Prior Selection for VAR Forecasts

Nayoung Lee *

October 26, 2026

Abstract

Selecting the shrinkage hyperparameter is crucial for forecasting with parameter-rich VAR models. However, the approach of Giannone, Lenza, and Primiceri (2015) raises two concerns for forecasting: sensitivity to the choice of hyperprior and excessive weight on in-sample fit relative to predictive variance in the marginal likelihood. To address these issues, this paper employs Relative Entropy Prior Sensitivity (REPS) developed by Ho (2023), which provides a systematic way to assess sensitivity to the hyperprior and to examine how the marginal likelihood balances its two components while remaining applicable to VARs of any size. I find that, relative to GLP, a REPS-based choice of the shrinkage hyperparameter improves forecast accuracy particularly (i) during the financial crisis, (ii) for GDP growth, inflation, and the federal funds rate, and (iii) at longer forecast horizons. In normal times, by contrast, it slightly reduces accuracy. These gains arise because REPS reduces in-sample fit in exchange for lower predictive variance. This trade-off translates into forecast accuracy improvements when the dynamics in the estimation sample do not persist out of sample, as during the structural break associated with the financial crisis

1 Introduction

Forecasting macroeconomic variables is important for policymakers because policy decisions depend not only on current economic conditions but also on expected future developments. Since Sims (1980), vector autoregressions (VARs) have been widely used for this purpose because they provide a flexible way to summarize the dynamic relations among macroeconomic variables.

A well-known difficulty of VAR models is that they are rich in parameters. This problem becomes more severe as the number of variables increases. Bayesian VARs address this issue by introducing shrinkage priors, which reduce estimation uncertainty by shrinking VAR coefficients toward more parsimonious specifications. This is especially important for medium- and large-sized VARs, where the number of coefficients can be large relative to the sample size. Prior shrinkage therefore plays a central role in allowing VARs to use larger information sets without losing forecast accuracy.

Crucially, the forecasting performance of Bayesian VARs depends on the choice of the hyperparameters governing the strength of shrinkage. In the Minnesota prior, the overall tightness parameter, λ, determines how strongly the VAR coefficients are shrunk toward their prior means. A larger λ allows the data to speak more, while a smaller λ imposes tighter shrinkage. If λ is too large, the model may retain unstable or weakly estimated dynamics. If λ is too small, the

*Department of Economics, Indiana University. E-mail: nayolee@iu.edu

model may suppress relationships that remain useful for forecasting. This makes hyperparameter selection a key part of Bayesian VAR forecasting.

Giannone, Lenza, and Primiceri (2015; hereafter GLP) propose a widely used approach to select these hyperparameters. Instead of fixing the hyperparameters subjectively, they treat them as unknown parameters and place prior distributions over them. The hyperparameters are then estimated by combining the hyperprior with the marginal likelihood. This approach has become standard practice because it replaces subjective hyperparameter choices with estimation and, under conjugate priors, is easy to implement.

However, the GLP approach raises two concerns for forecasting. First, the posterior estimate of the shrinkage hyperparameter depends on the choice of the hyperprior. Different hyperpriors can lead to different degrees of shrinkage, affecting forecasting performance. Second, the marginal likelihood balances in-sample fit with predictive variance, which may not align with the forecasting objective. An alternative value of the hyperparameter that reduces both in-sample fit and predictive variance may improve density forecasts by producing a sharper predictive distribution.

To address these concerns, this paper applies a Relative Entropy Prior Sensitivity (REPS)-based prior selection approach to Bayesian VAR forecasting. I focus on the Minnesota prior and its overall tightness hyperparameter, λ. Relative Entropy Prior Sensitivity, developed by Ho (2023), is a global prior sensitivity method that uses relative entropy to measure the distance between a baseline prior and an alternative prior. I use REPS to construct an entropy-bounded alternative hyperprior and obtain the corresponding REPS-based value of λ. I then compare forecasts generated using this value with forecasts generated using posterior draws of λ from GLP.

REPS is well suited to this analysis for three reasons. First, it provides a disciplined way to move away from the baseline hyperprior through relative entropy, without selecting an arbitrary alternative, and thus to assess prior sensitivity. Second, the REPS solution relies on the marginal likelihood, the exact same object GLP uses. This allows me to examine how rebalancing the in-sample fit and predictive variance components of the marginal likelihood affects forecasting performance. Third, it is computationally feasible for small, medium, and large VARs.

Empirically, I use the FRED-QD dataset and forecast four macroeconomic variables: real GDP growth, GDP deflator inflation, the unemployment rate, and the federal funds rate. I consider small, medium, and large VARs and forecast horizons from one to eight quarters. The evaluation periods include a pre-crisis sample, a longer sample containing the financial crisis and COVID, a financial-crisis period, and a COVID period. Point forecasts are evaluated using root mean squared forecast error (RMSFE), and density forecasts are evaluated using log predictive scores.

I find that REPS, by rebalancing the in-sample fit and predictive variance components of the marginal likelihood, improves both point and density forecasts, particularly (i) during the financial-crisis evaluation period, (ii) for GDP growth, inflation, and the federal funds rate, and (iii) at longer forecast horizons. Outside these cases, deviating from the GLP hyperparameter does not generally improve density forecasts. In the pre-crisis sample, the REPS-based hyperparameter performs slightly worse, particularly for inflation and the federal funds rate. During COVID, REPS modestly improves point forecasts, but these gains do not carry over to density forecasts.

To understand the underlying mechanism, I first examine how the REPS-based λ changes the marginal likelihood that GLP use to select λ. I decompose the marginal likelihood into an in-sample fit component and a predictive variance component. The decomposition shows that in every model size and evaluation period, the smaller REPS-based λ reduces the in-sample fit but also the predictive variance, generating a sharper predictive distribution. A sharper distribution, however, does not necessarily imply more accurate forecasts.

Whether this accuracy gain is realized depends on the forecasting value of the in-sample fit given up by REPS, that is, on whether the dynamics captured in the estimation sample remain useful during the forecast period. When these dynamics persist, as in the pre-crisis sample, sacrificing in-sample fit is costly and the REPS-based prior generally worsens point forecast accuracy. However, when they do not persist, as during the financial crisis, the simpler dynamics implied by stronger shrinkage can provide more accurate point forecasts, particularly at longer horizons.

The variable-level results corroborate how this mechanism operates. During the financial crisis, VAR dynamics estimated largely from the pre-crisis period implied a stronger rebound in GDP growth, greater disinflation, and an earlier exit from the effective lower bound than actually occurred. Tighter shrinkage by REPS attenuates these predictions and generates simpler paths that more closely match the persistent post-crisis period. Unemployment, however, is an exception. Its dynamics were smooth and persistent, and its relation to output remained stable through the crisis. The estimated dynamics, therefore, remained informative. By pulling the unemployment equation toward its prior, the REPS-based prior weakens these dynamics and reduces forecast accuracy.

These point forecast results determine whether the sharper predictive distributions under REPS improve density forecasts. A more concentrated distribution raises density forecast accuracy when it is centered on an accurate point forecast, but lowers it when centered on an inaccurate one. During the financial crisis, REPS produces sharper and well-centered predictive distributions, because the dynamics in the estimation sample do not persist through the structural break and the in-sample fit lost under REPS has little forecasting value. In the pre-crisis sample, by contrast, these dynamics persist out of sample, so REPS produces sharper but off-center distributions, and density forecasts deteriorate.

COVID is an exception. REPS improves point forecasts modestly, but these gains do not translate into better density forecasts. What distinguishes COVID from the financial crisis is how the economy evolved after the shock. After the financial crisis, output, inflation, and the policy rate departed from pre-crisis dynamics for an extended period and evolved slowly, close to the simple paths implied by the prior. During COVID, by contrast, unusually large shocks were followed by rapid reversals that neither the estimated dynamics nor the prior could track. Tighter predictive distributions therefore understate uncertainty and assign too little probability to these large realizations. Tighter shrinkage thus helps not whenever pre-crisis dynamics break down, but when the economy settles into a prolonged period that the prior approximates better than the estimated dynamics do.

Lastly, I document substantial heterogeneity in the size of the REPS adjustment across model sizes and periods and show that it reflects differences in the shape of the marginal likelihood.

This paper is related to three strands of literature. First,it contributes to the literature on hyperparameter selection in Bayesian VARs. While GLP provide a hierarchical method that estimates hyperparameters using the marginal likelihood, this paper asks whether the hyperparameter it selects is always appropriate for forecasting. The results show that moving toward tighter shrinkage can improve forecasts when the relationships captured by the estimation sample fail to persist, but can reduce accuracy when those relationships remain stable.

Second, it contributes to the literature on Bayesian VAR forecasting with shrinkage priors. Banbura et al. (2010) show that large Bayesian VARs forecast well when the degree of shrinkage increases with model size, while Koop (2013) compares medium and large Bayesian VARs under alternative prior specifications. More recent work develops richer, flexible shrinkage schemes, such as global-local and adaptive priors (e.g., Huber and Feldkircher, 2019; Cross, Hou, and Poon, 2020). While recent studies largely focus on comparing different prior families, this paper stays within the widely used Minnesota prior framework and systematically examines how alternative choices of its overall tightness affect forecasts across model sizes and economic periods.

Third, this paper contributes to the literature on prior sensitivity in Bayesian forecasting. Feldkircher (2012), Chan et al. (2019), and Jacobi et al. (2025) study how Bayesian forecasts and other posterior quantities respond to prior assumptions. Ho (2023) develops a global robust Bayesian analysis method that is computationally feasible in large models. This paper applies REPS to Bayesian VAR forecasting and uses its output to study both how sensitive forecasts are to the shrinkage hyperparameter and whether moving away from the GLP choice improves forecast accuracy.

The paper makes three main contributions. First, it applies an entropy-bounded prior distortion to the selection of the Minnesota tightness parameter and evaluates the resulting forecasts across model sizes, horizons, and economic periods. Second, it shows that the GLP-based hyperparameter need not be optimal for out-of-sample forecasting. The value of moving toward tighter shrinkage depends on whether the estimated dynamics persist out of sample, and it varies across forecasted periods, variables, and horizons. Third, it documents heterogeneity in the REPS-selected shrinkage hyperparameter across model sizes and periods, and shows that this variation is driven by the shape of the marginal likelihood.

The remainder of the paper is organized as follows. Section 2 introduces the VAR model and the Minnesota prior. Section 3 describes the forecasting methodology and forecast evaluation metrics. Section 4 presents the REPS framework. Section 5 describes the data. Section 6 presents the empirical results, including density forecasts, point forecasts, and the marginal-likelihood decomposition. Section 7 concludes.

2 Model

2.1 VAR Model

In this paper, I study the following VAR model with p lags:

$$
\begin{aligned}
y_t &= C + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \\
u_t &\sim \mathcal{N}(0, \Sigma)
\end{aligned}
$$

where y_t is a $N \times 1$ vector of endogenous variables, u_t is a $N \times 1$ vector of exogenous shocks, and C, B_j for $j = 1, \cdots, p$ and Σ are of dimensions $N \times 1$, $N \times N$, and $N \times N$, respectively, with unknown parameters.

Stacking over T observations, I obtain:

$$\underbrace{\begin{pmatrix} y_t' \\ \vdots \\ \vdots \\ y_{t+T-1}' \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} y_{t-1}' & \cdots & y_{t-p}' & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ y_{t+T-2}' & \cdots & y_{t+T-1-p}' & 1 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} B_1' \\ \vdots \\ B_p' \\ C' \end{pmatrix}}_{B} + \underbrace{\begin{pmatrix} u_t' \\ \vdots \\ \vdots \\ u_{t+T-1}' \end{pmatrix}}_{U}$$

$$Y = XB + U$$

where Y is a $T \times N$ matrix of endogenous variables, X is a $T \times (Np + 1)$ matrix of regressors, U is a $T \times N$ matrix of shocks, and B is a $(Np + 1) \times N$ matrix with unknown parameters.

I focus on prior distributions for VAR coefficients belonging to the following normal-inverse-Wishart family:

$$\begin{aligned} \Sigma &\sim \mathcal{IW}_N(\underline{S}, \underline{v}) \\ B|\Sigma &\sim \mathcal{MN}_{(Np+1)\times N}(\underline{B}, \underline{V}, \Sigma) \\ \text{vec}(B)|\Sigma &\sim \mathcal{N}_{(Np+1)N\times 1}(\text{vec}(\underline{B}), \Sigma \otimes \underline{V}) \end{aligned}$$

where $\underline{B}$ is a $(Np + 1) \times N$ prior mean matrix, $\underline{V}$ is a $(Np + 1) \times (Np + 1)$ matrix governing column-specific variances, $\underline{S}$ is an $N \times N$ scale matrix, and $\underline{v}$ is the degrees of freedom parameter.

Σ : For the prior on Σ, to guarantee the existence of its prior mean, I follow Kadiyala and Karlsson (1997) and set the degrees of freedom for the inverse-Wishart distribution to $\underline{\nu} = N+2$. Additionally, I take $\underline{S}$ to be a diagonal matrix with $\hat{\sigma}^2$ on the diagonal, where $\hat{\sigma}^2$ is estimated as the residual variance from a univariate AR(1) process.

B : For the prior on B, I adopt the *Minnesota prior*, originally proposed by Doan et al. (1984). This prior is motivated by the empirical observation that many macroeconomic time series behave like unit root processes. It imposes shrinkage that helps reduce the complexity of parameter-rich VAR models by shrinking them toward more parsimonious specifications. I discuss this prior in greater detail in Section 2.2, *Shrinkage priors.*

The posterior distribution is then given by:

$$\begin{aligned} \Sigma|Y &\sim \mathcal{IW}_N(\bar{S}, \bar{v}) \\ B|Y, \Sigma &\sim \mathcal{MN}_{(Np+1)\times N}(\bar{B}, \Sigma, \bar{V}) \end{aligned}$$

where:

$$\begin{aligned} \bar{B} &= \bar{V}(\underline{V}^{-1}\underline{B} + X^\top Y), \quad \bar{V} = (\underline{V}^{-1} + X^\top X)^{-1} \\ \bar{S} &= \underline{S} + Y^\top Y + \underline{B}^\top \underline{V}^{-1}\underline{B} - \bar{B}^\top \bar{V}^{-1}\bar{B}, \quad \bar{v} = \underline{v} + T \end{aligned}$$

This paper adheres to the definitions of small, medium, and large VARs provided by GLP. This is different from the classification in Banbura et al. (2010), which could be explored in future research.

1. *SMALL*

   This is a monetary VAR with three key variables —GDP, the GDP deflator, and the federal funds rate— plus the unemployment rate.

2. *MEDIUM*

   This medium-scale VAR includes the eight variables used in the estimation of the DSGE model by Smets and Wouters (2007) for the U.S. economy. It adds real consumption, real investment, hours worked, and real compensation to the small VAR

3. *LARGE*

   This large-scale VAR includes 23 variables that encompass the previous two specifications and additionally incorporate several important labor market, financial, and monetary variables.

2.2 Shrinkage Priors

VARs tend to be rich in parameters, and this problem intensifies with the size of the model. To mitigate this problem, researchers often introduce shrinkage priors that "shrink" the richly parameterized unrestricted VAR model toward a more parsimonious specification. Here, I introduce the most widely used shrinkage prior for the VAR coefficients—the Minnesota prior.

The *Minnesota prior* is motivated by the empirical observation that many macroeconomic time series behave approximately as unit root processes. Accordingly, the prior mean on the VAR coefficients is constructed to reflect this belief by centering the prior mean for the first own-lag of each variable around one, while shrinking all other coefficients toward zero.

$$\mathbb{E}(\mathrm{B}_\ell^{ij}) = \begin{cases} d_i & \text{if } i = j \text{ and } \ell = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $d_i = 1$ for variables modeled in levels and $d_i = 0$ for stationary variables (e.g., in differences). ℓ is the lag order, and $i, j = 1, \cdots, N$. This implies the following prior mean matrix for the system of VAR equations:

$$\underline{\mathbf{B}} = \left(I_N; 0_{N\times N(p-1)}; 0_{1\times N}\right)$$

The matrix $\underline{V}$ determines the prior covariance matrix of B and is assumed to be diagonal, reflecting the lack of strong prior beliefs about correlations between parameters.

$$\Sigma \otimes \underline{V} = \begin{pmatrix} \Sigma_{11}\underline{V} & \dots & \Sigma_{1N}\underline{V} \\ \vdots & \vdots & \vdots \\ \Sigma_{N1}\underline{V} & \dots & \Sigma_{NN}\underline{V} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11}\begin{pmatrix} \underline{V}_{11} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \underline{V}_{(Np+1)(Np+1)} \end{pmatrix} & \dots & \Sigma_{1N}\begin{pmatrix} \underline{V}_{11} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \underline{V}_{(Np+1)(Np+1)} \end{pmatrix} \\ \vdots & \vdots & \vdots \\ \Sigma_{N1}\begin{pmatrix} \underline{V}_{11} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \underline{V}_{(Np+1)(Np+1)} \end{pmatrix} & \dots & \Sigma_{NN}\begin{pmatrix} \underline{V}_{11} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \underline{V}_{(Np+1)(Np+1)} \end{pmatrix} \end{pmatrix}$$

The diagonal elements of $\underline{V}$ govern the prior variances of the VAR coefficients. In the Minnesota prior, the prior reflects beliefs about the relative importance of lags and cross-variable relationships. Specifically, it assigns smaller variances to coefficients on more distant lags, reflecting the belief that higher-order lags are less important. In addition, it rescales the prior variances of coefficients based on the relative variances of the explanatory and dependent variables.

Formally, the diagonal elements of $\underline{V}$ are given by:

$$\underline{\mathbf{V}}_{ii} = \frac{\lambda}{\ell^2 \hat{\sigma}_i^2} \text{ where } i = (\ell - 1)N + n + 1,\ \ell = 1, \cdots, p,\ n = 1, \cdots, N$$

where ℓ is the lag order, $\hat{\sigma}_i^2$ is the estimated residual variance from an AR(1) model for variable i, and λ is a hyperparameter governing overall tightness.

Combining this with the error variance Σ_{jj} from the jth equation, the implied prior variance for the coefficient on the pth lag of variable i in jth equation is:

$$\text{Var}(\text{B}_p^{ij}) = \frac{\lambda\, \hat{\sigma}_j^2}{p^2\, \hat{\sigma}_i^2}$$

The key hyperparameter is λ, which controls the overall tightness of the prior by scaling the variances of the coefficient distributions. As such, the choice of λ is critical.

Given the central role of λ in shrinking the VAR models, the main analysis focuses exclusively on cases where the Minnesota prior is used. The analysis incorporating the dummy-initial-observation priors alongside the Minnesota prior is presented in the appendix.

2.3 Prior Selection: Hyperparameter Estimation by GLP

Prior to GLP, economists typically adopted values of λ from previous studies, even when using different datasets or model sizes. Instead, GLP treat λ as an unknown parameter and estimate it by placing a prior distribution over it. Following their specification, I impose a Gamma hyperprior with a mode of 0.2 and a standard deviation of 0.4

When estimating hyperparameters using the Metropolis-Hastings algorithm, GLP use the following marginal distribution as the likelihood for the hyperparameters. This is referred to as the marginal likelihood of the data as a function of the hyperparameters, since it is obtained by integrating out the VAR coefficients.

$$\begin{aligned} p(y|\lambda) &= \int p(y|\theta, \lambda)\ p(\theta|\lambda)\ d\theta \\ &\propto \underbrace{\left| \left(V_\epsilon^{\text{posterior}}\right)^{-1} V_\epsilon^{\text{prior}} \right|^{\frac{T-p+d}{2}}}_{\text{In-Sample Fit}} \quad \cdot \quad \underbrace{\prod_{t=p+1}^{T} \left|V_{t|t-1}\right|^{-\frac{1}{2}}}_{\text{Predictive Variance}} \end{aligned}$$

where θ denotes the VAR coefficients, λ represents the tightness hyperparameter, $V_\epsilon^{\text{posterior}}$ and $V_\epsilon^{\text{prior}}$ are the posterior and prior means of the residual variance, and $V_{t|t-1}$ is the variance of the one-step-ahead forecast of y, averaged prior draws of Σ. They argue that this method effectively selects hyperparameters by balancing the trade-off between in-sample fit and predictive variance. Moreover, it admits a closed-form expression under conjugate priors, allowing for easy numerical implementation.

However, the GLP approach raises two concerns for forecasting. First, the posterior estimate remains sensitive to the choice of the hyperprior. Different hyperpriors can lead to different degrees of shrinkage, affecting forecasting performance. Second, the marginal likelihood balances in-sample fit with predictive variance, which may not align with the forecasting objective. An alternative value of the hyperparameter that reduces in-sample fit may improve density forecasts by producing a sharper predictive distribution. To examine whether alternative hyperparameters can improve such forecasts, albeit at the cost of in-sample fit, this paper uses the Relative Entropy Prior Sensitivity tool introduced by Ho (2023), which is discussed in detail in Sections 4 and 6.

I generated 10^5 hyperparameter prior and posterior draws for each estimation sample. Posterior draws were obtained using the Metropolis-Hastings algorithm, with 2×10^4 burn-in draws. This was followed by 10^6 additional draws, of which one in ten was retained.

For each draw of the posterior hyperparameters, I then estimated the VAR coefficients using closed-form expressions for their posterior distribution. Consequently, I generated 10^5 estimates of the VAR coefficients for each estimation sample. Alternatively, one can use the posterior mean of the hyperparameter draws to estimate the VAR coefficients. While this paper focuses on the former approach, results using the latter method are reported in Appendix A.

3 Forecasts

This section introduces the forecasting method and the metrics used to evaluate forecast performance. Based on the iterated forecasting approach, I generate one- to eight-quarter-ahead forecasts for real GDP growth, GDP-deflator inflation, the unemployment rate, and the federal funds rate—both point forecasts and density forecasts—using data from the past 20 years. Point forecasts and density forecasts are then evaluated using the root mean squared forecast error (RMSFE) and the log predictive score, respectively.

3.1 Forecasting Method

I forecast four macroeconomic indicators: real GDP growth, GDP deflator inflation, the unemployment rate, and the federal funds rate, using small ($n = 4$), medium ($n = 8$), and large ($n = 23$) VAR models with $p = 4$ lags. Specifically, I employ an iterated forecasting approach with a rolling window to generate a series of one- to eight-step-ahead forecasts for each variable.

Let $\mathbf{y}_{1:T}$ denote the estimation sample. The iterated forecasting method first constructs the companion form of the VAR model and then generates h-step-head forecasts iteratively as follows:

VAR:

$$y_t = C + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \text{ for } t = p+1, \cdots, T$$

Companion form:

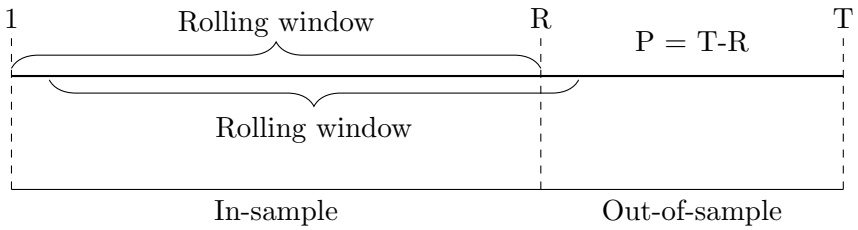
$$\underbrace{\begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p} \end{bmatrix}}_{X_{t+1}} = \begin{bmatrix} 1 & 0 & \dots 0 & \\ C & B_1 & \dots & B_p \\ 0 & I_N & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p-1} \end{bmatrix} + \begin{bmatrix} 0 \\ u_t \\ \vdots \\ 0 \end{bmatrix}$$

$$X_{t+1} = (B^+)X_t + \tilde{u}_t$$

Multi-step (h-step-ahead) forecast:

$$\hat{X}_{t+h} = (B^+)\hat{X}_{t+h-1} + \tilde{u}_{t+h} \text{ for } h = 1, \cdots, 8$$

The second row of X_{t+1} corresponds to a draw from the posterior predictive density of y_{T+h}. Given the 100,000 estimates of the VAR coefficients from the sample $\mathbf{y}_{1:T}$ in Section 2, I generate 100,000 draws from the predictive density of y_{T+h}. I then iterate this procedure by updating the estimation sample by shifting the sample forward one quarter to $\mathbf{y}_{2:T+1}$, re-estimating the VAR, and generating forecasts of y_{T+1+h}. This process is repeated one quarter at a time until the end of the sample.



More precisely, starting from a sample ranging from 1968Q2 to 1988Q1, I generate draws from the posterior predictive density of the model for 1988Q2 (one quarter ahead) and 1990Q1 (two years ahead). I then update the estimation sample, one quarter at a time, until the end of the sample, 2023Q4. At each rolling window, I re-estimate the model and produce F_{draws} from the posterior predictive density.

Table 1. Forecasting Periods Outline

Periods	t=1	t=R+1	t=T-7	t=T	P
Sample 1	Apr, 1968	Apr, 1988	Apr, 2006	Jan, 2008	73
Sample 2	Apr, 1968	Apr, 1988	Jan, 2023	Oct, 2024	140
Financial Crisis	Oct, 1987	Oct, 2007	Oct, 2015	Oct, 2017	33
COVID	Jan, 2000	Jan, 2020	Jan, 2022	Oct, 2023	9

I repeat this procedure for four different periods: Sample 1, Sample 2, the financial crisis, and COVID. This procedure yields F_{draw}, a time series of P density forecasts for each variable and forecast horizon. I use the median of each distribution as the point forecast.

(1) Entire Sample: Sample 1 excludes the financial crisis and COVID periods by ending before the financial crisis, whereas Sample 2 includes all available data.

(2) Unusual Periods: I define unusual periods (e.g. financial crisis, COVID) based on the Fed's policy decisions. Specifically, I set the first one-step-ahead forecast for these periods (t=R+1) as the time when the Fed began lowering the federal funds rate. Likewise, I set the last one-step-ahead forecast for these periods (t=T-7) as the time when the Fed started raising the federal funds rate. This definition reflects the role of forecasting in informing policy decisions.

F_{draw} is equal to the number of hyperparameter posterior draws, obtained from the estimation procedure following GLP in Section 2. For each draw, I sample the corresponding VAR coefficients and generate one- to eight-step-ahead forecasts using the iterated method. This results in 100,000 forecasts for each variable i and forecast horizon h across rolling windows. In contrast, Ho's approach provides a single hyperparameter value. Using this fixed value, I generate 100,000 draws of VAR coefficients and construct forecasts based on each draw.

As a result, I obtain the set of h-step ahead forecasts for variable i's at the t th rolling window, represented as $\left\{y_{i,h}^{t,f}\right\}_{f=1}^{100,000}$, for both Ho's method and GLP's method. The point forecast, $\hat{y}_{i,h}^{t}$, is defined as the median of this set. Let $y_{i,h}^{t}$ denote the realized value for $t = 1, \cdots, P$. Then, the forecast error is given by $e_{i,h}^{t} = \hat{y}_{i,h}^{t} - y_{i,h}^{t}$ for $t = 1, \cdots, P$.

Algorithm:

1. Generate F_{draw} draws of the hyperparameter λ using the approach of GLP.
2. For each draw of λ, compute the corresponding estimates of the VAR coefficients.
3. Given the VAR coefficient estimates, generate variable i's h-step-ahead forecast using the iterated method.

$\Rightarrow$ Obtain F_{draw} draws from the forecast density, from which point forecasts can be computed.

3.2 Measure of Forecasting Accuracy

3.2.1 Point Forecasts

I measure the accuracy of a point forecast using the root mean squared forecast error (RMSFE) for variable i and forecast horizon h where $i \in \{$real GDP growth, GDP-deflator inflation, unemployment rate, federal funds rate$\}$ and $h = \{1, \cdots, 8\}$. The RMSFE of variable i at horizon h is computed as follows:

$$\text{RMSFE}_{i,h} = \sqrt{\frac{1}{P}\sum_{t=1}^{P}(y_{t+h}^{i} - \hat{y}_{t+h}^{i})^2}$$

where y_{t+h}^{i} denotes the actual value of variable i at time $t+h$, $\hat{y}_{t+h}^{i}$ denotes the median of the predictive density of variable i at time $t+h$, P denotes the length of out-of-sample forecast period, and F_{draw} is the number of draws from the predictive density of variable i at horizon h. A lower RMSFE indicates a superior forecast performance.

To compare the point forecast accuracy of two models using different hyperparameter selection methods —REPS (**REPS**) versus Giannone et al.'s method (**GLP**)—, I compute the percentage change in RMSFE for variable i at horizon h as follows:

$$\left(1 - \frac{\text{RMSFE}_{i,h}^{\mathbf{REPS}}}{\text{RMSFE}_{i,h}^{\mathbf{GLP}}}\right) \times 100$$

A positive value indicates that the **REPS** method outperforms **GLP** method in forecast accuracy.

3.2.2 Density Forecasts

I measure the accuracy of density forecasts using the average log predictive score (SCORE) of variable i and horizon h for $i \in \{\text{real GDP growth, GDP-deflator inflation, unemployment rate, federal funds rate}\}$ and $h = \{1, \cdots, 8\}$. The SCORE of variable i at horizon h is computed as:

$$\text{SCORE}_{i,h} = \frac{1}{P} \sum_{t=1}^{P} s_t(y_{t+h}^i)$$

where the log predictive score $s_t(y_{t+h}^i)$ is computed using a quadratic approximation of the Gaussian predictive distribution, following Adolfson et al. (2007):

$$s_t(y_{t+h}^i) = -0.5 \left[\ln(2\pi) + \ln(V_{t+h|t}^i) + \frac{(y_{t+h}^i - \overline{y}_{t+h|t}^i)^2}{V_{t+h|t}^i} \right]$$

for $t = 1, \cdots, P$, where, given F_{draw} from the forecast distribution of y_{t+h}^i, $\left\{ y_{t+h,f}^i \right\}_{f=1}^{F_{\text{draw}}}$, the posterior mean $\overline{y}_{t+h|t}^i$ and the posterior variance $V_{t+h|t}^i$ are computed from the forecast distribution, and y_{t+h}^i denotes the actual value. Since the log score evaluates the predictive density at the actual value, a higher log score indicates superior forecast performance.

To compare the density forecast accuracy of two models—REPS (**REPS**) versus Giannone et al.'s method (**GLP**)—I compute the percentage gain in SCORE for variable i at horizon h as follows:

$$(\text{SCORE}_{i,h}^{\mathbf{REPS}} - \text{SCORE}_{i,h}^{\mathbf{GLP}}) \times 100$$

A positive value indicates that the **REPS** method outperforms the **GLP** method.

3.3 Tests of Equal Forecast Accuracy

Once I have obtained the forecast accuracy metrics, I need a method to assess whether the forecasting performance of the VAR model differs significantly depending on the hyperparameter selection method. To determine the appropriate testing procedure, it is important to note that this paper employs rolling window estimation to generate unconditional forecasts from potentially nested models. Based on this setup, I employ two tests for point forecast evaluation: (i) the Diebold-Mariano (1995) test and (ii) the Clark and West (2007) test. For density forecast evaluation, I adopt the test proposed by Amisano and Giacomini (2007).

3.3.1 Diebold and Mariano (1995) Test

For point forecasts, the use of rolling window estimation justifies the application of the forecast accuracy test proposed by Giacomini and White (2006), which accounts for parameter estimation uncertainty that may persist in this setting. Giacomini and White (2006) show that, for unconditional forecasts, their test statistic is equivalent to that of Diebold and Mariano (1995). Given this equivalence, I proceed with the Diebold-Mariano (1995) test to evaluate forecast accuracy.

Let $d_t = \left(e_{i,h}^{t,\mathbf{REPS}}\right)^2 - \left(e_{i,h}^{t,\mathbf{GLP}}\right)^2$ for $t = 1, \ldots, P$, where $e_{i,h}^{t,\cdot}$ denote the forecast error for variable i at horizon h. Let $\bar{d}$ denote the sample mean of d_t, and estimate the variance of d_t, $\text{var}(d)$, using a heteroskedasticity and autocorrelation consistent (HAC) estimator. The Diebold-Mariano test statistic is then given by:

$$\text{DM} = \frac{\bar{d}}{\sqrt{\frac{1}{P}\,\text{var}(d)}} \sim N(0,1).$$

If the null hypothesis of equal forecast accuracy is rejected, it indicates a statistically significant difference in the predictive performance of the two models.

3.3.2 Clark and West (2007) Test

Since the VAR models considered in this paper include the same set of variables and differ only in their hyperparameter values, the Clark and West (2007) test for nested models is also appropriate. I therefore include the Clark and West (2007) test as a complementary evaluation of forecast accuracy under nested models.

When comparing the mean squared forecast errors (MSFE) of nested models, the Clark-West test corrects for the upward bias in the MSFE of the larger model. To implement the test, define:

$$f_{i,h}^t = \left(e_{i,h}^{t,\mathbf{GLP}}\right)^2 - \left[\left(e_{i,h}^{t,\mathbf{REPS}}\right)^2 - (\hat{y}_{i,h}^{t,\mathbf{GLP}} - \hat{y}_{i,h}^{t,\mathbf{REPS}})^2\right] \text{ for } t = 1, \cdots, P$$

Then, regress $f_{i,h}^t$ on a constant and use the resulting t-statistic for the coefficient on the constant to test whether the MSFEs of two models are equal. Rejection of the null implies that the **REPS** model yields a significantly smaller MSFE than the **GLP** model.

3.3.3 Amisano and Giacomini (2007) Test

For density forecasts, this paper uses the Amisano and Giacomini (2007) t-test of equal means, utilizing the log score for each forecast relative to the benchmark.

In the Amisano and Giacomini (2007) test, the test statistic is defined as follows:

$$\text{WLR}_{i,t+h} = \omega(y_{t+h}^{i,\text{st}}) \cdot \left[s_t(y_{t+h}^i)^{\mathbf{REPS}} - s_t(y_{t+h}^i)^{\mathbf{GLP}}\right] \text{ for } t = 1, \cdots, P$$

where $y_{t+h}^{i,\text{st}} = (y_{t+h}^i - u_t)/\sigma_t$ is the realized value of variable i at period $t+h$ standardized using the mean and standard deviation of the estimation sample, ω is standard normal density function, $s_t(y_{t+h}^i)$ is a log predictive score of variable i at $t+h$ evaluated at t.

$$t_{i,h} = \frac{\overline{\text{WLR}}_{i,h}}{\hat{\sigma}_{i,h}/\sqrt{P}}$$

where $\overline{\text{WLR}}_{i,h}$ denotes the sample mean of $\overline{\text{WLR}}_{i,t+h}$ and $\hat{\sigma}_{i,h}$ is the variance estimate of $\text{WLR}_{i,t+h}$ for $t = 1, \cdots, P$ following the suggestion by Amisano and Giacomini (2007). If the null hypothesis of equal forecast accuracy is rejected, it indicates a statistically significant difference in the predictive performance of the two models.

4 Relative Entropy Prior Sensitivity (REPS)

To explore how VAR forecasts change across hyperparameter values and to examine whether alternative values improve forecast accuracy at the cost of in-sample fit, I use Relative Entropy Prior Sensitivity (REPS) developed by Ho (2023). This approach is particularly well-suited to our analysis for three key reasons: (i) its function as a tool for analyzing the sensitivity of posterior estimates to prior assumptions, (ii) its ability to analyze the role of the likelihood in estimating hyperparameters, and (iii) its applicability to models of any size.

REPS is a global prior sensitivity analysis tool that is computationally feasible for both small and large models, using relative entropy to measure the distance between distributions. This allows it to overcome the computational limitations of existing prior sensitivity methods. Accordingly, REPS enables us to analyze how sensitive posterior estimates—and resulting VAR forecasts—are to changes in prior hyperparameters across different model sizes. This feature of REPS aligns with the first two reasons it is well-suited for our analysis: its ability to assess posterior sensitivity to prior assumptions and its feasibility across model sizes.

The REPS method solves the following problem:

$$\tilde{\lambda} = \min_{M(\lambda):M>0,\mathbb{E}_\pi M(\lambda)=1} \mathbb{E}_p\left[\frac{M(\lambda)\cdot\lambda}{\mathbb{E}_p(M)}\right] \quad \text{s.t. } \mathbb{E}_\pi\left[M(\lambda)\log M(\lambda)\right] \leq \mathcal{R}.$$

Here, λ denotes the Minnesota tightness hyperparameter, $\pi(\lambda)$ is the original prior distribution, and $p(\lambda)$ denotes the corresponding posterior distribution. The alternative prior is defined as $\tilde{\pi}(\lambda) = M(\lambda)\pi(\lambda)$, where $M(\lambda)$ is a distortion function. I impose a constraint that ensures that the distance between the original prior and the alternative prior measured in relative entropy remains below a prespecified value $\mathcal{R}$, for which Ho (2023) provides a data-driven rule of thumb for choosing $\mathcal{R}$. Thus, the goal of REPS is to find the distortion function $M(\lambda)$ that minimizes the distorted posterior mean of λ, $\tilde{\lambda}$, subject to this entropy bound.

Ho (2023) shows that the solution to this problem, $M(\lambda)$, takes the simple exponential form:

$$M(\lambda) \propto \exp\left(\eta \cdot L(X|\lambda)\cdot(\lambda-\tilde{\lambda})\right)$$

where η is the Lagrange multiplier, $L(X|\lambda)$ is the likelihood of the data given λ, and $\tilde{\lambda}$ is the value of the objective function. Intuitively, the distortion function depends on the likelihood because greater distortion from using an alternative prior arises in regions where the likelihood is high.

By substituting the solution back into the objective function, it becomes:

$$\tilde{\lambda} = \mathbb{E}_p\left[\frac{\exp\left(\eta\cdot L(X|\lambda)\cdot(\lambda-\tilde{\lambda})\right)}{\mathbb{E}_p\left[\exp\left(\eta\cdot L(X|\lambda)\cdot(\lambda-\tilde{\lambda})\right)\right]}\lambda\right]$$

Given posterior draws of λ, denoted λ_i, this simplifies to:

$$\tilde{\lambda} \propto \sum_i \exp\left(L(X|\lambda_i)\cdot(\lambda_i-\tilde{\lambda})\right)\lambda_i$$

This formula shows that $\tilde{\lambda}$ is a weighted average of the posterior draws of λ, where the weights depend on the marginal likelihood. Note that $L(X|\lambda_i)$ corresponds to the marginal likelihood

used for hyperparameter estimation in GLP, which is central to our analysis. Because the REPS distortion operates through this likelihood, REPS allows me to examine how the marginal likelihood shapes both the selected hyperparameter and the trade-off between in-sample fit and predictive variance. This feature corresponds to the third reason it is well suited for our analysis: its ability to analyze how the likelihood shapes hyperparameter estimation.

To implement REPS, I first estimate the hyperparameter λ following GLP and obtain posterior draws, along with the corresponding marginal likelihood and prior density for each draw. Second, I solve the REPS minimization problem using these outputs to obtain the **distorted hyperparameter**, also referred to as the **REPS-based hyperparameter**. Third, I estimate the VAR model and generate point and density forecasts using both the posterior draws of λ from GLP and the REPS-based value of λ. Finally, I compare the resulting forecasts using the forecast metrics introduced earlier. This comparison allows me to assess whether the REPS-based hyperparameter improves forecast accuracy. Specifically, I explore improvements across different model sizes, forecast horizons, and time periods.

Algorithm:

1. Estimate λ following Giannone et al. (2015) and store the posterior draws of the hyperparameters, along with the corresponding marginal likelihood and the prior density for each draw.
2. Using the outputs from Step 1, apply REPS to obtain the distorted values of λ.
3. Estimate the VAR model and generate forecasts using the posterior draws of λ from Giannone et al. (2015) and REPS-based λ, respectively.
4. Compare the resulting point and density forecasts using forecast metrics.

In addition to the baseline comparison, I conduct two complementary robustness exercises in the Appendix. First, I replace the GLP posterior draws of λ with their posterior mean and compare the resulting GLP plug-in forecasts with the REPS plug-in forecasts. This exercise removes hyperparameter uncertainty from the GLP benchmark and isolates the role of the selected value of λ. Second, I implement REPS as a posterior reweighting scheme. In this exercise, I retain the GLP posterior draws of λ and the associated predictive distributions, but replace the equal mixture weights used by GLP with weights implied by the REPS distortion function. This preserves the GLP posterior-mixture structure and examines whether the REPS distortion has predictive content beyond the use of a single plug-in hyperparameter. The plug-in exercise yields results close to the baseline, indicating that the gains reflect the selected value of λ rather than the treatment of hyperparameter uncertainty. Under posterior reweighting, the gains become larger and broader in the small VAR. Across both exercises, the gains for GDP growth and inflation remain the most robust, whereas the results for the federal funds rate and the unemployment rate are more sensitive to the forecasting implementation.

5 Data

This paper uses a macroeconomic dataset for the US economy, FRED-QD, constructed by McCracken and Ng (2016) and maintained by the Federal Reserve Bank of St.Louis. The dataset covers the period from 1959Q1 to 2024Q4 and consists of 248 quarterly series. Since the variable

Capacity utilization is only available starting from 1967Q1, my sample begins from that period. I take the log difference and multiply by 400 for all variables except the unemployment rate and the federal funds rate, to make the variables approximately stationary and easier to interpret.

Table 2. Variables and Transformation by Model Size

Variables	**FRED Mnemonic**	**Transformation**	**VAR**	
			Small	**Medium**
Real GDP	GDPC1	400×log-diff	x	x
GDP deflator	GDPCTPI	400×log-diff	x	x
Unemployment rate	UNRATE	level	x	x
Federal funds rate	FEDFUNDS	level	x	x
Real consumption	PCECC96	400×log-diff		x
Real investment	GPDIC1	400×log-diff		x
Hours worked: Business sector	HOANBS	400×log-diff		x
Real compensation per hours	CPMPRNFB	400×log-diff		x
Consumer price index	CPIAUCSL	400×log-diff		
Commodity price	CUSR0000SAC	400×log-diff		
Industrial production	INDPRO	400×log-diff		
Employment	CE160V	400×log-diff		
Employment in the services sector	SRVPRD	400×log-diff		
Real residential investment	PRFIx	400×log-diff		
Nonresidential investment	PNFIx	400×log-diff		
Personal consumption expenditures, price index	PCECTPI	400×log-diff		
Gross private domestic investment, price index	GPDICTPI	400×log-diff		
Capacity utilization	TCU	400×log-diff		
Consumer expectations	UMCSENTx	400×log-diff		
One-year bond rate	GS1	400×log-diff		
Five-years bond rate	GS5	400×log-diff		
S&P500	S&P 500	400×log-diff		
Trade weighted US dollar index	TWEXAFEGSMTHx	400×log-diff		
Real M2	M2SL	400×log-diff		

6 Results

This section establishes the following main results. First, the forecasting gains under REPS are concentrated at longer horizons in the financial-crisis period, and for GDP growth, inflation, and the federal funds rate. Second, these gains arise mainly from forecasts issued after the economy entered a structural break during the financial crisis, a period characterized by a prolonged effective lower bound, a sluggish recovery, and missing disinflation. In this period, the in-sample fit that REPS sacrifices for lower predictive variance has little forecasting value, because the dynamics estimated from the pre-crisis sample no longer persist.

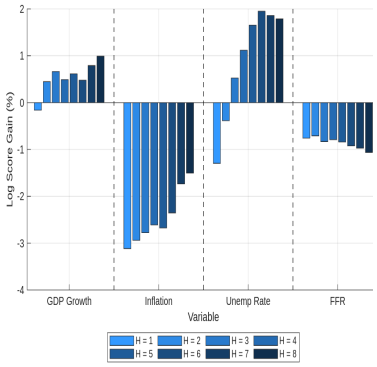
6.1 Density Forecast Accuracy

Because REPS changes the balance between in-sample fit and predictive variance embedded in the marginal likelihood, I begin with density forecast performance. Density forecasts provide the most direct measure of whether the reduction in predictive variance induced by tighter shrinkage is beneficial out of sample.

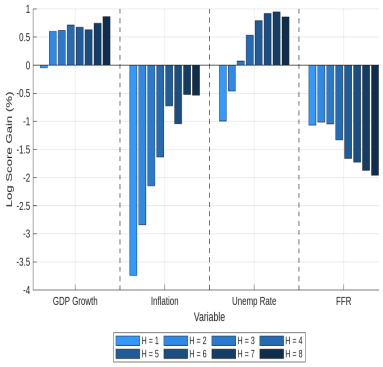
Figure 1. Log Score Difference

1. Sample 1

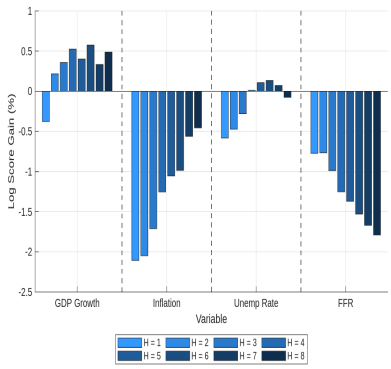
Small VAR



Medium VAR

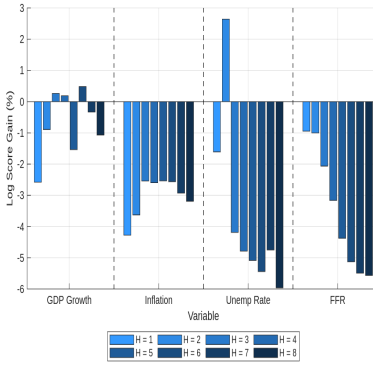


Large VAR

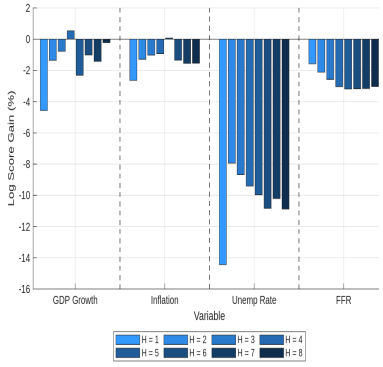


2. Sample 2

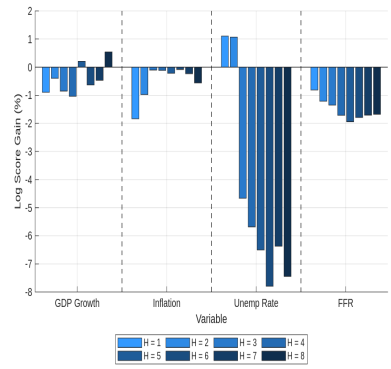
Small VAR



Medium VAR

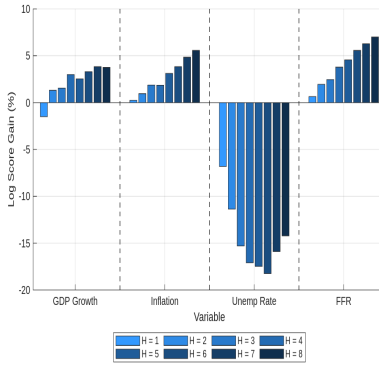


Large VAR

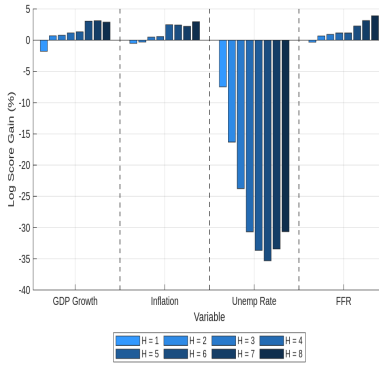


3. Financial Crisis

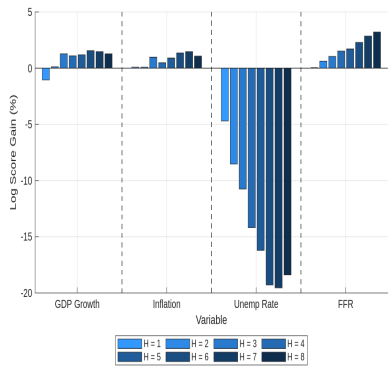
Small VAR



Medium VAR

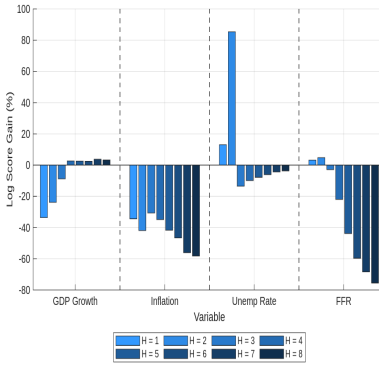


Large VAR

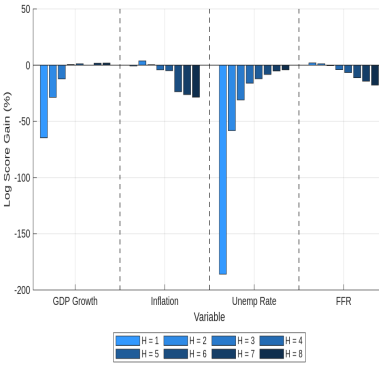


4. COVID

Small VAR



Medium VAR



Large VAR

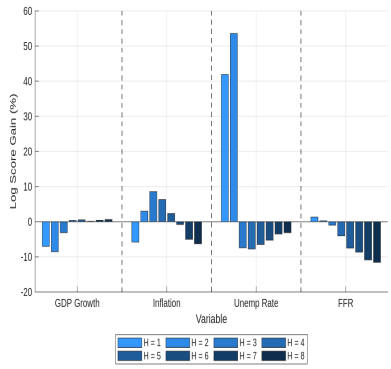


Figure 1 plots the change in log score times 100 from using the REPS-based λ instead of the posterior draws of λ from GLP. Positive values indicate that REPS-based λ yields more accurate density forecasts. Two results stand out. First, deviating from the GLP-based λ does not yield noticeable improvements in density forecasts during normal times. Second, the tighter REPS-based λ improves density forecasts during the financial crisis for GDP growth, inflation, and the federal funds rate.

In Sample 1, the REPS-based λ yields slightly more accurate density forecasts across all three model sizes, particularly for GDP growth (0–1%) and the unemployment rate (0–2%). Changes for inflation and the federal funds rate are negative. Sample 2 combines the pre-crisis, financial-crisis, and COVID periods. Its results for inflation and the federal funds rate resemble those of Sample 1. Importantly, during the financial crisis, the REPS-based prior improves density forecasts of GDP growth, inflation, and the federal funds rate in all three model sizes (up to 7%), and the gains increase with the forecast horizon. At $h = 8$, the gains reach 4–8% in the small VAR and decline with model size. The unemployment forecasts, in contrast, deteriorate, most strongly at short horizons (3–6% at $h = 1$). During COVID, the results change sign across variables, horizons, and model sizes and show no systematic pattern.

The Amisano and Giacomini (2007) tests of equal forecast accuracy in the financial crisis sample reject equal density accuracy for GDP growth and inflation mainly at longer horizons, and for unemployment at most horizons (Table 3). Combined with Figure 1, these rejections favor REPS for GDP growth and inflation and favor GLP for unemployment. Consequently, these findings imply that REPS not only yields statistically significant improvements in density forecasts, but also highlights that forecast performance is sensitive to the choice of the hyperprior on λ.

Table 3. Financial crisis: p-values from AG test

Horizon	1	2	3	4	5	6	7	8
Small								
- GDP growth	0.3707	0.1585	0.0234	0.0000	0.0076	0.0000	0.0000	0.0000
- Inflation	0.7895	0.3660	0.0537	0.1305	0.0261	0.0022	0.0000	0.0000
- Unemployment	0.0423	0.0364	0.0410	0.0486	0.0601	0.0535	0.0630	0.0691
- FFR	0.5982	0.1991	0.2008	0.0598	0.0454	0.0132	0.0035	0.0004
Medium								
- GDP growth	0.1723	0.5678	0.3799	0.1815	0.2927	0.0000	0.0000	0.0000
- Inflation	0.6595	0.8132	0.5294	0.5576	0.0126	0.0087	0.0128	0.0000
- Unemployment	0.0559	0.0493	0.0344	0.0266	0.0238	0.0182	0.0187	0.0199
- FFR	0.8405	0.7228	0.7110	0.6975	0.7356	0.4768	0.2862	0.1621
Large								
- GDP growth	0.2350	0.8210	0.0023	0.0709	0.0941	0.0002	0.0000	0.0000
- Inflation	0.9250	0.9235	0.1757	0.4447	0.2646	0.0177	0.0362	0.0260
- Unemployment	0.0198	0.0280	0.0145	0.0160	0.0080	0.0077	0.0125	0.0145
- FFR	0.9668	0.6363	0.4901	0.3107	0.2865	0.1172	0.0309	0.0097

Whether concentrated predictive distribution improves density forecast accuracy depends on the accuracy of the point forecast. When point forecasts are accurate, this increased concentration enhances the log score. However, when forecasts are less accurate, overly peaked density forecasts

can lead to a substantial reduction in the log score due to overconfidence in incorrect predictions. Figure 1 reflects both cases: log scores improve clearly during the financial crisis, whereas gains during the pre-crisis sample are modest or negative.

To illustrate this more clearly, note that the log predictive score, $s_t(y^i_{t+h})$, is computed based on the quadratic approximation for Gaussian predictive distribution,

$$s_t(y^i_{t+h}) = -0.5 \left[\ln(2\pi) + \ln(V^i_{t+h|t}) + \frac{(y^i_{t+h} - \overline{y}^i_{t+h|t})^2}{V^i_{t+h|t}} \right]$$

where, given F_{draw} from the forecast distribution of y^i_{t+h}, $\left\{ y^i_{t+h,f} \right\}_{f=1}^{F_{draw}}$, $\overline{y}^i_{t+h|t}$is the posterior mean of the simulated forecast distributions, $V^i_{t+h|t}$ is the posterior variance of the simulated forecast distributions, and y^i_{t+h} is the actual value. Under REPS, $V^i_{t+h|t}$ is substantially smaller than under GLP. While this greater concentration raises the log score when the realized outcome lies close to the predictive mean, it magnifies log-score losses when the distribution is poorly centered, due to the penalized error term $(y^i_{t+h} - \overline{y}^i_{t+h|t})^2 / V^i_{t+h|t}$. I next examine whether the sharper REPS predictive distributions are accompanied by more accurate point forecasts across different periods.

6.2 Point Forecast Accuracy

I next compare point forecast accuracy under REPS and GLP to determine whether the sharper REPS predictive distributions are accompanied by more accurate point forecasts. Figure 2 plots the percentage change in RMSFE from using the REPS-based λ instead of the posterior draws of λ from GLP. Positive values indicate that REPS-based λ yields more accurate point forecasts. Two results stand out. First, deviating from the GLP-based λ does not yield meaningful improvements in point forecasts in normal times. Second, the tighter REPS-based λ improves longer-horizon forecasts during the financial crisis.

In Sample 1, the REPS-based λ yields slightly less accurate forecasts in all three model sizes, particularly for inflation (up to about 6%) and the federal funds rate (1–3.5%). Changes for GDP growth and the unemployment rate are small and change sign across horizons. Sample 2 combines the pre-crisis, financial-crisis, and COVID periods. Its results for inflation and the federal funds rate resemble those of Sample 1. Importantly, during the financial crisis, the REPS-based prior improves forecasts of GDP growth, inflation, and the federal funds rate in all three model sizes, and the gains increase with the forecast horizon. At $h = 8$, the gains reach 4–8% in the small VAR and decline with model size. The unemployment forecasts, in contrast, deteriorate, most strongly at short horizons (3–6% at $h = 1$). During COVID, the results change sign across variables, horizons, and model sizes and show no systematic pattern.

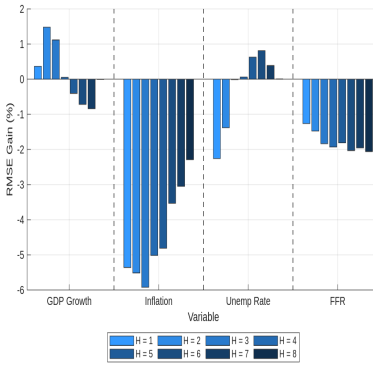
To summarize, gains from the REPS-based prior are concentrated in

1. **Period:** Financial crisis
2. **Horizon:** Longer forecast horizons
3. **Variable:** GDP growth, inflation, and the federal funds rate
4. **Model size:** All three model sizes, with gains declining as the model grows.

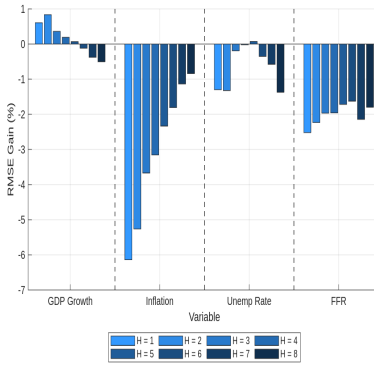
Figure 2. Percentage Change in RMSFE

1. Sample 1

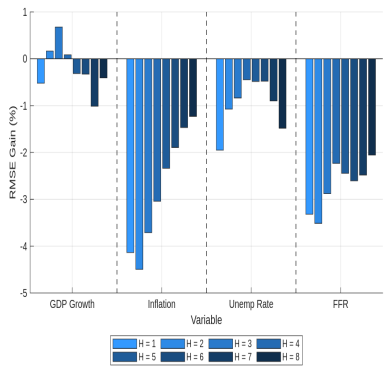
Small VAR



Medium VAR

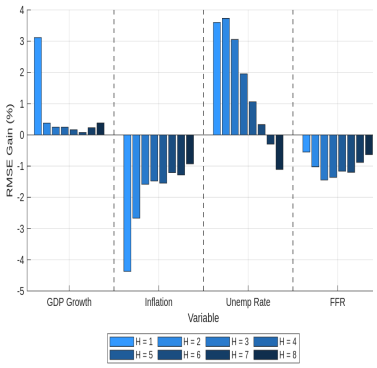


Large VAR

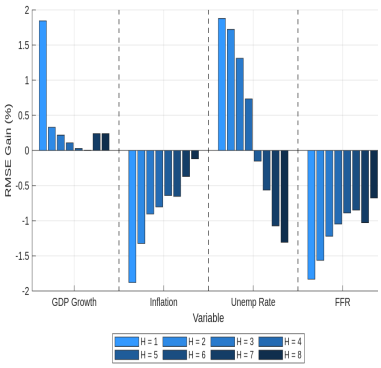


2. Sample 2

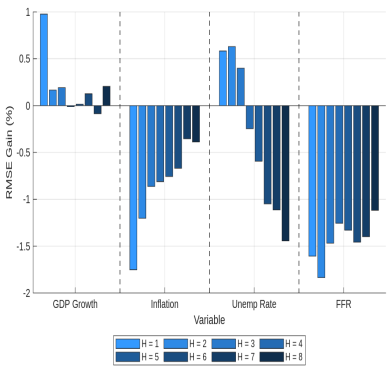
Small VAR



Medium VAR

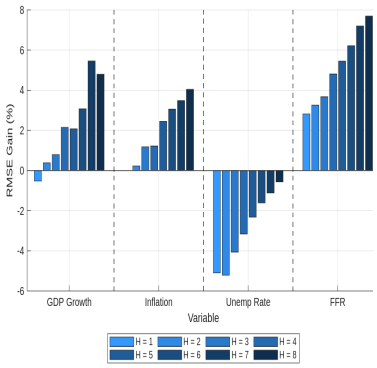


Large VAR

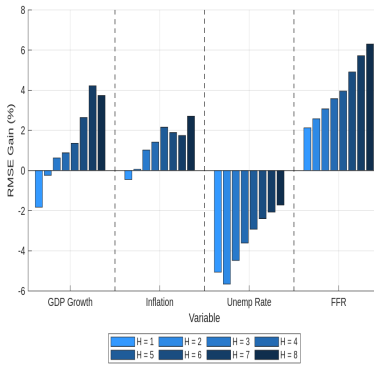


3. Financial Crisis

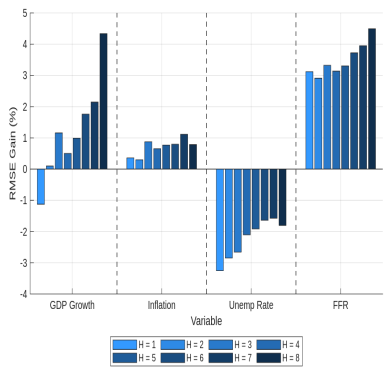
Small VAR



Medium VAR

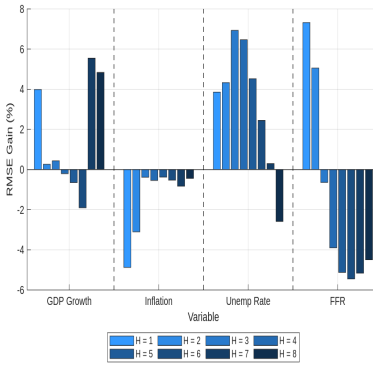


Large VAR

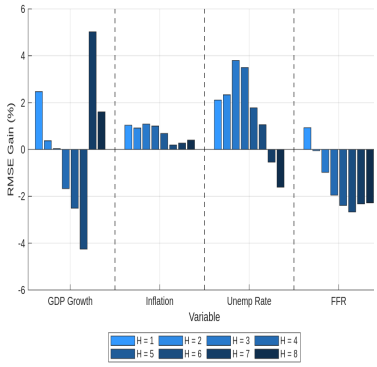


4. COVID

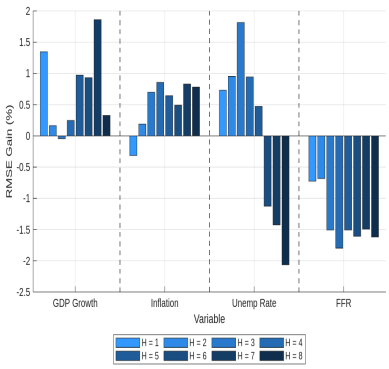
Small VAR



Medium VAR



Large VAR



Combined with the density forecast results, REPS generates more concentrated predictive distributions, but whether this improves density forecast accuracy depends crucially on point forecast accuracy improvements. During the financial crisis, point forecast accuracy improved. Hence, the resulting predictive distribution under REPS was sharper and centered around a more accurate mean, leading to density forecast gains. Conversely, in the pre-crisis period, point forecasts deteriorated. Although REPS produced a sharper predictive distribution, it was centered around an inaccurate mean, resulting in poor density forecasts. During the COVID period, the results are mixed. While point forecast accuracy improved for some variables and horizons, these improvements did not translate into better density forecasts for reasons discussed in Section 6.3.

Table 4. Financial crisis: p-values from DM test and CW test

Horizon	1	2	3	4	5	6	7	8
Small								
DM test								
- GDP growth	0.7515	0.3256	0.2050	0.0362	0.1231	0.1406	0.1189	0.1633
- Inflation	0.4994	0.3967	0.0670	0.1339	0.1453	0.0619	0.0149	0.0016
- Unemployment	0.9675	0.9363	0.9086	0.8698	0.8368	0.7837	0.6933	0.5848
- FFR	0.1938	0.1991	0.1813	0.1186	0.1275	0.1210	0.0996	0.0760
CW test								
- GDP growth	0.6436	0.2680	0.1343	0.0080	0.0648	0.0692	0.0415	0.0603
- Inflation	0.4357	0.3043	0.0290	0.0486	0.0413	0.0053	0.0003	0.0000
- Unemployment	0.9222	0.9436	0.9349	0.9086	0.8540	0.7624	0.6375	0.5005
- FFR	0.1660	0.1371	0.1119	0.0487	0.0381	0.0353	0.0306	0.0275
Medium								
DM test								
- GDP growth	0.9057	0.5863	0.2390	0.1292	0.1386	0.1503	0.1196	0.1604
- Inflation	0.6816	0.4713	0.1218	0.0785	0.0966	0.0070	0.0169	0.0013
- Unemployment	0.9448	0.9067	0.8906	0.8755	0.8671	0.8536	0.8613	0.8604
- FFR	0.2660	0.2998	0.2638	0.2318	0.2162	0.1973	0.1803	0.1626
CW test								
- GDP growth	0.8348	0.5493	0.1768	0.0623	0.0785	0.0706	0.0416	0.0403
- Inflation	0.6410	0.4183	0.0685	0.0256	0.0202	0.0016	0.0017	0.0000
- Unemployment	0.8869	0.9228	0.9221	0.9263	0.9273	0.9115	0.8923	0.8389
- FFR	0.2289	0.2210	0.1825	0.1458	0.1129	0.0824	0.0649	0.0589
Large								
DM test								
- GDP growth	0.9198	0.4432	0.0339	0.1829	0.0570	0.1084	0.2033	0.1217
- Inflation	0.3456	0.3307	0.1148	0.2131	0.0824	0.1210	0.0886	0.2417
- Unemployment	0.9868	0.9665	0.9710	0.9799	0.9830	0.9970	0.9999	1.0000
- FFR	0.1340	0.2209	0.1905	0.1596	0.1284	0.1476	0.1727	0.1627
CW test								
-GDP growth	0.8834	0.3878	0.0149	0.1271	0.0279	0.0504	0.0968	0.0212
-Inflation	0.3323	0.2662	0.0704	0.1244	0.0369	0.0518	0.0241	0.1214
-Unemployment	0.9554	0.9741	0.9881	0.9946	0.9973	0.9997	1.0000	1.0000
-FFR	0.1514	0.1489	0.1272	0.1108	0.0666	0.0525	0.0641	0.0627

Tests of equal forecast accuracy provide statistically significant evidence only for selected variable–horizon combinations. Table 4 reports one-sided p-values for the null hypothesis of equal accuracy against the alternative that the REPS-based prior is more accurate during the financial crisis. The Diebold–Mariano test rejects the null most clearly for inflation at the longest horizons in the small and medium VARs ($p < 0.02$ at $h = 7$ and 8). For GDP growth and the federal funds rate, rejections occur only at isolated horizons and mostly at the 10% level, and in the large VAR the evidence is weak throughout. For unemployment, the p-values are close to one: reversing the alternative, the GLP forecasts are significantly more accurate at short horizons in the small VAR and at all horizons in the large VAR. With 33 evaluation windows and gains of a few percent, the tests have limited power, so the statistical evidence for the REPS gains is modest and concentrated where the RMSFE gains are largest.

These results raise two questions. Why does the REPS-based prior improve forecast accuracy only during the financial crisis? Why do the gains differ across variables, horizons, and model sizes? To answer them, I examine how REPS reweights the two components of the marginal likelihood that GLP uses to select λ.

6.3 Trade-off Between In-Sample Fit and Predictive Variance: REPS vs GLP

Recall that GLP use the following marginal likelihood, expressed as a function of the hyperparameter, as the likelihood function for hyperparameter estimation:

$$p(y|\lambda) \propto \underbrace{\left|\left(V_\varepsilon^{\text{posterior}}\right)^{-1} V_\varepsilon^{\text{prior}}\right|^{\frac{T-p+d}{2}}}_{\text{In-sample fit}} \cdot \underbrace{\prod_{t=p+1}^{T} \left|V_{t|t-1}\right|^{-\frac{1}{2}}}_{\text{Predictive Variance}}$$

Here, λ denotes the Minnesota prior tightness hyperparameter, and $V_{t|t-1} = \mathbb{E}_\Sigma[\text{Var}(y_t|y^{t-1}, \Sigma)]$ is the variance of the one-step-ahead forecast of y_t (conditional on Σ), averaged over all prior realizations of Σ.

The marginal likelihood balances two in-sample components: how well the VAR fits the estimation sample and how concentrated its one-step predictive distribution is. Moving to the smaller REPS-based λ changes this balance mechanically. Stronger shrinkage worsens the fit component but reduces predictive variance, thereby increasing the predictive-variance contribution to the marginal likelihood.

To quantify how REPS changes this balance, I compute each component of the marginal likelihood and evaluate how the REPS-based λ affects these components relative to the λ chosen by GLP. After taking logarithms, the differences in marginal likelihood between REPS and GLP can be rewritten as:

$$\log p(y|\lambda) \propto \underbrace{\frac{T-p+d}{2} \cdot \log\left|\left(V_\varepsilon^{\text{posterior}}\right)^{-1} V_\varepsilon^{\text{prior}}\right|}_{\text{Fit}(\lambda)} \underbrace{-\frac{1}{2}\sum_{t=p+1}^{T} \log\left|V_{t|t-1}\right|}_{\text{PV}(\lambda)}$$

where T is the window size, p is the lag order, and $d = N + 2$ is the degrees of freedom in the inverse-Wishart prior for Σ. I compute each component separately under both methods and calculate the change relative to GLP to assess how REPS shifts the balance between in-sample fit and forecast accuracy.

$$\Delta \log p = \left[\text{Fit}(\lambda_{\text{REPS}}) - \text{Fit}(\bar{\lambda}_{\text{GLP}})\right] + \left[\text{PV}(\lambda_{\text{REPS}}) - \text{PV}(\bar{\lambda}_{\text{GLP}})\right]$$

Table 5. Change in the log marginal likelihood and its components (REPS - GLP)

	Small			Medium			Large		
Period	$\Delta \log p$	ΔFit	ΔPV	$\Delta \log p$	ΔFit	ΔPV	$\Delta \log p$	ΔFit	ΔPV
Sample 1	-0.82	-4.54	3.72	-0.48	-6.14	5.66	-0.34	-10.36	10.02
Sample 2	-0.89	-4.88	3.99	-0.46	-6.39	5.93	-0.33	-10.63	10.30
Fin. Crisis	-0.94	-5.07	4.13	-0.58	-7.17	6.60	-0.35	-11.60	11.25
COVID	-1.12	-5.97	4.86	-0.04	-4.06	4.02	-0.25	-7.69	7.44

Table 5 measures how much the in-sample trade-off shifts when REPS tightens the prior. Across all periods, the tighter REPS-based specification sacrifices in-sample fit in exchange for a sharper one-step predictive distribution (ΔFit, ΔPV). The net change in the log marginal likelihood remains slightly negative because the latter does not fully offset the former ($\Delta \log p$). These variations in the changes of fit and predictive variance across sample periods and model sizes are driven by the shape of the marginal likelihood, which I study in Appendix B.

However, this reallocation does not necessarily imply an improvement in forecast accuracy. In the small VAR, for example, COVID produces the largest fit loss and predictive-variance gain, yet no systematic forecast accuracy improvement. Whether the movement is beneficial depends on whether the dynamics suppressed by tighter shrinkage remain useful in the forecast period.

Normal times. In the pre-crisis sample, the relationships among the variables in the estimation sample largely persist out of sample. The fit given up by REPS reflects dynamics that remain useful for forecasting, so the tighter prior reduces forecast accuracy, particularly for inflation, whose persistence REPS attenuates.

Financial crisis. The estimation sample for the financial-crisis forecasts is dominated by the Great Moderation, while the forecast period features a prolonged effective lower bound, a slow recovery, and inflation that fell less than pre-crisis relationships implied. Standard VARs show substantial parameter instability after the crisis (Aastveit et al. (2017)). The fit given up by REPS therefore largely reflects dynamics that did not persist out of sample. By shrinking these dynamics, REPS generates simpler forecast paths that better match the prolonged post-crisis period, particularly at longer horizons.

The variable-level results line up with this mechanism.

- **Federal funds rate.** Linear VARs that ignore the effective lower bound tend to predict an early exit from it (Carriero et al., (2025)). By shrinking the feedback from output and inflation and pulling the own-lag coefficient toward one, the tighter prior keeps the predicted policy rate near its current level and delays the predicted normalization.
- **Inflation.** A VAR estimated on pre-crisis data transmits the sharp rise in unemployment into a large predicted decline in inflation, which did not materialize (the missing disinflation; Coibion and Gorodnichenko (2015)). Tighter shrinkage weakens these Phillips-curve dynamics and pulls inflation toward a low-persistence path governed by the prior.
- **GDP growth.** After a deep contraction, VARs estimated on pre-crisis business cycles tend to predict a stronger rebound than the slow post-crisis recovery delivered. Tighter

shrinkage attenuates the cross-variable and higher-lag mechanisms that generate this rebound and pulls the growth forecast toward a low-persistence path governed by the prior.

- **Unemployment rate.** Unlike the relationships governing inflation and the policy rate, the output–unemployment relationship remained comparatively stable (Ball, Leigh, and Loungani, 2017). In addition, unemployment trended persistently up in 2008–09 and down in 2010–15. The tighter prior pulls the unemployment equation toward a random walk: it weakens the momentum in its changes and the link to output growth. This costs forecast accuracy mainly at short horizons, where quarter-to-quarter momentum is most informative. At longer horizons, extrapolating that momentum across turning points is less useful, and the loss fades.

To examine this interpretation further, I examine the financial-crisis sample in greater detail. Although treated as a single evaluation period above, it spans the transition into the crisis, the prolonged ELB regime, and the recovery that followed. Examining these episodes separately helps determine whether the REPS gains arise during the disruption itself or after pre-crisis dynamics have become misleading due to the structural break. Because a forecast made in one episode may be evaluated against an outcome in another, I classify the financial-crisis evaluation sample by both the date at which it is made (the forecast origin) and the date for which it is evaluated (the forecast period).

The sample contains forecast origins from 2007Q3, the quarter of the first cut in the federal funds rate, through 2015Q3, the quarter immediately preceding liftoff. The forecast horizons span three phases: the crisis transition through 2008Q4, the ELB period from 2009Q1 through 2015Q4, and, at longer horizons, the post-liftoff period beginning in 2016Q1. Table 6 cross-classifies the forecasts along these two dimensions.

Table 6. Decomposition of the Financial-Crisis Sample

Forecast Origin	**Forecasted Periods**		
	Transition, $\leq$ 2008Q4	**ELB, 2009Q1–2015Q4**	**Post-liftoff, $\geq$ 2016Q1**
Transition, 2007Q3–2008Q4	A1	A2	–
ELB, 2009Q1–2015Q3	–	B1	B2

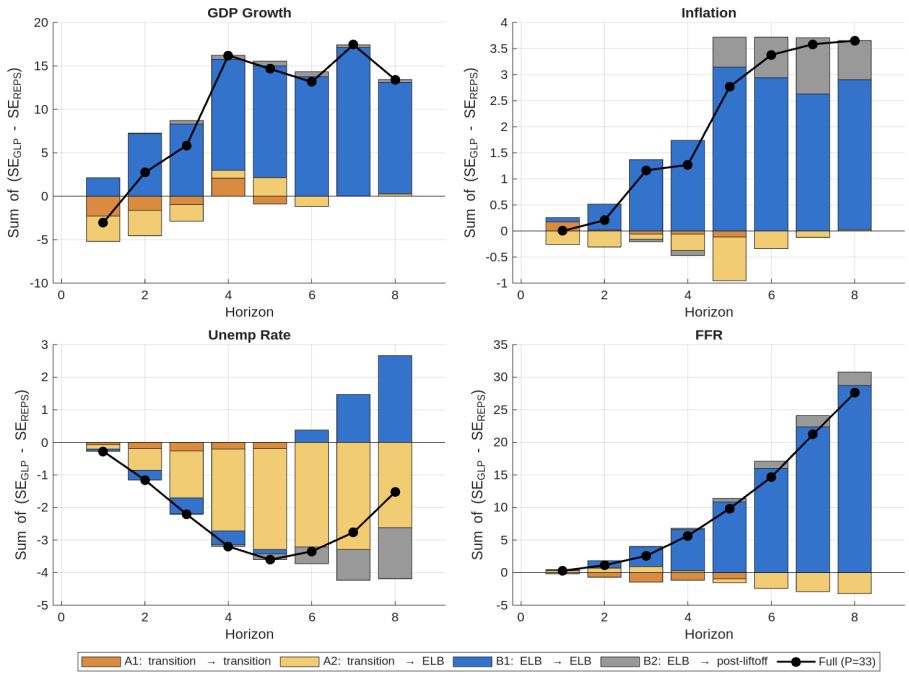
Note: The forecast origin is the last quarter of the estimation window. At $h = 1$, A1, A2, B1, and B2 contain 5, 1, 27, and 0 forecasts; at $h = 8$, they contain 0, 6, 20, and 7.

The classification yields four groups based on the phase in which the forecast is made and the phase being forecast: transition-to-transition (A1), transition-to-ELB (A2), ELB-to-ELB (B1), and ELB-to-post-liftoff (B2). Because the loss differentials are additive, the four groups sum to the full-sample loss differential at every horizon.

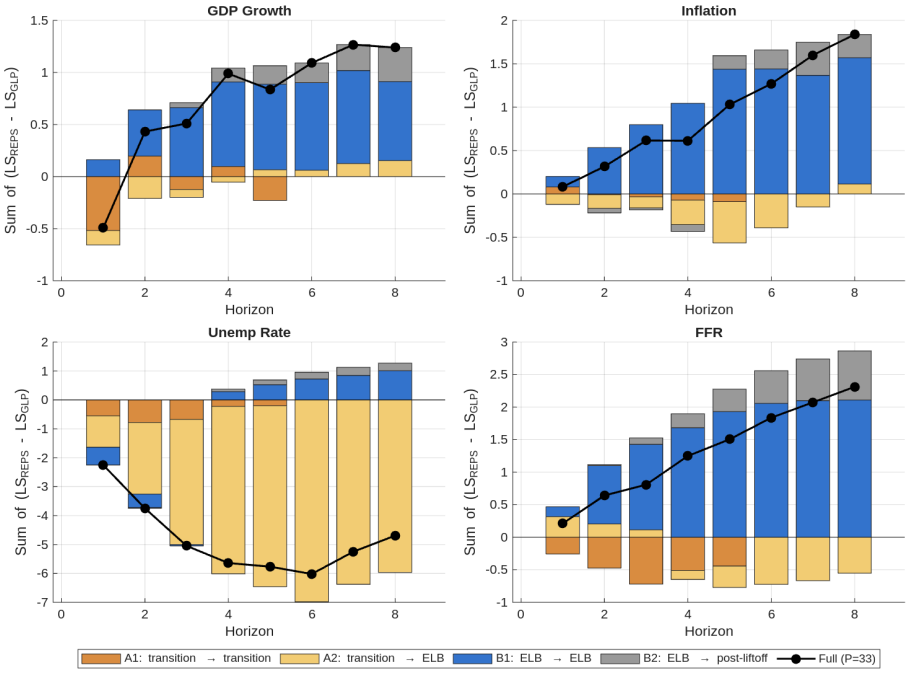
Figure 3 shows that, in the small VAR, the gains come almost entirely from B1. For GDP growth, inflation, and the federal funds rate, B1 accounts for essentially all of the cumulative gains beyond the shortest horizons. The medium and large VARs show the same pattern (Appendix C). Within B1, the REPS-based prior lowers RMSFE by up to 6% for GDP growth and 5% for inflation in the small VAR, while producing significant log-score gains at most horizons.

Figure 3. Decomposition of the forecasting performance (REPS-GLP): (Financial crisis, small VAR)

(a) Point forecasts



(b) Density forecasts



Notes:

(i) For variable i, horizon h, and forecast origin k, the loss differential in panel (a) is $d^{\mathrm{SE}}_{i,h,k} = \left(y^i_{k+h} - \hat{y}^{i,\mathrm{GLP}}_{k+h|k}\right)^2 - \left(y^i_{k+h} - \hat{y}^{i,\mathrm{REPS}}_{k+h|k}\right)^2$, where $\hat{y}$ is the point forecast. In panel (b) it is $d^{\mathrm{LS}}_{i,h,k} = s^{\mathrm{REPS}}_k(y^i_{k+h}) - s^{\mathrm{GLP}}_k(y^i_{k+h})$, where $s_k(\cdot)$ is the Gaussian log predictive score defined in Section 3.2

(ii) Each bar shows $\sum_{k \in G_h} d_{i,h,k}$ for group $G \in \{\mathrm{A1}, \mathrm{A2}, \mathrm{B1}, \mathrm{B2}\}$ defined in Table 6. The line shows the sum over all 33 forecasts, which equals the sum of the bars at every horizon. Positive values favor REPS.

(iii) Results for the medium and large VARs are in Appendix C.

The concentration of the gains in B1 indicates that the benefits of tighter shrinkage emerge after the economy has entered the ELB regime rather than during the transition into it. For earlier B1 origins, forecasts start from initial conditions that already reflect the ELB, whereas the coefficients are estimated largely using pre-crisis observations. Propagated through these pre-crisis dynamics, the ELB starting point implies a faster return toward historical norms, such as a stronger rebound in GDP growth and an earlier liftoff, than actually occurred. By dampening these dynamics, the tighter REPS prior keeps forecasts closer to their starting point, which better matches the persistent ELB period.

By contrast, forecasts generated during the transition (A1 and A2) favor GLP on balance, even when their forecast periods fall in the early ELB period. For these origins, the economy had not yet reached the ELB and was still deteriorating, and propagating this starting point through the estimated dynamics correctly extrapolated the subsequent movements: further rate cuts through the end of 2008 and rising unemployment through 2009. By keeping forecasts closer to their starting point, the tighter REPS prior understates these movements and yields less accurate forecasts for these variables. Post-liftoff forecasts in B2 contribute little to the aggregate results.

Consistent with this interpretation, the forecasting gains of REPS within B1 decline over time. When B1 is split at 2012Q2, the eight-quarter-ahead gains for GDP growth disappear for later forecast origins (2012Q3–2013Q4), while those for the federal funds rate fall to roughly one-quarter of their earlier magnitude across all three model sizes. This decline reflects the gradual adaptation of the GLP estimates: as the rolling windows incorporate more ELB observations, the estimated dynamics themselves imply more persistence from the ELB starting point, leaving less scope for the tighter REPS prior to correct.

Taken together, REPS improves forecasts during the financial crisis by dampening estimated dynamics that no longer hold after the structural break, and these gains are largest when forecasts are propagated from an origin that already reflects the ELB.

Why do the gains decline with model size? In large VARs, the GLP choice is already tighter, so their coefficients are already shrunk heavily toward the prior mean (Appendix B). Therefore, further tightening through REPS has less room to change the forecasts. Furthermore, Table 5 shows that the net cost of moving from the GLP choice to the REPS-based λ shrinks with model size, from 0.94 log points per window in the small VAR to 0.35 in the large VAR during the financial crisis ($\Delta \log p$). The data distinguish the two values of λ less in larger models, and the two priors therefore produce more similar forecasts. Both effects reduce the difference between the REPS and GLP forecasts, and hence both the gains and the losses, as the model grows.

Why do the gains grow with the horizon? Multi-step forecasts are obtained by iterating the estimated VAR, so the estimated dynamics are applied repeatedly and their effects compound with the horizon. When these dynamics do not persist out of sample, as after the financial crisis, the resulting errors also grow with the horizon. The tighter prior dampens the propagation of shocks through the system and pulls the forecasts toward a prior mean. This matters little one quarter ahead but increasingly at longer horizons.

COVID. COVID provides the clearest evidence that neither a larger hyperparameter distortion nor a lower predictive variance is sufficient for better forecasts. Both are the largest of all periods (Table 5, see Appendix B for why the adjustment is largest in this period), yet the

point forecast gains show no systematic pattern. As during the financial crisis, the pre-COVID dynamics broke down, but the economy did not settle into a prolonged, slowly evolving period. The lower-bound episode was short-lived, and output, unemployment, inflation, and the policy rate reversed rapidly. Neither the estimated dynamics nor the simpler paths implied by the prior could track these reversals across horizons, so tighter shrinkage helps at some horizons and hurts at others. With only nine evaluation windows, these estimates are imprecise, but the sign reversals in the RMSFE gains are consistent with this interpretation.

These results clarify what REPS does and does not do. Because REPS solves a minimization problem[1], it selects a value of λ below the GLP posterior mean, providing a disciplined way to move toward tighter shrinkage. It does not, however, determine whether that move is beneficial. As the decomposition and the period-level results show, the move improves forecasts only when the dynamics given up in the estimation sample do not persist out of sample.

6.4 Implications

What do these findings imply for a forecaster in practice? Two implications follow.

First, REPS can serve as a diagnostic tool. By moving λ away from the GLP estimate within the relative entropy bound, a forecaster can readily gauge the sensitivity of predictive performance to prior tightness. Substantial shifts in forecasts signal that the standard data-driven hyperparameter estimate warrants closer scrutiny.

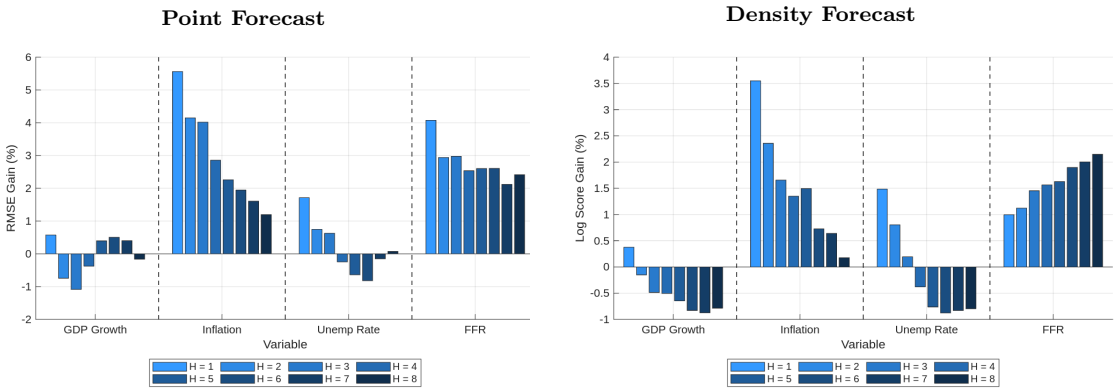
Second, the results indicate when a tighter prior is likely to help. The gains are concentrated in forecasts made after the economy entered the ELB, while the estimation window was still dominated by pre-crisis observations, and they fade as the window incorporates more ELB data. Such situations are often recognizable in real time, when the economy moves into conditions rarely observed in the estimation sample. In these situations, a prior tighter than the GLP choice is a reasonable adjustment until the estimates adapt.

The appropriate direction of adjustment, however, depends on the period. The relative entropy bound equally allows λ to move toward a looser prior when the REPS problem is formulated as a maximization rather than a minimization. In the pre-crisis sample with the medium VAR, the results mirror those of the financial crisis: inflation and the federal funds rate, which deteriorate under the tighter prior, improve by up to 5% in both point and density forecasts under the looser prior (Figure 4). The GLP choice thus lies between the two directions.

Thus, a looser prior performs better in the pre-crisis sample, whereas a tighter prior is advantageous during the financial crisis. While the GLP choice lies between these two directions, the direction of beneficial adjustment differs systematically across periods. Translating this period-specific evidence into a real-time rule for choosing the direction of adjustment remains a key challenge for future research.

[1]Alternatively, one can solve the corresponding maximization problem, which yields a value of λ above the GLP posterior mean and therefore looser shrinkage.

Figure 4. Forecast Accuracy Comparison (REPS-GLP) under Looser Prior



7 Conclusion

Selecting the shrinkage hyperparameter is crucial for forecasting with parameter-rich VAR models. The standard approach of Giannone, Lenza, and Primiceri (2015) selects the overall tightness of the Minnesota prior by combining a hyperprior with the marginal likelihood. For forecasting purposes, however, this approach raises two concerns: the selected hyperparameter depends on the choice of hyperprior, and the marginal likelihood balances in-sample fit with predictive variance in a way that need not align with the forecasting objective. To address these concerns, this paper uses Relative Entropy Prior Sensitivity (REPS) to move away from the GLP hyperprior within an entropy bound and examines when the resulting prior improves forecasts.

REPS rebalances the trade-off between in-sample fit and predictive variance in the marginal likelihood. Evaluating the resulting forecasts across model sizes, variables, horizons, and periods, I find that the gains are concentrated in the financial crisis, in GDP growth, inflation, and the federal funds rate, and at longer forecast horizons, and that they decline as the model grows. In the pre-crisis sample, the REPS-based prior slightly reduces accuracy, particularly for inflation and the federal funds rate, and during the financial crisis it reduces the accuracy of unemployment forecasts. During COVID, the results show no systematic pattern.

A decomposition of the marginal likelihood explains these patterns. REPS gives up in-sample fit in exchange for lower predictive variance at a small net cost in marginal likelihood, producing sharper predictive distributions. Whether this trade-off pays off depends on whether the dynamics given up in the estimation sample persist out of sample. When they persist, as in the pre-crisis sample, the sacrificed fit is costly, and sharper distributions centered on less accurate point forecasts lower the log score. When they do not, as after the financial crisis, the tighter prior improves both point and density forecasts. Classifying the financial-crisis forecasts by forecast origin and forecast period sharpens this interpretation: the gains arise almost entirely from forecasts made after the economy entered the effective lower bound (ELB), while the estimation windows were still dominated by pre-crisis observations, and they fade as the windows incorporate more ELB observations. COVID shows that a breakdown of past dynamics is not sufficient. The adjustment in λ and the reduction in predictive variance are the largest of all periods, yet neither the estimated dynamics nor the simpler paths implied by the prior could track the rapid reversals that followed the shock.

These results suggest that no single degree of shrinkage dominates across periods: a looser prior

than the GLP choice improves forecasts of inflation and the federal funds rate in the pre-crisis sample, whereas a tighter prior helps after a shock moves the economy into a prolonged period that the estimation sample rarely contains. REPS therefore provides a disciplined way to assess prior sensitivity and to examine how forecasts respond to tighter or looser shrinkage, although translating this period-specific evidence into real-time forecasting remains an open question.

Several extensions follow. First, the REPS objective could be tailored to forecasting, for example by targeting multi-step predictive performance rather than the hyperparameter itself, at additional computational cost. Second, REPS could be applied to richer prior structures, such as a separate cross-variable shrinkage hyperparameter. Lastly, a rule that chooses the direction and degree of adjustment in real time, based on signals that estimated relationships are breaking down, would turn the insights of this paper into a practical forecasting tool.

Reference

Aastveit, Knut Are., Carriero, Andrea., Clark, Todd E., Marcellino, Massimiliano., 2017, Have Standard VARs Remained Stable Since the Crisis?, Journal of Applied Econometrics 32(5): 931-951.

Adolfson, Malin., Linde, Jesper., Villani, Mattias., 2007, Forecasting Performance of an Open Economy DSGE Model, Econometric Reviews 26(2-4): 289-328.

Amisano, Gianni., Giacomini, Raffaella., 2007, Comparing Density Forecasts via Weighted Likelihood Ratio Tests, Journal of Business & Economic Statistics 25(2): 177-190.

Ball, Laurence., Leigh, Daniel., Loungani, Prakash., 2017, Okun's Law: Fit at 50?, Journal of Money, Credit and Banking 49(7): 1413-1441.

Banbura, Marat., Giannone, Domenico., Reichlin, Lucrezia., 2010, Large Bayesian Vector Autoregression, Journal of Applied Econometrics 25: 71-92.

Carriero, Andrea., Clark, Todd E., Marcellino, Massimiliano., Mertens, Elmar., 2025, Forecasting with Shadow Rate VARs, Quantitative Economics 16(3): 795-822.

Carriero, Andrea., Clark, Todd E., Marcellino, Massimiliano., 2019, Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors, Journal of Econometrics 212(1): 137-154.

Chan, Joshua., Jacobi, Liana., Zhu, Can., 2019, How sensitive are VAR forecasts to prior hyperparameters? An automated sensitivity analysis, Advances in Econometrics, 40A: 229–248.

Clark, Todd E., West, Kenneth D., 2007, Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, Journal of Econometrics 138: 292-311.

Coibion, Olivier., Gorodnichenko, Yuriy., 2015, Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation, American Economic Journal: Macroeconomics 7(1): 197-232.

Cross, Jamie L., Hou, Chenghan., Poon, Aubrey., 2020, Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity, International Journal of Forecasting 36(2): 899-915.

Diebold, Francis., Mariano, Roberto., 1995, Comparing Predictive Accuracy, Journal of Business & Economic Statistics 13: 253-263.

Feldkircher, Martin., 2011, Forecast Combination and Bayesian Model Averaging: A Prior Sensitivity Analysis, Journal of Forecasting, 31(4): 361-376.

Giacomini, Raffaella., White, Halbert., 2006, Tests of Conditional Predictive Ability, Econometrica 74(6): 1545-1578.

Giannone, Domenico., Lenza, Michele., Primiceri, Giorgio., 2015, Prior Selection for Vector Autoregressions, The Review of Economics and Statistics 97: 436-451.

Ho, Paul, 2023, Global Robust Bayesian Analysis in Large Models, Journal of Econometrics 235: 608-642.

Huber, Florian., Feldkircher, Martin ., 2017, Adaptive Shrinkage in Bayesian Vector Autoregressive Models, Journal of Business & Economic Statistics 37(1): 27-39.

Jacobi, Liana., Zhu, Dan., Joshi, Mark., 2025, Estimating Posterior Sensitivities with Application to Structural Analysis of Bayesian Vector Autoregressions, Journal of Business & Economic Statistics, 43(1): 134-149.

Karlsson, Sune, 2013, Forecasting with Bayesian Vector Autoregression, Handbook of Economic Forecasting, vol.2, 791-897.

Koop, Gary., 2013, Forecasting with Medium and Large Bayesian VARs, Journal of Applied Econometrics 28: 177-203.

McCracken, Michael.,Ng, Serena., 2021, Fred-Qd: A Quarterly Database for Macroeconomic Research, Federal Reserve Bank of St.Louis Review 103: 1-44.

Sims, Christopher., 1980, Macroeconomics and Reality, Econometrica 48(1): 1-48.

Smets, Frank., Wouters, Rafael., 2007, Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, American Economic Review 97(2): 586-606.

Appendices

A Alternative Forecast Specifications

On the main analysis, I evaluate whether using a single REPS-based value of the Minnesota tightness parameter, $\tilde{\lambda}$, improves forecast accuracy relative to the standard GLP approach, which generates forecasts by integrating over posterior draws of λ. This comparison evaluates the practical REPS plug-in forecasting rule proposed in this paper. However, it is not a pure comparison of hyperparameter values, because the two approaches differ in their treatment of hyperparameter uncertainty: GLP averages over posterior draws of λ, whereas the REPS plug-in forecast conditions on a single distorted value.

To assess whether the baseline results reflect the selected value of λ or the different treatment of λ-uncertainty, I conduct two complementary robustness exercises. First, I replace the GLP posterior draws of λ with their posterior mean and compare this GLP plug-in forecast with the REPS plug-in forecast (Appendix A.1). This comparison places both methods on equal footing as plug-in forecasts and isolates the role of the selected hyperparameter value. Second, I implement REPS as a posterior reweighting scheme (Appendix A.2). In this exercise, I retain the GLP posterior draws of λ, but replace the equal mixture weights used by GLP with REPS distortion weights. This preserves the posterior-mixture structure of GLP and examines whether the REPS distortion has predictive content beyond the use of a single plug-in hyperparameter.

A.1 Forecast Accuracy Improvements Using Posterior Mean Instead of Posterior Draws of the Hyperparameter

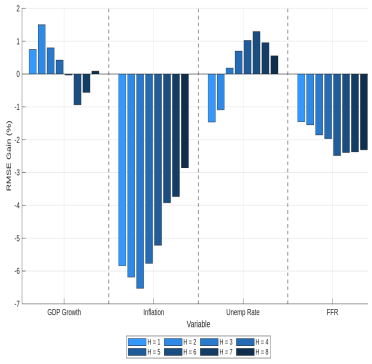
First, I use the posterior mean of λ. In this case, the comparison between GLP and REPS reflects only the uncertainty arising from the VAR coefficients, whereas using posterior draws captures uncertainty from both the VAR coefficients and the hyperparameter. If the baseline gains arose mainly from this different treatment of hyperparameter uncertainty, they should shrink once both forecasts condition on a single value of λ.

Figure 5 shows that this is not the case. The point forecast results closely resemble the baseline results in Section 6. During the financial crisis, the REPS-based λ improves forecasts of GDP growth, inflation, and the federal funds rate, and the gains increase with the forecast horizon and decline with model size. At $h = 8$, the gains reach 5–10% in the small VAR and 2–5% in the large VAR. The unemployment rate forecasts, in contrast, deteriorate. In the pre-crisis sample (Sample 1), the REPS-based λ again yields less accurate forecasts of inflation and the federal funds rate, by up to 6.5% and 3.6%, respectively, while changes for GDP growth are small. The Diebold–Mariano tests in Table 7 confirm this pattern: the null of equal accuracy is rarely rejected in favor of REPS in Sample 1, and the p-values close to one for inflation and the federal funds rate indicate that GLP is significantly more accurate for these variables. Because the results are largely unchanged, the forecasting gains under REPS reflect the selected value of λ rather than the treatment of hyperparameter uncertainty.

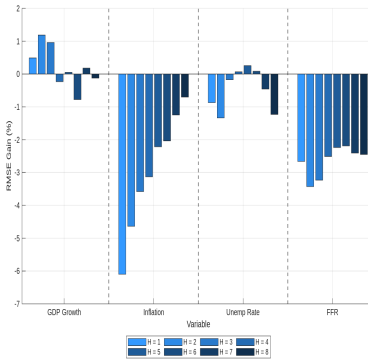
Figure 5: Percentage Change in RMSFE

1. Sample 1

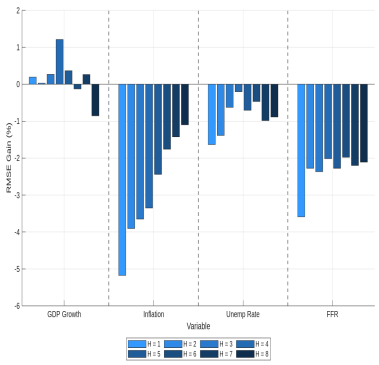
Small VAR



Medium VAR

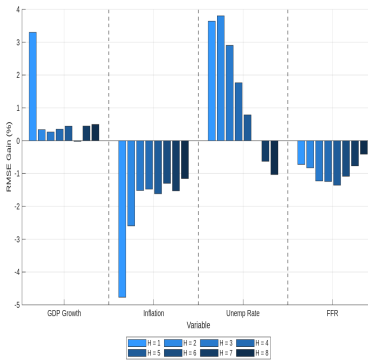


Large VAR

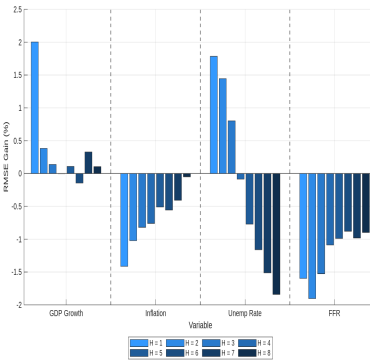


2. Sample 2

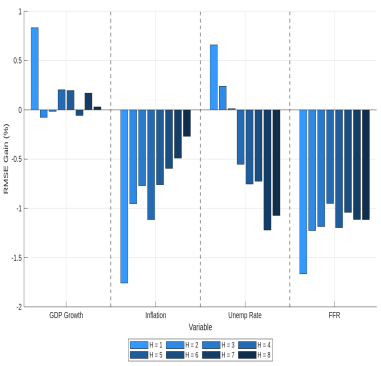
Small VAR



Medium VAR

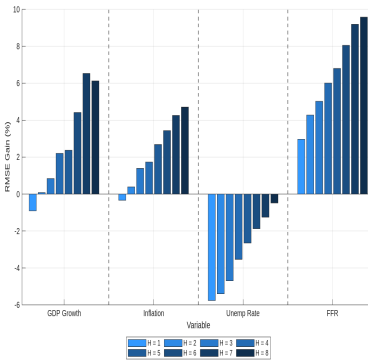


Large VAR

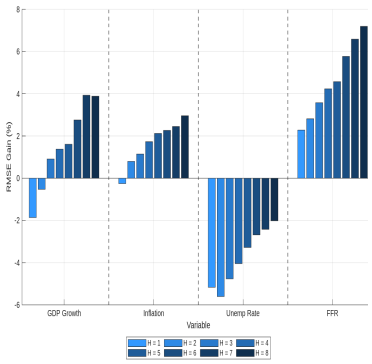


3. Financial Crisis

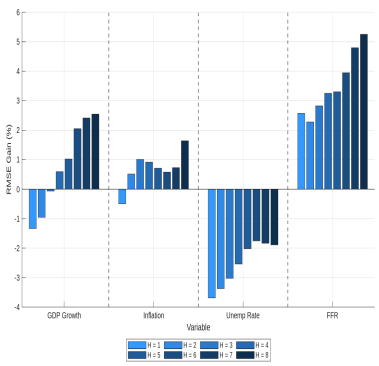
Small VAR



Medium VAR

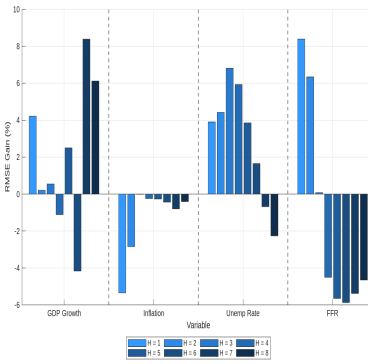


Large VAR

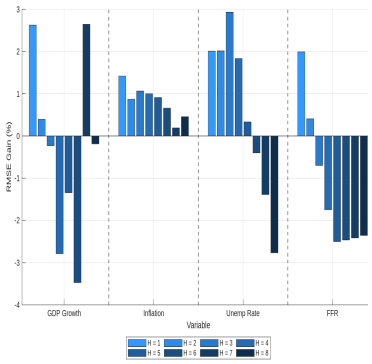


4. COVID

Small VAR



Medium VAR



Large VAR

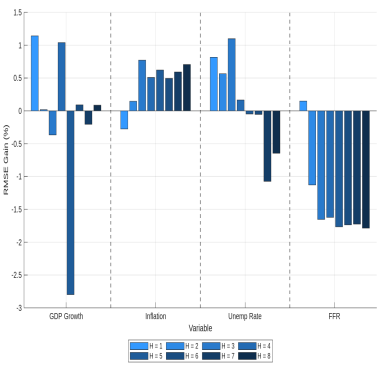


Table 7. Sample 1: p-values from DM test (posterior-mean λ)

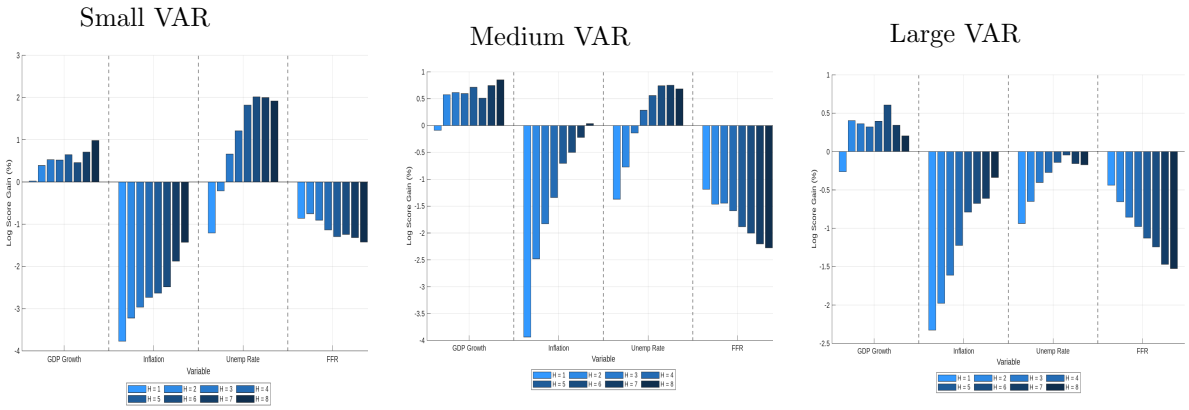
Horizon	1	2	3	4	5	6	7	8
Small								
- GDP growth rate	0.1560	0.0050	0.1891	0.3158	0.5329	0.9183	0.8604	0.4566
- Inflation	1.0000	0.9999	0.9996	0.9984	0.9965	0.9957	0.9870	0.9792
- Unemployment rate	0.9514	0.7863	0.4548	0.3503	0.2891	0.2659	0.3364	0.4090
- Federal funds rate	0.9261	0.9265	0.9196	0.9260	0.9412	0.9492	0.9659	0.9957
Medium								
- GDP growth rate	0.1856	0.0204	0.0667	0.6148	0.4636	0.9884	0.3556	0.5713
- Inflation	1.0000	1.0000	0.9997	0.9982	0.9837	0.9740	0.9031	0.7469
- Unemployment rate	0.8388	0.8054	0.5363	0.4877	0.4499	0.4843	0.5885	0.7417
- Federal funds rate	0.9983	0.9979	0.9980	0.9909	0.9840	0.9854	0.9935	0.9929
Large								
- GDP growth rate	0.3838	0.4802	0.3265	0.0344	0.0771	0.5642	0.2634	0.9852
- Inflation	1.0000	1.0000	0.9998	0.9984	0.9861	0.9760	0.9604	0.9310
- Unemployment rate	0.9863	0.9560	0.7314	0.5688	0.6965	0.6374	0.7969	0.7806
- Federal funds rate	0.9975	0.9859	0.9856	0.9844	0.9953	0.9923	0.9927	0.9995

Note: One-sided p-values of the Diebold–Mariano test with the Harvey, Leybourne, and Newbold (1997) small-sample correction; H_1: the REPS plug-in forecast is more accurate than the GLP plug-in forecast (posterior mean of λ). Squared errors of the predictive median. Sample 1 (forecast origins 1988Q1–2006Q1, 73 windows).

The density forecast results are also similar to the baseline (Figure 6). During the financial crisis, the REPS-based λ raises log scores for GDP growth, inflation, and the federal funds rate at longer horizons, and these gains are statistically significant in most cases, whereas log scores for the unemployment rate deteriorate significantly. In Sample 1, REPS yields small but significant gains for GDP growth and significant losses for inflation and the federal funds rate . Removing hyperparameter uncertainty from the GLP forecasts therefore leaves the relative density performance of the two methods largely unchanged. This suggests that, relative to uncertainty about the VAR coefficients and future shocks, uncertainty about λ contributes little to the dispersion of the GLP predictive distribution.

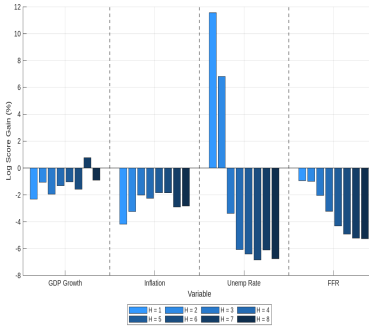
Figure 6: Percentage Change in Log Score

1. Sample 1

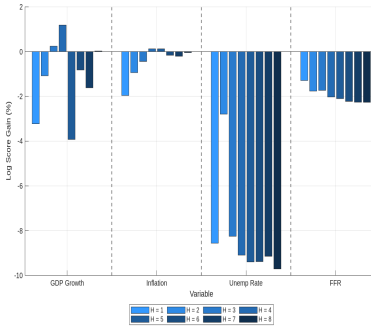


2. Sample 2

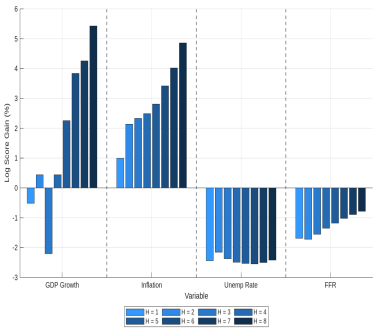
Small VAR



Medium VAR

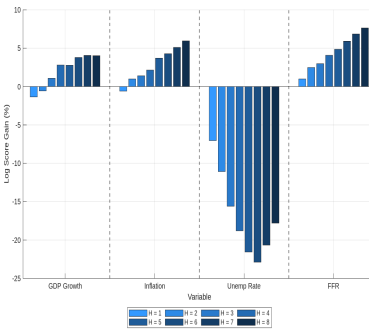


Large VAR

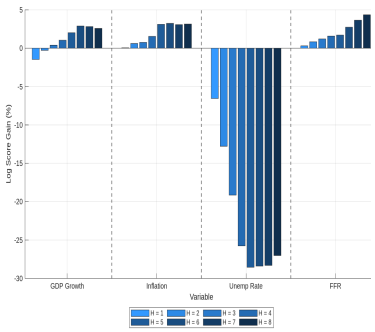


3. Financial Crisis

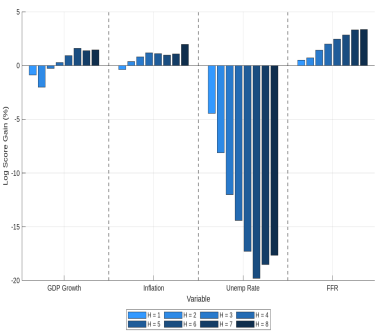
Small VAR



Medium VAR

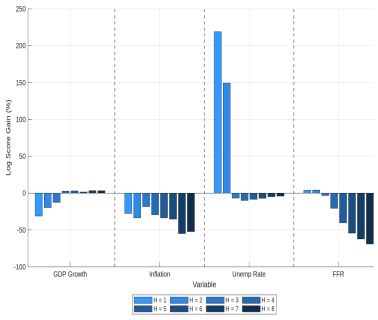


Large VAR

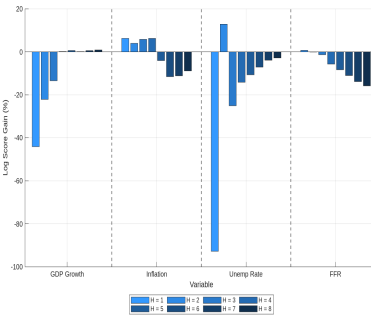


4. COVID

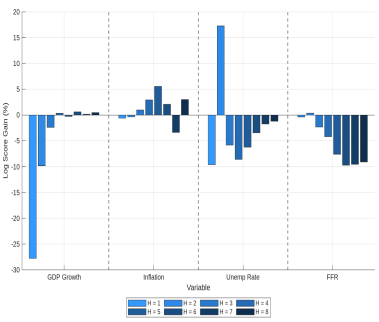
Small VAR



Medium VAR



Large VAR



A.2 Forecast Accuracy Under Alternative Forecast Aggregation Scheme

As an alternative to conditioning on a single REPS-based value of λ, I implement REPS as a posterior reweighting scheme. I retain the GLP posterior draws of λ and the associated predictive distributions, but replace the equal mixture weights used by GLP with REPS distortion weights. This preserves the posterior-mixture structure of GLP and assesses whether the REPS distortion has predictive content beyond the use of a single plug-in hyperparameter.

Let $\{\lambda_s\}_{s=1}^{S}$ denote the posterior draws of the tightness parameter obtained from GLP. The standard GLP predictive density is an equal-weight mixture,

$$p^{GLP}(y^{T+h}|Y) = \frac{1}{S}\sum_{s=1}^{S} p(y^{T+h}|Y,\lambda_s).$$

The REPS-reweighted predictive density keeps the same posterior draws but replaces the equal weights with REPS distortion weights,

$$p^{REPS\text{-}w}(y^{T+h}|Y) = \sum_{s=1}^{S} w_s^{REPS}\, p(y^{T+h}|Y,\lambda_s), \qquad w_s^{REPS} = \frac{\exp\left[\eta \bar{L}_s(\lambda_s - \bar{\lambda}^{REPS})\right]}{\sum_{j=1}^{S}\exp\left[\eta \bar{L}_j(\lambda_j - \bar{\lambda}^{REPS})\right]},$$

Table 8. Sample 1: p-values from AG test (posterior-mean λ)

Horizon	1	2	3	4	5	6	7	8
Small								
- GDP growth rate	0.9005	0.0144	0.0045	0.0006	0.0000	0.0083	0.0001	0.0000
- Inflation	0.0000	0.0000	0.0000	0.0000	0.0003	0.0005	0.0077	0.0568
- Unemployment rate	0.0000	0.5589	0.1707	0.0216	0.0004	0.0001	0.0001	0.0003
- Federal funds rate	0.0002	0.0065	0.0050	0.0042	0.0047	0.0111	0.0095	0.0049
Medium								
- GDP growth rate	0.6205	0.0005	0.0001	0.0004	0.0000	0.0006	0.0000	0.0000
- Inflation	0.0000	0.0000	0.0000	0.0001	0.0796	0.2037	0.4951	0.9018
- Unemployment rate	0.0000	0.0630	0.7660	0.5645	0.2460	0.1117	0.0984	0.1319
- Federal funds rate	0.0000	0.0000	0.0005	0.0015	0.0014	0.0017	0.0014	0.0008
Large								
- GDP growth rate	0.1372	0.0016	0.0063	0.0247	0.0015	0.0000	0.0109	0.2147
- Inflation	0.0000	0.0000	0.0000	0.0000	0.0161	0.0098	0.0170	0.1901
- Unemployment rate	0.0000	0.0016	0.1013	0.3237	0.6303	0.8778	0.5817	0.5562
- Federal funds rate	0.0322	0.0098	0.0060	0.0061	0.0024	0.0016	0.0003	0.0003

Note: Two-sided p-values of the Amisano–Giacomini (2007) test of equal log predictive scores, REPS plug-in versus GLP plug-in (posterior mean of λ), Sample 1 (forecast origins 1988Q1–2006Q1, 73 windows).

where $\bar{L}_s$ denotes the normalized marginal likelihood evaluated at λ_s, and η is the Lagrange multiplier associated with the entropy constraint. Since the REPS problem considered here minimizes the distorted posterior mean of λ, the multiplier is negative. As a result, posterior draws with relatively small values of λ_s receive larger weights when they are also supported by the marginal likelihood. The reweighted forecast therefore preserves the posterior-mixture structure of GLP while tilting posterior mass toward tighter priors.

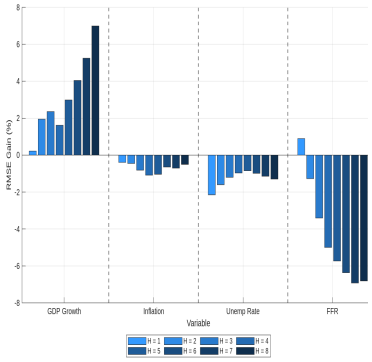
Figures 7 and 8 report the point and density forecast results of the reweighted implementation. The gains are largest in the small VAR, where reweighting improves forecasts of nearly all variables in every evaluation period, and most strongly in the crisis periods. During the financial crisis, the RMSFE falls by 6–13% at $h = 8$ for all four variables, including the unemployment rate, and log scores rise significantly at most horizons. During COVID, the gains for GDP growth, inflation, and the federal funds rate are larger still, although point forecasts of the unemployment rate deteriorate sharply and the estimates rest on only nine windows. Unlike the plug-in implementation, reweighting also improves forecasts in Sample 1, most notably for GDP growth, by up to 7% at $h = 8$, and for inflation density forecasts.

In the medium and large VARs, the gains are smaller and concentrated in GDP growth and inflation. Density forecasts of these variables improve significantly during the financial crisis and in Sample 1, whereas forecasts of the unemployment rate and the federal funds rate generally deteriorate, most clearly during COVID.

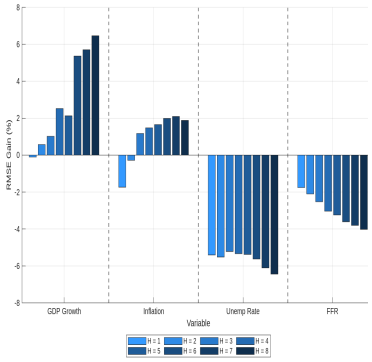
Figure 7: Percentage Change in RMSFE

1. Sample 1

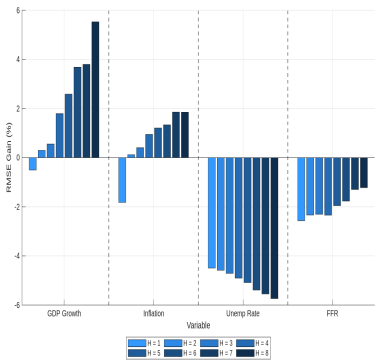
Small VAR



Medium VAR

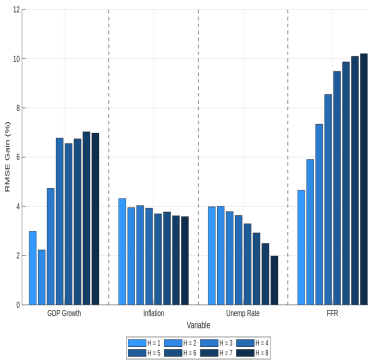


Large VAR

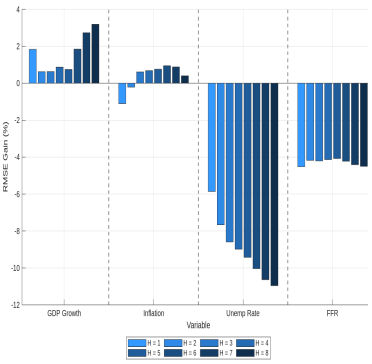


2. Sample 2

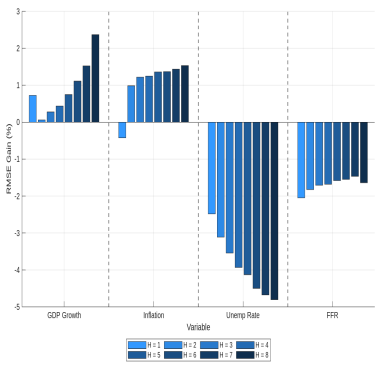
Small VAR



Medium VAR

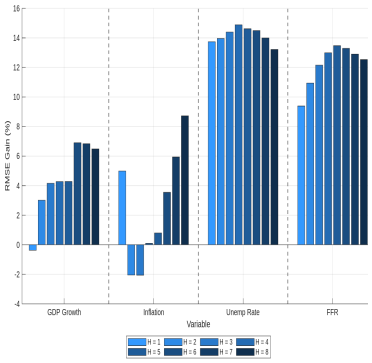


Large VAR

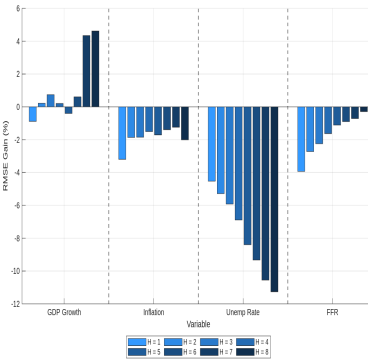


3. Financial Crisis

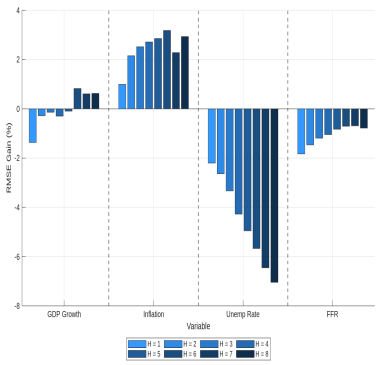
Small VAR



Medium VAR

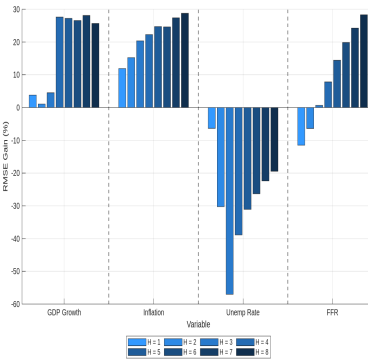


Large VAR

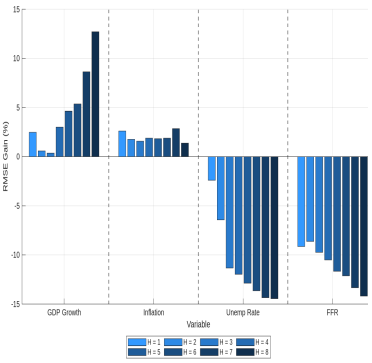


4. COVID

Small VAR



Medium VAR



Large VAR

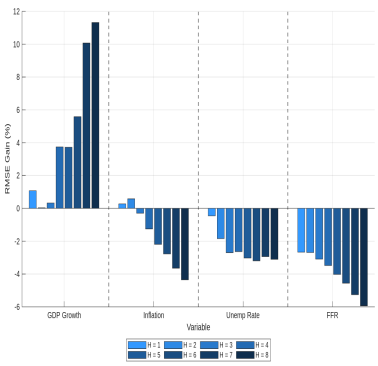
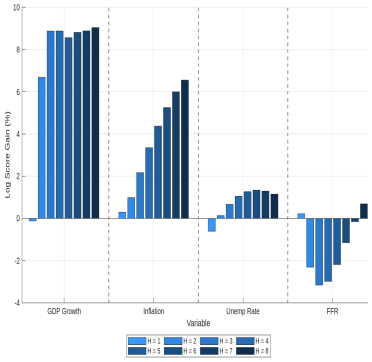


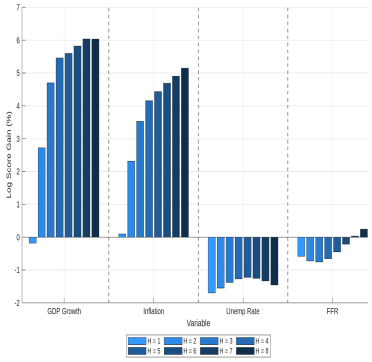
Figure 8. Percentage Change in Log Socre

1. Sample 1

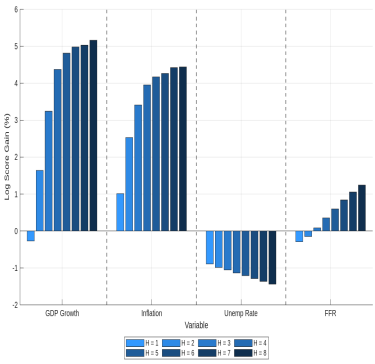
Small VAR



Medium VAR

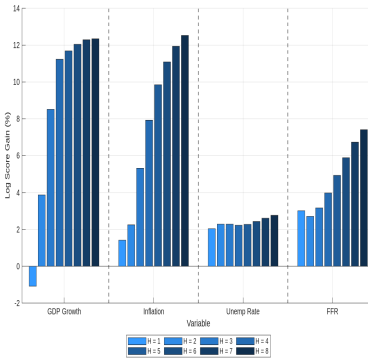


Large VAR

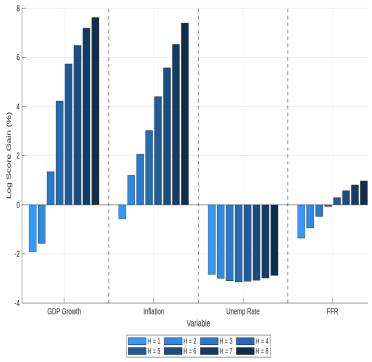


2. Sample 2

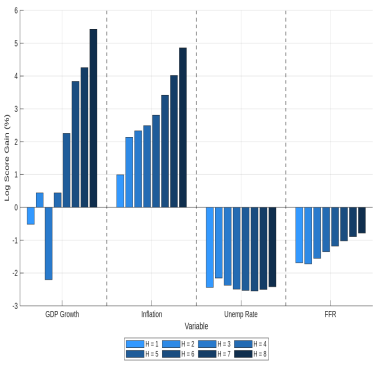
Small VAR



Medium VAR

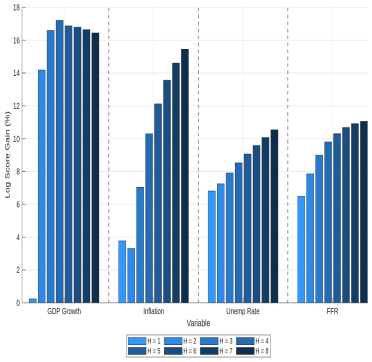


Large VAR

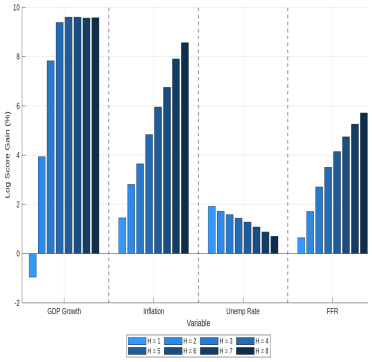


3. Financial Crisis

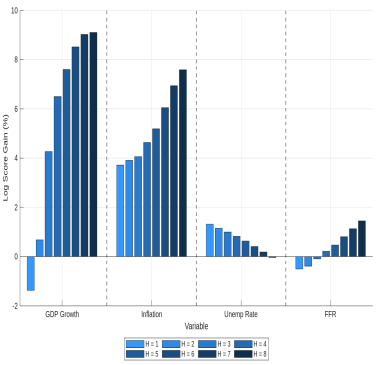
Small VAR



Medium VAR

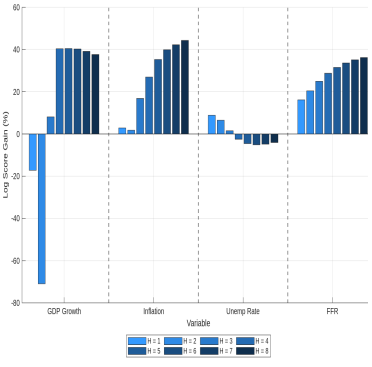


Large VAR

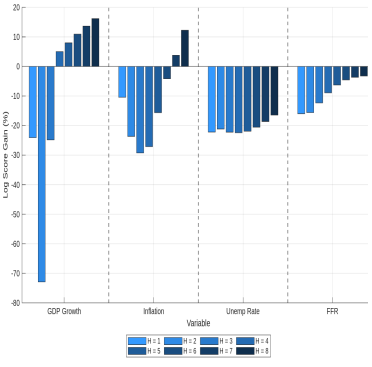


4. COVID

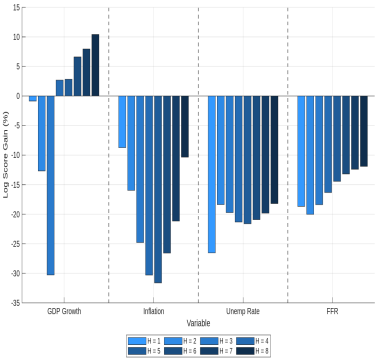
Small VAR



Medium VAR



Large VAR



One interpretation of the stronger results is that the two implementations use the REPS distortion differently. The plug-in implementation places all weight on a single tighter value of λ. The reweighted implementation instead downweights posterior draws with large values of λ, which generate more volatile forecasts, while retaining a range of tighter values supported by the marginal likelihood. Tightening the prior in this softer way avoids the cost of over-shrinking that the plug-in implementation incurs in periods such as Sample 1.

Overall, the reweighting exercise shows that the predictive content of the REPS distortion is not tied to the plug-in implementation. When the REPS weights tilt the GLP posterior toward tighter priors instead of collapsing it to a single value, the gains become larger and broader in the small VAR, while in larger models they remain concentrated in GDP growth and inflation.

B Drivers of the REPS Hyperparameter Distortion

Sections 6.1 through 6.3 examine whether the tighter shrinkage induced by REPS improves forecast accuracy and why its forecasting value differs across variables, horizons, and evaluation periods. This subsection addresses a distinct question: why does the magnitude of the REPS adjustment itself vary across model sizes and periods?

I document substantial heterogeneity in the REPS-based adjustment to λ. I then show that this heterogeneity can be understood through the interaction between the posterior support of λ and the likelihood-based REPS weights. This mechanism determines how far REPS moves the hyperparameter away from the GLP posterior mean. It does not, by itself, determine whether that movement improves out-of-sample forecast accuracy.

Observed hyperparameter distortions. Table 9 reports the percentage difference between the REPS-based value of λ and the posterior mean under GLP:

$$100 \times \frac{\lambda_{\text{REPS}} - \bar{\lambda}_{\text{GLP}}}{\bar{\lambda}_{\text{GLP}}}. \tag{1}$$

Table 9. Difference Between the REPS-Based λ and the GLP-based λ

Period	Small VAR	Medium VAR	Large VAR
Sample 1	-18.01	-14.29	-8.51
Sample 2	-19.48	-16.08	-10.50
Financial crisis	-19.01	-15.93	-35.03
COVID	-34.80	-30.15	-27.15

Note: Negative entries indicate that REPS selects a smaller value of the overall tightness parameter and therefore imposes stronger shrinkage. The GLP benchmark is the posterior mean of λ. Since the baseline GLP forecasts in the main exercise integrate over posterior draws of λ, the reported differences measure the magnitude of the hyperparameter adjustment rather than providing an exact decomposition of the forecast differences.

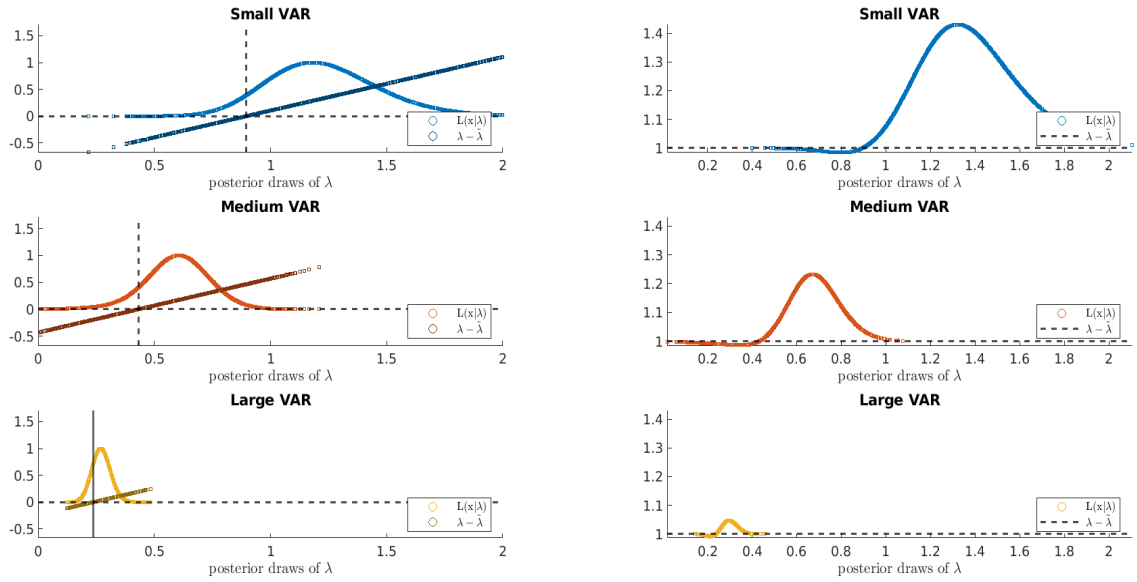
REPS selects a lower value of λ in every period and model size, but the magnitude of the adjustment differs substantially. In the small and medium VARs, the largest reductions occur during COVID. In the large VAR, the largest reduction occurs during the financial crisis. The

adjustment therefore does not follow a simple monotonic ordering by model size or evaluation period.

The table also shows that the size of the hyperparameter adjustment is not monotonically related to out-of-sample forecast performance. Sample 1 exhibits meaningful reductions in λ even though the forecasts of inflation and the federal funds rate generally deteriorate. During COVID, the reductions are larger, but the RMSFE effects remain unstable across variables and horizons. By contrast, during the financial crisis, tighter shrinkage produces systematic longer-horizon gains for GDP growth, inflation, and the federal funds rate. The magnitude of the distortion therefore measures how strongly REPS changes the forecasting model, not whether the change is beneficial.

Model-size and period heterogeneity. The percentage changes reported in Table 9 document substantial heterogeneity in the REPS adjustment across model sizes and evaluation periods. To illustrate the mechanism behind this heterogeneity, Figure 9 plots the marginal likelihood, the deviation term $(\lambda - \widetilde{\lambda})$, and the corresponding REPS weighting function for a selected rolling window in Sample 1.

Figure 9. Marginal Likelihood and REPS Distortion by Model Size



(a) $L(X \mid \lambda)$ and $(\lambda - \widetilde{\lambda})$ (b) REPS weighting function

Note: The horizontal line in Panel (a) denotes zero, and the vertical line denotes $\lambda = \widetilde{\lambda}$. The figure is constructed using the 70th rolling window in Sample 1. It illustrates the mechanism in one selected window rather than establishing a general ordering across model sizes.

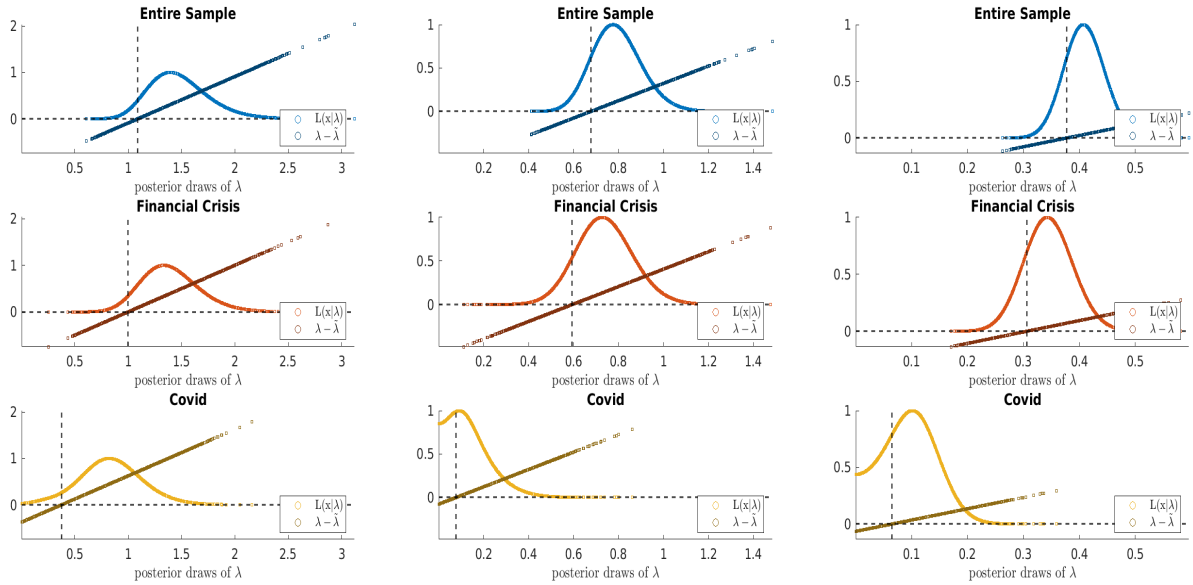
In this selected Sample 1 window, the small VAR exhibits support over a relatively broad range of λ values, whereas the corresponding distribution for the large VAR is more concentrated. The broader support provides greater scope for the REPS weighting function to move the selected value of λ. This pattern is consistent with the larger percentage adjustments observed for the small and medium VARs in Sample 1.

Differences across evaluation periods. Figure 9 provides a complementary illustration across Sample 1, the financial-crisis period, and COVID.

The period comparison illustrates that the posterior support and the likelihood shape relevant for the REPS adjustment change over time. In the selected COVID windows, a relatively broad range of tightly shrunk specifications receives non-negligible support. This pattern is consistent with the comparatively large reductions in λ reported for the small and medium VARs during COVID. The corresponding period-specific weighting functions are reported in Appendix A.2.

Because Figures 9 and 10 represent selected rolling windows, they should be interpreted as illustrations of the mechanism rather than as evidence of a universal ordering of likelihood concentration across all models and periods. Nevertheless, together with Table 9, they show how differences in posterior support and likelihood shape can generate heterogeneous REPS adjustments under the same procedure.

Figure 10. Marginal Likelihood and REPS Distortion Across Periods



(a) Small VAR (b) Medium VAR (c) Large VAR

Note: The horizontal line denotes zero, and the vertical line denotes $\lambda = \widetilde{\lambda}$. The figure uses the 36th, 85th, and 134th rolling windows, corresponding to Sample 1, the financial-crisis period, and COVID, respectively. The panels are selected-window illustrations.

Magnitude versus forecasting value. The magnitude of the REPS adjustment and its forecasting value are distinct. The posterior support, the marginal likelihood, and the REPS weighting function determine how far λ moves away from the GLP benchmark. Whether the resulting forecast improves depends on how tighter shrinkage changes the variable-specific forecast dynamics and whether those dynamics match the subsequently realized dynamics.

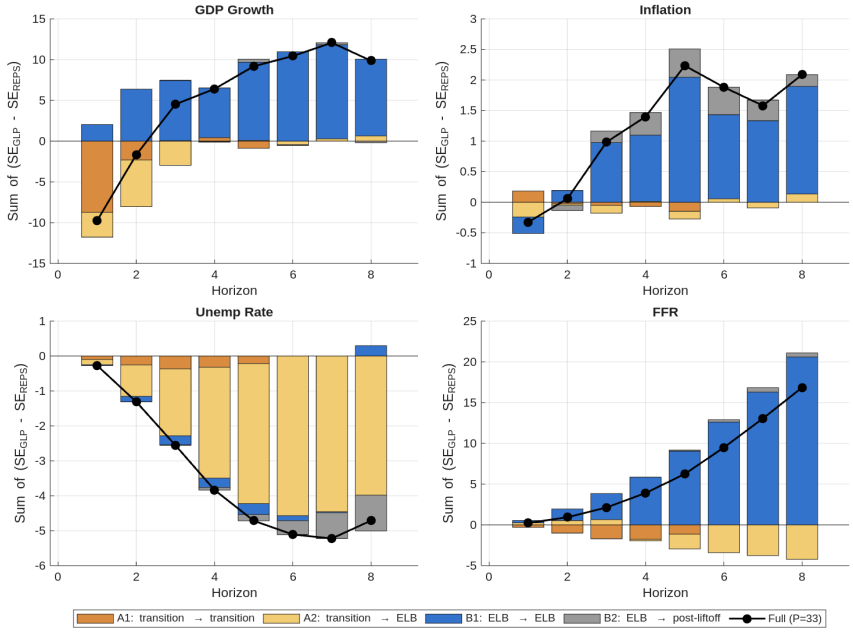
Sample 1 and COVID provide direct examples of this distinction. Both periods exhibit economically meaningful reductions in λ, but neither produces systematic RMSFE gains. During the financial crisis, in contrast, the stronger shrinkage generates simpler forecast paths that better match the prolonged post-crisis period for GDP growth, inflation, and the federal funds rate. Thus, a large hyperparameter distortion can produce a material change in the forecast without necessarily improving forecast accuracy.

C Decomposition of the forecasting performance (REPS-GLP)

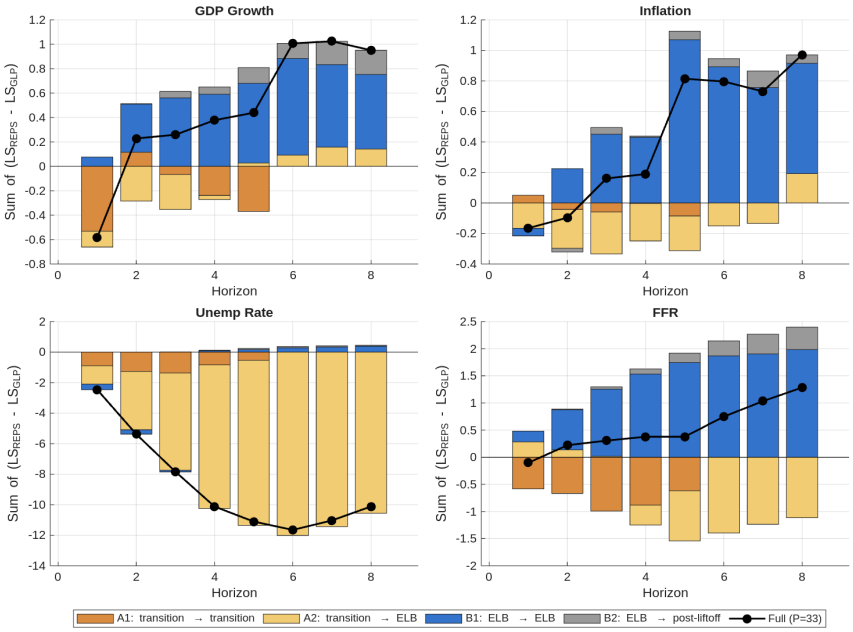
As in Section 6.3, the medium and large VARs show the same pattern, with smaller gains; in the large VAR, B1 also contributes to the unemployment loss.

Figure 11. Decomposition of the forecasting performance (REPS-GLP): (Financial crisis, medium VAR)

(a) Point forecasts



(b) Density forecasts



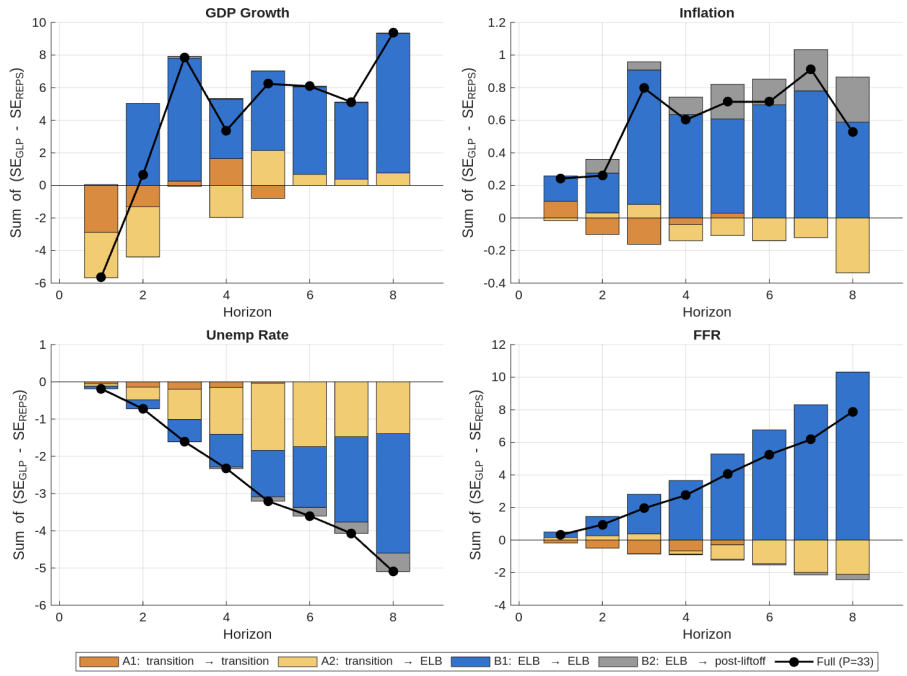
Notes:

(i) For variable i, horizon h, and forecast origin k, the loss differential in panel (a) is $d^{\mathrm{SE}}_{i,h,k} = \left(y^i_{k+h} - \hat{y}^{i,\mathrm{GLP}}_{k+h|k}\right)^2 - \left(y^i_{k+h} - \hat{y}^{i,\mathrm{REPS}}_{k+h|k}\right)^2$, where $\hat{y}$ is the point forecast. In panel (b) it is $d^{\mathrm{LS}}_{i,h,k} = s^{\mathrm{REPS}}_k(y^i_{k+h}) - s^{\mathrm{GLP}}_k(y^i_{k+h})$, where $s_k(\cdot)$ is the Gaussian log predictive score defined in Section 4.

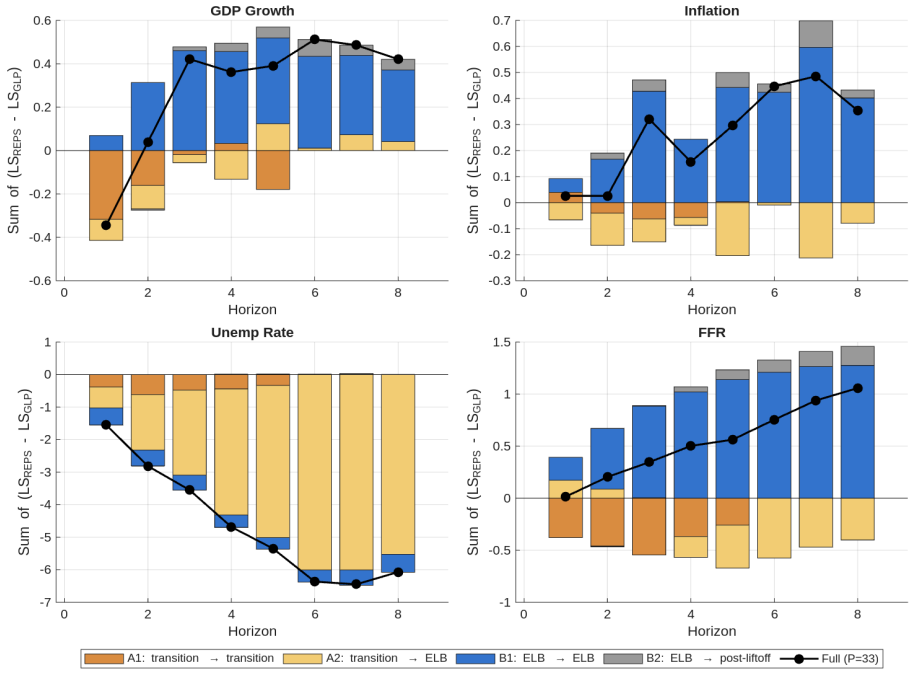
(ii) Each bar shows $\sum_{k \in G_h} d_{i,h,k}$ for group $G \in \{\mathrm{A1}, \mathrm{A2}, \mathrm{B1}, \mathrm{B2}\}$ defined in Table 6. The line shows the sum over all 33 forecasts, which equals the sum of the bars at every horizon. Positive values favor REPS.

Figure 12. Decomposition of the forecasting performance (REPS-GLP): (Financial crisis, large VAR)

(a) Point forecasts



(b) Density forecasts



Notes:

(i) For variable i, horizon h, and forecast origin k, the loss differential in panel (a) is $d^{\mathrm{SE}}_{i,h,k} = \left(y^i_{k+h} - \hat{y}^{i,\mathrm{GLP}}_{k+h|k}\right)^2 - \left(y^i_{k+h} - \hat{y}^{i,\mathrm{REPS}}_{k+h|k}\right)^2$, where $\hat{y}$ is the point forecast. In panel (b) it is $d^{\mathrm{LS}}_{i,h,k} = s^{\mathrm{REPS}}_k(y^i_{k+h}) - s^{\mathrm{GLP}}_k(y^i_{k+h})$, where $s_k(\cdot)$ is the Gaussian log predictive score defined in Section 4.

(ii) Each bar shows $\sum_{k \in G_h} d_{i,h,k}$ for group $G \in \{\mathrm{A1}, \mathrm{A2}, \mathrm{B1}, \mathrm{B2}\}$ defined in Table 6. The line shows the sum over all 33 forecasts, which equals the sum of the bars at every horizon. Positive values favor REPS.