has been computed in the past and is available to the user at the time when the estimation sample is extended by the observations $y_{T_1+1:T}$. For both models, the accuracy of the log MDD approximation remains roughly the same for a given α when using generalized tempering, but the reduction in runtime is substantial. To put it differently, generalized tempering shifts the SMC time-accuracy curve to the left, which of course is the desired outcome.

Scenario 2. Rather than adding observations in a single block, we now add four observations at a time. This corresponds to a setting in which the DSGE model is re-estimated once a year, which is a reasonable frequency in central bank environments. More formally, we partition the sample into the subsamples $y_{1:T_1}$, $y_{T_1+1:T_2}$, $y_{T_2+1:T_3}$, ..., where $T_s - T_{s-1} = 4$. After having approximated the posterior based on observations $y_{1:T_1}$, we use generalized tempering to compute the sequence of densities $p(y_{1:T_s})$ for $s = 2, \ldots, S$. At each step s we initialize the SMC algorithm with the particles that represent the posterior $p(\theta|y_{1:T_{s-1}})$. To assess the accuracy of this computation, we repeat it $N_{run} = 200$ times. We fix the tuning parameter for the adaptive construction of the tempering schedule at $\alpha = 0.95$. The first sample, $y_{1:T_1}$ ranges from 1966:Q4 to 1991:Q3, and the last sample ends in 2016:Q3.

Figure 4 depicts the time series of the mean and the standard deviation of the log MDD

Figure 4: AS Model: Log MDD Increments $\log(p(y_{T_{s-1}+1:T_s}|y_{1:T_{s-1}}))$



Notes: Results are based on $N_{run} = 200$ runs of the SMC algorithm with $\alpha = 0.95$.

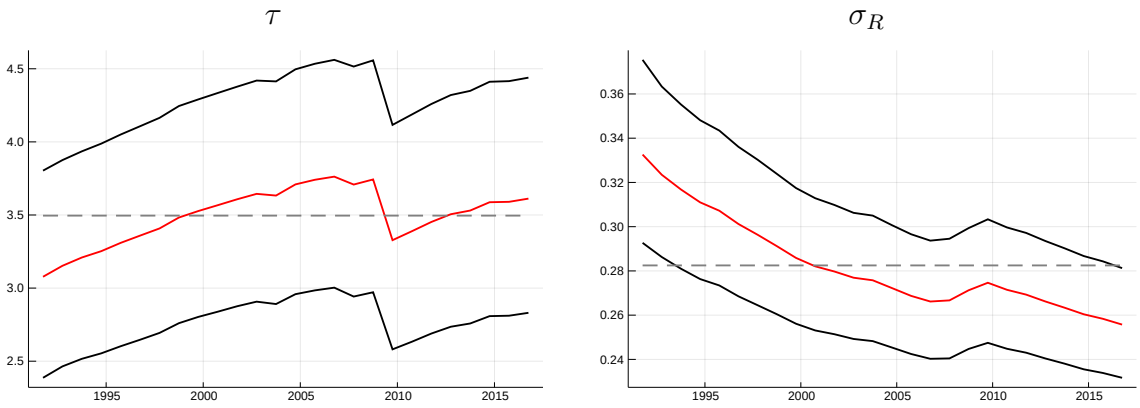
increments

$$\log \hat{p}(y_{T_{s-1}+1:T_s}|y_{1:T_{s-1}}) = \log \hat{p}(y_{1:T_s}) - \log \hat{p}(y_{1:T_{s-1}})$$

across the 200 SMC runs. The median (across time) of the average (across repetitions) log MDD increment is -17.6. The median standard deviation of the log MDD increments is 0.04, and the median of the average run time is 0.1 minutes, or 6 seconds. The largest deviation from these median values occurred during the Great Recession when we added the 2009:Q4 to 2010:Q3 observations to the sample. During this period the log MDD increment was only -40.6 and the standard deviation jumped up to 0.4 because four observations lead to a substantial shift in the posterior distribution. In this period, the run-time of the SMC algorithm increased to 0.6 minutes, or 36 seconds.

Figure 5 depicts the evolution of posterior means and coverage intervals for two parameters, τ and σ_R. The τ sequence exhibits a clear blip in 2009, which coincides with the increased run time of the algorithm. Most of the posteriors exhibit drifts rather than sharp jumps and the time-variation in the posterior mean is generally small compared to the overall uncertainty captured by the coverage bands. Overall, generalized tempering provides a framework online estimation of DSGE models that is substantially more acccurate than simple data tempering, at little additional computational cost, even when the additional data substantially changes the posterior. Importantly, generalized tempering can also seemlessly handle the data revisions inherent in macroeconomic time series.

Figure 5: AS Model: Evolution of Posterior Means and Coverage Bands



Notes: Sequence of posterior means (red line) and 90% coverage bands black lines. The dashed line indicates the temporal average of the posterior means. We use $\alpha = 0.95$ for the SMC algorithm.

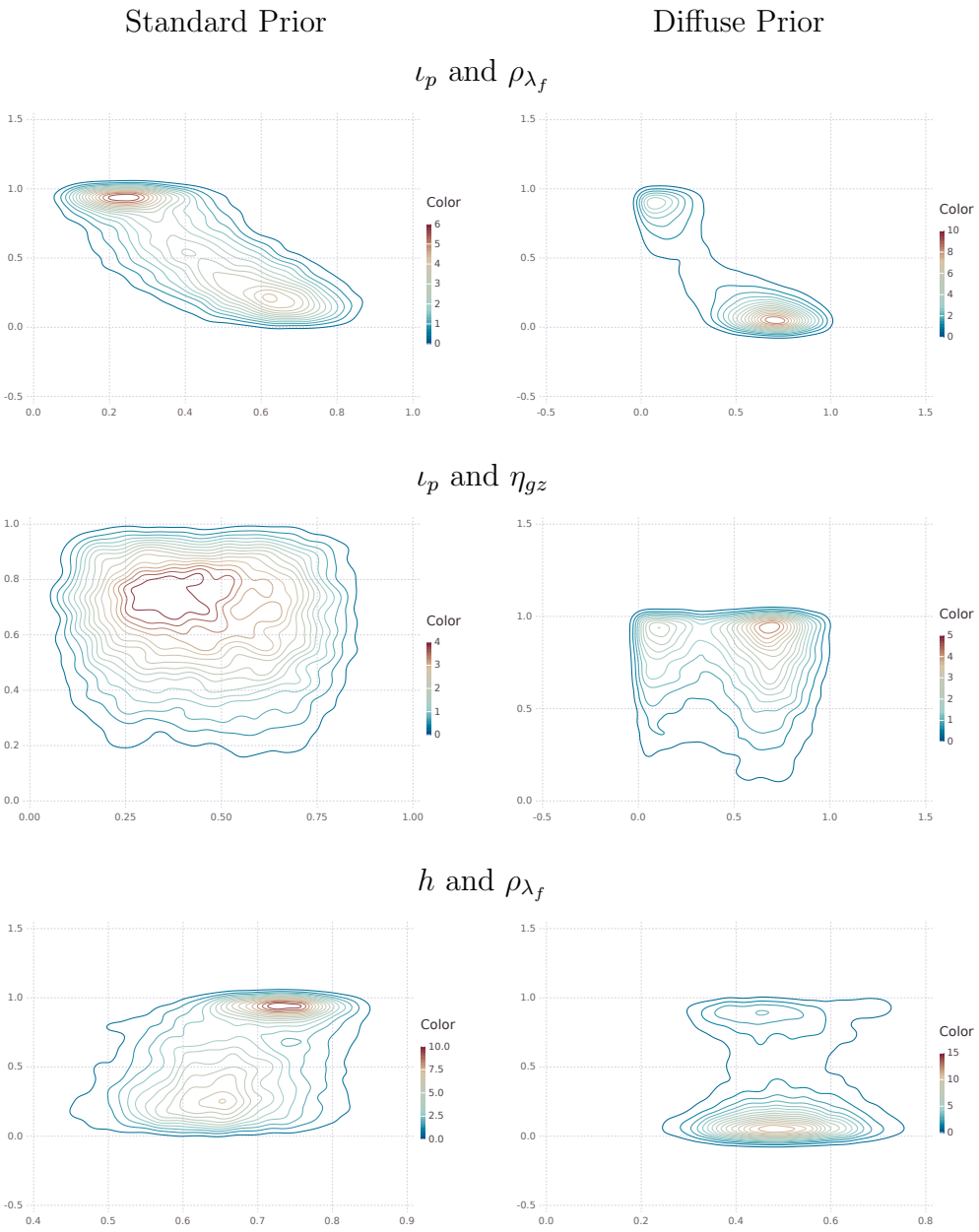
4.3 Exploring Multimodal Posteriors

An important advantage of SMC samplers over standard RWMH samplers is their ability to characterize multimodal posterior distributions. Multimodality may arise because the data are not informative enough to be able to disentangle internal versus external propagation mechanisms, e.g., Calvo price and wage stickiness and persistence of exogenous price and wage markup shocks. Measuring the relative contribution of various structural shocks included in a DSGE model is also often difficult based on the sample information. Herbst and Schorfheide (2014) provided an example of a multimodal posterior distribution obtained in a SW model that is estimated under a diffuse prior distribution. Below, we document that a multimodal posterior may also arise if the SW model is estimated on a shorter sample with the informative prior used by Smets and Wouters (2007) originally. Capturing this bimodality correctly will be important for the accurate computation of predictive densities that are generated as part of real-time forecast applications in Section 5.

Figure 6 depicts various marginal bivariate posterior densities for parameters of the SW model estimated based on a sample from 1960:Q1 to 1991:Q3.[8] The plots in the left column of the figure corresponds to the "standard" prior for the SW model. The joint posteriors for the parameters ι_p and ρ_{λ_f} (weight on the backward-looking component in the New Keynesian

[8]For these results we match the sample used in Section 5, where we discuss the predictive ability of the various DSGE models. See footnote 11 for a description of why the samples used in section 4 and 5 differ slightly.

Figure 6: SW Model: Posterior Contours for Selected Parameter Pairs



Notes: Estimation sample is 1960:Q1 to 1991:Q3. We use $\alpha = 0.98$ for the SMC algorithm.

price Phillips curve and persistence of price markup shock) on the one hand, and h and ρ_{λ_f} (the former determines the degree of habit formation in consumption) on the other hand, exhibit clear bimodal features, albeit one mode dominates the other. For the parameter pair ι_p and η_{gz} (loading government spending on technology shock innovations), the bimodality

is less pronounced.

The right column of Figure 6 shows posteriors for the same estimation sample but the "diffuse" prior of Herbst and Schorfheide (2014). This prior is obtained by increasing the standard deviations for parameters with marginal Normal and Gamma distributions by a factor of three and using uniform priors for parameters defined on the unit interval.[9] Under the diffuse prior, the multimodal shapes of the bivariate posteriors are more pronounced and for the first two parameter pairs both modes are associated with approximately the same probability mass. Thus, using a sampler that correctly captures the non-elliptical features of the posterior is essential for valid Bayesian inference and allows researchers to estimate DSGE models under less informative priors that have been traditionally used in the literature. In Section 5.3 we will examine the forecasting performance of the SW model under the diffuse prior.

5 Predictive Density Evaluations

In this section we compare the forecast performance of two DSGE models, one without (SW) and one with (SWFF) financial frictions, based on real-time data. The models are estimated recursively, using the generalized tempering approach described in Section 4.2. Our results complement existing work on the evaluation of DSGE model forecasts, e.g., Adolfson et al. (2007), Edge and Gürkaynak (2010), Christoffel et al. (2011), Del Negro and Schorfheide (2013), Wolters (2015), Diebold et al. (2017), Cai et al. (forthcoming). Rather than reporting root-mean-squared errors (RMSEs), we will focus on log predictive density scores, which are a widely-used criterion to compare density forecasts across models; see, for instance, Del Negro et al. (2016) and Warne et al. (2017) in the context of DSGE model forecasting.

Section 5.1 provides details regarding the construction of the real-time dataset and of the timing and *ex post* evaluation of the forecast. For the sake of comparison with the literature, Section 5.2 evaluates the predictive ability of the DSGE models using what we have referred to as the "standard" prior, that is, the prior used in previous work (which for most parameters amounts to the prior used in Smets and Wouters, 2007). DSGE models estimated with

[9] The prior on the shocks standard deviations is the same for the diffuse and standard prior, as this prior is already very loose under the standard specification (an inverse Gamma with only 2 degrees of freedom). A summary is provided in Table A-2 of the Appendix.

Table 4: Data series used in each model

Variable	SW	SWπ	SWFF
GDP growth	X	X	X
Consumption growth	X	X	X
Investment growth	X	X	X
Real wage growth	X	X	X
Hours worked	X	X	X
GDP deflator inflation	X	X	X
Federal funds rate	X	X	X
10y inflation expectations		X	X
Spread			X

Bayesian methods have often been criticized however for using overly informative priors; see, for instance Romer (2016)'s critique of the Smets and Wouters (2007) priors. It is therefore a legitimate question to ask whether loosening the prior substantially changes the forecasting performance of these models. This is the question addressed in Section 5.3. It is clear from the results in Section 4.3 that, given the severe multimodalities present in the posterior distribution under the diffuse prior, SMC techniques are necessary to accurately characterize predictive densities.

5.1 Real-Time Dataset and DSGE Forecasting Setup

In this section, we first discuss the data series used for the DSGE model estimation, summarized in Table 4, and the process of constructing a real-time dataset, which follows the approach of Del Negro and Schorfheide (2013, Section 4.1) and Edge and Gürkaynak (2010). Subsequently, we will discuss the DSGE forecast setup, which mirrors that of Cai et al. (forthcoming).

Data on nominal GDP (GDP), the GDP deflator (GDPDEF), nominal personal consumption expenditures (PCEC), and nominal fixed private investment (FPI) are produced at a quarterly frequency by the Bureau of Economic Analysis, and are included in the National Income and Product Accounts (NIPA). Average weekly hours of production and non-supervisory employees for total private industries (AWHNONAG), civilian employment (CE16OV), and the civilian non-institutional population (CNP16OV) are produced by the

Bureau of Labor Statistics (BLS) at a 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. Since our models are estimated on quarterly data, we take averages of the monthly data. Compensation per hour for the non-farm business sector (COMPNFB) is obtained from the Labor Productivity and Costs release, and produced by the BLS at a quarterly frequency.

The federal funds rate (henceforth FFR) is obtained from the Federal Reserve Board's H.15 release at a business day frequency. Long-run inflation expectations (average CPI inflation over the next 10 years) are available from the SPF from 1991:Q4 onward. 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.[10] Since the BCEI and the SPF measure inflation expectations in terms of the average CPI inflation and we instead use the GDP deflator and/or core PCE inflation as observables for inflation, as in Del Negro and Schorfheide (2013) we subtract 0.5 from the survey measures, which is roughly the average difference between CPI and GDP deflator inflation across the whole sample. We measure interest-rate spreads as the difference between the annualized Moody's Seasoned Baa Corporate Bond Yield and the 10-Year Treasury Note Yield at constant maturity. Both series are available from the Federal Reserve Board's H.15 release. All data series start in 1960:Q1 or the first quarter in which the series is first available.[11] Section B.5 in the Appendix provides additional details on the construction of the observables.

We consider two samples for our forecast evaluation. The first sample starts in January 1992 (this is the beginning of the sample used in Edge and Gürkaynak, 2010, and Del Negro and Schorfheide, 2013) and ends in January 2017, the last quarter for which eight period-ahead forecasts could be evaluated against realized data. The second sample is the one studied in Cai et al. (forthcoming), which covers the recovery from the Great Recession. This sample starts in April 2011 and ends in April 2016. Within these samples, we construct

[10] 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.

[11] This differs from the estimation sample start date in Section 4. The reason for this difference is primarily computational. In order to feasibly run the large number of simulations required for Section 4, we used the Chandrasekhar recursions (see Herbst (2015)) which can only be used when all data series are entirely nonmissing. Since our hours series is not available until 1964:Q1, we chose to match the estimation sample used by Herbst and Schorfheide (2014) (presample 1965:Q4 to 1966:Q3 and main sample starting in 1966:Q4). However in this section, we start all series in 1960:Q1 in order to match Cai et al. (forthcoming).

real-time datasets using data vintages available on the 10th of January, April, July, and October of each year.

We use the St.Louis Fed's ALFRED database as our primary source of vintaged data. Hourly wage vintages are only available on ALFRED beginning in 1997; we use pre-1997 vintages from Edge and Gürkaynak (2010). The GDP, GDP deflator, investment, hours, and employment series have vintages available for the entire sample. The financial variables and the population series are not revised.[12] Our convention, which follows Del Negro and Schorfheide (2013), is to call the end of sample T the last quarter for which we have NIPA data, that is, GDP, GDP deflator, *et cetera*. So for instance T = 2010:Q4 for forecasts generated using data available on April 10, 2011. Note that in light of our timing the $T+1$ information for the financial data is available, given that this data is produced in real time and that, say, in April 2011 the first quarter of 2011 is completed. We therefore always incorporate this $T+1$ financial information in our forecasts.

In our exercise we also consider projections that are conditional on external interest rate forecasts following Del Negro and Schorfheide (2013, Section 5.4). We do so both because these conditional projections were used in central banks during the zero-lower bound (henceforth, ZLB) period, as documented in Cai et al. (forthcoming), and because it is in any case interesting to find out whether such conditioning helps or hinders the DSGE models' forecasting performance. In order to construct the conditional projections, we augment the measurement equation to add

$$R^e_{t+k|t} = R_* + \mathbb{E}_t[R_{t+k}], \qquad k = 1, \ldots, K$$

where $R^e_{t+k|t}$ is the observed k-period-ahead interest rate forecast, $\mathbb{E}_t[R_{t+k}]$ is the model-implied interest rate expectation, and R_* is the steady-state interest rate.

For any given quarter t, the interest rate expectations observables $R^e_{t+1|t}, \ldots, R^e_{t+K|t}$ come from the BCFF forecast released in the first month of quarter $t+1$.[13] For example, for t = 2008:Q4, we use the January 2009 BCFF forecasts of interest rates. Differently from Del Negro and Schorfheide (2013), we use the expanded dataset containing interest rate forecasts

[12] For each real-time vintage, we use the Hodrick-Prescott filter on the population data observations available as of the forecast date.

[13] We take the number of anticipated shocks K to be 6, which is the maximum number of BCFF forecast quarters (excluding the observed quarterly average that we impute in the first forecast period). Note that since the BCFF survey is released during the first few days of the month, the information set of BCFF forecasters is effectively t – that is, they have no information about quarter $t+1$.

Table 5: Summary of $T+1$ conditioning information

Variable	Source	Conditioning on			
		Neither	FFR Exp.	Nowcasts	Nowcasts and FFR Exp.
Spread	Observed Data	X	X	X	X
R_{T+1}	Observed Data	X	X	X	X
$R_{T+2\|T+1}$	$R^e_{T+2\|T+1}$		X		X
⋮	⋮		⋮		⋮
$R_{T+K+1\|T+1}$	$R^e_{T+K+1\|T+1}$		X		X
GDP growth$_{T+1\|T+1}$	BCEI forecast of $T+1$ GDP growth			X	X
GDP defl. inflation$_{T+1\|T+1}$	BCEI forecast of $T+1$ GDP deflator inflation			X	X

also in the estimation of the model's parameters beginning in 2008:Q4—the start of the ZLB period—reflecting the post-financial crisis era of central bank forward guidance.[14] In order to provide the model with the ability to accommodate federal funds rate expectations, the policy rule in the model was augmented with anticipated policy shocks; see Section B in the Appendix for additional details.

Some of the projections discussed below are conditional on nowcasts —that is, forecasts of the current quarter — of GDP growth, GDP deflator inflation, and financial variables, following Del Negro and Schorfheide (2013, Section 5.3). We accomplish this by appending an additional period of partially observed data for period $T+1$ (the current quarter, given our timing convention).[15] Specifically, for each real-time forecast vintage, we condition on the corresponding BCEI release's mean forecasts of GDP growth and GDP deflator inflation in period $T+1$. Our choice of forecast origin months means that the entire first forecast quarter has already elapsed by the time the forecast is made, so quarterly averages of financial variables have been observed in their entirety. Finally, when we also condition the projections on interest rate expectations we use the BCFF interest rate forecast $R^e_{T+2:T+K+1|T+1}$ as observed expectations of future interest rates in quarter $T+1$. Table 5 summarizes the $T+1$

[14] Rather than estimating a separate standard deviation $\sigma_{r^m,k}$ for each of the K anticipated shocks, we impose the restriction $\sigma^2_{r^m,k} = \sigma^2_{r^m}/K$, which implies that the sum of the variances of the anticipated shocks equals the variance of the contemporaneous shock $\sigma^2_{r^m}$. We do so because at the beginning of the ZLB period, we have too few observations to estimate these variances independently.

[15] Unlike in Del Negro and Schorfheide (2013), we treat the nowcast for $T+1$ as a perfect signal of y_{T+1}, a specialization of both of the *Noise* and *News* assumptions in that paper in which we set $\eta_{T+1} = 0$.

conditioning information.[16]

Finally, the forecasts are evaluated using the latest data vintage, following much of the existing literature on DSGE forecast evaluation. Specifically, for the results shown below we use the vintage downloaded on April 18, 2019. Note that the predictive densities for real GDP growth are computed on per capita data.

5.2 Log Predictive Density Scores with Standard Prior

Figure 7 shows the logarithm of the predictive densities for real GDP growth, GDP deflator inflation, and both variables jointly (left, middle, and right column, respectively) computed over the January 1992 to January 2017 sample for the two DSGE models we are considering: SW (blue solid lines) and SWFF (red solid lines). At each time t, the predictive densities are computed for $h = 2$, 4, 6, and 8 quarters ahead, and for different information sets $\mathcal{I}_t^m$, where m denotes the model. For each model, the information set always includes the history of observables outlined in Table 4 up to time t (and hence differs across models), but can also include additional information, as described in the previous section (see Table 5). Specifically, the rows of Figure 7 show the predictive densities when $\mathcal{I}_t$ includes nowcasts for period $t+1$ and interest rate projections (first row), interest rate projections only (second row), nowcasts only (third row), or neither (last row).

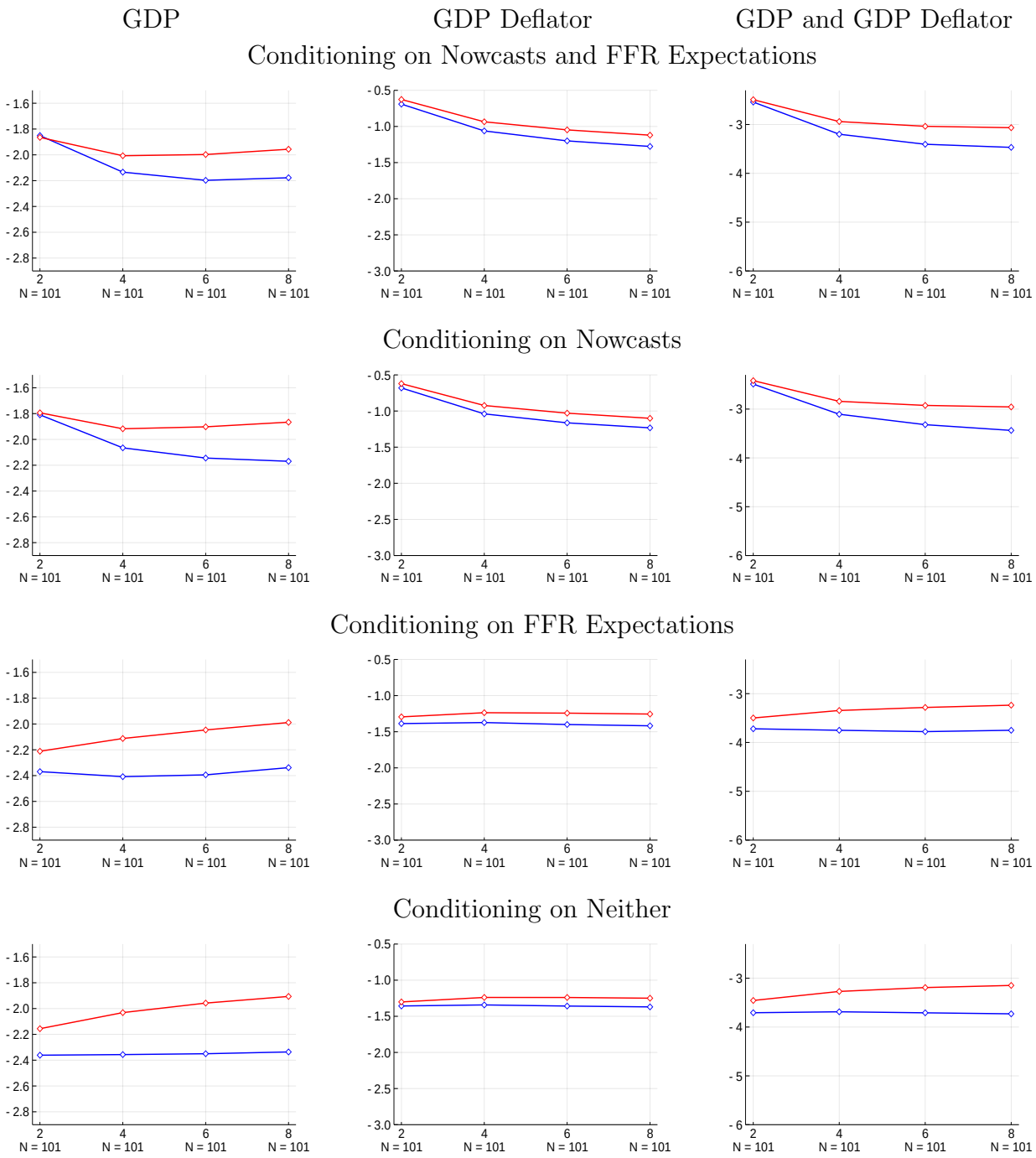
Before commenting on the results a few details on the computation of the predictive densities are in order. First of all, the objects being forecasted are the h-period averages of the variables of interest j,

$$\bar{y}_{j;t+h,h} = \frac{1}{h}\sum_{s=1}^{h} y_{j;t+s}.$$

While the previous literature on forecasting with DSGE models generally focused on h-period ahead forecasts of $y_{j;t+h}$, we choose to focus on averages as they arguably better capture the relevant object for policy-makers: accurately forecasting inflation behavior over the next two years is arguably more important than predicting inflation eight quarters ahead. Conditional on a parameter vector θ, the time t h-period-ahead posterior predictive density for model m

[16] Note that we do not use any of this $T+1$ information in estimating the model parameters—that is, the models are estimated only using time T information. In fact, for the predictive density evaluation exercise, we do not reestimate the DSGE model in every quarter, but only once a year using the January vintage.

Figure 7: Average Log Predictive Scores for SW vs SWFF



Note: These panels compare the log predictive densities from the SW DSGE Model (blue diamonds) with the SWFF DSGE model (red diamonds) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations.

is approximated by

$$p(\bar{y}_{j;t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m) = \sum_{i=1}^{N} p(\bar{y}_{j;t+h,h}|\theta^i, \mathcal{I}_t^m, \mathcal{M}_m),$$

where N is the number of SMC particles and $p(\bar{y}_{j;t+h,h}|\theta^i, \mathcal{I}_t^m, \mathcal{M}_m)$ is the predictive density conditional on the particle θ^i (Section A.1 in the Appendix provides the computational details). The objects plotted in Figure 7 are average of the time t log predictive densities across the sample $[T_0, T_1]$, namely

$$\frac{1}{T_1 - T_0} \sum_{t=T_0}^{T_1} \log\, p(\bar{y}_{j;t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m).$$

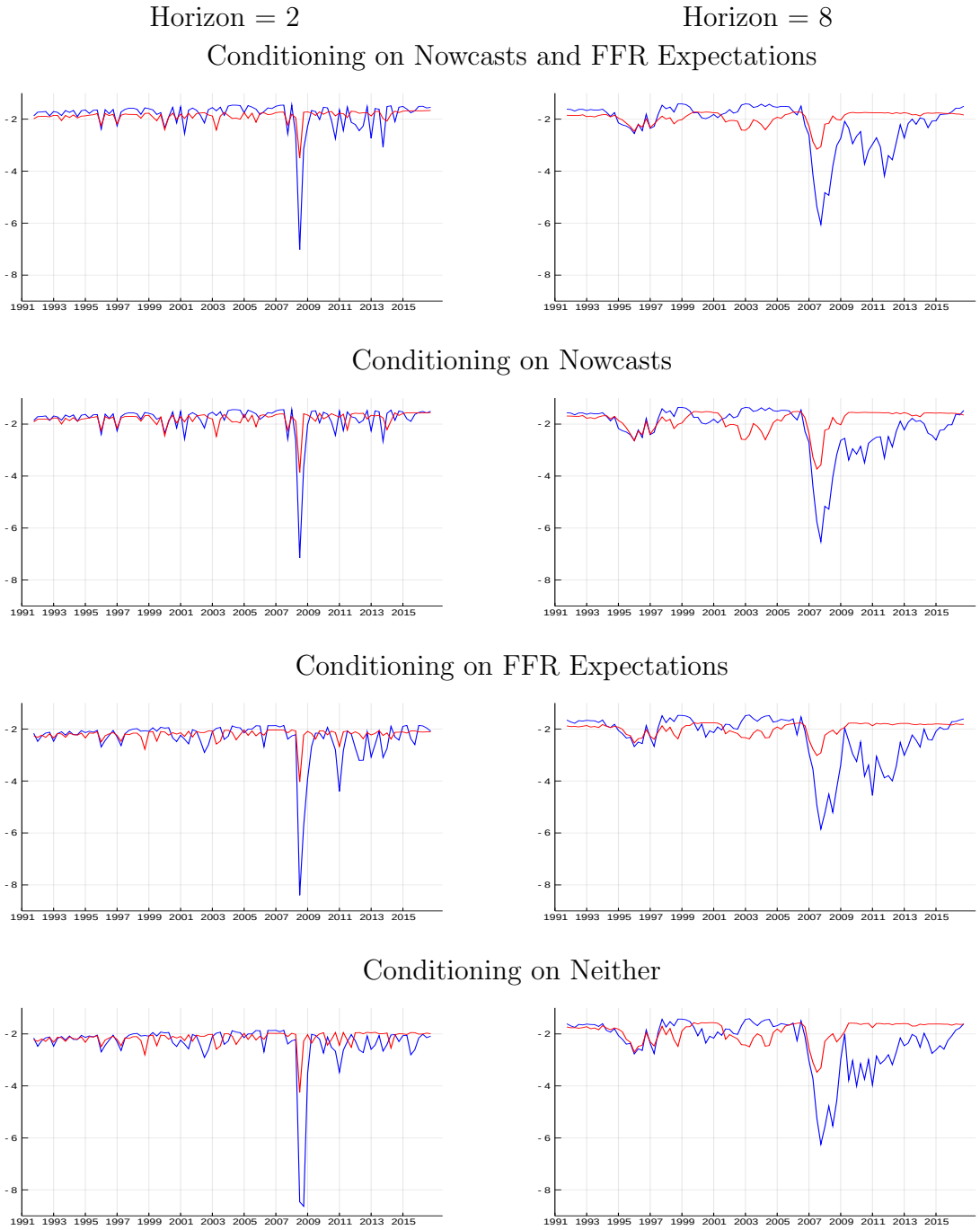
Figure 7 highlights three results. First, regardless of the variable being forecasted, the forecast horizon, and the conditioning assumptions the SWFF model performs better than the SW model: the log predictive scores for the SWFF model (red lines) are never lower than those for SW (blue lines), and most of the times are higher. Second, conditioning on nowcasts improves the forecasting performance of both models, regardless of whether one also conditions on interest rates expectations or not. This can be appreciated by comparing the first and the second row to the third and the fourth row. Not surprisingly the improvements are more pronounced for shorter horizons.

Third, the benefits of conditioning on FFR expectations are less clear cut. Comparing the first with the third row, and the second with the fourth row, the average log predictive density scores are generally a bit lower when conditioning on interest rate forecasts. Of course, from the policy perspective conditioning on a plausible path for interest rates is often an institutional requirement, especially when forward guidance is in place, regardless of whether this helps or hinders the forecasting performance.

Figure 8 shows the time series of the log predictive scores for GDP growth for 2 and 8 periods ahead forecasts (Section C.2 in the Appendix shows the log predictive scores over time for all variables and forecast horizons). Recall that the main difference between the two models is that the SWFF model incorporates financial frictions and, correspondingly, uses spreads between corporate and Treasury yields as an additional observable. It is clear from this figure that the superior forecasting performance of SWFF is due to better forecasting accuracy from the Great Recession onward. Before the Great Recession, the red and the blue line often cross one another. From 2009 on, the red line is almost always above the blue line, particularly for eight period ahead forecasts.

Figure 9 confirms these results by showing the average log predictive scores for the post Great Recession period, that is, from 2011 through 2016 (again, this is the same sample as

Figure 8: Log Predictive Scores for GDP Growth Over Time — SW vs SWFF

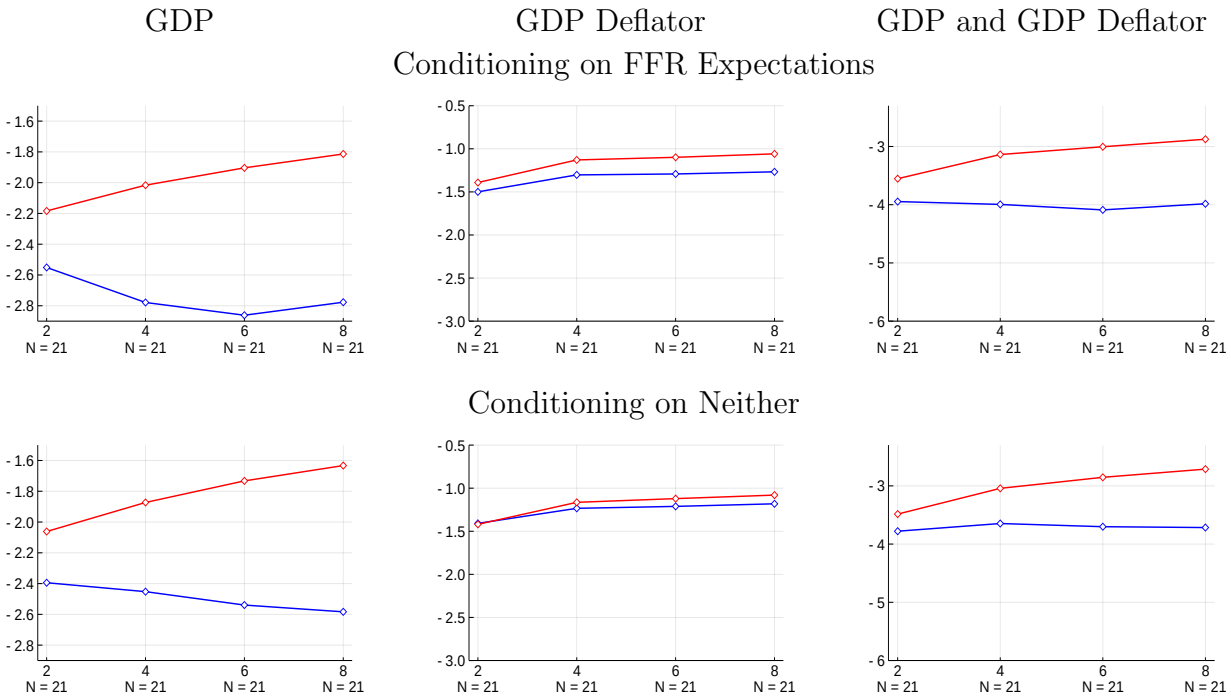


Note: These panels compare $\log p(\bar{y}_{t+h,h}|\mathcal{I}_t^m, \mathcal{M}_m)$ from the SW DSGE model (blue) with the SWFF DSGE model (red) averaged over two, four, six, and eight quarter horizons for output growth. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Cai et al., forthcoming).[17] It is clear from Figure 9 that the gap in forecasting accuracy

[17] Figure 9 only shows the results for two conditioning assumptions, "Conditioning on FFR Expectations"

Figure 9: Average Log Predictive Density Scores for SW vs SWFF: Post-Recession Sample



Note: These panels compare the log predictive densities from the SW DSGE model (blue diamonds) and with the SWFF DSGE model (red diamonds) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.

between the two models is larger for the post-Great Recession period than for the whole sample, especially for output growth. It is also clear that conditioning on FFR expectations widens the gap. In Cai et al. (forthcoming) we explain these results in terms of (i) the failure of the SW model to explain the Great Recession, which affects its forecasting performance thereafter, and (ii) the so-called "forward guidance puzzle" (Del Negro et al., 2012; Carlstrom et al., 2015). The latter refers to the fact that rational expectations representative agents model tend to overestimate the impact of forward guidance policies. As we discuss in Cai et al. (forthcoming), this problem is particularly severe for the SW model, leading to projections that were overoptimistic during the recovery from the Great Recession, when forward guidance was in place.

5.3 Log Predictive Density Scores with Diffuse Priors

As illustrated in Section 4.3, the prior used in the estimation of DSGE models is often quite informative, in the sense that it affects the posterior distribution. While from an economic

and "Conditioning on Neither". Figure A-1 presents results for all conditioning assumptions.

or econometric point of view informative priors are not necessarily problematic, in that they incorporate a priori information gleaned from other studies (see for instance the discussion in Del Negro and Schorfheide, 2008), it is interesting to examine the effect of these relatively tight priors on the forecasting performance of the model. We therefore compute the log predictive scores also under what we refer to as the "diffuse" prior specification.[18]

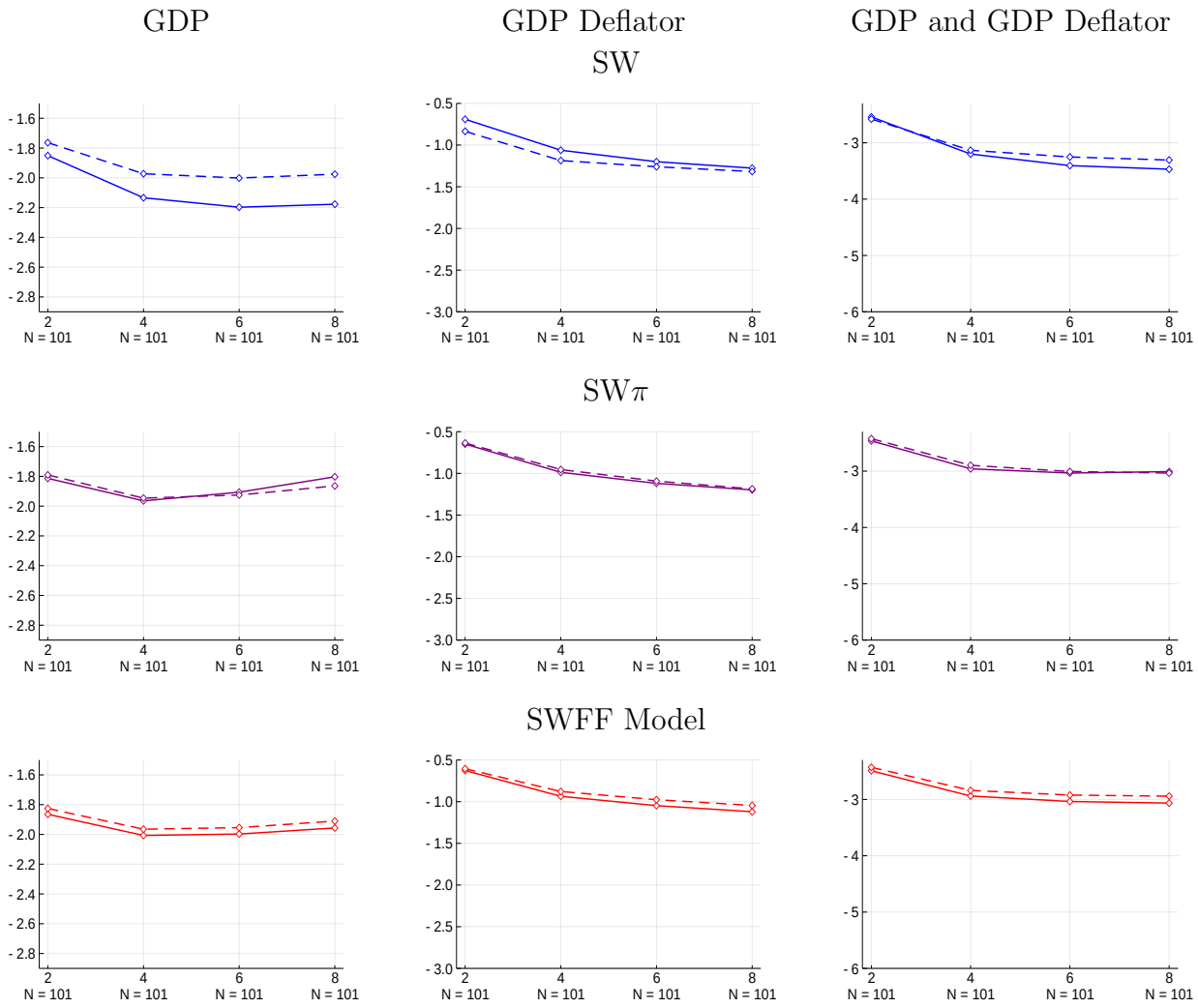
From a frequentist perspective, increasing the prior variance reduces the bias of the Bayes estimator while increasing its variability. The net effect on the mean-squared estimation (and hence forecast) error is therefore ambiguous. In models in which not all parameters (or functions thereof) are identified, the prior serves as a "tie-breaker," and introduces curvature into the posterior in directions in which the likelihood function is flat. If a prior mainly adds information where the likelihood is uninformative and parameters are not identified based on the sample information, then making the prior more diffuse will not have a noticeable effect on the forecasting performance of DSGE models because it mainly selects among parameterizations that track the data equally well.

Figure 10 compares the average log predictive scores obtained with the standard (solid line) and the diffuse (dashed line) prior for the SW (blue, top row), SWπ (purple, middle row), and SWFF (red, bottom row) DSGE models. For brevity, we only show results for forecasts generated conditioning on nowcasts and FFR expectations, but the results are broadly similar under other conditioning assumptions (these are shown in Section C.3 of the Appendix). The panels in Figure 10 show that the differences in average log predictive scores between the standard and the diffuse prior are generally small.

The largest differences arise in the SW model. While the GDP forecast improves noticably, the inflation forecast deteriorates as the prior becomes more diffuse. The latter result is consistent with previous findings in the literature. A tight prior on the constant inflation target around 2% in the SW model is key for estimating a target that is consistent with the low inflation in the latter part of the sample. If the prior variance is increased, then the posterior mean shifts closer to the average inflation rate during the estimation period, which is

[18]For the sake of comparison, under the diffuse prior we only increase the standard deviation for the parameters that are common across the three medium-scale DSGE models, which is the vast majority of the parameters. So for instance we do not change the standard deviations of the parameters pertaining to the financial frictions in the SWFF model. We have, however, computed predictive densities for the SWFF where the standard deviation of all parameters are increased, and we obtained results very similar to those shown below. Finally, note that the prior on the π_* parameter is already very loose under the standard specification for the SWπ and SWFF models, so we leave it unchanged in the diffuse specification.

Figure 10: Comparison of Predictive Densities under Standard and Diffuse Priors



Note: These panels compare the predictive densities estimated with the standard (solid line) and the diffuse (dashed line) prior for the SW (blue, top row), SWπ (purple, middle row), and SWFF (red, bottom row) DSGE models. The predictive densities are averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations. The forecasts associated with these predictive densities are generated conditioning on nowcasts and FFR expectations.

substantially larger than inflation during and after the mid 1990s. When the constant target is replaced by a time-varying one, as in model SWπ, and long-run inflation expectations are added to the list of observables, then the predictive scores under the standard and diffuse prior are essentially identical. For the SWFF model the diffuse prior leads to a uniform, albeit small, improvement in the log predictive score.

To the extent the a tighter prior amplifies and downweighs competing modes in the likelihood function, which are apparent from the multi-modal posterior densities plotted Figure 6, the fact that a model estimated with a tight and a loose prior has similar forecast

performance does not imply that the two versions of the model are indistinguishable in all dimensions. The posterior distributions for some of the parameters determining the relative importance of the endogenous and exogenous propagation mechanism may be very different. In turn, this could result to different impulse response functions to structural shocks and the predicted effect of policy interventions may well vary across estimations.

6 Conclusion

As the DSGE models used by central banks become more complex, improved algorithms for Bayesian computations are necessary. This paper provides a framework for performing parallel and online estimation of DSGE models using SMC techniques. Rather than starting from scratch each time a DSGE model has to be re-estimated, the SMC algorithm makes it possible to mutate and re-weight posterior draws from an earlier estimation so that they approximate a new posterior based on additional observations that have become available since the previous estimation. The same approach could also be used to transform posterior draws for one model into posterior draws for another model that shares the same parameter space, e.g., a linear and a nonlinear version of a DSGE model.

References

Adolfson, Malin, Jesper Lindé, and Mattias Villani, "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," *Econometric Reviews*, 2007, *26* (2-4), 289–328.

An, Sungbae and Frank Schorfheide, "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 2007, *26* (2-4), 113–172.

Aruoba, S. Borağan and Frank Schorfheide, "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, 2010, *forthcoming*.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist, "The Financial Accelerator in a Quantitative Business Cycle Framework," in John Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1, Amsterdam: North Holland, 1999, chapter 21, pp. 1341–1393.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The Financial Accelerator in a Quantitative Business Cycle Framework," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1C, Amsterdam: North-Holland, 1999, chapter 21, pp. 1341–93.

Cai, Michael, Marco Del Negro, Marc P. Giannoni, Abhi Gupta, Pearl Li, and Erica Moszkowski, "DSGE Forecasts of the Lost Recovery," *International Journal of Forecasting*, forthcoming.

Cappé, Olivier, Eric Moulines, and Tobias Ryden, *Inference in Hidden Markov Models*, Springer Verlag, New York, 2005.

Carlstrom, Charles T., Timothy S. Fuerst, and Matthias Paustian, "Inflation and Output in New Keynesian Models with a Transient Interest Rate Peg," *Journal of Monetary Economics*, December 2015, *76*, 230–243.

Chopin, Nicolas, "A Sequential Particle Filter for Static Models," *Biometrika*, 2002, *89* (3), 539–551.

_ , "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Annals of Statistics*, 2004, *32* (6), 2385–2411.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113*, 1–45.

_ **, Roberto Motto, and Massimo Rostagno**, "The Great Depression and the Friedman-Schwartz Hypothesis," *Journal of Money, Credit and Banking*, 2003, *35*, 1119–1197.

_ **,** _ **, and** _ , "Financial Factors in Economic Fluctuations," *Manuscript, Northwestern University and European Central Bank*, 2009.

_ **,** _ **, and** _ , "Risk Shocks," *American Economic Review*, 2014, *104* (1), 27–65.

Christoffel, Kai, Günter Coenen, and Anders Warne, "Forecasting with DSGE Models," in Michael Clements and David Hendry, eds., *Oxford Handbook on Economic Forecasting*, Oxford University Press, 2011.

Creal, Drew, "Sequential Monte Carlo Samplers for Bayesian DSGE Models," *Manuscript, University Chicago Booth*, 2007.

_ , "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 2012, *31* (3), 245–296.

Del Negro, Marco and Frank Schorfheide, "Forming Priors for DSGE Models (and How it Affects the Assessment of Nominal Rigidities)," *Journal of Monetary Economics*, 2008, *55* (7), 1191–1208.

_ **and** _ , "DSGE Model-Based Forecasting," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting, Volume 2*, Elsevier, 2013.

_ **and Stefano Eusepi**, "Fitting Observed Inflation Expectations," *Journal of Economic Dynamics and Control*, 2011, *35*, 2105–2131.

_ **, Marc P. Giannoni, and Christina Patterson**, "The Forward Guidance Puzzle," *FRBNY Staff report*, 2012.

_ **, Marc P Giannoni, and Frank Schorfheide**, "Inflation in the Great Recession and New Keynesian Models," *American Economic Journal: Macroeconomics*, 2015, *7* (1), 168–196.

Diebold, Francis X., Frank Schorfheide, and Minshul Shin, "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," *Journal of Econometrics*, 2017, *201* (2), 322–332.

Durham, Garland and John Geweke, "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," in Ivan Jeliazkov and Dale Poirier, eds., *Advances in Econometrics*, Vol. 34, Emerald Group Publishing Limited, West Yorkshire, 2014, chapter 6, pp. 1–44.

Edge, Rochelle and Refet Gürkaynak, "How Useful Are Estimated DSGE Model Forecasts for Central Bankers," *Brookings Papers of Economic Activity*, 2010, *41* (2), 209–259.

Gali, Jordi, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press, 2008.

Geweke, John and B. Frischknecht, "Exact Optimization By Means of Sequentially Adaptive Bayesian Learning," *Mimeo*, 2014.

Gordon, Neil, D.J. Salmond, and Adrian F.E. Smith, "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *Radar and Signal Processing, IEE Proceedings F*, 1993, *140* (2), 107–113.

Graeve, Ferre De, "The External Finance Premium and the Macroeconomy: US Post-WWII Evidence," *Journal of Economic Dynamics and Control*, 2008, *32* (11), 3415 – 3440.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 1998, *87* (3), 342–36.

Herbst, Edward, "Using the "Chandrasekhar Recursions" for Likelihood Evaluation of DSGE Models," *Computational Economics*, 2015, *45* (4), 693–705.

_ **and Frank Schorfheide**, "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 2014, *29* (7), 1073–1098.

_ **and** _ , *Bayesian Estimation of DSGE Models*, Princeton University Press, 2015.

Jasra, Ajay, David A. Stephens, Arnaud Doucet, and Theodoros Tsagaris, "Inference for Levy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo," *Scandinavian Journal of Statistics*, 2011, *38*, 1–22.

Laseen, Stefan and Lars E.O. Svensson, "Anticipated Alternative Policy-Rate Paths in Policy Simulations," *International Journal of Central Banking*, 2011, *7* (3), 1–35.

Liu, Jun S, *Monte Carlo Strategies in Scientific Computing*, Springer Verlag, New York, 2001.

Moral, Pierre Del, Arnaud Doucet, and Ajay Jasra, "An AdaptiveSequential Monte Carlo Method for Approximate Bayesian Computation," *Statistical Computing*, 2012, *22*, 1009–1020.

Negro, Marco Del, Raiden B Hasegawa, and Frank Schorfheide, "Dynamic prediction pools: an investigation of financial frictions and forecasting performance," *Journal of Econometrics*, 2016, *192* (2), 391–405.

Romer, Paul, "The trouble with macroeconomics," *The American Economist*, 2016, *20*, 1–20.

Schäfer, Christian and Nicolas Chopin, "Sequential Monte Carlo on Large Binary Sampling Spaces," *Statistical Computing*, 2013, *23*, 163–184.

Smets, Frank and Raf Wouters, "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 2003, *1* (5), 1123 – 1175.

_ **and** _ , "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 2007, *97* (3), 586 – 606.

Towns, J., T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr, "XSEDE: Accelerating Scientific Discovery," *Computing in Science & Engineering*, Sept.-Oct. 2014, *16* (5), 62–74.

Warne, Anders, Günter Coenen, and Kai Christoffel, "Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGE-VAR, and VAR Models," *Journal of Applied Econometrics*, 2017, *32* (1), 103–119.

Wolters, Maik, "Evaluating Point and Density Forecasts of DSGE Models," *Journal of Applied Econometrics*, 2015, *30* (1), 74–96.

Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, 2003.

Zhou, Yan, Adam M. Johansen, and John A.D. Aston, "Towards Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach," *arXiv Working Paper*, 2015, *1303.3123v2.*

Online Appendix

A Computational Details

A.1 Predictive Density Formulas

This Appendix focuses on the computation of h-step predictive densities $p(y_{t:t+h}|\mathcal{I}_{t-1}^m, \mathcal{M}_m)$ as well as their average over time, $p(y_{t:t+h}|\mathcal{M}_m)$. The starting point is the state-space representation of the DSGE model. The transition equation

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

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

$$y_t = \mathcal{Z}(\theta)s_t + \mathcal{D}(\theta), \tag{A-2}$$

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.

   (c) Conditional forecasts (using GDP, GDP deflator, 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 four 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

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

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

$$p(y^o_{t:t+h}|\theta, \mathcal{I}_{t-1}) = p_N(y^o_{t:t+h}; \tilde{\mathcal{D}} + \tilde{\mathcal{Z}}\hat{s}_{t:t+h|t-1}, \tilde{\mathcal{Z}}\hat{P}_{t:t+h|t-1}\tilde{\mathcal{Z}}'), \tag{A-3}$$

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

$$p(Fy^o_{t:t+h}|\theta, \mathcal{I}_{t-1}) = p_N(Fy^o_{t:t+h}; F\tilde{\mathcal{D}} + F\tilde{\mathcal{Z}}\hat{s}_{t:t+h|t-1}, F\tilde{\mathcal{Z}}\hat{P}_{t:t+h|t-1}\tilde{\mathcal{Z}}'F'). \tag{A-4}$$

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

$$p(\bar{y}^o_{t+h,h}|\mathcal{I}_{t-1}) = \frac{1}{N}\sum_{i=1}^{N} p(\bar{y}^o_{t+h,h}|\theta^i, \mathcal{I}_{t-1}). \tag{A-5}$$

Further, we show these densities averaged over a time horizon from T_0 to T.

$$p(\bar{y}^o_{t+h,h}) = \frac{1}{N_T}\sum_{t=T_0}^{T} p(\bar{y}^o_{t+h,h}|\mathcal{I}_{t-1}). \tag{A-6}$$

A.2 Adaptive Tempering Schedule with Incremental Upper Bounds

One practical concern is the roots of the function in the early stages of SMC are located very close to the lower bound of the proposed interval. This makes the algorithm inefficient at finding a root, depending on the specified tolerance of convergence. One amendment we can make to the adaptive algorithm is to find a smart way of proposing "incremental" upper bounds for this interval so that the roots are more equidistant from each interval bound and are thus more easily discoverable by the root-finding algorithm.

Algorithm 1 Adaptive Tempering Schedule (with incremental upper bounds)

$j = 2, n = 2$
$\phi_1 = 0$
$\tilde{\phi} = \phi_1$
$\vec{\hat{\phi}} = \{\hat{\phi}_1, ...\hat{\phi}_{N-1}, \hat{\phi}_N\}$ ▷ Where $\hat{\phi}_1 = 0$ and $\hat{\phi}_N = 1$.
while $\phi_n < 1$ **do**
    $f(\phi) = ESS(\phi) - \alpha ESS_{n-1}$
    while $f(\tilde{\phi}) \geq 0$ and $j \leq N$ **do**
        $\tilde{\phi} = \hat{\phi}_j$
        $j = j + 1$
    end
    if $f(\tilde{\phi}) < 0$ **then**
        $\phi_n = root(f, [\phi_{n-1}, \tilde{\phi}])$
    else
        $\phi_n = 1.$
    end
    n = n + 1
end

Notation:
$\vec{\hat{\phi}}$ is the grid of "proposed bounds".
j is the index of the current "proposed bound".
N is the number of elements in the grid of "proposed bounds".
$\tilde{\phi}$ is the current proposed upper bound.

In this algorithm, we generate a grid of proposed bounds, $\vec{\hat{\phi}}$, which could be either a uniform grid from $(0, 1]$, or some other kind of grid, e.g. a fixed schedule generated from some N_ϕ and λ for a similar model/dataset. A bound, $\tilde{\phi}$, is valid if $f(\tilde{\phi}) < 0$. Starting from

the previous valid upper bound $\tilde{\phi} = \hat{\phi}_j$, the inner loop finds the next valid upper bound for the interval $[\phi_{n-1}, \tilde{\phi}]$ in $\{\hat{\phi}_{j+1}, ..., \hat{\phi}_N\}$. $\tilde{\phi}$ could remain unchanged ($\tilde{\phi} = \hat{\phi}_j$) if the bound still remains valid (so the loop will never be entered) or could increment however many times is necessary for $\tilde{\phi} = \hat{\phi}_k$ for some $k > j$ to be a valid upper bound.

The reason for the last conditional "if $f(\tilde{\phi}) < 0$" is if $j > N$, then there are no more elements of the grid that are valid to propose as an upper bound, i.e. $\tilde{\phi} = \hat{\phi}_N = 1$. However, it could still be the case that there are valid upper bounds between $\tilde{\phi} = \hat{\phi}_{N-1}$ and $\tilde{\phi} = \hat{\phi}_N = 1$ that would cause $\Delta ESS = \alpha$.

B DSGE Model Descriptions

This section of the appendix contains the model specifications for AS, SW, SWπ, and SWFF, along with a description of how we construct our data, and a table with the priors on the parameters of the various models.

B.1 An-Schorfheide Model (AS)

B.1.1 Equilibrium Conditions

We write the equilibrium conditions of the three equation model from An and Schorfheide (2007), by expressing each variable in terms of percentage deviations from its steady state value. Let $\hat{x}_t = \ln(x_t/x)$ and write

$$1 = \beta \mathbb{E}_t \left[e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{R}_t - \hat{z}_{t+1} - \hat{\pi}_{t+1}} \right] \tag{A-7}$$

$$\begin{aligned} 0 = {} & \left(e^{\hat{\pi}_t} - 1\right) \left[\left(1 - \frac{1}{2\nu}\right) e^{\hat{\pi}_t} + \frac{1}{2\nu} \right] \\ & - \beta \mathbb{E}_t \left[\left(e^{\hat{\pi}_{t+1}} - 1\right) e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{y}_{t+1} - \hat{y}_t + \hat{\pi}_{t+1}} \right] \\ & + \frac{1-\nu}{\nu \phi \pi^2} \left(1 - e^{\tau \hat{c}_t}\right) \end{aligned} \tag{A-8}$$

$$e^{\hat{c}_t - \hat{y}_t} = e^{-\hat{g}_t} - \frac{\phi \pi^2 g}{2} \left(e^{\hat{\pi}_t} - 1\right)^2 \tag{A-9}$$

$$\begin{aligned} \hat{R}_t = {} & \rho_R \hat{R}_{t-1} + (1-\rho_R)\psi_1 \hat{\pi}_t \\ & + (1-\rho_R)\psi_2 \left(\hat{y}_t - \hat{g}_t\right) + \epsilon_{R,t} \end{aligned} \tag{A-10}$$

$$\hat{g}_t = \rho_g \hat{g}_{t-1} + \epsilon_{g,t} \tag{A-11}$$

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \epsilon_{z,t}. \tag{A-12}$$

The equilibrium law of motion of consumption, output, interest rates, and inflation has to satisfy the expectational difference equations (A-7) to (A-12).

Log-linearization and straightforward manipulation of Equations (A-7) to (A-9) yield the following representation for the consumption Euler equation, the New Keynesian Phillips

curve, and the monetary policy rule:

$$\begin{aligned}\hat{y}_t &= \mathbb{E}_t[\hat{y}_{t+1}] - \frac{1}{\tau}\left(\hat{R}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] - \mathbb{E}_t[\hat{z}_{t+1}]\right) \\ &\quad +\hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}] \\ \hat{\pi}_t &= \beta\mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t) \\ \hat{R}_t &= \rho_R\hat{R}_{t-1} + (1-\rho_R)\psi_1\hat{\pi}_t + (1-\rho_R)\psi_2\left(\hat{y}_t - \hat{g}_t\right) + \epsilon_{R,t}\end{aligned} \tag{A-13}$$

where

$$\kappa = \tau\frac{1-\nu}{\nu\pi^2\phi}. \tag{A-14}$$

B.1.2 Measurement Equation

The AS model is estimated using quarterly output growth, and annualized CPI inflation and nominal federal funds rate, whose measurement equations are given below:

$$\begin{aligned}\textit{Output growth} &= \gamma + 100(y_t - y_{t-1} + z_t) \\ \textit{CPI} &= \pi_* + 400\pi_t \\ \textit{FFR} &= R_* + 400R_t\end{aligned} \tag{A-15}$$

where all variables are measured in percent. π_* and R_* measure the steady-state levels of net inflation and short term nominal interest rates, respectively. They are treated as parameters in the estimation.

B.2 Smets-Wouters Model (SW)

We include a brief description of the log-linearized equilibrium conditions of the Smets and Wouters (2007) model to establish the foundation for explaining the later models. We deviate from the original Smets-Wouters specification by detrending the non-stationary model variables by a stochastic rather than a deterministic trend. This is done in order to express the equilibrium conditions in a flexible manner that accommodates both trend-stationary and unit-root technology processes. The model presented below is the model referred to in the paper as the SW model.

Let $\tilde{z}_t$ be the linearly detrended log productivity process, defined here as:

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \sigma_z\epsilon_{z,t},\ \epsilon_{z,t} \sim N(0,1) \tag{A-16}$$

All non-stationary variables are detrended 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, which is denoted by z_t, follows the process:

$$z_t = \ln(Z_t/Z_{t-1}) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z \epsilon_{z,t} \tag{A-17}$$

All of the variables defined below will be given in log deviations from their non-stochastic steady state, where the steady state values will be denoted by *-subscripts.

B.2.1 Equilibrium Conditions

The *optimal allocation of consumption* satisfies the following Euler equation:

$$\begin{aligned} c_t = &-\frac{(1-he^{-\gamma})}{\sigma_c(1+he^{-\gamma})}\left(R_t - I\!E_t[\pi_{t+1}] + b_t\right) + \frac{he^{-\gamma}}{(1+he^{-\gamma})}\left(c_{t-1} - z_t\right) \\ &+ \frac{1}{(1+he^{-\gamma})} I\!E_t\left[c_{t+1} + z_{t+1}\right] + \frac{(\sigma_c - 1)}{\sigma_c(1+he^{-\gamma})}\frac{w_* L_*}{c_*}\left(L_t - I\!E_t[L_{t+1}]\right). \end{aligned} \tag{A-18}$$

where c_t is consumption, L_t denotes hours worked, 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, $R_t - I\!E_t[\pi_{t+1}]$, and follows an AR(1) process with parameters ρ_b and σ_b. The parameters σ_c and h capture the relative degree of risk aversion and the degree of habit persistence in the utility function, respectively.

The *optimal investment decision* comes from the optimality condition for capital producers and satisfies the following relationship between the level of investment i_t and the value of capital, q_t^k, both measured in terms of consumption:

$$q_t^k = S'' e^{2\gamma}(1+\beta e^{(1-\sigma_c)\gamma})\left(i_t - \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}(i_{t-1} - z_t) - \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}} I\!E_t[i_{t+1} + z_{t+1}] - \mu_t\right) \tag{A-19}$$

This relationship is affected by investment adjustment costs (S'' is the second derivative of the adjustment cost function) and by the marginal efficiency of investment μ_t, an exogenous process which follows an AR(1) with parameters ρ_μ and σ_μ, and that affects the rate of transformation between consumption and installed capital (see Greenwood et al. (1998)).

The *installed capital*, which we also refer to as the capital stock, evolves as:

$$\bar{k}_t = \left(1 - \frac{i_*}{\bar{k}_*}\right)\left(\bar{k}_{t-1} - z_t\right) + \frac{i_*}{\bar{k}_*} i_t + \frac{i_*}{\bar{k}_*} S'' e^{2\gamma}(1 + \beta e^{(1-\sigma_c)\gamma})\mu_t \tag{A-20}$$

where $\frac{i_*}{\bar{k}_*}$ is the steady-state ratio of investment to capital. The parameter β captures the intertemporal discount rate in the utility function of the households.

The *arbitrage condition between the return to capital and the riskless rate* is:

$$\frac{r^k_*}{r^k_* + (1-\delta)} I\!E_t[r^k_{t+1}] + \frac{1-\delta}{r^k_* + (1-\delta)} I\!E_t[q^k_{t+1}] - q^k_t = R_t + b_t - I\!E_t[\pi_{t+1}] \tag{A-21}$$

where r^k_t is the rental rate of capital, r^k_* its steady-state value, and δ the depreciation rate.

The relationship between $\bar{k}_t$ and the *effective capital* rented out to firms k_t is given by:

$$k_t \;=\; u_t - z_t + \bar{k}_{t-1}. \tag{A-22}$$

where capital is subject to variable capacity utilization, u_t.

The optimality condition determining the *rate of capital utilization* is given by:

$$\frac{1-\psi}{\psi} r^k_t = u_t. \tag{A-23}$$

where ψ captures the utilization costs in terms of foregone consumption.

From the optimality conditions of goods producers it follows that all firms have the same *capital-labor ratio*:

$$k_t \;=\; w_t - r^k_t + L_t. \tag{A-24}$$

Real marginal costs for firms are given by:

$$mc_t \;=\; (1-\alpha)\; w_t + \alpha\; r^k_t. \tag{A-25}$$

where α is the income share of capital (after paying markups and fixed costs) in the production function.

All of the equations mentioned above have the same form regardless of whether or not technology has a unit root or is trend-stationary. A few small differences arise for the following two equilibrium conditions.

The *production function* under trend stationarity is:

$$y_t \;=\; \Phi_p\left(\alpha k_t + (1-\alpha)L_t\right) + \mathcal{I}\{\rho_z < 1\}(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t. \tag{A-26}$$

The last term $(\Phi_p - 1)\frac{1}{1-\alpha}\tilde{z}_t$ drops out if technology has a stochastic trend because then one must assume that the fixed costs are proportional to the trend.

The *resource constraint* is:

$$y_t = g_t + \frac{c_*}{y_*}c_t + \frac{i_*}{y_*}i_t + \frac{r^k_* k_*}{y_*}u_t - \mathcal{I}\{\rho_z < 1\}\frac{1}{1-\alpha}\tilde{z}_t, \tag{A-27}$$

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:

$$g_t = \rho_g g_{t-1} + \sigma_g \epsilon_{g,t} + \eta_{gz}\sigma_z \epsilon_{z,t} \tag{A-28}$$

The *price and wage Phillips curves* respectively are:

$$\begin{aligned}\pi_t = \frac{(1-\zeta_p\beta e^{(1-\sigma_c)\gamma})(1-\zeta_p)}{(1+\iota_p\beta e^{(1-\sigma_c)\gamma})\zeta_p((\Phi_p-1)\epsilon_p+1)}mc_t \\ + \frac{\iota_p}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}}\pi_{t-1} + \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\iota_p\beta e^{(1-\sigma_c)\gamma}}I\!E_t[\pi_{t+1}] + \lambda_{f,t}\end{aligned} \tag{A-29}$$

$$\begin{aligned}w_t = \frac{(1-\zeta_w\beta e^{(1-\sigma_c)\gamma})(1-\zeta_w)}{(1+\beta e^{(1-\sigma_c)\gamma})\zeta_w((\lambda_w-1)\epsilon_w+1)}\left(w^h_t - w_t\right) \\ - \frac{1+\iota_w\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}\pi_t + \frac{1}{1+\beta e^{(1-\sigma_c)\gamma}}\left(w_{t-1} - z_t - \iota_w\pi_{t-1}\right) \\ + \frac{\beta e^{(1-\sigma_c)\gamma}}{1+\beta e^{(1-\sigma_c)\gamma}}I\!E_t\left[w_{t+1} + z_{t+1} + \pi_{t+1}\right] + \lambda_{w,t}\end{aligned} \tag{A-30}$$

where ζ_p, ι_p, and ϵ_p are the Calvo parameter, the degree of indexation, and the curvature parameters in the Kimball aggregator for prices, with the equivalent parameters with subscript w corresponding to wages.

The variable w^h_t corresponds to the *household's marginal rate of substitution between consumption and labor* and is given by:

$$\frac{1}{1-he^{-z^*_*}}\left(c_t - he^{-z^*_*}c_{t-1} + he^{-z^*_*}z_t\right) + \nu_l L_t = w^h_t. \tag{A-31}$$

where η_l is the curvature of the disutility of labor (equal to the inverse of the Frisch elasticity in the basence of wage rigidities).

The mark-ups $\lambda_{f,t}$ and $\lambda_{w,t}$ follow exogenous ARMA(1, 1) processes:

$$\lambda_{f,t} = \rho_{\lambda_f}\lambda_{f,t-1} + \sigma_{\lambda_f}\epsilon_{\lambda f,t} + \eta_{\lambda_f}\sigma_{\lambda_f}\epsilon_{\lambda_f,t-1} \tag{A-32}$$

$$\lambda_{w,t} = \rho_{\lambda_w}\lambda_{w,t-1} + \sigma_{\lambda_w}\epsilon_{\lambda w,t} + \eta_{\lambda_w}\sigma_{\lambda_w}\epsilon_{\lambda_w,t-1} \tag{A-33}$$

Lastly, the monetary authority follows a *policy feedback rule*:

$$R_t \;=\; \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1\pi_t + \psi_2(y_t - y_t^f)\right) + \psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m \tag{A-34}$$

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

The exogenous component of the policy rule r_t^m evolves according to the following process:

$$r_t^m = \rho_{r^m} r_{t-1}^m + \epsilon_t^R + \sum_{k=1}^{K} \epsilon_{k,t-k}^R \tag{A-35}$$

where ϵ_t^R is the usual contemporaneous policy shock and $\epsilon_{k,t-k}^R$ is a policy shock that is known to agents at time $t-k$, but affects the policy rule k periods later — that is, at time t. As outlined in Laseen and Svensson (2011), these *anticipated policy shocks* allow us to capture the effects of the zero lower bound on nominal interest rates, as well as the effects of forward guidance in monetary policy.

B.2.2 Measurement Equations

The SW model is estimated using seven quarterly macroeconomic time series, whose measurement equations are given below:

$$\begin{aligned}
&\textit{Output growth} && = \gamma + 100(y_t - y_{t-1} + z_t)\\
&\textit{Consumption growth} && = \gamma + 100(c_t - c_{t-1} + z_t)\\
&\textit{Investment growth} && = \gamma + 100(i_t - i_{t-1} + z_t)\\
&\textit{Real Wage growth} && = \gamma + 100(w_t - w_{t-1} + z_t)\\
&\textit{Hours} && = \bar{l} + 100 l_t\\
&\textit{Inflation} && = \pi_* + 100\pi_t\\
&\textit{FFR} && = R_* + 100 R_t\\
&FFR_{t,t+j}^e && = R_* + I\!E_t[R_{t+j}], j = 1, ..., 6
\end{aligned} \tag{A-36}$$

where all variables are measured in percent, π_* and R_* measure the steady-state levels of net inflation and short term nominal interest rates, respectively, and $\bar{l}$ represents the mean of the hours (this variable is measured as an index).

The priors for the DSGE model parameters are the same as in Smets and Wouters (2007) and are summarized in Panel I of the priors table listed in the SW^{++} section.

B.3 Smets-Wouters Model with Time-Varying Inflation Target (SWπ)

The SWπ model builds on SW by allowing the inflation target to be time-varying. The time-varying inflation target, π_t^*, allows us to capture the dynamics of inflation and interest rates in the estimation sample.

The time-varying *inflation target* evolves according to

$$\pi_t^* = \rho_{\pi^*}\pi_{t-1}^* + \sigma_{\pi^*}\epsilon_{\pi^*,t} \tag{A-37}$$

where $0 < \rho_{\pi^*} < 1$ and $\epsilon_{\pi^*,t}$ is an i.i.d. shock. π_t^* is a stationary process, although the prior on ρ_{π^*} forces this process to be highly persistent.

B.3.1 Measurement Equations

As in Aruoba and Schorfheide (2010) and Del Negro and Eusepi (2011), we use data on long-run inflation expectations in the estimation of SWπ. This allows us to pin down the target inflation rate to the extent that long-run inflation expectations contain information about the central bank's objective.

Thus there is an additional measurement equation for 10 year inflation expectations that augments (A-36), given by

$$\textit{10y Infl. Expectations} = \pi_* + \mathbb{E}_t\left[\frac{1}{40}\sum_{j=0}^{39}\pi_{t+j}\right] \tag{A-38}$$

B.4 Smets-Wouters Model with Financial Frictions (SWFF)

Financial frictions are incorporated into the SW model following the work of Bernanke et al. (1999a) and Christiano et al. (2009).

B.4.1 Equilibrium Conditions

SWFF replaces (A-21) with the following equation for the *excess return on capital* — that is, the spread between the expected return on capital and the riskless rate — and the definition of the *return on capital*, $\tilde{R}_t^k$, respectively:

$$\mathbb{E}_t\left[\tilde{R}_{t+1}^k - R_t\right] = -b_t + \zeta_{sp,b}(q_t^k - \bar{k}_t - n_t) + \tilde{\sigma}_{\omega,t} \tag{A-39}$$

and

$$\tilde{R}_t^k - \pi_t = \frac{r_*^k}{r_*^k + (1-\delta)} r_t^k + \frac{(1-\delta)}{r_*^k + (1-\delta)} q_t^k - q_{t-1}^k \tag{A-40}$$

where $\tilde{R}_t^k$ 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. (2009)) and follows an AR(1) process with parameters ρ_{σ_ω} and σ_{σ_ω}.

The following equation outlines the *evolution of entrepreneurial net worth*:

$$\hat{n}_t = \zeta_{n,\tilde{R}_t^k}\left(\tilde{R}_t^k - \pi_t\right) - \zeta_{n,\tilde{R}_t^k}(R_{t-1} - \pi_t) + \zeta_{n,qK}(q_{t-1}^k + \bar{k}_{t-1}) + \zeta_{n,n} n_{t-1} - \frac{\zeta_{n,\sigma_\omega}}{\zeta_{sp,\sigma_\omega}} \tilde{\sigma}_{\omega,t-1} \tag{A-41}$$

B.4.1.1 Measurement Equations

SWFF's additional measurement equation for the spread (given below) augments the standard set of SW measurement equations (A-36) along with (A-38).

$$Spread = SP_* + 100\mathbb{E}_t\left[\tilde{R}_{t+1}^k - R_t\right] \tag{A-42}$$

where SP_* measures the steady-state spread. Priors are specified for the parameters SP_*, $\zeta_{sp,b}$, ρ_{σ_ω}, σ_{σ_ω}, and the parameters $\bar{F}_*$ and γ_* (the steady-state default probability and the survival rate of entrepreneurs, respectively), are fixed.

B.5 Data Transformation

The data are transformed following Smets and Wouters (2007), with the exception of the civilian population data, which are filtered using the Hodrick-Prescott filter to remove jumps around census dates. For each financial variable, we take quarterly averages of the annualized daily data and divide by four. Let Δ denote the temporal difference operator. Then:

GDP growth	=	$100 * \Delta LN((GDP/GDPDEF)/CNP16OV)$
Consumption growth	=	$100 * \Delta LN((PCEC/GDPDEF)/CNP16OV)$
Investment growth	=	$100 * \Delta LN((FPI/GDPDEF)/CNP16OV)$
Real wage growth	=	$100 * \Delta LN(COMPNFB/GDPDEF)$
Hours worked	=	$100 * LN((AWHNONAG * CE16OV/100)/CNP16OV)$
GDP deflator inflation	=	$100 * \Delta LN(GDPDEF)$
FFR	=	$(1/4) *$ FEDERAL FUNDS RATE
$FFR^e_{t+k\|t}$	=	$(1/4) *$ BLUE CHIP k-QUARTERS AHEAD FFR FORECAST
10y inflation exp	=	(10-year average CPI inflation forecast $- 0.50)/4$
Spread	=	$(1/4) *$ (Baa Corporate $-$ 10 year Treasury)

In the long-term inflation expectation transformation, 0.50 is the average difference between CPI and GDP annualized inflation from the beginning of the sample to 1992.

B.6 Prior Specifications

We estimate the model using Bayesian techniques. This requires the specification of a prior distribution for the model parameters. For most parameters common with Smets and Wouters (2007), we use the same marginal prior distributions. As an exception, we favor a looser prior than Smets and Wouters (2007) for the quarterly steady-state inflation rate π_*; it is centered at 0.75% and has a standard deviation of 0.4%. Regarding the financial frictions, we specify priors for the parameters SP_*, $\zeta_{sp,b}$, ρ_{σ_ω}, and σ_{σ_ω}, while we fix the parameters corresponding to the steady-state default probability and the survival rate of entrepreneurs, respectively. In turn, these parameters imply values for the parameters of (A-41). Information on the priors is provided in the subsequent tables.

Table A-1: Prior Definitions: An Schorfheide Model

	Type	Mean	Std Dev		Type	Mean	Std Dev
e_R	-	0.45	0.00	ρ_g	U	0.50	0.29
e_y	-	0.12	0.00	ρ_z	U	0.50	0.29
e_π	-	0.29	0.00	σ_R	IG	0.40	4.00
rA	G	0.50	0.50	σ_g	IG	1.00	4.00
γ_Q	N	0.40	0.20	σ_z	IG	0.50	4.00
κ	U	0.50	0.29	τ	G	2.00	0.50
$\pi*$	G	7.00	2.00	ψ_1	G	1.50	0.25
ρ_R	U	0.50	0.29	ψ_2	G	0.50	0.25

Note: The table shows the prior specifications of each of the model parameters in the An and Schorfheide (2007, AS) model. The table specifies the parameter name, the distribution type, where B, N, G, and IG stand for Beta, Normal, Gamma, Inverse Gamma, and the parameter means and standard deviations (written in parentheses). The Inverse Gamma prior is characterized by the mode and degrees of freedom.

Table A-2: Standard and Diffuse Priors for Medium Scale DSGE Models

		Standard Prior			Diffuse Prior		
Parameter	Type	SW	Common	SWFF and SWπ	SW	Common	SWFF and SWπ
				Policy			
ψ_1	N		1.500 (0.250)			1.500 (0.750)	
ψ_2	N		0.120 (0.050)			0.120 (0.150)	
ψ_3	N		0.120 (0.050)			0.120 (0.150)	
ρ	B		0.750 (0.100)			0.500 (0.289)	
ρ_{rm}	B		0.500 (0.200)			0.500 (0.289)	
σ_{rm}	IG		0.100 (2.000)			0.100 (2.000)	
				Nominal Rigidities			
ζ_p	B		0.500 (0.100)			0.500 (0.289)	
ι_p	B		0.500 (0.150)			0.500 (0.289)	
ε_p	-		10.000			10.000	
ζ_w	B		0.500 (0.100)			0.500 (0.289)	
ι_w	B		0.500 (0.150)			0.500 (0.289)	
ε_w	-		10.000			10.000	
				Other Endogenous Propagation and Steady State			

Table A-2: Standard and Diffuse Priors for Medium Scale DSGE Models

		Standard Prior			Diffuse Prior		
Parameter	Type	SW	Common	SWFF and SWπ	SW	Common	SWFF and SWπ
100γ	N		0.400 (0.100)			0.400 (0.300)	
α	N		0.300 (0.050)			0.300 (0.150)	
$100(\beta^{-1}-1)$	G		0.250 (0.100)			0.250 (0.300)	
σ_c	N		1.500 (0.370)			1.500 (1.110)	
h	B		0.700 (0.100)			0.500 (0.289)	
ν_l	N		2.000 (0.750)			2.000 (2.250)	
δ	-		0.025			0.025	
Φ	N		1.250 (0.120)			1.250 (0.360)	
S''	N		4.000 (1.500)			4.000 (4.500)	
ψ	B		0.500 (0.150)			0.500 (0.289)	
$\bar{L}$	N		-45.00 (5.00)			-45.00 (15.00)	
λ_w	-		1.500			1.500	
π_*	G	0.620 (0.100)		0.750 (0.400)	0.620 (0.300)		0.750 (0.400)
g_*	-		0.180			0.180	
			Financial Frictions (SWFF Only)				
$F(\omega)$	-			0.030			0.030
spr_*	G	-		2.000 (0.100)			2.000 (0.100)
ζ_{spb}	B	-		0.050 (0.005)			0.050 (0.005)
γ_*	-			0.990			0.990
			Exogenous Process				
ρ_g	B		0.500 (0.200)			0.500 (0.289)	
ρ_b	B		0.500 (0.200)			0.500 (0.289)	
ρ_μ	B		0.500 (0.200)			0.500 (0.289)	
ρ_z	B		0.500 (0.200)			0.500 (0.289)	
ρ_{λ_f}	B		0.500 (0.200)			0.500 (0.289)	
ρ_{λ_w}	B		0.500 (0.200)			0.500 (0.289)	
η_{λ_f}	B		0.500 (0.200)			0.500 (0.289)	
η_{λ_w}	B		0.500 (0.200)			0.500 (0.289)	
σ_g	IG		0.100 (2.000)			0.100 (2.000)	
σ_b	IG		0.100 (2.000)			0.100 (2.000)	

Table A-2: Standard and Diffuse Priors for Medium Scale DSGE Models

		Standard Prior			Diffuse Prior		
Parameter	Type	SW	Common	SWFF and SWπ	SW	Common	SWFF and SWπ
σ_μ	IG		0.100 (2.000)			0.100 (2.000)	
σ_z	IG		0.100 (2.000)			0.100 (2.000)	
σ_{λ_f}	IG		0.100 (2.000)			0.100 (2.000)	
σ_{λ_w}	IG		0.100 (2.000)			0.100 (2.000)	
η_{gz}	B		0.500 (0.200)			0.500 (0.289)	
		Financial Frictions Exogenous Process (SWFF Only)					
ρ_{σ_ω}	B	-	0.750 (0.150)		-	0.500 (0.289)	
ρ_{π^*}	-		0.990			0.990	
σ_{σ_ω}	IG	-	0.050 (4.000)			0.050 (4.000)	
σ_{π^*}	IG	-	0.030 (6.000)			0.030 (6.000)	

Note: The table shows the prior specifications of each of the model parameters in the SW, SWπ, and SWFF models for both a "standard prior" and "diffuse prior". The diffuse priors specification follows Herbst and Schorfheide (2014) Table A-3. The table specifies the parameter name, the distribution type, where B, N, G, and IG stand for Beta, Normal, Gamma, Inverse Gamma, and the parameter means and standard deviations (written in parentheses). The Inverse Gamma prior is characterized by the mode and degrees of freedom. The priors for the parameters that are common across models are listed under the "Common" column.

C Additional Results

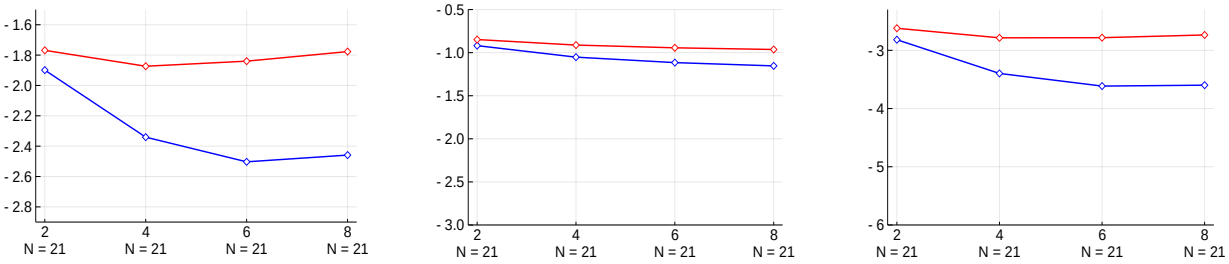
This section contains various robustness exercises that complement the results reported in the main paper. In Section C.1 we show average log predictive scores for various samples and model specifications. In Section C.2 we show the evolution of log predictive scores over time. Finally, in Section C.3 we compare forecasts obtained under the standard and diffuse priors.

C.1 Average Log Predictive Scores

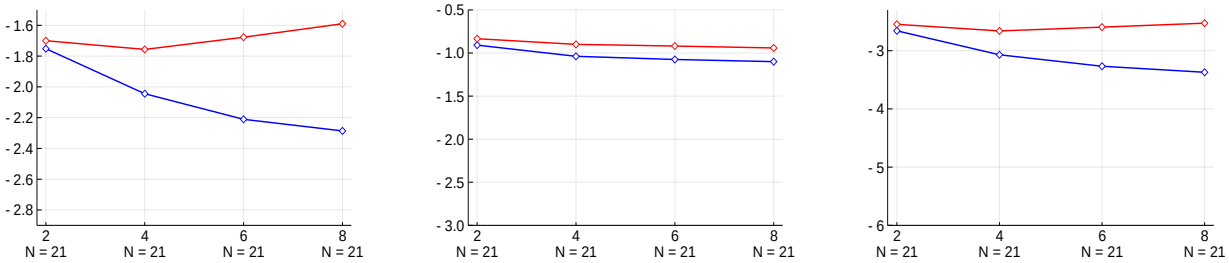
Figure A-1: Average Log Predictive Scores for SW vs SWFF: Post-Recession Sample

GDP | GDP Deflator | GDP and GDP Deflator

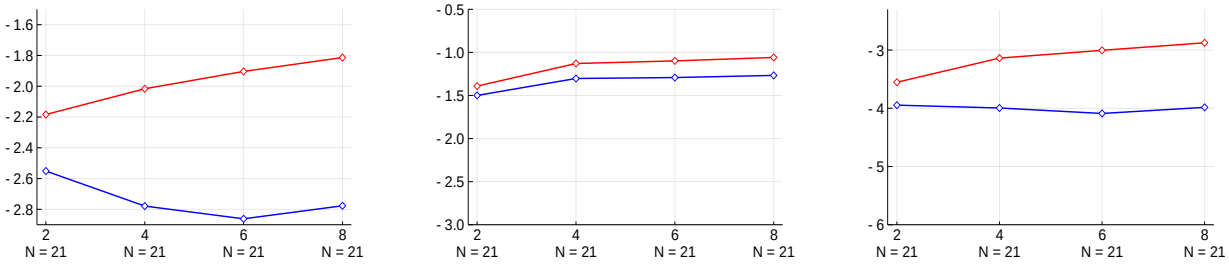
Conditioning on Nowcasts and FFR Expectations



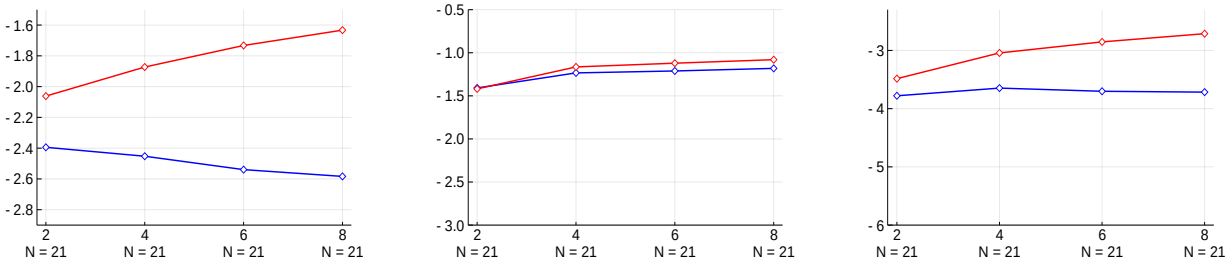
Conditioning on Nowcasts



Conditioning on FFR Expectations

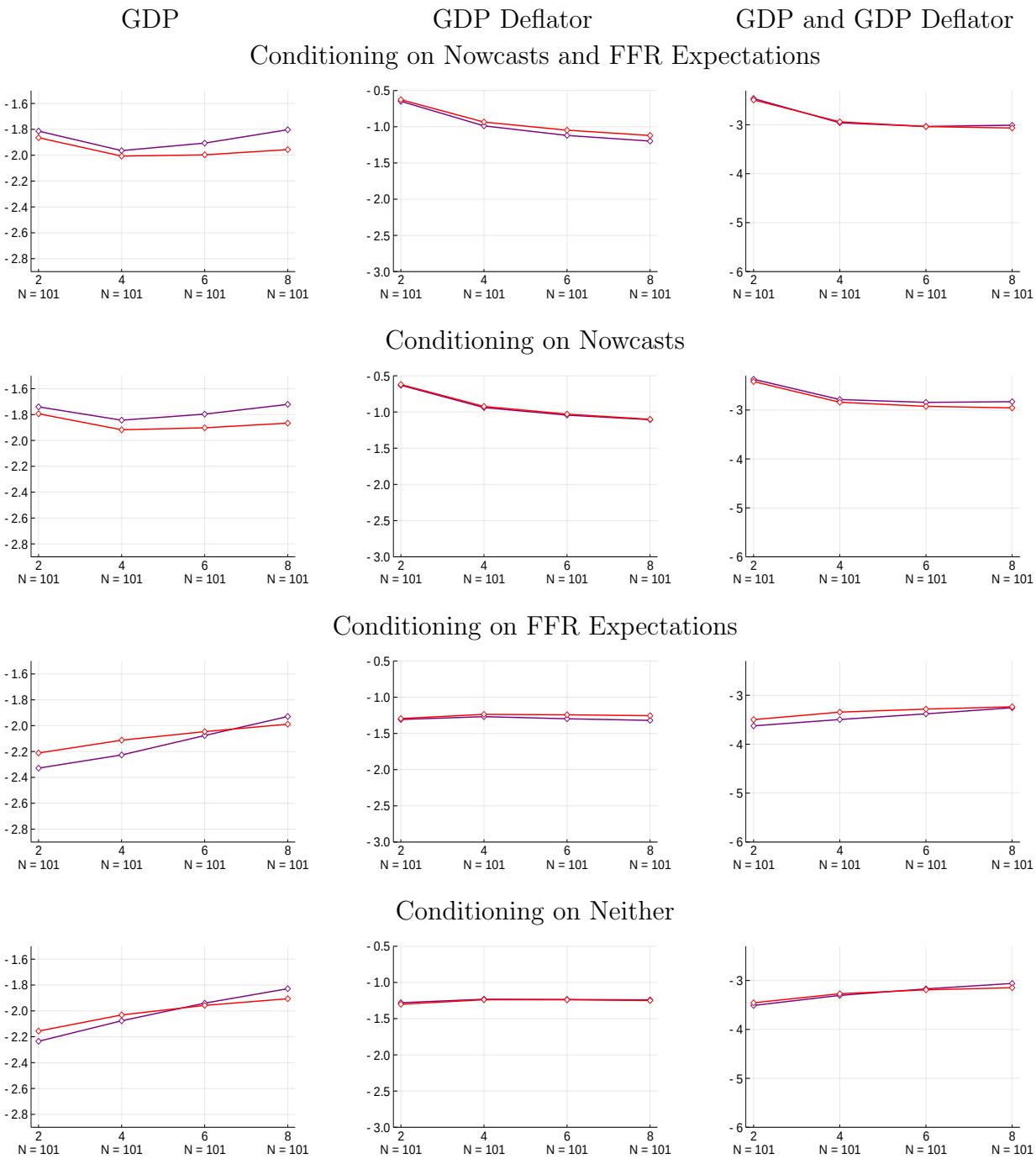


Conditioning on Neither



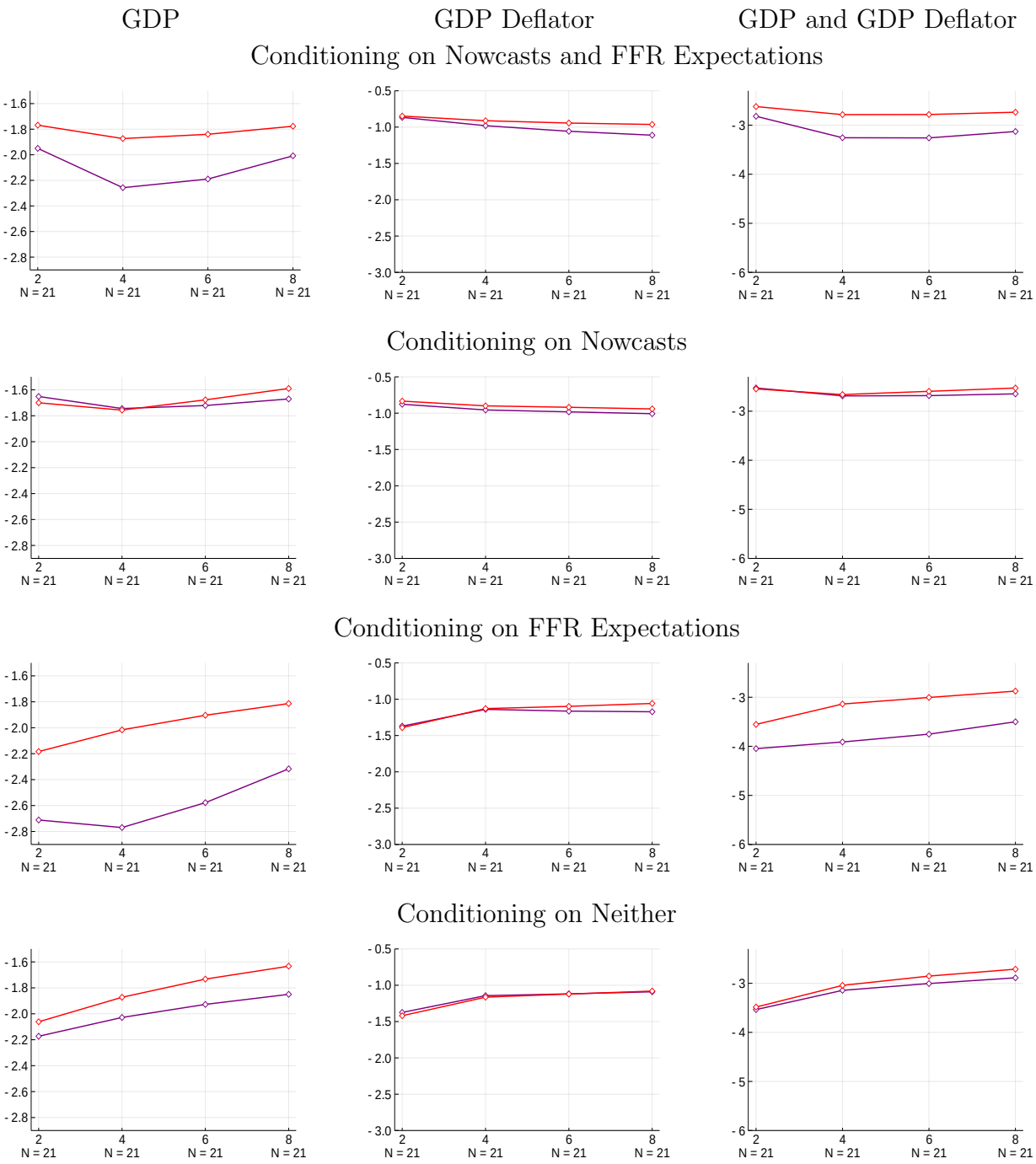
Note: These panels compare the log predictive scores from the SW DSGE Model (blue diamonds) and with the SWFF DSGE model (red diamonds) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.

Figure A-2: Average Log Predictive Scores for SWπ vs SWFF



Note: These panels compare the log predictive scores from the SWπ DSGE Model (blue diamonds) with the SWFF DSGE model (red diamonds) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations.

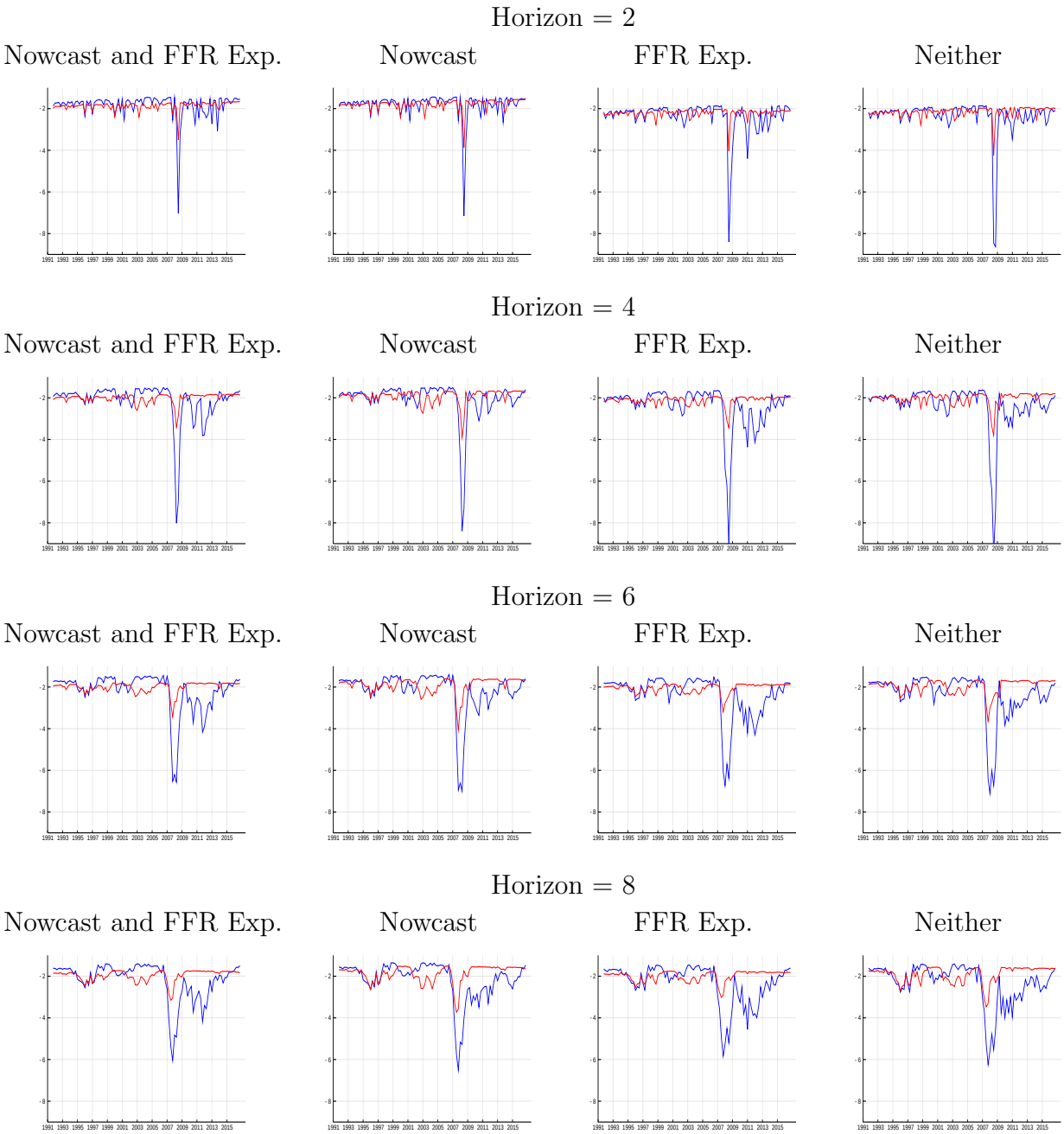
Figure A-3: Average Log Predictive Scores for SWπ vs SWFF: Post-Recession Sample



Note: These panels compare the log predictive scores from the SWπ DSGE Model (blue diamonds) and with the SWFF DSGE model (red diamonds) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.

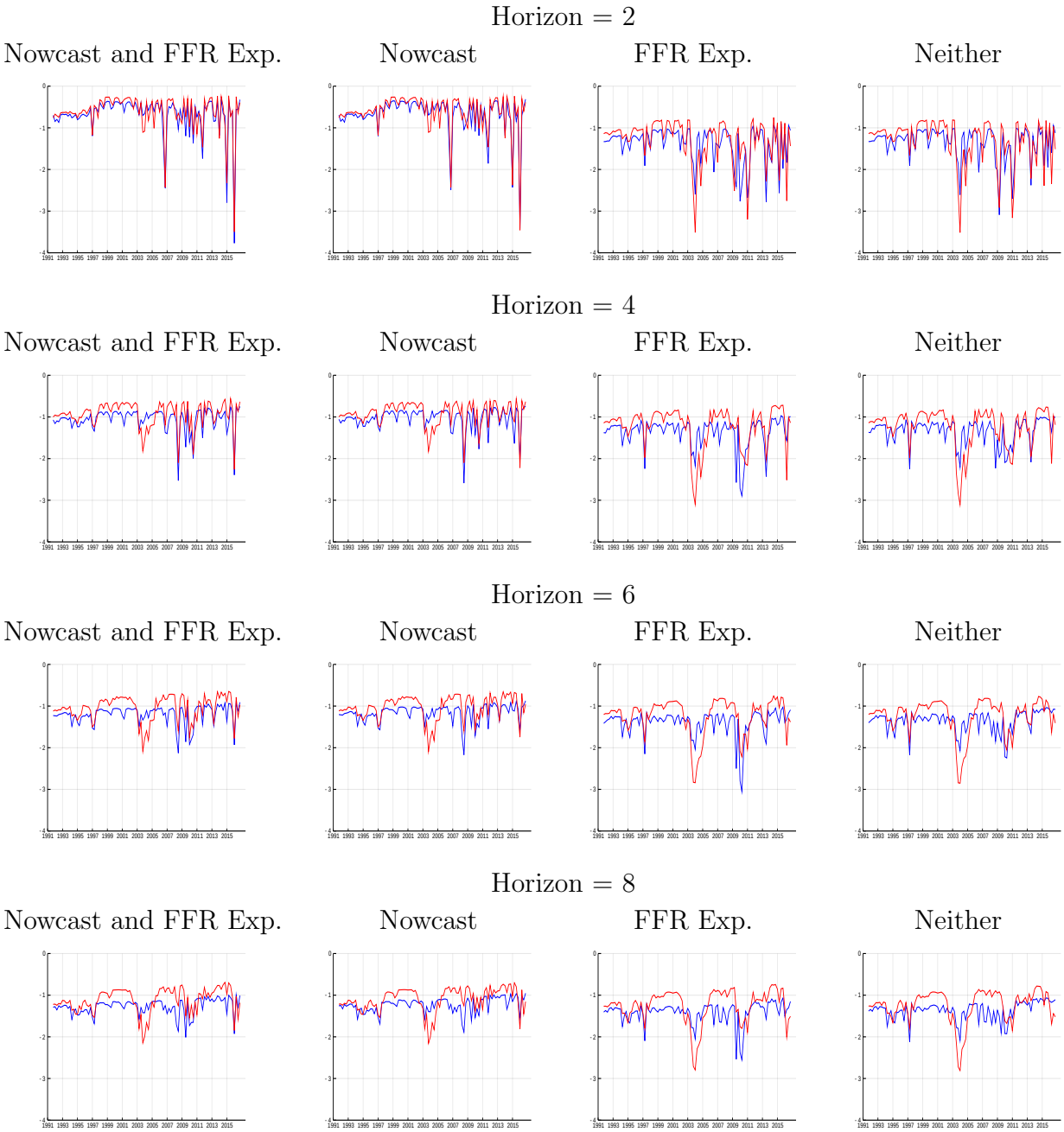
C.2 Log Predictive Scores Over Time

Figure A-4: Log Predictive Scores Over Time for SW vs SWFF—Predicted Variable: GDP



Note: The four panels show the predictive density comparisons across time for the SW (blue line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for output growth. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Figure A-5: Log Predictive Scores Over Time for SW vs SWFF—Predicted Variable: GDP Deflator

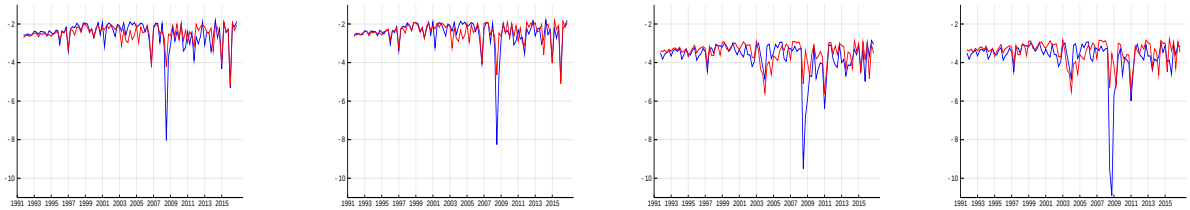


Note: The four panels show the predictive density comparisons across time for the SW (blue line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for inflation. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Figure A-6: Log Predictive Scores Over Time for SW vs SWFF—Predicted Variables: GDP and GDP Deflator

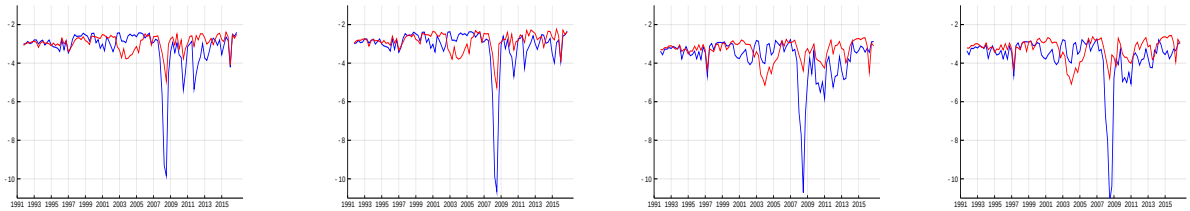
Horizon = 2

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



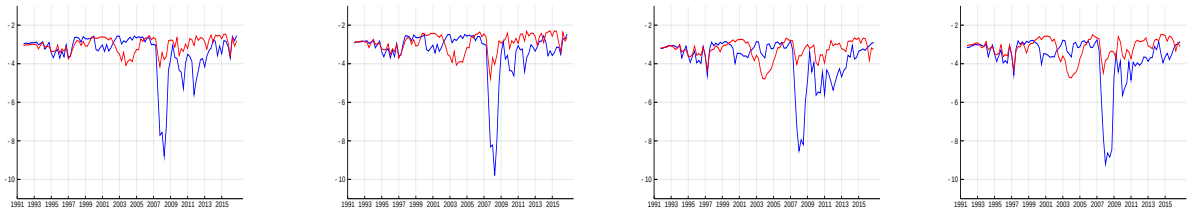
Horizon = 4

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



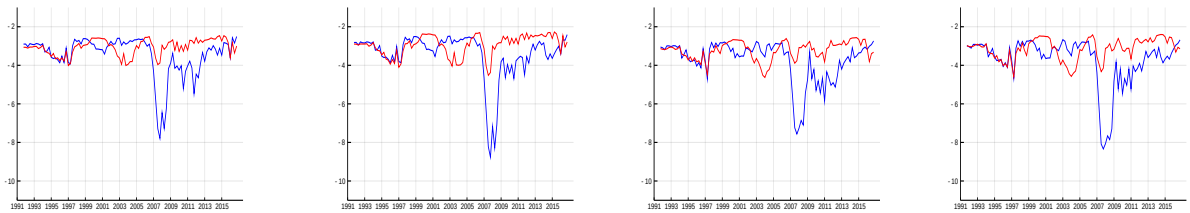
Horizon = 6

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



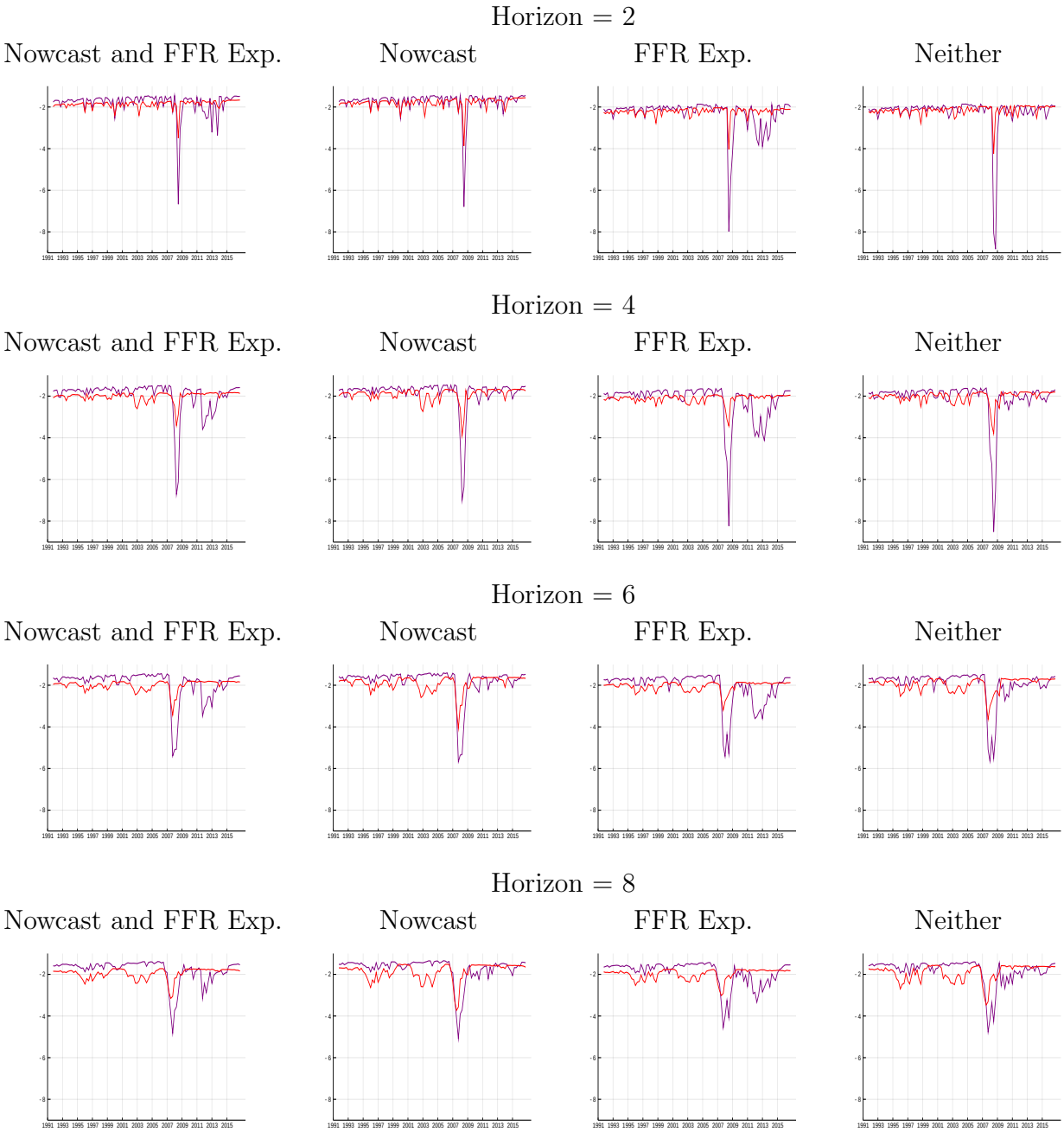
Horizon = 8

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



Note: The four panels show the predictive density comparisons across time for the SW (blue line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for output growth and inflation. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Figure A-7: Log Predictive Scores Over Time for SWπ vs SWFF—Predicted Variable: GDP

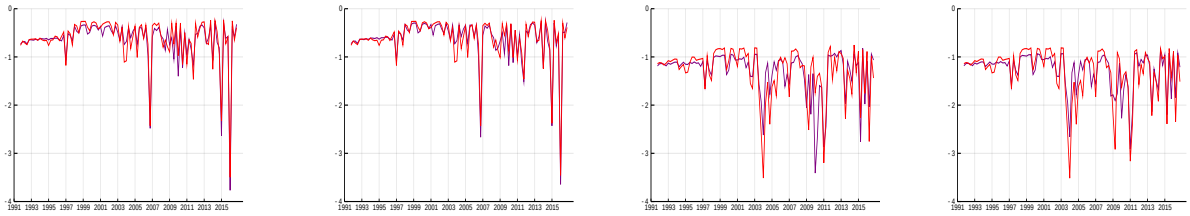


Note: The four panels show the predictive density comparisons across time for the SWπ (purple line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for output growth. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Figure A-8: Log Predictive Scores Over Time for SWπ vs SWFF—Predicted Variable: GDP Deflator

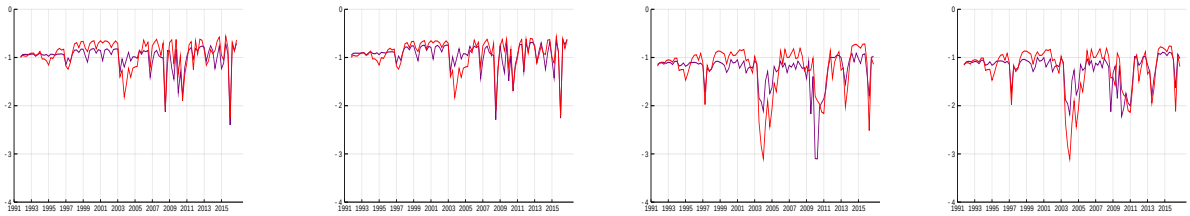
Horizon = 2

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



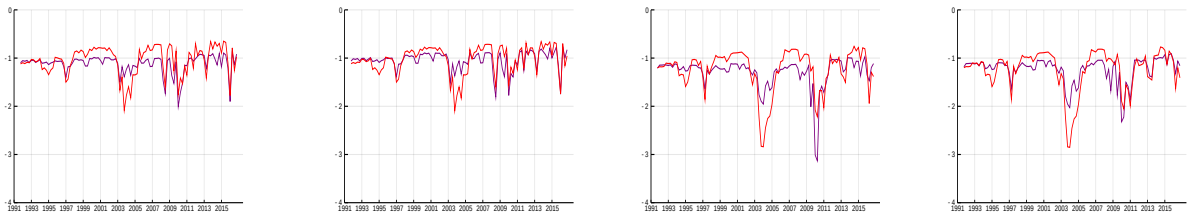
Horizon = 4

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



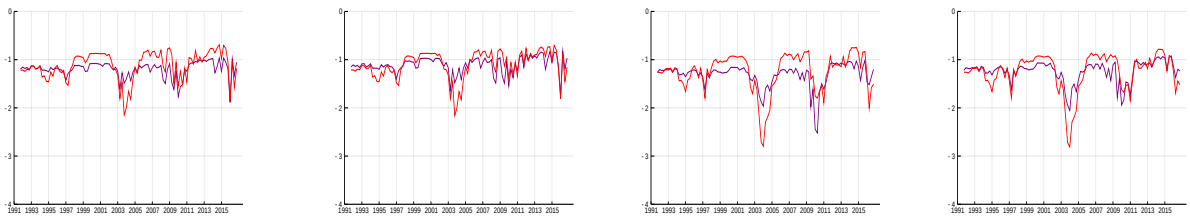
Horizon = 6

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



Horizon = 8

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither

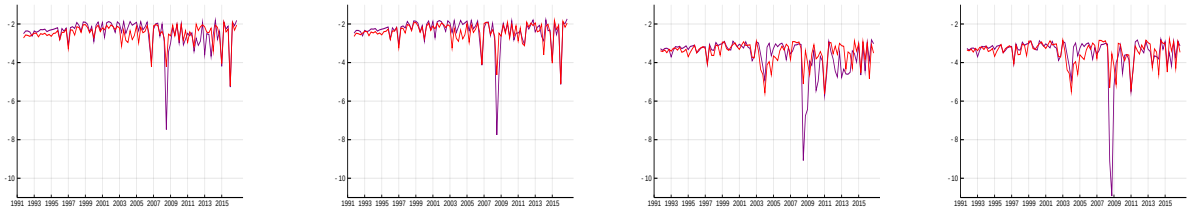


Note: The four panels show the predictive density comparisons across time for the SWπ (purple line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for inflation. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

Figure A-9: Log Predictive Scores Over Time for SWπ vs SWFF—Predicted Variables: GDP and GDP Deflator

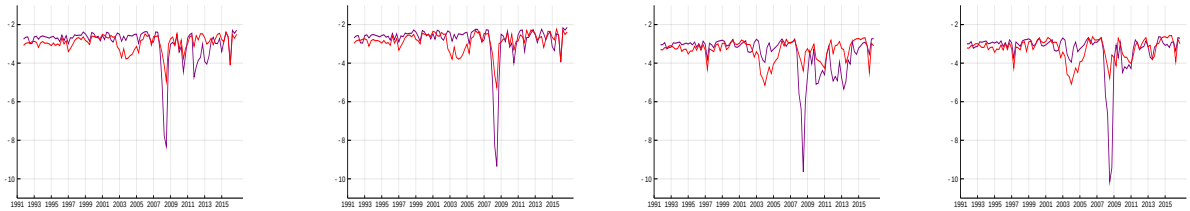
Horizon = 2

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



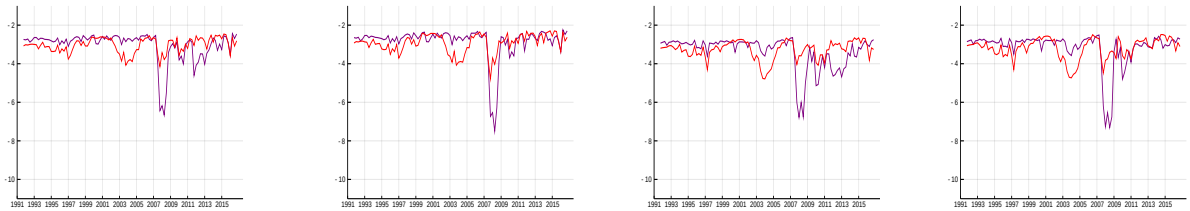
Horizon = 4

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



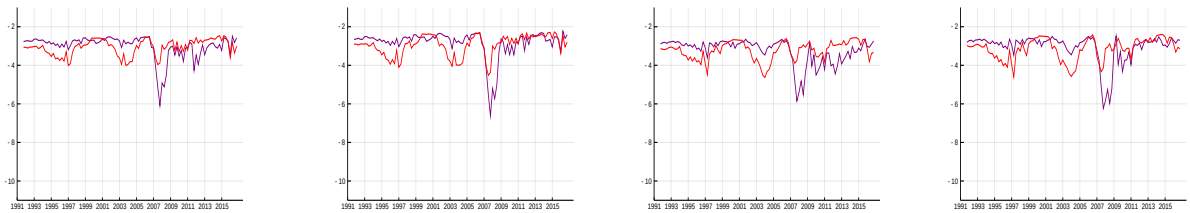
Horizon = 6

Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



Horizon = 8

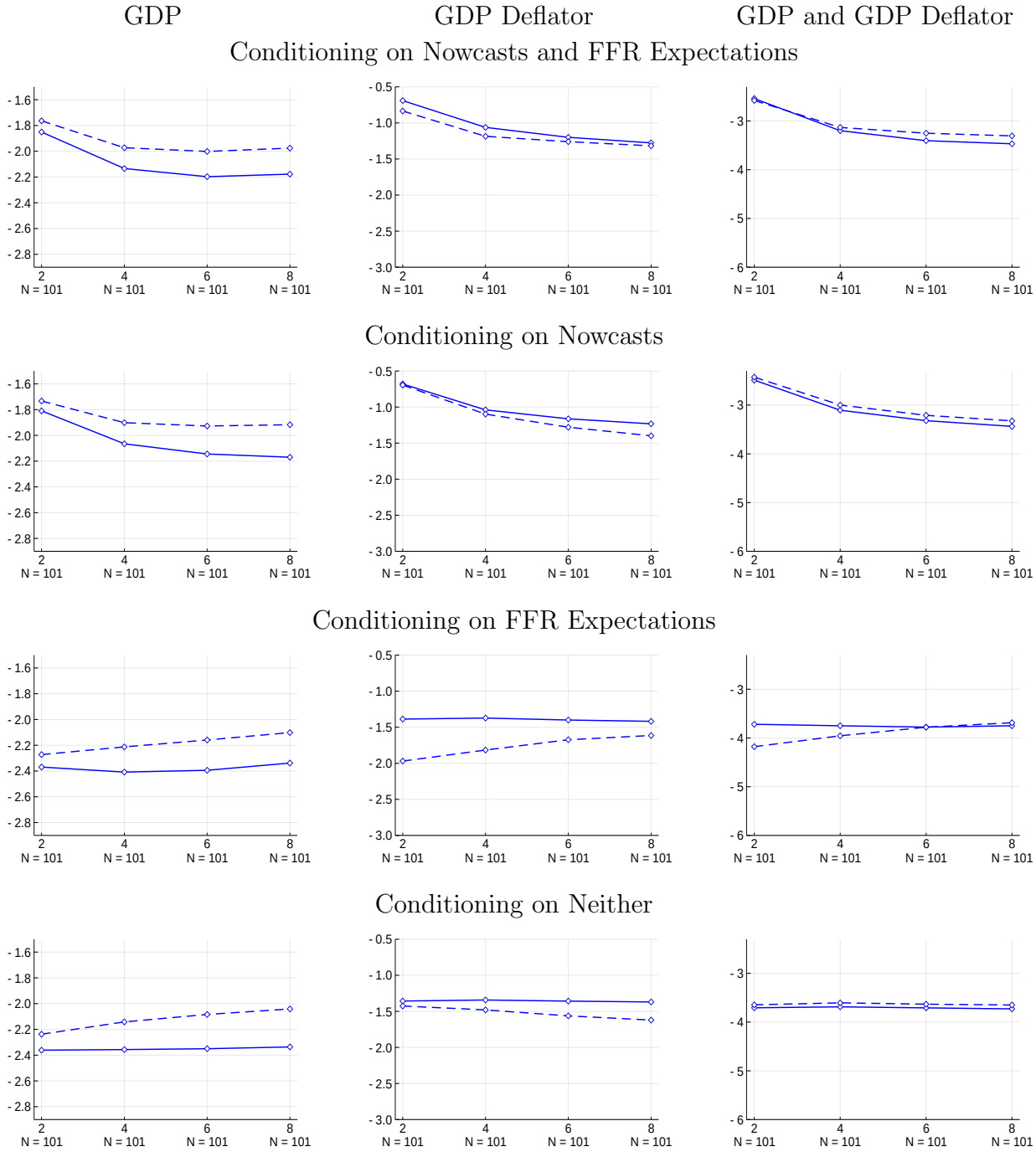
Nowcast and FFR Exp. | Nowcast | FFR Exp. | Neither



Note: The four panels show the predictive density comparisons across time for the SWπ (purple line) and SWFF (red line) DSGE models averaged over 2, 4, 6, and 8 quarter horizons for output growth and inflation. Forecast origins from January 1992 to January 2017 only are included in these calculations. The x-axis shows the quarter in which the forecasts were generated (time $T+1$ in the parlance of section 5.1).

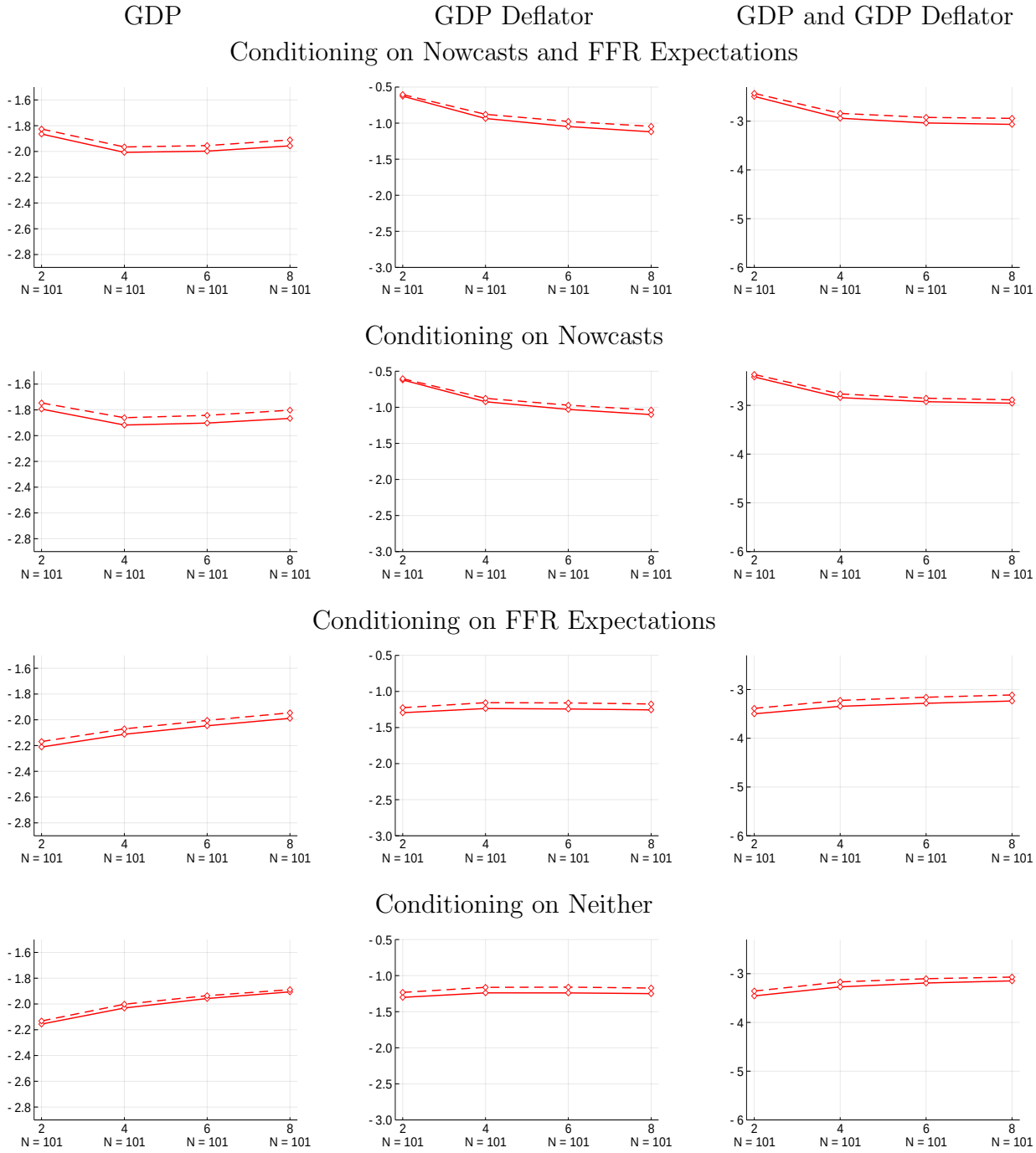
C.3 Comparing Forecasts with Standard and Diffuse Priors

Figure A-10: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SW model



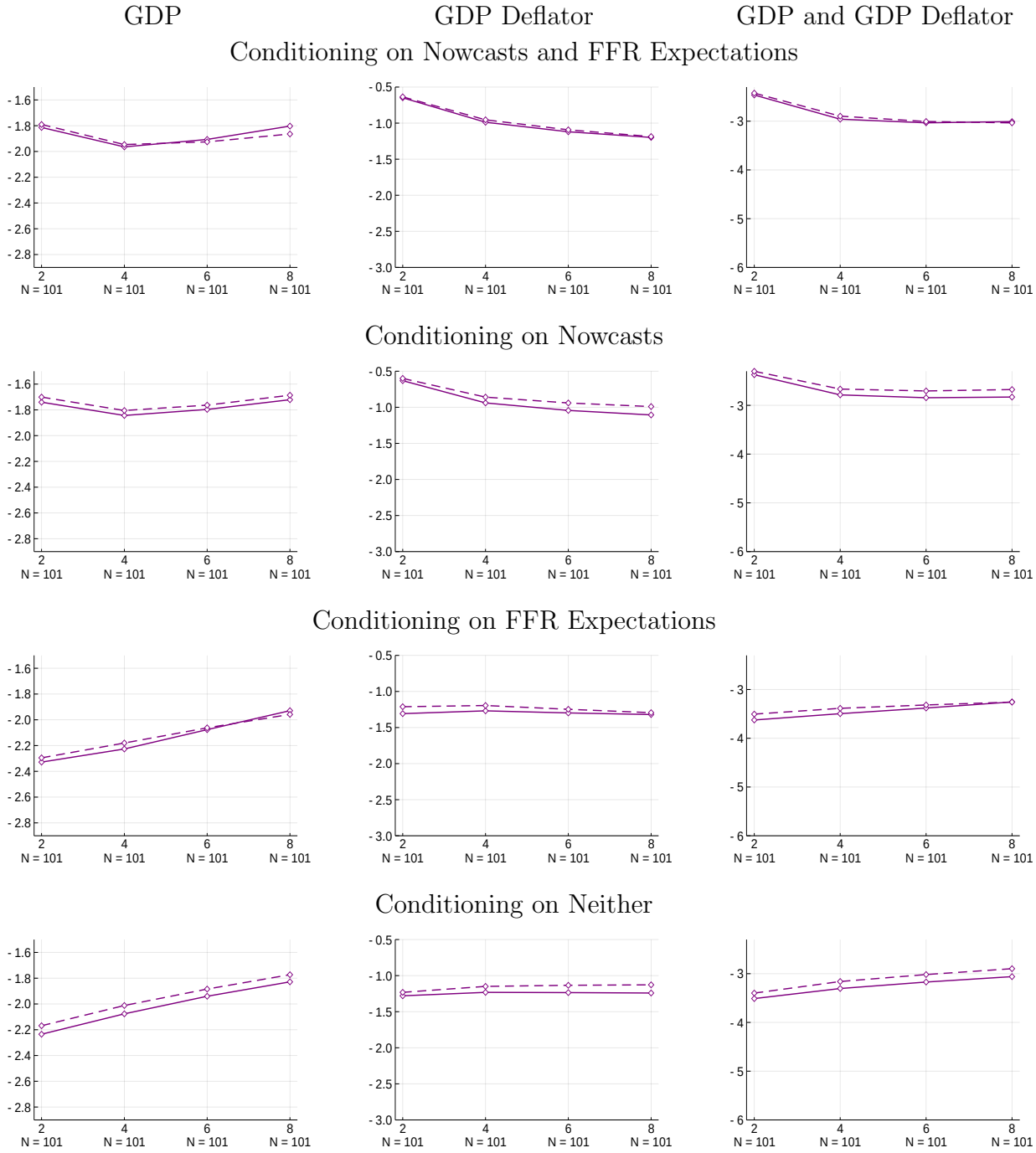
Note: These panels compare the log predictive scores from the SW DSGE model estimated with a standard prior (blue solid) with the same model estimated with a diffuse prior (blue dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations.

Figure A-11: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SWFF model



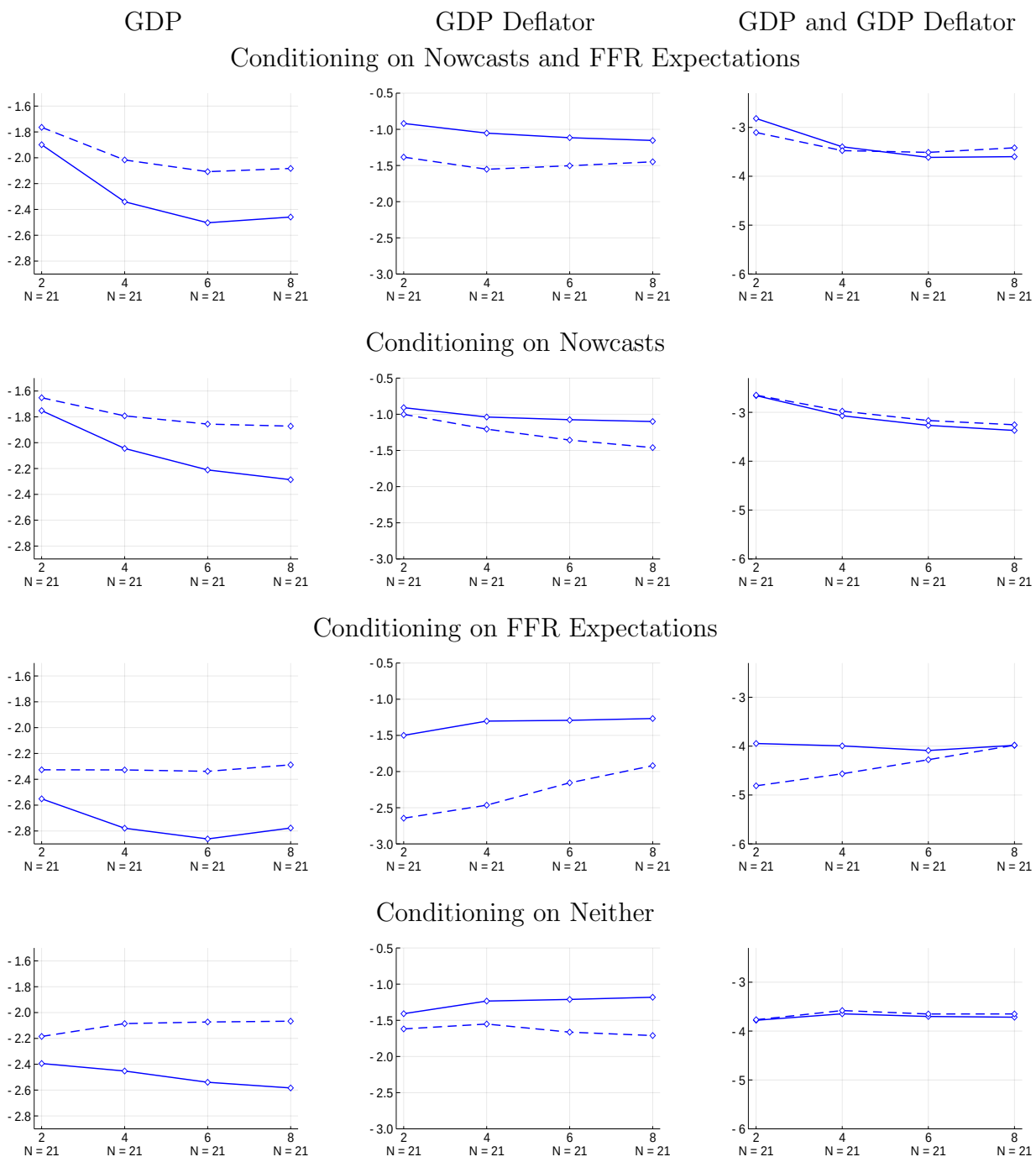
Note: These panels compare the log predictive scores from the SWFF DSGE model estimated with a standard prior (red solid) with the same model estimated with a diffuse prior (red dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations.

Figure A-12: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SWπ model



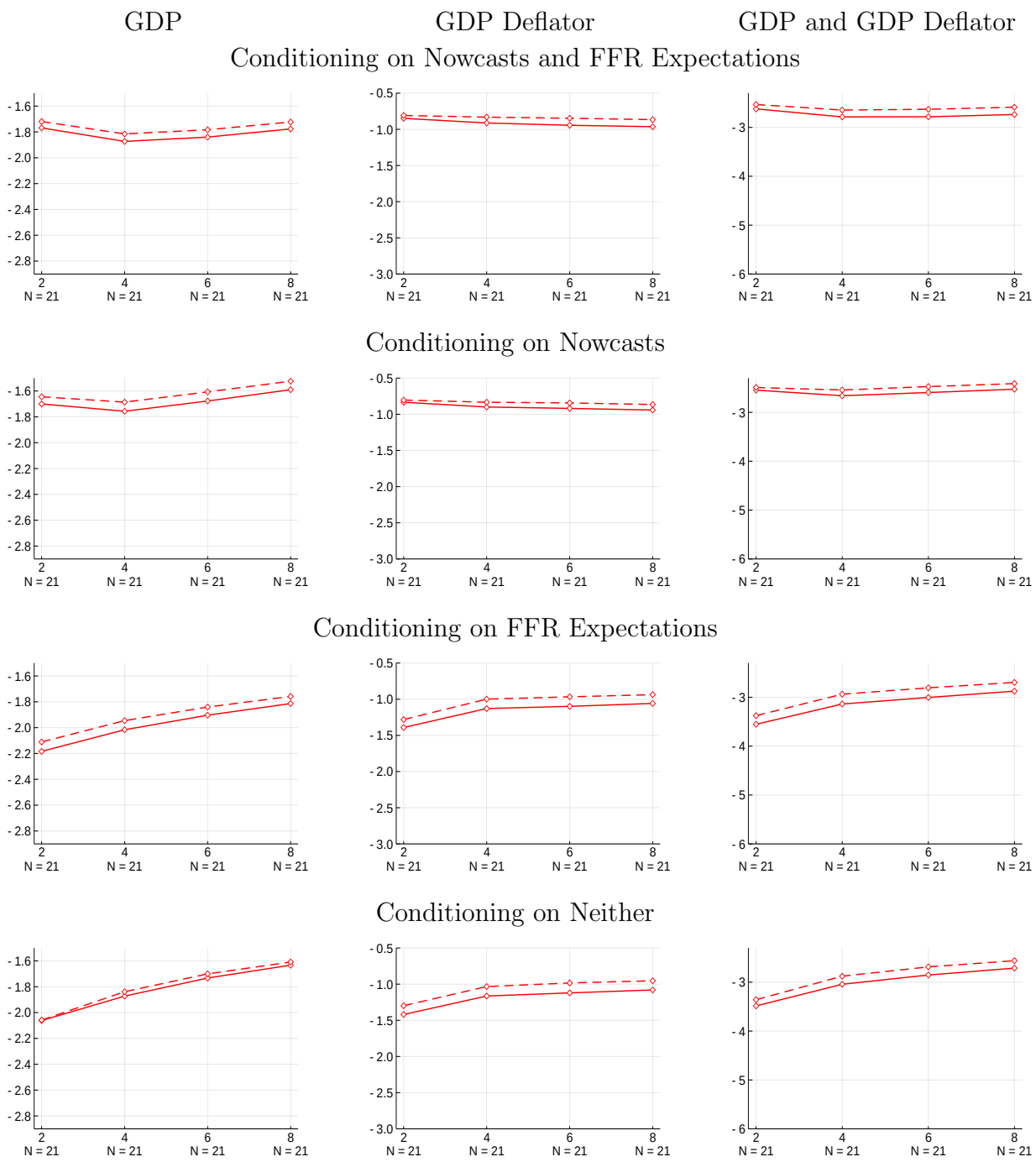
Note: These panels compare the log predictive scores from the SWπ DSGE model estimated with a standard prior (purple solid) with the same model estimated with a diffuse prior (purple dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from January 1992 to January 2017 only are included in these calculations.

Figure A-13: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SW model: Post-Recession Sample



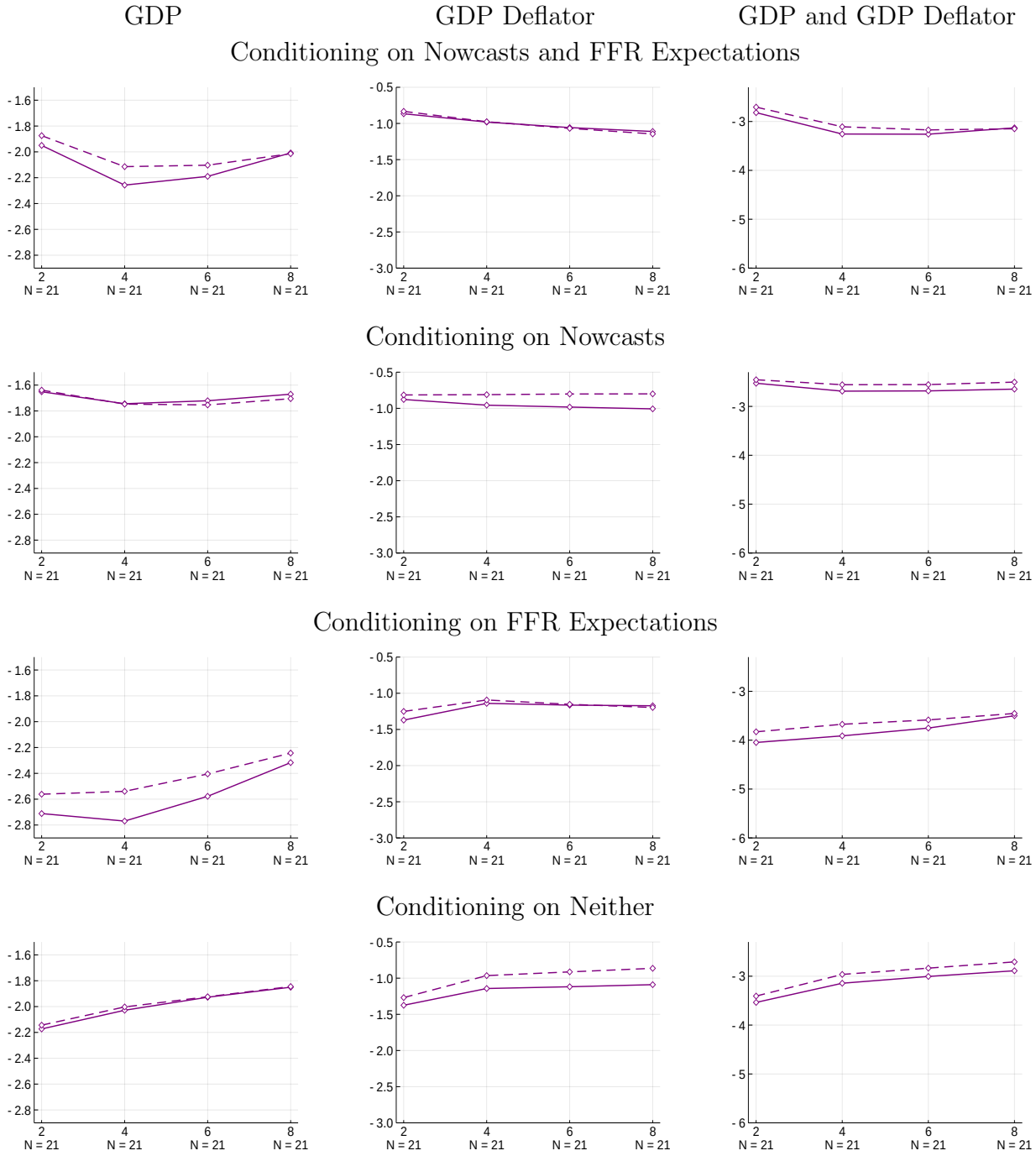
Note: These panels compare the log predictive scores from the SW DSGE Model estimated with a standard prior (blue solid) with the same model estimated with a diffuse prior (blue dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.

Figure A-14: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SWFF model: Post-Recession Sample



Note: These panels compare the log predictive scores from the SWFF DSGE Model estimated with a standard prior (red solid) with the same model estimated with a diffuse prior (red dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.

Figure A-15: Comparison of Log Predictive Scores under "Standard" and "Diffuse" Priors, SWπ model: Post-Recession Sample



Note: These panels compare the log predictive scores from the SWπ DSGE Model estimated with a standard prior (purple solid) with the same model estimated with a diffuse prior (purple dashed) averaged over two, four, six, and eight quarter horizons for output growth and inflation individually, and for both together. Forecast origins from April 2011 to April 2016 only are included in these calculations.