Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy*

Ellis W. Tallman[†] Saeed Zaman[‡]

This version: March 11, 2019
Original version: June 21, 2018

Abstract

This paper constructs hybrid forecasts that combine forecasts from vector-autoregression (VAR) model(s) with both short and long-term expectations from surveys. Specifically, we use relative entropy to tilt one-step ahead and long-horizon VAR forecasts to match the nowcast and long-horizon forecast from the Survey of Professional Forecasters. We consider a variety of VAR models ranging from simple fixed-parameter to time-varying parameters. The results across models indicate meaningful gains in multi-horizon forecast accuracy relative to model forecasts that do not incorporate long-term survey conditions. Accuracy improvements are achieved for range of variables including those that are not directly tilted but are affected through spillover effects from tilted variables. The accuracy gains for hybrid inflation forecasts from simple VARs are substantial, statistically significant, and competitive to time-varying VARs, univariate benchmarks, and survey forecasts. We view our proposal as an indirect approach to accommodating structural change and moving end points.

Keywords: relative entropy, survey forecasts, structural change, density forecasts
JEL Classification Number: E17, C53, C11, C32

*We gratefully acknowledge comments from editor Mike McCracken, Associate editor, two anonymous referees, Domenico Giannone, Gary Koop, Michele Modugno, Aubrey Poon, Tatevik Sekhposyan, Benjamin Wong, our discussant Sharada Davidson, the participants at the 11th International Conference on Computational and Financial Econometrics, Post-Graduate Research Away Day at the University of Strathclyde, International Institute of Forecasters 38th International Symposium on Forecasting, 2nd Central Bank Forecasting Conference at the Bank of England, 2018 Midwest Econometrics Group Meeting, Fall 2018 Midwest Macroeconomics Meetings. We are grateful to Andrea Carriero and Todd Clark for sharing Matlab code estimating large BVAR with stochastic volatility. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Contact information: ellis.tallman@clev.frb.org and saeed.zaman@clev.frb.org.

[†]Federal Reserve Bank of Cleveland

[‡]Federal Reserve Bank of Cleveland and University of Strathclyde

1 Introduction

Macroeconomic forecasters often use atheoretical models for forecasting. Banbura, Giannone, and Reichlin (2010) show that large fixed-parameter VARs containing more than 100 variables can work effectively, a finding that has contributed to resurgence in the use of VARs in forecasting and policy analysis by both central banks and private forecasters.

In this paper, we propose a technique to adjust the forecasts of the implied trends from a VAR toward (plausible) values proposed by judgmental forecasters. Specifically, we utilize the technique of relative entropy to alter the medium-term to long-horizon VAR forecast to match that of the real-time survey long-horizon forecast.

Long-run forecasts contained in published surveys of professional forecasters are reasonable proxies for underlying trends because they adjust faster than the unrestricted long-run model forecasts in response to any exogenous and/or underlying shifts in the economy (e.g. Kozicki and Tinsley, 1998; Faust and Wright, 2013; Wright, 2013). The quicker adjustment of the survey expectations stems from the fact the survey participants have at their disposal indicators typically not included in the information set fed into the models, such as information about the value of the inflation target, central bank communication, and demographic factors.

In fixed-parameter VAR models, the unrestricted long-run forecasts converge to or close to the unconditional mean of the estimation sample (i.e implied trend or end points in the terminology of Kozicki and Tinsley, 1998), which at times differ substantially with economists' view of the long-run values for particular variables. Contributing to this divergence in views is the use of history for model estimation that may reflect an outdated characterization of the macroeconomic relationships including the unconditional means. Furthermore, the entire forecast trajectory would be biased in the direction of the implied trend estimated in the model over the full sample. This is because beyond five quarters the forecasts are increasingly influenced by the implied trend of the model and therefore a badly estimated trend is the primary source of forecast errors for medium-term forecasts (as emphasized in Clements and Hendry, 1999; Kozicki and Tinsley, 2001a,b; Clark and McCracken, 2008).

To address the misspecification issues arising from structural changes, econometricians have introduced various innovations such as time-varying parameters to standard VAR models (e.g. Cogley and Sargent, 2005; Primiceri, 2005). These models appear to do well in forecasting variables that exhibit notable structural shifts (e.g. D'Agostino et al, 2013). These models have become popular as computational demands required to operate them have become less binding due both to availability of greater computing power and also the introduction of newer methods. The latter advantage has also allowed the potential to estimate these models with larger information sets (e.g. Koop and Korobilis, 2013). However, these models are complex (as there are so many moving parts involved) requiring a certain level of sophistication thereby limiting their wider use.

Some practitioners in the US and other advanced economies estimate VAR models with data starting in 1985, i.e. after the well-documented structural change. Empirical evidence documenting the good forecasting ability of VAR estimated with shorter sample (e.g. Aastveit, et al, 2017), supports this practice.

To account for the numerous models in use, we consider the efficacy of our proposal

in a range of VAR models. Specifically, we consider a constant parameter VAR estimated with a longer sample (starting 1959), shorter sample (post-1985), and time-varying VAR. We consider model specifications both with and without stochastic volatility and perform our assessment on both small-scale and medium-scale VARs. All in all, our model-space consists of 10 models.

We use relative entropy to combine the VAR forecasts with the forecasts reported in the Survey of Professional Forecasters (SPF). Since its application to economic forecasting by Robertson et al (2005), the technique has gained widespread usage in combining model-based forecasts with external information. This increasing usage mainly stems from its ease of use, computational ease, and flexibility; it allows the forecaster to combine appropriately both the mean condition and the modeler’s confidence in that mean condition (i.e., variance) as illustrated in Krüger, Clark, and Ravazzolo (2017) denoted KCR from hereafter.

KCR uses relative entropy to generate conditional forecasts with moment conditions that match survey forecasts using the short-term forecast from the survey as the mean condition on a one-step-ahead VAR forecast. They construct variance conditions around their mean conditions using ex-post real-time survey nowcast errors (using a rolling window over a pre-forecast evaluation sample). Our paper extends KCR: in addition to tilting the VAR forecast in the nowcast quarter, we tilt the medium- to long-horizon forecast coming out of the VARs to match the long-horizon forecast reported in the external survey of forecasters, the SPF. Our methodology parallels that of Altavilla, Giacomini, and Ragusa (2017), who also use relative entropy to tilt the segments of the yield curve forecasts from the term structure models to match survey expectations.¹

Two popular approaches to directly influence VAR forecasts with survey long-run forecasts are: (1) steady-state BVAR (developed by Villani, 2009), Wright (2013) uses it to show that prior beliefs on the unconditional mean of the variable informed by a survey’s (Blue Chip) long-run forecasts lead to systematic improvements in forecast accuracy for a range of U.S. macroeconomic variables, especially for inflation; (2) modeling some or all VAR variables in ‘gaps’ where gaps are computed as deviation from trends either informed from survey long-run projections or univariate moving average methods (e.g. Kozicki and Tinsley, 2001a,b; Clark and McCracken, 2010; Zaman, 2013). The crucial difference between our approach and these existing approaches is that our approach influences forecasts post-model estimation. The insights in Wright (2013) motivate our study, and the results in this paper echo many of those reported in Wright.²

Our main question of interest is whether we can achieve any meaningful gains in the forecast accuracy of the VAR variables over the forecast horizon of interest to policymakers by forcing the medium-term to long-horizon forecast of the select number of VAR variables to match the published survey forecasts. Essentially the forecast of interest is

¹The empirical application in this paper fits within the broader context in the use of relative entropy for ‘theory-coherent forecasting’ as proposed in Giacomini and Ragusa (2014). In our case, the use of survey long-run forecasts as proxy for trends can be viewed as adjusting the forecasts from atheoretic VAR models toward long-run equilibrium values that are partly informed from economic theory.

²In a related work, Giannone, Lenza, and Primiceri (2018) propose a natural conjugate prior to influence the joint long-run behavior of the VAR variables modeled in levels (denote the prior PLR).

a hybrid forecast consisting of a *survey nowcast*,³ a *BVAR forecast*, and a *survey long-horizon forecast*.

Our empirical forecast evaluation results provide evidence of notable improvements in both the point and density forecast accuracy of hybrid VAR forecasts for several macroeconomic variables. All models in our set benefit from combination with survey information; not surprisingly, the accuracy gains are largest for the model specifications estimated with a longer sample and smallest for time-varying VARs but gains are statistically significant. This result indicates that tilting helps mitigate misspecification issues more in models that are thought to have a higher degree of misspecification (caused in part by badly estimated unconditional means). Relatedly, we find that all models benefit from tilting in the post-crisis period, a period associated with structural change (see Aastveit et al, 2017).

The accuracy gains are achieved for variables that are directly tilted and importantly also for variables that are indirectly influenced through the spillover effects of the tilted variables (e.g. wage inflation and payroll employment). Over the forecast horizon of interest to monetary policymakers (i.e., 1 quarter to 12 quarters ahead), the gains in forecast accuracy are strongest for inflation and the federal funds rate. We infer that the apparent structural break in inflation and the extended experience of the federal funds rate near the zero lower bound (ZLB) are the crucial challenges especially in VAR models estimated with fixed coefficients.

We summarize additional findings as follows. First, we show that the point forecast accuracy of hybrid CPI inflation forecasts from our modeling approach is competitive with tough to beat univariate benchmark models. The forecasts are also competitive to SPF and the Federal Reserve’s Greenbook. Our results offer policymakers a practical contribution. Monetary policymakers desire to use multivariate model(s) which allow for feedback effects from policy to the real economy and inflation. But often these multivariate models are unable to match the forecasting performance of the univariate forecasting models. The application in this paper provides one potential path.

Second, we show that hybrid forecasts generated from simple VARs estimated with post-World War II data are competitive with the forecasts generated from computationally intensive time-varying VARs. For inflation, hybrid forecasts are more accurate and the gains are statistically significant. This result is of practical importance for those practitioners who are reluctant to use time-varying VARs for variety of reasons including computational reasons, complexity issues, or shorter data sample (a primary concern in the case of developing and emerging market economies).

The paper is structured as follows. The next section describes the data and empirical models. Section 3 compares the survey long-run forecasts with the BVAR model long-run forecasts. Section 4 details our methodology of generating the hybrid forecasts. Section 5 reports the forecasting results. Section 6 concludes. Supplementary sets of results are reported in the companion Online Appendix.

³There is a long list of papers documenting the usefulness of nowcasts in helping improve the forecasting accuracy for future horizons (e.g., KCR; Knotek and Zaman 2019).

2 Data and the Empirical Model

2.1 Data

Our empirical examination uses real-time data at a quarterly frequency. Our model space contains small sized VARs (denoted Small VAR) consisting of five variables and medium sized VARs (denoted Medium VAR) consisting of 10 variables. The Small VAR consists of real GDP, the CPI, the unemployment rate, the effective federal funds rate, and the credit spread (defined as the difference between the yield on Baa corporate bond and the yield on the 10-year Treasury note). Both real GDP and the CPI enter in annualized quarterly rates and the remaining three variables, unemployment rate, effective federal funds rate, and credit spread are defined in units of percentage points. We use a Small VAR because several papers on VAR forecasting employ it as a benchmark VAR containing core variables of interest to central bankers. The Medium VAR adds to the Small VAR variables shown to be useful in improving forecasts of the core variables. Forecasts of the additional variables may be of their own interest to central bankers, like for example, productivity growth and wage inflation measures. Specifically, these additional five variables include real personal consumption expenditures, nonfarm business productivity, the employment cost index-wage and salary of private workers (ECI), nonfarm payroll employment, and the core CPI (i.e., the CPI excluding food and energy). All these variables are transformed to quarterly annualized growth rates. To compute the growth rates, we use 400 times the log difference formula.

We use both the Federal Reserve Bank of Philadelphia’s real-time data set for macroeconomists and the Federal Reserve Bank of Saint Louis’s ALFRED database to construct our real-time quarterly data set. Quarterly data on financial variables, which are real-time by construction, are downloaded from Haver Analytics.

All vintages of real-time quarterly data coincide with the survey date of the SPF, which is a quarterly survey released approximately in the middle of the second month of the quarter. That is, each forecast origin coincides with the SPF survey release date and so whatever quarterly data is available is utilized for estimation. For example, forecast origin in February 2016, for real GDP would use the advance estimate for 2015.Q4 (reported by the Bureau of Economic Analysis in the last week of January 2016); for monthly variables such as unemployment rate, it will be the second estimate for 2015.Q4 (reported by the Bureau of Labor Statistics in early February 2016). The real-time vintages start in 1994.Q1 and end in 2016.Q4. In each real-time vintage, data sample begins in 1959.Q4; in the 1994.Q1 vintage the data sample goes through 1993.Q4; in the last vintage (i.e. 2016.Q4) the data sample ends 2016.Q3. The first real-time forecasting run is performed with estimation data ending in 1993.Q4, and out-of-sample forecasts generated one to twelve quarters-ahead, i.e. for 1994.Q1 through 1996.Q4.

For the purpose of forecast evaluation, we treat the ‘truth’ as the third quarterly estimate (as in Tulip 2009).⁴

⁴For example, in the case of real GDP, for 2015.Q4, the third estimate would correspond to the actual value available as of early 2016.Q2. Similarly, for unemployment rate which is a monthly variable, the third quarterly estimate for the reference quarter (e.g. 2015.Q4) would coincide with the third revision to the final month of the quarter (e.g. December 2015) which will roughly be the first week of

We collect the SPF nowcasts and SPF long-horizon forecasts for real GDP growth, CPI inflation, and the unemployment rate. The SPF does not report nowcast and long-run forecast for the federal funds rate but it does report the long-run forecast for 3-month Treasury bill. Accordingly, we treat the long-run forecast for 3-month bill as the long-run estimate for the federal funds rate.⁵ For the nowcast, we use the intra-quarterly daily data on the federal funds rate along with the simple daily random walk model of Knotek and Zaman (2019) to construct an estimate.⁶ We collect the daily data starting from January 1, 1994 through December 31, 2016 using Haver Analytics. For illustrative purposes, we also collect the long-run projections from Blue Chip Economic Indicators (and Blue Chip Financial Forecasts), Federal Open Market Committee’s (FOMC) Summary of Economic Projections (SEP). For the formal forecast evaluation exercises reported in the paper, we use nowcasts and long-horizon forecasts from the SPF.

Finally, to assess the accuracy of Federal Reserve Board’s Greenbook forecasts, we collect forecasts for real GDP, CPI inflation, and the unemployment rate (starting January 1994 FOMC meeting through December 2012 meeting) from Philadelphia Fed’s website (Real-Time Data Research Center).

2.2 Bayesian VAR Models

In this paper, all empirical examinations use VAR models estimated with Bayesian methods. We consider both fixed-parameter and time-varying parameter models (with and without stochastic volatility). Following a long-list of papers on forecasting with VARs, the lag order (p) is set to 4 in the case of fixed-parameter VARs and to 2 in the case of time-varying VARs. Our model specifications follow Clark and Ravazzolo (2015).

2.2.1 VAR with Constant Volatility

$$Y_t = A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \quad (1)$$

where $t = 1, \dots, T$, Y_t is an $n \times 1$ vector of n observed variables, A_0 is an $n \times 1$ vector of intercepts, $A_1, \dots, A_p$ are $n \times n$ matrices of coefficients, ε_t is an $n \times 1$ vector of error terms distributed normally with zero mean and variance-covariance matrix, $\Sigma = E\varepsilon_t \varepsilon_t'$.

For estimation details please refer to the Technical appendix.

second quarter following the reference quarter (e.g. 2016.Q2). We emphasize if instead we would have treated the ‘truth’ as latest available estimate, our results for relative scores are both qualitatively and quantitatively similar. Please refer to the results for Medium VAR in the working paper version of the paper (Tallman and Zaman, 2018). We use Haver Analytics to collect the most revised quarterly data for forecast evaluation (i.e. vintage available as of August 2017).

⁵Historically there is a small gap between the two with the federal funds rate averaging roughly 30 basis points higher on quarterly basis compared to the 3-month Treasury bill.

⁶The procedure involves using the available daily reading as of the SPF survey date and using that to fill the missing trading days of the quarter. The average of the daily readings (which includes the daily data and the random walk forecast) within the quarter constitutes our nowcast estimate.

2.2.2 VAR with Stochastic Volatility

The stochastic volatility process is modeled using the estimation procedure of Carriero, Clark, and Marcellino (2016). The algorithm assumes a Kronecker structure for the multivariate stochastic volatility and therefore makes it computationally feasible to estimate large VAR models.

$$\begin{aligned}
Y_t &= A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t \\
\varepsilon_t &= B^{-1} \Lambda_t^{0.5} \mu_t, \quad \mu_t \sim N(0, I_n), \quad \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{n,t}) \\
\log(\lambda_{j,t}) &= \log(\lambda_{j,t-1}) + e_{j,t}, \quad j = 1, n \\
e_t &\equiv (e_{1,t}, \dots, e_{n,t})' \sim N(0, \Phi) \\
\Sigma_t &\equiv \text{Var}(\varepsilon_t) = B^{-1} \Lambda_t B'^{-1}
\end{aligned} \tag{2}$$

where B is a lower triangular matrix with ones on the main diagonal and nonzero elements below it; $\lambda_{1,t}, \dots, \lambda_{n,t}$ are the diagonal elements of the matrix Λ_t representing the time-varying variances of the shocks and are assumed to evolve according to a geometric random walk; the variance-covariance matrix Φ of innovations e_t is assumed to be full rank, i.e., correlation among innovations of different equations is permitted; the Σ_t is the variance-covariance matrix of the reduced form residuals ε_t .

For complete estimation details, please refer to Carriero, Clark, and Marcellino (2016).

2.2.3 Time varying parameters VAR with Constant Volatility

The specification follows the setup as in Cogley and Sargent (2005) but allowing only regression coefficients (including the intercepts) to be time-varying and so does not allow stochastic volatility.

$$\begin{aligned}
Y_t &= X_t' A_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \\
A_t &= A_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, Q)
\end{aligned} \tag{3}$$

where $X_t = I_n \otimes [1, Y_{t-1}', \dots, Y_t - p']$; $A_t = \text{vec}(A_0, A_1, \dots, A_p)$ is a $n(k) \times 1$ column vector. The VAR coefficients stacked in the vector A_t are assumed to evolve independently according to a random walk with shocks ϵ_t that are permitted to be correlated across equations (i.e. Q is full rank).

For estimation details, please refer to Koop and Korobilis (2010).

2.2.4 Time varying parameters VAR with Stochastic Volatility

The specification follows exactly the setup laid out in Primiceri (2005) but the estimation procedure follows Del Negro and Primiceri (2013).

$$\begin{aligned}
Y_t &= X_t' A_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \Sigma) \\
A_t &= A_{t-1} + \epsilon_t, & \epsilon_t &\sim N(0, Q) \\
\beta_t &= \beta_{t-1} + \nu_t, & \nu_t &\sim N(0, R) \\
\varepsilon_t &= B_t^{-1} \Lambda_t^{0.5} \mu_t, & \mu_t &\sim N(0, I_n), \quad \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{n,t}) \\
\log(\lambda_{j,t}) &= \log(\lambda_{j,t-1}) + e_{j,t}, & j &= 1, n \\
e_t &\equiv (e_{1,t}, \dots, e_{n,t})' \sim N(0, \Phi) \\
\Sigma_t &\equiv \text{Var}(\varepsilon_t) = B_t^{-1} \Lambda_t B_t^{-1}
\end{aligned} \tag{4}$$

where B_t is a lower triangular matrix with ones on the main diagonal and nonzero elements below it; β_t is a column vector that stacks (by row) off-diagonal and non-zero elements of matrix B_t and is assumed to evolve according to random walk with innovations that are permitted to be correlated across equations; VAR coefficients stacked in the vector A_t are assumed to evolve independently according to a random walk with shocks that are permitted to be correlated across equations; $\lambda_{1,t}, \dots, \lambda_{n,t}$ are the diagonal elements of the matrix Λ_t representing the time-varying variances of the shocks and are assumed to evolve according to a geometric random walk; the variance-covariance matrix Φ of innovations e_t is assumed to be full rank, i.e., correlation among innovations of different equations is permitted; the Σ_t is the variance-covariance matrix of the reduced form residuals ε_t .

2.2.5 Model set

VAR models are commonly estimated either using longer history (e.g. post World War II data) or with shorter history (e.g. post-1985). Accordingly, we estimate our set of VAR models using a longer sample (starting 1959.Q4 onwards) and shorter sample (starting 1985.Q1 onwards). As discussed earlier, we consider both smaller-sized (i.e. 5 variables) and medium-sized (i.e. 10 variables) specifications. Accordingly, we estimate VAR with constant volatility and VAR with stochastic volatility for both smaller and medium sized versions. For time-varying VARs only a small-scale three variable version (containing real GDP growth, CPI inflation, and unemployment rate) estimated over a longer sample is examined. All in all, our model space consists of 10 VAR models:⁷

1. Small VAR est 1960: fixed-parameter small VAR est. with data 1959.Q4 onwards
2. Small VAR SV est 1960: fixed-parameter small VAR with SV est. 1959.Q4 onwards
3. Small VAR est 1985: fixed-parameter small VAR est. with data 1985.Q1 onwards
4. Small VAR SV est 1985: fixed-parameter small VAR with SV est. 1985.Q1 onwards

⁷We abstract from examining our proposal on nonlinear VAR models (e.g. Threshold VARs and Markov-Switching VARs) for three reasons: (1) Several recent papers have empirically shown that the forecasting performance of these models are at best competitive to time-varying VARs (e.g. Barnett et al, 2014; Alessandri and Mumtaz, 2017; Aastveit et al, 2017); (2) They are less popular for forecasting compared to the models we consider; (3) To keep our results manageable.

5. Medium VAR est 1960: fixed-parameter med VAR est. with data 1959.Q4 onwards
6. Medium VAR SV est 1960: fixed-parameter med VAR with SV est. 1959.Q4 onwards
7. Medium VAR est 1985: fixed-parameter med VAR est. with data 1985.Q1 onwards
8. Medium VAR SV est 1985: fixed-parameter med VAR with SV est. 1985.Q1 onwards
9. TVP VAR: time-varying small VAR est. with data 1959.Q4 onwards
10. TVP VAR SV: time-varying small VAR with SV est. with data 1959.Q4 onwards

For TVP VAR and TVP VAR SV, the first 10 years of data (i.e. 1959.Q4 to 1969.Q3) are used as a training sample for prior elicitation.

2.2.6 Forecast Evaluation Metrics

The models defined above are estimated using standard Markov Chain Monte Carlo (MCMC) methods. For the details on the precise algorithms we refer the readers to the papers cited under the description of each model.

At each forecast origin, for each model, the estimation involves simulating the model with ‘D’ draws.⁸ For each posterior draw, the simulated forecast path is constructed by iterating the model h quarters forward (where $h=1$ to 40). The ‘D’ forecast paths constitute multivariate predictive density. The point forecast is simply the (posterior) mean of the empirical predictive density.⁹

We perform a real-time out-of-sample forecasting evaluation using a recursively expanding estimation window. The forecast evaluation sample spans 1994.Q1 through 2016.Q4. We evaluate point forecasts using the widely used metric of mean squared forecast error (MSE),

$$MSE_{i,h} = \frac{\sum_{t=T_0}^{T_1-h} (Y_{i,t+h} - \hat{Y}_{i,t+h})^2}{T_1 - h - T_0 + 1} \quad (5)$$

where i corresponds to the macroeconomic variable of interest (e.g., real GDP growth), T_0 denotes one quarter prior to the start of the evaluation period (e.g., 1993.Q4), T_1 the end of the evaluation period (2016.Q4), $Y_{i,t+h}$ is the actual realization, and $\hat{Y}_{i,t+h}$ is the forecast.

We evaluate the performance of density forecasts using the metric continuous ranked probability score (CRPS) as proposed by Gneiting and Raftery (2007). CRPS measures the closeness between the actual realization and the predictive density: the closer the distribution to the actual realization, the smaller is the CRPS value and the more accurate is the predictive density.¹⁰ Accordingly, it is defined as

⁸D=20,000 for time-invariant VAR model specifications; D=40,000 + 4000 (burn-in) for all model specifications with stochastic volatility and or time-varying parameters; results are very similar if we instead use 20,000 for burn-in.

⁹In the case of time-varying VARs, following the common practice, the point forecast is defined as the median of the predictive density.

¹⁰Assessment based on log-score metric largely confirms the inference obtained from CRPS, and the results are reported in the online appendix A3.

$$CRPS_t^{i,h}(Y_{i,t+h}) = \int_{-\infty}^{\infty} (F(z) - 1 \{Y_{i,t+h} \leq z\})^2 dz = E_p |Y_{i,t+h}^* - Y_{i,t+h}| - 0.5 E_p |Y_{i,t+h}^* - Y_{i,t+h}^+| \quad (6)$$

where $Y_{i,t+h}$ is the actual realization, F is the cumulative distribution function corresponding to the predictive density f , $1 \{Y_{i,t+h} \leq z\}$ is an indicator function that takes a value of 1 if $Y_{i,t+h} \leq z$ and a value of 0 otherwise, and $Y_{i,t+h}^*$ and $Y_{i,t+h}^+$ are independent random draws from $p(Y^{T+1,T+H}, \theta | Y^T)$.

The CRPS metric favors predictive densities that have higher probabilities near the actual realization. As defined above, a lower CRPS would be preferable to a higher score. We report the average of the CRPS, computed over our forecast evaluation period.

To assess the statistical significance of the differences in forecast accuracy between the baseline and the hybrid forecasts, we follow KCR and Altavilla, Giacomini, and Ragusa (2017). We use the Diebold and Mariano (1995) and West (1996) test of equal predictive accuracy for pairwise comparisons of the RMSE using the two-sided tests of standard normal. In computing the test, we use the HAC variance estimator (an input into the test statistic) with the lag $h - 1$ truncation parameter and adjust the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); see Clark and McCracken (2013). As emphasized in KCR the use of our test statistics based on standard normal critical values are likely to be on the conservative side and should be treated as an approximation that, in our case, deals with issues such as the nesting of forecasts and conditional forecasting (see Clark and McCracken, 2017).

The density calibration (i.e. absolute accuracy) of the density forecasts is assessed using interval forecasts (i.e. 70% prediction intervals), and assessments of probability integral transforms (PITS) via battery of statistical tests: Knüppel (2015), Berkowitz (2001), and Kolmogorov-Smirnov. For the latter, all three statistical tests generally point to similar inference therefore we just report the results from Knüppel test. For space limitations, the density calibration results are reported in the online appendix A2.

3 Real-time Long-horizon survey forecasts versus BVAR forecasts

In the United States, the SPF and the Blue Chip Economic Indicators are the two most widely known and easily available forecast surveys routinely published. The SPF is a quarterly survey released in the middle of the second month of the quarter. Each SPF survey release contains the forecasts of the macroeconomic and financial variables for the current quarter (i.e., nowcasts) and forecasts up to four quarters ahead. For the survey carried out in the third quarter of the year, the SPF asks respondents for their estimates of the natural rate of unemployment. Similarly, for the surveys carried out in the first quarter, the SPF asks respondents their projections of long-run values (defined as 10 year annual average) of real GDP growth, the short-term interest rate (i.e., yield on the

3-month Treasury bill), and a few other variables. The SPF contains the projections for all the core set of variables of interest for this paper; for our purposes, we treat the SPF's projections of the natural rate of unemployment as the long-run forecast of the unemployment rate. Following the forecasting literature, for the SPF projections, we use the median projection, and for the Blue Chip we use the mean projection. The evolution of the long-run forecasts across the two surveys is fairly similar therefore for the sake of brevity, in this section we only focus on the SPF and estimates from the Blue Chip are relegated to the online appendix for interested readers.¹¹

Our forecast evaluation exercises in this paper will use the nowcasts and long-run forecasts from the SPF.¹²

The four panels in Figure 1 plot the real-time evolution of the macroeconomic long-run forecasts for real GDP growth (upper left), unemployment rate (upper right), CPI inflation (lower left), and short-term interest rate (lower right) respectively. Each panel, plots forecasts from Small VAR estimated using longer history (1960+), Small VAR estimated over shorter sample (1985+), a small time-varying VAR with SV, and the SPF. The sample period spans 1994.Q1 to 2016.Q4, matching our forecast evaluation sample.

The real-time data that we use to estimate our VAR models would be a subset of the information set available to the professional forecasters. The professional forecasters would likely rely on larger information set, including judgmental opinions of their own and of the subject matter experts along with their own econometric methodologies to come up with their forecasts. The use of forward guidance by the central bank and, more generally, the era of more predictable monetary policy (through central bank communications) since the beginning of the financial crisis are other examples of important information at the disposal of the survey participants. As a result, survey participants (collectively) are likely to have more timely and more informed long-run projections. Figure 1 provides some evidence in support of this claim.

The figure, indicates two notable observations. First, the long-run forecasts from VAR estimated with longer history adjust very sluggishly. In contrast, survey projections fluctuate considerably more. Second, time-varying VAR that is built to explicitly accommodate structural change (and uses all the available history for estimation) on average appears to be adjusting more rapidly compared to its time-invariant counterpart. However, it adjusts slower compared to SPF.

Starting with real GDP growth, the movements in forecasts from TVP-VAR-SV and Small-VAR(1985) are generally similar to forecasts from the SPF, with SPF projections on the lower side. At the beginning of 1994, all three, SPF, TVP-VAR-SV and Small-VAR(1960) were forecasting the underlying growth in the range of 2.7 to 2.9 percent,

¹¹In the online appendix, also plotted alongside the Blue Chip estimates are the estimates of the long-run values from the FOMC's SEP.

¹²Our choice of the SPF is motivated in part because it is publicly available. Relatedly, Croushore (2010) documents the good inflation forecasting properties of long-run forecasts of CPI inflation from the SPF and suggests using them as a proxy for inflation expectations.

roughly four-tenths lower than that of the Small-VAR(1985). As time rolled forward from 1994 through 1996, professional forecasters gradually lowered their estimate while forecasts from the VARs remained steady. In early 1997, professional forecasters had revised their forecasts back up by a couple of tenths to 2.5 percent and it remained stable at that level through the end of 1999. Over this same period, forecasts from all three VARs were also revised up. Moving into 2000, while the VARs' projection held steady in the range between 3.1 and 3.4 percent, professional forecasters strongly revised up their projection by roughly six-tenths, to 3.1 percent. This upward revision was in response to stronger growth data the prior two years averaging more than 4 percent. The upward revision continued through 2005, at which point long-run forecasts by both survey forecasters and the VAR were roughly in agreement at 3.4 percent. Beginning in 2006 and onward, the survey forecasters gradually lowered their growth forecasts, reaching 2.3 percent by the end of 2016. This rate of growth roughly matches the US economy's average growth rate since the start of the post-crisis recovery. Over this same period, forecasts from all three VARs also edged lower but by a smaller margin (3.3 to 3.0 for Small-VAR(1960), and 3.0 to 2.6 percent for both TVP-VAR-SV and Small-VAR(1985)) compared to professional forecasters (from 3.4 to 2.3 percent).

In the case of the unemployment rate, while the forecasts implied from the VAR models have fluctuated in a narrow range around 6 percent, the estimate of the natural rate of professional forecasters has evolved in line with the movements in the business cycle. For example, beginning in 1994 through the start of 2001, the estimate trended lower, but at the onset of the 2001 recession, it reversed and began to trend up until the beginning of the recovery. Thereupon it trended lower until the onset of the Great Recession. In response to a large upward spike in the unemployment rate reflecting the severity of the recession and the associated disruptions to the labor market, the professional forecasters rapidly adjusted upward their projections of the natural rate of unemployment. By the end of 2011, the professional forecasters' estimate of the natural rate of unemployment increased to 6.0 percent, close to that implied by the VAR. Thereafter, as the recovery picked up pace, professional forecasters adjusted their estimates downward reaching 4.8 percent by the end of 2016.

We next examine forecasts for the nominal variables, CPI inflation and short-term interest rate. First, over our forecast evaluation sample, there is a noticeable downward trend in the projections. Second, compared to the plot for real variables, there is a sizable gap between the forecasts from the professional forecasters and the VAR forecasts. For CPI inflation, even though all three VAR projections continue to gradually trend lower, the SPF projection is relatively stable from 1999 onwards. The projection from the TVP-VAR-SV is very similar to that from Small-VAR(1985) especially since the start of the recovery in late 2009. However, there is a sizable but a declining gap between the SPF projection and projections from each of the VARs. Since 2009.Q1, the SPF projection is on average 1.70 percentage points lower than the projection from Small-VAR(1960), 0.50 percentage point lower compared to Small-VAR(1985) and 0.70 percentage point lower compared to TVP-VAR-SV. In the case of the short-term interest rate, since 2009.Q1, the SPF projection is on average 288 basis points lower than the projection from Small-

VAR(1960), and 160 basis points lower compared to Small-VAR(1985).

Overall, these charts provide suggestive evidence that professional forecasters are quick to adjust inflation expectations and to recalibrate their estimates of underlying trend growth. In reality, it is difficult to distinguish between transitory fluctuations and fluctuations associated with changes to the underlying trend. As a result, forecasters gradually learn about shifts in the underlying trend. Even then, their expectations adjust more rapidly than implied by the statistical models.

4 Methodology for Tilting Forecasts

4.1 Relative Entropy

The technique of relative entropy, applied by Robertson, Tallman and Whiteman (2005) to economic forecasting, consists of modifying a given predictive distribution to a new predictive distribution such that it satisfies a given set of moment conditions while minimizing the relative entropy between two predictive distributions.

Let's begin with an unrestricted predictive distribution, $p(Y^{T+1,T+H} | Y^T)$, corresponding to an n -dimensional random variable Y obtained from a VAR model. Assume this predictive density consists of D draws $\{Y_i, i = 1, \dots, D\}$, then the corresponding weights are $\{w_i = 1/D, i = 1, \dots, D\}$. Suppose now that the modeler wants to impose moment conditions contained in the matrix $\bar{g}$ on this predictive distribution $p(Y^{T+1,T+H} | Y^T)$ such that $\sum_{i=1}^D w_i p(Y_i^{T+1,T+H}) \neq \bar{g}$, i.e., the mean of the predictive distribution $p(\cdot)$ is not equal to the mean condition required by the modeler (denote it as "new" information). For the predictive distribution to satisfy the new information requires modifying the original weights $\{w_i, i = 1, \dots, D\}$. The new weights $\{w_i^*, i = 1, \dots, D\}$ that satisfy this new information is equivalent to finding a new predictive distribution that is as close as possible to the original predictive density in the information-criterion sense.

Specifically, the relative entropy or the Kullback-Leibler Information Criterion (KLIC) of w^* to w is

$$K(w^* : w) = \sum_{i=1}^D w^* \log\left(\frac{w_i^*}{w_i}\right) \quad (7)$$

We solve for new weights that minimize $K(w^* : w)$ and satisfy the following constraints

$$w_i^* \geq 0, \quad \sum_{i=1}^D w_i^* = 1, \quad \sum_{i=1}^D w_i^* p(Y_i^{T+1,T+H}) = \bar{g} \quad (8)$$

The first and second terms reflect the fact that weights are probabilities and so should be non-negative and sum to one. The third term represents the new moment conditions.

The solution to the above minimization problem using the method of Lagrange is

$$w_i^* = \frac{w_i \exp(\gamma' p(Y_i^{T+1, T+H}))}{\sum_{i=1}^D w_i \exp(\gamma' p(Y_i^{T+1, T+H}))} \quad (9)$$

where γ is the vector of Lagrange multipliers associated with the constraints. According to this, the original weights w have been "exponentially" tilted to produce the new weights w^* .

The vector of Lagrange multipliers (i.e. tilting parameters) can be obtained as a solution to the following minimization problem,

$$\gamma = \arg \min_{\gamma} \sum_{i=1}^D w_i \exp(\gamma' [p(Y_i^{T+1, T+H}) - \bar{g}]) \quad (10)$$

Then using the newly computed weights, the updated expectation of other functions of interest can be computed simply as

$$\sum_{i=1}^D w_i^* h(Y_i^{T+1, T+H}) \quad (11)$$

If the interest is in the modified probabilistic density $g(Y^{T+1, T+H})$, which will be the case in our density forecast evaluation exercises, then as discussed in Cogley, Morogov, and Sargent (2005) importance sampling techniques could be used to redraw $Y_i^{T+1, T+H}$ from the original density $p(Y^{T+1, T+H})$ using the newly found weights, w^* , which can be achieved using the multinomial resampling algorithm of Gordon, Salmond, and Smith (1993). The steps of the algorithm (taken from Cogley, Morogov, and Sargent, 2005) are detailed in the online appendix A13.

In our forecasting exercises, the current quarter (median) forecast from the SPF is used as the mean condition on the one-step ahead VAR predictive density. Similarly, the long-horizon (median) forecast from the SPF acts as the mean condition on the VAR predictive density at the horizon determined for combining the VAR with survey long-run (detailed in the next section).¹³ In addition, following KCR we also restrict the variance of the one-step ahead VAR predictive density (i.e. uncertainty around the current/nowcast quarter) with the variance condition computed as the variance of the SPF forecast errors over a fixed-length rolling window preceding the forecast origin. A variance condition constructed via this approach is defined as an ex-post forecast uncertainty measure (see Clements, 2014; KCR).

¹³SPF also reports forecasts for four subsequent quarters beyond nowcast quarter (i.e. h=2Q to h=5Q). One could use these additional survey forecasts as the moment conditions for the respective VAR predictive densities to obtain more accurate hybrid forecasts for the remaining forecast horizons of interest to policy makers (h=6Q to h=12Q). We explore the usefulness of these additional conditions in the online appendix A6 ("Are there benefits to utilizing survey information for additional horizons?")

Specifically, if we treat $\hat{Y}_{t,h}^{SPF}$ as the SPF forecast for indicator Y_t , then the variance condition is formed as follows,

$$\sum_{q=0}^{15} (Y_{t-Delay-q} - \hat{Y}_{t-Delay-q,h}^{SPF})^2 \quad (12)$$

where $h=1Q$, q reflects the number of past forecasts used to compute the variance of errors, and Delay indicates the number of quarters it takes the forecaster to learn about the actual realization. To remain consistent with our measure of ‘truth’ defined earlier (see Data section), Delay is set equal to 2, for macroeconomic variables (real GDP growth, CPI inflation, and the unemployment rate). For financial variable, the federal funds rate, the Delay is set equal to 1, reflecting the fact that actual quarterly value is available immediately preceding the last day of the quarter (i.e. first day of the next quarter). For example, at forecast origin, 1997.Q1, the variance conditions for macroeconomic variables are based on the variance of the SPF nowcast errors computed over the preceding period 1992.Q4 through 1996.Q3. Similarly, for the federal funds rate, it is variance of the errors over the period 1993.Q1 through 1996.Q4.

In a VAR, conditioning or tilting on some future horizon will influence the forecast starting from the jumping-off point all the way to the conditioned forecast horizon. For example, if we tilt real GDP growth at forecast horizon $h=6Q$, then tilting it will likely impact the forecast trajectory from $h=1Q$ to $h=5Q$ for all the variables.¹⁴ Simultaneously conditioning on multiple variables (in a system such as VAR) would result in forecast trajectories that reflect the cumulative effect of those conditions.

We use relative entropy as opposed to other approaches to conditional forecasting because of its ease of use, computational simplicity, and flexibility. Relative entropy is an effective and flexible conditional forecasting methodology, because it allows us to combine both the mean condition and the modeler’s confidence in that mean condition (i.e., variance) effectively. This is an important advantage if the interest is in density forecasts in addition to point forecasts. Furthermore, specifying the mean condition only would not result in the automatic shrinkage of the variance around the mean condition to zero; relative entropy will assume that the variance around that mean condition is the same as the unconditional.¹⁵ In addition, as discussed in Giacomini and Ragusa (2014), relative entropy does not require Gaussian assumptions for the original densities or the modified tilted densities. This latter advantage makes possible the application of our proposal on nonlinear VARs and on density forecasts generated from a combination of several component density forecasts because both these approaches generate non-normal

¹⁴In the online appendix A12 we provide an intuition behind this spillover feature using an analytical Gaussian example.

¹⁵Alternative approaches to constructing conditional forecasts include Waggoner and Zha’s (1999) soft conditioning which is an extension of Doan et al. (1984); Andersson et al. (2010); and the Kalman filter approach as in Banbura et al. (2015). All these three approaches can also allow for both mean and variance conditions. Antolin-Diaz et al (2019) formally prove the equivalence between VAR conditional forecasting and relative entropy for the Gaussian case. If we instead use the approach of Doan, Litterman and Sims (1984) to impose our conditions, we get very similar results for the point forecast evaluation.

densities.¹⁶

4.2 Determining the Forecast Horizon for Tilting

In combining the survey long-run projections with the VAR forecast, the initial inclination would be to combine the survey projections with the VAR forecast at some very distant horizon. This assumption is valid, because the terminology “long-run” projection by definition suggests many years into the future. That said, several macroeconomic variables (transformed to growth rates) display little persistence and therefore tend to move back rapidly to their respective (unconditional) mean—the unrestricted long-run model forecast. Real GDP growth fits into this category. The forecasts of real GDP growth from VAR models typically tend to move back toward the estimated mean within a year.¹⁷ On the other extreme are series such as the unemployment rate, which are very persistent. Depending on the starting point, it may take them several years to move back toward the model-implied long run.

Based on the work of Clements and Hendry (1999) and Kozicki and Tinsley (2001a,b) we also know that in covariance-stationary VAR models, forecasts beyond 4 quarters are heavily influenced by the model’s implied equilibrium value (i.e. unconditional mean). Clements and Hendry (1999) illustrate that a poorly estimated mean of the variable is the dominant source of the forecast errors beyond four quarters (e.g. higher estimate of trend than what is thought to be reasonable will result in forecasts that are persistently biased upwards).¹⁸ This suggests, to influence the trend estimate implied from the model with the one informed from survey we need to begin doing that as soon as the model’s implied trend is expected to dominate the forecast values. For real GDP growth, it suggests targeting the horizon somewhere at the forecast horizon $h=4Q$ or $h=5Q$. If imposed late in the forecast horizon, the trend is influenced too far out to have a meaningful influence on the forecast horizons of interest. Hence the forecast remains biased or corrupted from the influence of the model’s implied steady-state.¹⁹

Accordingly, we propose that the horizon for tilting should be variable specific, and we suggest the following²⁰

At each forecast origin t , retrieve the persistence estimates (i.e., slope parameters), corresponding to variable i from equation i of the VAR system in (1).

$$\rho_{i,t}^{+,BVAR} = \sum_{l=1}^p \bar{A}_{i,l}^{(i,i)} \quad (13)$$

¹⁶Both these avenues are left for future research.

¹⁷See Domit et al (2019) for UK GDP growth; KCR for US GDP growth.

¹⁸See technical appendix 7.2 for an illustration of the implied steady-state on the forecast trajectory.

¹⁹We illustrate this with two empirical exercises reported in the online appendix A15. In the first exercise, the horizon for combination is set dogmatically at $h=25Q$ (i.e 7 years out), and in the second exercise, it is set at $h=40Q$ (10 years out). The results indicate reduced gains in forecast accuracy for the horizons of interest. The reduced effects are most notable in the specifications that include SV.

²⁰The suggestion is roughly equivalent to generating unconditional forecasts far into the future (e.g. 40 quarters out) from the VAR model and then determining for each variable the precise forecast horizon at which it converges to its equilibrium value (i.e. implied steady state).

where $\bar{A}_{i,t}^{(i,i)}$ represents the posterior mean estimate of the slope coefficient of variable i in equation i of the VAR system in (1). It reflects an estimate of variable i 's persistence conditional on the VAR system.²¹

The corresponding metric that roughly determines the number of quarters it takes to revert back to the VAR's implied steady state is

$$h_{i,t}^{+,VAR} = \frac{1}{1 - \rho_{i,t}^{+,VAR}} \quad (14)$$

The horizon, $h_{i,t}^*$ at which the survey long-run forecast is combined with the VAR forecast for variable i is set as

$$h_{i,t}^* = \max \{P_t^Q, h_{i,t}^{+,VAR}\} \quad (15)$$

where $P_t^Q - 1$ specifies the minimum number of quarters prior to which the long-horizon survey forecast takes over the VAR forecast. The max operator insures that at least for the $P_t^Q - 2$ number of quarters following the nowcast quarter the hybrid forecast uses the VAR forecast. In our exercises we set $P_t^Q = 5$ to reflect our preference to have a dynamic and informative forecast in the short to medium term. We note that the choice of P_t^Q does not influence our results; setting $P_t^Q = 0$ gives us very similar forecasting results because this choice binds only on real GDP growth (for just one or two quarters) but not other variables.²²

In our empirical forecasting exercises, over the forecast evaluation sample, the horizon at which survey long-run takes over has ranged between h=5Q and h=17Q for CPI inflation; ranged between h=9Q and h=30Q for the federal funds rate; ranged between h=10Q and h=27Q for the unemployment rate; and for Real GDP growth the horizon for combination has remained steady at h=5Q. In the interest of brevity, the figures plotting the evolution of tilting horizon by variable and for each VAR model are relegated to the online appendix A14.²³

²¹Our results are robust if the forecast horizon for tilting uses an estimate (of persistence) obtained by recursively (i.e. at each forecast origin) estimating a univariate regression AR(4) for the variable of interest. The results are reported in tables A19 and A20 in the online appendix A15.

²²Over our forecast evaluation sample, for most forecast origins, on average it takes roughly three quarters for real GDP growth to move back to trend growth. By setting h=5Q, we delay by two quarters the takeover by the survey long-run forecast over the model forecast, and so permit the possibility to impose conditions on additional VAR forecast quarters (i.e. h=2Q through h=4Q) for which survey expectations are available.

²³With the exception of TVP-VAR SV and TVP-VAR, for all the remaining VAR models, the forecast horizon for combination is determined based on the specification with constant variance, and the same value of the combination horizon is used for the counterpart specification that allows for SV. This facilitates more direct comparison between specifications in assessing the role of SV in forecast accuracy. However, our results are qualitatively similar if we relax this restriction.

5 Results

5.1 Forecasting Exercise

Our main question of interest pertains to whether we achieve meaningful improvements in the forecast accuracy of the variables of interest (to monetary policymakers) by tilting the model-based forecasts to match the modeler’s long-run value. In our examination, the modeler’s long-run value equates to the (median of the) long-horizon SPF projections.

To answer this we perform a real-time out-of-sample forecasting evaluation over the period 1994.Q1 to 2016.Q4. We begin by estimating our VAR models using real-time data beginning in 1959.Q4 through 1993.Q4 and iteratively generate unconditional forecast one to forty quarters out. Then we re-estimate the models using an additional data point and again generate forecasts up to 40 quarters out. We repeat this recursive exercise until 2016.Q3. That is, the last estimation sample uses data from 1959.Q4 to 2016.Q3, and the forecasts span the period 2016.Q4 to 2026.Q3; with data for evaluation available through 2016.Q4, we would be able to evaluate the one-step-ahead forecast only. The forecasts generated through this recursive exercise are denoted ‘raw’ VAR forecasts. Next, using the technique of relative entropy we tilt the one-step-ahead ‘raw’ VAR forecasts (i.e. predictive densities) generated in the previous step to match the SPF nowcasts for real GDP growth, CPI inflation (and core CPI inflation²⁴ in the medium VAR), the unemployment rate, and the federal funds rate. We denote the resulting tilted forecasts (corresponding to all variables) as the ‘baseline’ forecasts. The ‘baseline’ forecasts tilt on the nowcasts only (i.e. both the nowcast mean and variance). Next, we generate another set of forecasts, but this time tilting the ‘raw’ VAR forecasts toward both the nowcasts and the survey long-run projections for the same set of variables (as in the ‘baseline’ forecasts). We denote these forecasts as ‘hybrid’ forecasts. Finally, for each VAR model under consideration, we evaluate and compare both the point forecast accuracy and density forecast accuracy among the raw VAR, baseline, and the hybrid forecasts in a pairwise fashion.

All the tables (unless specified) are formatted so to facilitate quick comparisons across raw, baseline, and hybrid forecasts for each model, and comparison of raw forecasts across models. The comparison of raw forecasts between small and medium VAR models help assess the usefulness of additional variables in improving the accuracy of the core set of variables. In addition, formatting of the results helps assess whether tilting helps more models that are inferior to begin with.²⁵ Tables 1 through 6 (and tables A1 through A4 in the online appendix A1) report the real-time forecast accuracy of real GDP growth, CPI inflation, the unemployment rate, the federal funds rate, and credit spreads (the tables A1 through A4 also report accuracy for the additional five variables), and allow a quick assessment on the usefulness of stochastic volatility on forecast accuracy. Each table is split into two panels. The left panel reports accuracy results of the VAR model

²⁴For core CPI inflation, the SPF nowcasts are not available until 2007.Q1. The inflation nowcasting models of Knotek and Zaman (2017) or Modugno (2013) could be used to produce core CPI inflation nowcasts prior to 2007.Q1. For the sake of consistency and simplicity, we use core CPI nowcasts from the SPF starting in 2007.Q1.

²⁵To conveniently assess the effectiveness of tilting please refer to online appendix A4.

with constant variance, and the right panel reports results for VAR model with stochastic volatility. For each variable, the first row reports the MSE (CRPS for density forecasts) for the raw VAR forecast. The subsequent three rows report the relative MSE (relative CRPS for density forecasts): MSE of the baseline forecast relative to raw forecast, MSE of the hybrid forecast relative to raw forecast, and MSE of the hybrid forecast relative to the baseline forecast. The accuracy results are reported for the forecast horizons: one quarter out (i.e., nowcast quarter), four, eight, and 12 quarters out respectively. The forecast evaluation is based on the full sample spanning 1994.Q1 to 2016.Q4.

5.1.1 Results: Small VAR estimated with longer sample

Table 1 reports the real-time point forecast accuracy from the Small VAR estimated with data going back to 1959.Q4. A couple of things immediately stand out. First, allowing for SV helps improve point forecast accuracy, however, the magnitude of improvements and persistence in those gains vary across variables. The gains are strongest for CPI inflation and persist throughout the forecast horizon; for real GDP growth improvements are achieved at least through four quarters out. The improvements for unemployment rate are marginal and persist throughout. These findings are in line with Clark (2011) and D’Agostino (2013). For the federal funds rate and credit spread the improvements are short-lived as by four quarters out and beyond allowing for SV appears to worsen the forecast accuracy. Secondly, across all variables, tilting towards the survey nowcasts significantly improves the forecast accuracy in the nowcast quarter (i.e. $h=1Q$). The relative MSE is substantially below one for rows labelled ‘Baseline/Raw’ and ‘Hybrid/Raw’ respectively. These large improvements suggest that the SPF nowcasts are significantly more accurate than the VAR model’s one-step-ahead forecast, which is consistent with the results in KCR and several other studies documenting the usefulness of external nowcasts for the models estimated purely with quarterly data (e.g. Knotek and Zaman, 2019). Looking across all the rows labeled ‘Baseline/Raw’, the spillover effects (on subsequent forecast horizons) from more accurate nowcasts are longer lasting for persistent variables, CPI inflation, the unemployment rate, and the federal funds rate. For real GDP growth, the gains are relatively short-lived.

In all rows labeled ‘Hybrid/Raw’, the ratios are less than one and are generally smaller than those reported in the row immediately above (‘Baseline/Raw’) with few exceptions. This suggests that tilting the VAR forecasts to match the survey long-horizon forecasts in addition to the survey nowcasts leads to further improvement in accuracy. A notable difference across the two rows (i.e. Hybrid/Raw vs. Baseline/Raw) is the significantly improved accuracy in forecast horizons further out.

To get a sense of the marginal gains in accuracy from tilting towards survey long-horizon above and beyond from tilting towards the survey nowcasts only, the rows labeled ‘Hybrid/Baseline’ facilitates that assessment. For example, eight quarters out, the hybrid forecast for real GDP growth (in both VAR specifications) is on average 15 percent more accurate than the baseline forecast as indicated by a ratio of 0.85. Digging deeper into error evaluation, the improvements in the average accuracy of the real GDP growth forecast is coming mainly from the Great Recession and the subsequent recovery. This

is evident in Figure 2, which plots the time line of the cumulative sum of squared forecast errors for horizon $h=8Q$, beginning around the height of the Great Recession and through the end of evaluation sample, the baseline forecast consistently under performs hybrid forecast. This is evident by the plot corresponding to hybrid forecast lying below the baseline forecast beginning Great Recession, and with divergence between the two increasing over the remaining evaluation sample. The pattern of improved accuracy for hybrid forecast around the crisis period and beyond fits with the formal statistical assessment of a structural change over that period (see Aastveit et al, 2017). This results suggests that the hybrid approach is adjusting the forecasts to accommodate for the possible structural change. This is further confirmed by inspecting recursive forecast trajectories (not shown). The baseline forecast calls for stronger long-run projections, and as a result, the recursive baseline forecast trajectories continuously over-predict growth. However, the recursive trajectories from the hybrid forecast track the actual data relatively better because they rely on professional forecasters' assessment of a lower growth potential of the economy, perhaps drawn from demographics or specific assessments of technological change.

For CPI inflation, the forecast accuracy gains are substantially higher, statistically significant, and persist throughout; the hybrid forecast for CPI inflation is roughly 25 percent (11 percent for specification with SV) more accurate two years out, and 30 percent (17 percent for specification with SV) more accurate three years out compared to baseline forecast. This is not surprising because it is well known that the inflation process has exhibited pronounced changes in the underlying trend since the 1950s, and so accounting for those changes turns out to be very important in achieving improved accuracy. Just like in the case of real GDP growth, Figure 2 shows that since the post-crisis period hybrid forecasts are substantially more accurate compared to baseline forecasts. Unlike in the case of real GDP growth, the superior accuracy of the hybrid CPI inflation forecasts is evident over the entire forecast evaluation sample preceding the Great Recession. This improved accuracy prior to Great Recession has been possible because tilting is helping reduce the bias in the forecasts introduced by failing to account for structural break of the mid-80s (i.e. estimating with longer history). The sizable divergence between the hybrid and baseline forecasts since the post-crisis period partly reflects the fact that tilting is helping hybrid forecasts mitigate misspecification issues from failure to account structural breaks of both mid-80s and post-crisis period.

The improvements in accuracy for the effective federal funds rate are of similar magnitude to those for CPI inflation. The combination of improved inflation forecasts and tilting on judgmental survey-based long-run values of the federal funds rate results in highly accurate forecasts of the federal funds rate. The improvements in the forecast accuracy by the end of the third year are 30 percent (40 percent in VAR specification with SV) on average over the baseline forecast. Figure 2 confirms the conjecture that tilting is helping hybrid forecasts better handle structural change.

Credit spread is a variable that is not directly tilted but its accuracy is affected indirectly through the spillover effects of variables that are directly tilted. Not surprisingly, the combination of improved accuracy of federal funds rate, inflation, and real GDP growth is associated with improved accuracy of credit spread. The magnitude of the gains is similar to that seen for the federal funds rate.

In the case of the unemployment rate, the gains from tilting towards survey long-horizon forecasts are small on average. As evident in Figure 2, the marginal gains in accuracy reflect the slightly more accurate hybrid forecasts over the Great Recession and post-recovery period.

Table 2 reports the corresponding accuracy of the density forecasts. A lower value for CRPS is preferred, as such negative entry for row labeled ‘Baseline - Raw’ suggests that the baseline density forecast is more accurate on average compared to the raw density forecast. Negative entry for ‘Hybrid - Raw’ suggests that hybrid forecast is more accurate compared to raw forecast. Similarly, negative entry in the case of ‘Hybrid - Baseline’ suggests that hybrid density forecast is more accurate compared to baseline forecast on average. The results of the density forecast evaluation echo the results of the point forecast evaluation reported in Table 1. First, adding SV helps improve density forecasts with pattern of improvements quite similar to point forecast assessment. Second, baseline forecasts are more accurate compared to the raw VAR forecasts, and the hybrid forecasts are more accurate compared to both the baseline and the raw VAR forecasts. Third, the results illustrate that gains in the density forecast accuracy of the hybrid forecasts for CPI inflation, the federal funds rate, and credit spreads are substantial, statistically significant, and persist far into the future. Table A5 in the online appendix reports the density calibration diagnostics. For the VAR with constant volatility, with the exception of nowcast quarter (and subsequent quarter) for both hybrid and baseline forecasts, we find that density forecasts are badly calibrated (consistent with findings in Sekhposyan and Rossi, 2014). SV helps improve calibration of the density forecasts via improved coverage (i.e. predictions intervals close to the nominal coverage of 70%) consistent with findings in Clark (2011) and D’Agostino et al. (2013). However, it generally remains the case (for forecast horizons beyond the short-term) that density forecasts with SV are unable to pass all the necessary statistical tests (of PITS) so to be categorized as correctly calibrated.

5.1.2 Results: Small VAR estimated with post-1985 sample

Table 3 reports the real-time point forecast accuracy from the Small VAR estimated with data going back to 1985. The results indicate similar pattern of accuracy improvements as those seen in the case of VAR estimated with longer sample with the magnitude of gains smaller but statistically significant in many instances. SV helps improve point forecast accuracy but the gains are substantially smaller compared to what was seen in the case of VAR estimated with longer sample. Just like in the previous case for VAR with longer sample, hybrid approach helps improve accuracy in both economically meaningful and statistically significant way for real GDP growth, CPI inflation, the federal funds rate, and credit spread. One exception is the unemployment rate, in that hybrid approach appears to worsen the forecast accuracy but those deteriorations are not flagged as statistically significant. In that regard, results for unemployment rate across two VARs (longer sample and shorter sample) are similar as accuracy gains in VAR with longer sample were not statistically significant.

The plots in Figure 3 indicate that most of the accuracy improvements (or losses in the case of unemployment rate) for the hybrid approach (compared to baseline) are com-

ing from the evaluation period beginning 2010; i.e. coinciding with the sample period thought to have undergone structural change. This is exactly where we would expect to see gains for the hybrid forecasts from VARs estimated with data beginning after the well documented structural breaks of the mid-80s. It helps support our conjecture of hybrid helping accommodate structural breaks. Recall, in beginning 2012, Federal Reserve adopted inflation targeting framework, and it is since then that CPI inflation forecasts derived from the hybrid approach outperforms the baseline, though the gains are marginal but statistically significant. We also point out that these two VAR specifications outperform VAR specifications estimated with longer sample (see Table A12 in the online appendix A4 for the formal rankings based on forecast performance).

Table 4 reports the corresponding accuracy of the density forecasts. The results of the density forecast evaluation echo the results of the point forecast evaluation reported in Table 3. One exception is the unemployment rate, the worsening in density forecast accuracy of the hybrid forecasts are marginally significant whereas in the case of point forecast accuracy the deteriorations in accuracy of hybrid forecasts were generally not statistically significant. Table A6 in the online appendix A2 reports the density calibration diagnostics. Interestingly, the raw forecasts for real GDP growth from the VAR with constant variance are correctly calibrated. SV significantly helps improve the calibration of the density forecasts for CPI inflation; for unemployment rate and the federal funds rate, SV improves the coverage rate but density forecasts generally fail the statistical tests to be categorized as correctly calibrated. SV helps widen the prediction intervals (width of the 70% bands), consistent with the discussion in Clark (2011). Interestingly, we find generally that if the underlying (raw or baseline) forecast to be tilted is better calibrated, only then hybrid forecast helps improves calibration via improved coverage (i.e. prediction intervals close to the nominal coverage of 70%). There are exceptions, though. CPI inflation, for all three forecasts (Raw, Baseline, and Hybrid) from the VAR specification with SV appear to be well calibrated (based on p-values from the Knüppel test being greater than 0.10). A closer inspection reveals that of these three forecasts, hybrid forecast could be characterized slightly better calibrated as the coverage rates are generally closer to nominal rate of 70% compared to other two.

For the VAR specification with constant variance, the evidence of hybrid approach helping improve calibration is more prominent. As can be seen, the raw CPI inflation forecast is poorly calibrated because the null hypothesis of well calibrated density is rejected at the 10% significance level for all horizons (p-values from the Knüppel test are below 0.10). This is further supported by the corresponding empirical coverage rates that are well below the nominal value of 70%. Beyond the short-term, the baseline forecast also fails the tests of properly calibrated densities. In contrast, the hybrid forecast appears well calibrated for all the forecast horizons shown as the p-values (from the Knüppel test) are all above 0.10 and the empirical coverage rates are much closer to the nominal value of 70%.

5.1.3 Results: Time varying parameters VAR

Table 5 reports the real-time point forecast accuracy from the small-scale time-varying VAR estimated with data going back to 1959:Q4. The specification with SV helps im-

prove point forecast accuracy for both real GDP growth, and CPI inflation. For CPI inflation, results point to economically meaningful and statistical significant accuracy improvements similar to the improvements for time-invariant VARs (with and without SV). Specifically, for time-varying VAR with constant variance the magnitude of accuracy gains for hybrid forecasts is similar to that seen earlier for small VAR estimated with longer sample (e.g. three year out hybrid is 30% more accurate than the baseline). For the time-varying VAR specification with SV, the gains are relatively smaller and are similar to that seen for small VAR estimated with post-1985 data (e.g. three years out hybrid forecast is on average 18% more accurate compared to the baseline). We find that time-varying VAR with SV is among the most accurate of the models considered in this paper (see Table A12 online appendix A4). In the case of real GDP growth, and the unemployment rate, time-varying VAR with constant variance is a close competitor. However, for CPI inflation, it is the combination of time-varying parameters and stochastic volatility that helps position it as a good forecasting model. We suspect that the superior accuracy of SV specification for CPI inflation compared to constant variance specification is in part due to much tighter priors required for SV specification. Our results echo findings in D’Agostino et al (2013).²⁶ For real GDP growth, results indicate smaller benefit from hybrid approach, even though they appear economically meaningful they are not statistically significant. As in the case of VAR models discussed earlier, Figure 4 illustrates that most of the gains for hybrid approach are coming from the post-crisis period. In the case of unemployment rate, results are very similar to that of small VAR estimated with post-1985 sample. On average, compared to baseline, the hybrid approach leads to inferior forecasts for unemployment rate but the losses are statistically not significant. These losses are primarily coming from the post-crisis period.

Table 6 reports the corresponding accuracy of the density forecasts. The results of the density forecast evaluation echo the results of the point forecast evaluation reported in Table 5. Table A7 in the online appendix reports the density calibration diagnostics. Beyond the nowcast quarter, the density forecasts from the specification with constant variance are badly calibrated. Allowing for SV helps improve the calibration of density forecasts for real GDP growth and CPI inflation. For real GDP growth, density forecasts from SV specification are correctly calibrated. For CPI inflation, SV specification leads to improvements in the coverage rates but generally fail the statistical tests to be classified as correctly calibrated.

5.1.4 Results: Medium VAR

In the interest of brevity, results for medium VAR are relegated to an online appendix A1. Here we briefly summarize our findings. Firstly, patterns of both point and density forecast accuracy improvements for hybrid forecasts are generally similar to that of small VAR. Specifically, results for medium VAR estimated with longer sample echo the results reported for small VAR with longer sample. One difference is that the magnitude of improvements for hybrid forecasts is slightly smaller than those reported for small VAR.

²⁶They do not report results for time-varying VAR without SV but one can make a rough assessment by comparing TVP-AR model with time-varying VAR with SV (denoted TVP-VAR in their paper) for inflation. TVP-AR perform significantly worse compared to time-varying VAR with SV.

Secondly, the additional variables in medium VAR are helping improve the accuracy of the core variables of interest (compared to small VAR) therefore with more accurate raw forecasts tilting is helping slightly less. The finding that medium VAR generates more accurate forecasts compared to small VAR are in line with Banbura et al. (2010) and Koop (2013). Results for medium VAR estimated with shorter sample echo the results reported for small VAR with shorter sample. Unlike in the case of estimation with longer sample favoring medium-sized VAR over small-sized, we do not find this pattern for VAR models estimated with shorter sample.

Thirdly, the most useful aspect of the results for medium VAR is the strong positive spillover effects on the accuracy of the variables that are not directly tilted. Impressive and statistical significant gains in the accuracy of forecasts derived from hybrid approach are achieved for core CPI inflation, wage compensation, nonfarm payroll employment, and credit spread. It is also worth highlighting that adding core CPI inflation and SV in the medium VAR greatly helps improve the accuracy of the raw CPI forecasts. This implies tilting is less effective for CPI inflation for those specifications, but for core CPI inflation it is very effective in all medium VAR specifications.

5.1.5 Results: Hybrid vs. Federal Reserve’s Greenbook

The results shown so far largely support our conjecture that hybrid approach is helping mitigate misspecification issues by helping accommodate possible structural change (due to changing trends) to the extent it is detected in real-time by professional forecasters. As a final check in support of our conjecture we performed an additional exercise comparing the accuracy of our hybrid forecasts to that of Federal Reserve’s Greenbook (GB). GB forecast can be thought of as a combination of model and judgement; therefore it will be expected to handle structural change better than standard VARs. We confirm that indeed this is the case, and strikingly, the hybrid forecasts for real GDP growth and CPI inflation from our simple VARs are competitive to the GB forecasts. In the case of unemployment rate, hybrid forecast under performs GB during the Great Recession but on average is competitive to GB. For space constraints, results are presented in the online appendix A7.

Overall, our point and density forecasting results using real-time data provide compelling evidence that tilting VAR forecasts to match the long-run forecasts from the Survey of Professional Forecasters systematically leads to improved forecast accuracy for most variables over the forecast horizon of interest to monetary policymakers. Generally it is the case that our proposal is helping more models that are performing worse in raw form (i.e. raw VAR forecasts). Interestingly, rankings are generally maintained post-tilting, i.e. models ranked low prior to tilting continues to rank low post-tilting but the differences across accuracy are much smaller (see online appendix A4).²⁷

²⁷As a further check on robustness of this statement, we estimate Small VAR with loose priors and assess to what extent our proposal helps improve its accuracy. We find that this VAR specification ranks the lowest (as would be expected) prior to tilting, gains the most in terms of accuracy improvement from tilting, and remains ranked at the bottom post-tilting but with significantly reduced margin when compared to next best model. The results of this exercise are reported in online appendix A5.

5.2 Inflation Forecast Accuracy of Tilted VAR Compared to Univariate Benchmarks

Tilting the VAR forecasts to match the long-horizon survey forecasts leads to meaningful gains in forecast accuracy for most variables, and the gains in point forecast accuracy are substantial for nominal variables such as price inflation and wage inflation. Given these results, we investigate how the accuracy of inflation forecasts from the tilting approach compares to hard-to-beat univariate benchmark models. Accordingly, we next compare the inflation forecast accuracy from our medium VAR (with and without stochastic volatility) using the three most well-known univariate benchmarks: a random walk model (Atkeson and Ohanian, 2001), univariate unobserved component with stochastic volatility (UCSV) model of Stock and Watson (2007), and Faust and Wright (2013) inflation in gap.²⁸

Random walk model of Atkeson and Ohanian (2001). For our forecasting exercise, the forecasts of CPI inflation into the future is computed by averaging the previous four available quarterly annualized readings of the CPI . To construct a fair horse race, we set $\hat{\pi}_{t+1}$ equal to the *survey nowcast*

$$\hat{\pi}_{t+h} = 0.25(\hat{\pi}_{t+1} + \pi_t + \pi_{t-1} + \pi_{t-2}) \quad \text{for } h \geq 2 \quad (16)$$

Univariate unobserved component with stochastic volatility (UCSV) model of Stock and Watson (2007). The superior accuracy of this model in forecasting inflation is well documented in numerous studies. The model decomposes inflation into two components – a stochastic trend component and a transitory component – and assumes time-varying variances of the respective shocks to these two components. The specification of this model is as follows (for ease of exposition we retain the notation used in Stock and Watson (2007)):

$$\pi_t = \tau_t + \eta_t, \quad \text{where } \eta_t = \sigma_{\eta,t} \zeta_{\eta,t} \quad \zeta_{\eta,t} \text{ is i.i.d. } N(0, I_1) \quad (17)$$

$$\tau_t = \tau_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t = \sigma_{\varepsilon,t} \zeta_{\varepsilon,t} \quad \zeta_{\varepsilon,t} \text{ is i.i.d. } N(0, I_1) \quad (18)$$

$$\ln(\sigma_{\eta,t}^2) = \ln(\sigma_{\eta,t-1}^2) + \nu_{\eta,t}, \quad \text{where } \nu_{\eta,t} \text{ is i.i.d. } N(0, \gamma_1) \quad (19)$$

$$\ln(\sigma_{\varepsilon,t}^2) = \ln(\sigma_{\varepsilon,t-1}^2) + \nu_{\varepsilon,t}, \quad \text{where } \nu_{\varepsilon,t} \text{ is i.i.d. } N(0, \gamma_2) \quad (20)$$

The model forecast for inflation infinite quarters into the future is simply the model's current estimated trend inflation rate.²⁹

To construct a fair horserace, we set $\hat{\pi}_{t+1}$ equal to the *survey nowcast*. For forecasts $h \geq 2$, we estimate the model through $t+1$, treating the survey nowcast as data and

²⁸See Tallman and Zaman (2017) for a broader examination of the forecasting properties of other models as well as these two models.

²⁹The scalar parameters γ_1 and γ_2 determine the smoothness of the stochastic volatility process. Following Stock and Watson (2007) we fix both at 0.2.

computing the updated trend estimate.

$$\hat{\pi}_{t+h} = \hat{\tau}_{t:t+1} \quad \text{for } h \geq 2 \quad (21)$$

Inflation in Gap model of Faust and Wright (2013). It is patterned along the lines of Faust and Wright univariate AR1 inflation in gap model. According to this model, CPI inflation (quarterly annualized rate) is modeled as a gap, where gap is defined as a deviation from a trend which is taken to be the SPF long-horizon forecast for CPI inflation

$$\pi_t^{Gap} = \alpha_0 + \alpha_1 \pi_{t-1}^{Gap} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (22)$$

where $\pi_t^{Gap} = \pi_t - \pi_t^{SPF \text{ LR}}$

The estimated model is iteratively used to produce forecasts h periods ahead. The forecasts are of the gap (i.e. CPI inflation gap) and to it we add the latest available estimate of the SPF trend (as of forecast origin t) to get implied forecast of the CPI inflation.

Table 7 reports the results comparing out-of-sample CPI inflation forecasting performance (both point and density) across five models: Medium VAR est 1960, Medium VAR SV est 1960, random walk model, UCSV model, and Faust and Wright inflation in gaps model. Also, added is the forecasting performance of the SPF survey itself. Given the SPF cover only forecast horizons $h=1Q$ through $h=5Q$, entries in the table for $h=8Q$ and $h=12Q$ are blank. The top portion of the table reports the results corresponding to the full sample of evaluation, and the lower portion to the pre-crisis sample (i.e. 1994.Q1 to 2006.Q4). The left panel reports the accuracy comparison of Medium VAR (with constant variance) to the univariate forecasts and right panel reports the accuracy comparison of Medium VAR with SV to those same univariate forecasts.

As shown by the numbers reported in the table, the hybrid CPI point forecasts are competitive to univariate forecasts and in many cases more accurate, with several of those gains flagged as statistically significant. This result holds for each of the forecast evaluation samples.

For density forecast accuracy, the UCSV forecast is substantially superior to the baseline forecast obtained from Medium VAR with constant volatility (as evidenced by positive values) and the gains are statistically significant throughout. However, compared to the hybrid forecast, the UCSV forecast is competitive through most of the forecast horizons but slightly more accurate in the latter horizon (with the gain marginally significant). In the case of Medium VAR specification with SV, now both baseline and hybrid density forecasts are competitive to the UCSV. These results suggest that combining the long-horizon survey forecasts with the VAR forecasts improves the density forecasts notably. Adding SV helps further improve the density forecast accuracy of the hybrid forecasts.

The evidence that the multivariate model (VAR) estimated using a sample going back to 1960 tilted towards survey long-horizon forecasts generates inflation forecasts that rival both the point and density forecast accuracy of forecasts from tough benchmarks is

a worthwhile result because policymakers at central banks faced with inflation targeting may benefit from using models that allow for feedback between policy and the real economy and inflation. Such models have underperformed simple univariate approaches in terms of forecasting inflation and at times significantly so. The examination of our exercises suggests that the modeling technique we propose and illustrate may have some useful payoffs for policymakers.

6 Conclusions

In this paper, we propose a technique to adjust the medium-term and long-horizon forecasts from a VAR toward plausible values proposed by judgmental forecasters. We construct a hybrid forecast of survey nowcast, VAR forecast, and long-horizon survey projection. Specifically, applying flexible and powerful technique of relative entropy, we tilt the VAR forecast both in the near term with the survey nowcast and in the long run with the survey long-run projection. The horizon at which the long-run survey projection is combined with the VAR forecast is variable specific and determined by the variable’s estimated persistence at forecast origin.

We consider the efficacy of our proposal on a variety of VAR models estimated using Bayesian methods. We find that all models examined benefit through improved accuracy, with the largest gains in forecast accuracy seen in models that are estimated with longer history and smallest gains for models that attempt to accommodate structural changes. We find that tilting VAR forecasts to match the long-run forecasts from the Survey of Professional Forecasters systematically leads to improved forecast accuracy as indicated by lower MSEs and lower CRPS for most variables over the forecast horizon of interest (i.e., 1 to 3 years out) to monetary policymakers. The greatest gains in accuracy are achieved for variables believed to have undergone marked shifts in their permanent components (e.g., inflation and the federal funds rate). We also show substantial forecast accuracy improvements for a host of variables (such as compensation, payroll employment, credit spread) that are not directly tilted but that are affected through the spillover effects of the tilted variables.

We also show that hybrid inflation forecasts from simple VAR models rival those of relatively accurate univariate benchmark models. We view this as a useful practical contribution because among the many frustrations of monetary policymakers is the inability of multivariate models – which allow for feedback effects from policy to the real economy and inflation – to match the forecasting performance of the univariate forecasting models.

Finally, we show that all models considered display demonstrable improvements in forecast accuracy of hybrid forecasts for real GDP growth and CPI inflation over the past decade (i.e. Great Recession through present), coinciding with the time period associated with possible structural change. These results lead us to view our proposal as a low cost method to mitigate model instability issues arising from structural shifts caused by moving end points.

References

- [1] Aastveit, Knut Are, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino (2016) Have Standard VARs Remained Stable Since the Crisis? *Journal of Applied Econometrics* doi:10.1002/jae.2555
- [2] Altavilla, Carlo, Raffaella Giacomini, and Giuseppe Ragusa (2017) "Anchoring the Yield Curve Using Survey Expectations," *Journal of Applied Econometrics* 32(6): 1055-1068 doi:10.1002/jae.2588
- [3] Alessandri, Piergiorgio, and Haroon Mumtaz (2017) "Financial conditions and density forecasts for US output and inflation," *Review of Economic Dynamics* 24: 66-78
- [4] Andersson, Michael, Stefan Palmqvist, and Daniel Waggoner (2010) "Density-Conditional Forecasts in Dynamic Multivariate Models," Sveriges Riksbank Working paper 247
- [5] Antolin-Diaz, Juan, Ivan Petrella, and Juan F. Rubio-Ramirez (2019) "Structural Scenario Analysis with SVARs," manuscript
- [6] Atkeson, Andrew and Lee E. Ohanian (2001) Are Phillips Curves Useful for Forecasting Inflation? Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 211. Retrieved from <https://ideas.repec.org/a/fip/fedmqr/y2001iwinp2-11nv.25no.1.html>
- [7] Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010) Large Bayesian Vector Auto Regressions, *Journal of Applied Econometrics* 25: 71-92 doi:10.1002/jae.1137
- [8] Banbura, Marta, Domenico Giannone, and Michele Lenza (2015) Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-Sections, *International Journal of Forecasting*, 31 (3), pages 739-756 doi:10.1016/j.ijforecast.2014.08.013
- [9] Barnett, Alina, Haroon Mumtaz, and Konstantinos Theodoridis (2014) Forecasting UK GDP Growth and Inflation under Structural Change: A Comparison of Models with Time-Varying Parameters, *International Journal of Forecasting*, 30: 129-143 doi:10.1016/j.ijforecast.2013.06.002
- [10] Bauwens Luc, Michel Lubrano, and Jean Francois Richard (1999) Bayesian Inference in Dynamic Econometric Models, *Oxford University Press* doi:10.1093/acprof:oso/9780198773122.001.0001
- [11] Berkowitz, J. (2001) Testing density forecasts, with applications to risk management *Journal of Business and Economic Statistics*, 19(4): 465-474
- [12] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2015) Bayesian VARs: Specification Choices and Forecast Accuracy, *Journal of Applied Econometrics*, 30(1), pp 46-73 doi:10.1002/jae.2315

- [13] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016) Large Vector Autoregressions with Stochastic Volatility and Flexible Priors, Federal Reserve Bank of Cleveland Working Paper, no. 16-17. doi:/10.26509/frbc-wp-201617
- [14] Clark, Todd E. (2011) Real-Time Density Forecasts from Bayesian VARs with Stochastic Volatility, *Journal of Business Economics and Statistics*, 29: 327-341 doi:10.1198/jbes.2010.09248
- [15] Clark, Todd E., and Francesco Ravazzolo (2015) Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility, *Journal of Applied Econometrics*, 30: 551-575
- [16] Clark, Todd E., and Michael W. McCracken (2008) Forecasting with Small Macroeconomic VARs in the Presence of Instabilities, in *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, D. Rapach and M. Wohar, eds., Emerald Publishing, pp. 93-147 doi:10.1016/s1574-8715(07)00203-5
- [17] Clark, Todd E., and Michael W. McCracken (2010) Averaging Forecasts from VARs with Uncertain Instabilities, *Journal of Applied Econometrics*, 5-29 doi:10.1002/jae.1127
- [18] Clark, Todd E., and Michael W. McCracken (2013) Advances in Forecast Evaluation, in *Handbook of Economic Forecasting (vol. 2)*, eds. G. Elliott, and A. Timmermann, Amsterdam, the Netherlands: Elsevier, pp. 1107-1201 doi:10.1016/B978-0-444-62731-5.00020-8
- [19] Clark, Todd E., and Michael W. McCracken (2017) Tests of Predictive Ability for Vector Autoregressions Used For Conditional Forecasting, *Journal of Applied Econometrics*, 32: 533-553 doi:10.1002/jae.2529
- [20] Clements, Michael P., (2014) Forecast Uncertainty-Ex Ante and Ex Post: U.S. Inflation and Output Growth, *Journal of Business and Economic Statistics* 32: 206-216 doi:10.1080/07350015.2013.859618
- [21] Clements, Michael P., and David F. Hendry (1999) Forecasting Non-stationary Economic Time Series, *MIT Press: London*
- [22] Cogley, Timothy, and Thomas J. Sargent (2005) Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S., *Review of Economic Dynamics* 8: 262-302 doi:10.1016/j.red.2004.10.009
- [23] Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005) Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System, *Journal of Economic Dynamics and Control* 29: 1893-1925 doi:10.1016/j.jedc.2005.06.005
- [24] Croushore, Dean (2010) An Evaluation of Inflation Forecasts from Surveys Using Real-time Data, *BE Journal of Macroeconomics* 10(1):1-32 doi:10.2202/1935-1690.1677

- [25] D’Agostino, Antonello, Domenico Giannone, and Luca Gambetti (2013) Macroeconomic Forecasting and Structural Change, *Journal of Applied Econometrics*, 28 (1), pages 81-101 doi:10.1002/jae.1257
- [26] Del Negro, Marco, and Giorgio Primiceri (2013) Time Varying Structural Vector Autoregressions And Monetary Policy: A Corrigendum
- [27] Diebold, Francis X., and Roberto S. Mariano (1995) Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263 doi:10.2307/1392185
- [28] Doan Tom, Robert Litterman, and Chris Sims (1984) Forecasting and Conditional Projection Using Realistic Prior Distributions, *Econometric Reviews*, 3: 1-100 doi:10.1080/07474938408800053
- [29] Domit, Silvia, Francesca Monti, and Andrej Sokol (2019) Forecasting the UK economy with a medium-scale Bayesian VAR, *International Journal of Forecasting*, in press
- [30] Faust, Jon, and Jonathan Wright (2013) Inflation Forecasting, *In Handbook of Economic Forecasting, Elliot G. Timmermann A (eds)*, Vol. 2. North-Holland: Amsterdam; 2-56 doi:10.1016/b978-0-444-53683-9.00001-3
- [31] Giacomini, Raffaella and Giuseppe Ragusa (2014) "Theory-coherent forecasting," *Journal of Econometrics* 182: 145-155
- [32] Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2018) Priors For The Long Run, *Journal of the American Statistical Association* 147 doi:10.1080/01621459.2018.1483826
- [33] Gneiting, Tilmann, and Adrian E. Raftery (2007) Strictly Proper Scoring Rules, Predictions, and Estimations, *Journal of the American Statistical Association*, 102: 359-378. doi:10.1198/016214506000001437
- [34] Gordon, Neil J., D.J. Salmond, and A.F.M. Smith (1993) A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, *IEEE Proceedings-F140*, pp.107-113 doi:10.1049/ip-f-2.1993.0015
- [35] Harvey, David., Stephen Leybourne, and Paul Newbold (1997) Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting*, 13, 281-291 doi:10.1016/s0169-2070(96)00719-4
- [36] Knotek II, Edward S., and Saeed Zaman (2017) Nowcasting U.S. Headline and Core Inflation, *Journal of Money, Credit and Banking*, 49(5): 932-968 doi:10.1111/jmcb.12401
- [37] Knotek II, Edward S., and Saeed Zaman (2019) Financial Nowcasts and Their Usefulness in Macroeconomic Forecasting, *International Journal of Forecasting*, in press doi:/10.1016/j.ijforecast.2018.10.012

- [38] Knüppel, Malte. (2015) Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments, *Journal of Business Economics and Statistics*, 33(2): 270-281
- [39] Koop, Gary (2013) Forecasting with Medium and Large Bayesian VARs, *Journal of Applied Econometrics* 28, 177-203 doi:10.1002/jae.1270
- [40] Koop, Gary and Dimitris Korobilis (2010) Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics* 3 267-358
- [41] Koop, Gary and Dimitris Korobilis (2013) Large Time-Varying Parameter VARs, *Journal of Econometrics* 177: 185-198 doi:10.1016/j.jeconom.2013.04.007
- [42] Kozicki, Sharon and P.A. Tinsley (1998) Moving Endpoints and the Internal Consistency of Agents' Ex Ante Forecasts, *Computational Economics* 11: 21-40
- [43] Kozicki, Sharon and P.A. Tinsley (2001a) Shifting endpoints in the term structure of interest rates, *Journal of Monetary Economics* 47: 613-652
- [44] Kozicki, Sharon and P.A. Tinsley (2001b) Term structure views of monetary policy under alternative models of agent expectations, *Journal of Economic Dynamics and Control* 25: 149-184
- [45] Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo (2017) Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts, *Journal of Business and Economic Statistics* doi:10.1080/07350015.2015.1087856
- [46] Litterman, R.B., (1986) Forecasting With Bayesian Vector Autoregressions – Five Years of Experience, *Journal of Business and Economic Statistics* 4, 25-38. doi:10.1080/07350015.1986.10509491
- [47] Modugno, Michele (2013) Now-casting Inflation Using High Frequency Data, *International Journal of Forecasting*, 29(4): 664-675 doi:10.1016/j.ijforecast.2012.12.003
- [48] Primiceri, Giorgio E. (2005) Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, 821-852 doi:10.1111/j.1467-937x.2005.00353.x
- [49] Robertson, John C., and Ellis W. Tallman (1999) Vector Autoregressions: Forecasting and Reality, Federal Reserve Bank of Atlanta *Economic Review*, Vol. 84:1 Retrieved from <https://ideas.repec.org/a/fip/fedaer/y1999iq1p4-18nv.84no.1.html>
- [50] Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2005) Forecasting Using Relative Entropy, *Journal of Money, Credit and Banking*, 37(3): 383-401 doi:10.1353/jmcb.2005.0034
- [51] Rossi, Barbara, and Tatevik Sekhposyan (2014) Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set, *International Journal of Forecasting*, 30: 662-682

- [52] Sims, Christopher A., and Tao Zha (1998) Bayesian Methods for Dynamic Multivariate Models, *International Economic Review*, 39: 949-968 doi:10.2307/2527347
- [53] Stock, James H. and Mark W. Watson (2007) Why Has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking*, 39: 3-33 doi:10.1111/j.1538-4616.2007.00014.x
- [54] Tallman, Ellis W., and Saeed Zaman (2017) Forecasting Inflation: Phillips Curve Effects on Services Price Measures, *International Journal of Forecasting*, 33(2): 442-457 doi:10.1016/j.ijforecast.2016.10.004
- [55] Tallman, Ellis W., and Saeed Zaman (2018) Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy, Federal Reserve Bank of Cleveland working paper 18-09
- [56] Tulip, P (2009) Has the economy become more predictable? Changes in Greenbook forecast accuracy, *Journal of Money, Credit, and Banking*, 41(6): 1217-1331
- [57] Villani, Mattias (2009) Steady-State Priors for Vector Autoregressions, *Journal of Applied Econometrics*, 24: 630-650 doi:10.1002/jae.1065
- [58] Waggoner, Daniel F., and Tao Zha (1999) Conditional Forecasts in Dynamic Multivariate Models, *Review of Economics and Statistics*, 81: 639-651 doi:10.1162/003465399558508
- [59] West, Kenneth D. (1996) Asymptotic Inference About Predictive Ability, *Econometrica*, 64: 1067-1084 doi:10.2307/2171956
- [60] Wright, Jonathan (2013) Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior, *Journal of Applied Econometrics*, 28: 762-776 doi:10.1002/jae.2268
- [61] Zaman, Saeed (2013) Improving Inflation Forecasts in the Medium to Long Term, Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2013-16 Retrieved from <https://ideas.repec.org/a/fip/fedcec/y2013inov15n2013-16.html>
- [62] Zellner, Arnold (1971) An Introduction to Bayesian Inference in Econometrics, *Wiley: New York*

7 Technical Appendix

Specifically, we impose our prior beliefs about the coefficient estimates in $A_1, \dots, A_p$ and Σ using Normal inverse-Wishart (N-IW) conjugate priors.³⁰ The prior beliefs for the mean and variances of the coefficient matrices are as follows:

³⁰Natural conjugate priors such as N-IW have computational advantages and at the same time have competitive forecasting properties (see Koop, 2013)

$$E[A_l^{(i,j)}] = \begin{cases} \delta_i & \text{if } i = j, l = 1 \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

$$Var[A_l^{(i,j)}] = \lambda^2 \frac{1}{l^2} \frac{\sigma_i^2}{\sigma_j^2}, \quad l = 1, \dots, p \quad (24)$$

The scale factor $\frac{1}{l^2}$ helps impose the prior belief that recent lags play a more influential role compared to the distant ones by proportionally shrinking the variances on the more distant lags (centered on the prior mean of zero). The prior parameter σ_i equals the standard deviation of the residuals obtained from regressing the variable y_i on its own p lags and a constant over the sample period up to time t . We set δ_i equal to the sum of the autoregressive coefficients obtained from regressing the variable y_i on its own p lags and a constant over a pre-forecast evaluation sample.³¹ The hyperparameter λ governs the tightness of our prior beliefs. As $\lambda \rightarrow 0$ the prior dominates and so posterior equals prior, i.e., the data have no say. On the other hand as $\lambda \rightarrow \infty$, the prior's influence diminishes and so posterior estimates converge to OLS estimates.

The VAR model in (1) can be rewritten in compact matrix notation as

$$Y = XA + E \quad (25)$$

where $Y = [y_1, \dots, y_T]'$ is a $T \times n$ matrix, $X = [x_1, \dots, x_T]'$ is a $T \times k$ matrix co, $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]$ is a $1 \times k$ vector. $A = [A_c \ A_1 \ \dots \ A_p]$ is a matrix of VAR coefficients of size $k \times n$, $E = [\varepsilon_1 \ \dots \ \varepsilon_T]$ is a $T \times n$ matrix of innovation terms.

The conjugate Normal inverse-Wishart (N-IW) prior is:

$$vec(A) | \Sigma \sim N(vec(A_0), \Sigma \otimes \Omega_0), \quad \Sigma \sim iW(S_0, v_0) \quad (26)$$

where A_0 , Ω_0 , S_0 , and v_0 are the prior parameters whose values are set based on the VAR model in (1) that satisfies the prior moment conditions specified in (2).

Since our prior N-IW is conjugate, the resulting conditional posterior distribution (i.e., the product of the prior and likelihood function) is also N-IW.

$$vec(A) | \Sigma, Y \sim N(vec(\bar{A}), \Sigma \otimes \bar{\Omega}), \quad \Sigma \sim iW(\bar{S}, \bar{v}). \quad (27)$$

where

$$\bar{A} = (\Omega_0^{-1} + X'X)^{-1} (\Omega_0^{-1}A_0 + X'Y) \quad (28)$$

$$\bar{\Omega} = (\Omega_0^{-1} + X'X)^{-1} \quad (29)$$

³¹Since all the variables entering the VAR are stationary, if we instead set $\delta_i = 0$ for all i , we get qualitatively similar results. Studies such as Clark (2011) and Carriero et al. (2015) have used a value of 0.8 for variables that are known to exhibit persistence (e.g., the unemployment rate, inflation and the interest rate). Our results are very similar if we instead set $\delta_i = 0.8$

$$\bar{v} = v_0 + T. \quad (30)$$

are the respective posterior mean estimates of the VAR model.

And,

$$\bar{S} = A_0 + \hat{\varepsilon}'\hat{\varepsilon} + \hat{A}'X'X\hat{A} + A_0'\Omega_0^{-1}A_0 - \hat{A}'\bar{\Omega}^{-1}\hat{A} \quad (31)$$

where $\hat{A} = (X'X)^{-1}X'Y$ equals the OLS estimate of A and $\hat{\varepsilon} = Y - X\hat{A}$ are the OLS residuals (Zellner, 1971). We use the mixed estimation method of Litterman (1986) to implement the N-IW prior, which equates to appending the data matrices with dummy observations.

Previous research (e.g., Robertson and Tallman, 1999; Banbura et al., 2010) has documented further gains in forecast accuracy by imposing a 'sum of coefficients' (SOC) prior on the equations of the VAR. The hyper-parameter μ will govern the tightness of this prior.

7.0.1 Optimal values of Hyper-parameters using marginal likelihood

The values of λ and μ are set by maximizing the marginal likelihood of the model over a predefined two dimensional discrete grid of λ and μ . The optimization is performed over the pre-forecast evaluation sample, and optimal values obtained are kept fixed over the forecast evaluation sample.³²

$$[\lambda^+, \mu^+] = \arg \max_{[\lambda^g, \mu^g]} \ln p(Y) \quad (32)$$

where

$$p(Y) = \int p(Y|\theta)p(\theta)d(\theta) \quad (33)$$

is the marginal likelihood, and θ is the set of all model coefficients. Given we are using an N-IW prior, the marginal density $p(Y)$ can be solved in closed form as³³

$$\begin{aligned} p(Y) &= \left(\frac{1}{\pi}\right)^{\frac{nT}{2}} \times |(I + X\Omega_0X')^{-1}|^{\frac{n}{2}} \times |S_0|^{\frac{v_0}{2}} \\ &\times \frac{\Gamma_n(\frac{v_0+T}{2})}{\Gamma_n(\frac{v_0}{2})} \times |S_0 + (Y - XA_0)'(I + X\Omega_0X')^{-1}(Y - XA_0)|^{-\frac{v_0+T}{2}}. \end{aligned}$$

To evaluate the log marginal likelihood, we use the two-dimensional grid of discrete values defined as follows: $\lambda^g = [0.050, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50]$; and $\mu^g = [0.1, 0.15, 0.2, 0.25, 0.50, 1, 1.5, 2, 2.5, 3]$; For the Small BVAR estimated with data starting in 1959, the optimal values obtained are $\lambda = 0.4$ and $\mu = 0.2$ and for the Medium VAR, optimal values are $\lambda = 0.3$ and $\mu = 0.25$. The values we obtain are close to values

³²That said, if we instead fix the $\lambda = 0.2$ and $\mu = 1$ (widely used values) the results are qualitatively similar.

³³Derivation details can be found in Bauwens et al. (1999), Carriero et al. (2015) and Giannone et al. (2015).

that other researchers have obtained through a grid search optimization.

Note that prior specification for each equation is symmetric in its treatment of own lags of the dependent variable and lags of other variables. As such, we have a prior that is a natural conjugate (Normal inverse-Wishart prior), which proves to be convenient when solving for the model, because these priors can be implemented easily by augmenting the data matrices with dummy variables, permitting OLS to estimate the model equation by equation (for details see Banbura et al., 2010 and Carriero et al., 2015).

7.0.2 Illustrating the influence of estimated steady-state on the forecasts

In this section we illustrate two important concepts: the influence of model's estimated mean (i.e. steady-state) on the forecasts, and the direct role of persistence in determining the forecast horizon at which the convergence to the steady-state takes place. To keep things simple, let's consider a covariance-stationary AR(1) model,

$$y_t = c + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (34)$$

Covariance-stationary implies $|\rho| < 1$, and

$$\mathbb{E}(y_t) = \mathbb{E}(y_{t-1}) = \mu \quad (35)$$

where μ denotes the steady-state of the model. In long-samples it matches the mean of estimation sample.

Next taking the expectations of both sides of equation 34 gives us,

$$\mathbb{E}(y_t) = c + \rho \mathbb{E}(y_{t-1}) + \mathbb{E}(\varepsilon_t) \quad (36)$$

Plugging eq. 35 into 36, yields

$$\mu = c + \rho \mu = \frac{c}{1 - \rho} \quad (37)$$

Rearranging the above equation,

$$c = \mu(1 - \rho) \quad (38)$$

Rewrite equation 34 by substituting 38 into 34,

$$\begin{aligned} y_t &= \mu(1 - \rho) + \rho y_{t-1} + \varepsilon_t \\ y_t - \mu &= \rho(y_{t-1} - \mu) + \varepsilon_t \end{aligned} \quad (39)$$

ρ is the persistence of y and it determines the speed of convergence back to the steady-state. $1 - \rho$ is the pace of adjustment towards steady-state each period. $\frac{1}{1-\rho}$ is the number

of periods it takes y to converge back to its steady-state value. Higher values of ρ imply longer lived departure of y from the steady-state.

Denoting $\hat{\rho}$ and $\hat{\mu}$ as estimates obtained via estimation of available data through time t ,

The one-step ahead forecast,

$$\hat{y}_{t|t+1} - \hat{\mu} = \hat{\rho}(\hat{y}_t - \hat{\mu}) \quad (40)$$

Similarly, h -step ahead forecast,

$$\begin{aligned} \hat{y}_{t|t+h} - \hat{\mu} &= \hat{\rho}^h(\hat{y}_t - \hat{\mu}) \\ \hat{y}_{t|t+h} &= \hat{\mu} + \hat{\rho}^h(\hat{y}_t - \hat{\mu}) \end{aligned} \quad (41)$$

A closer inspection of above equation reveals that the h -step ahead forecast is the sum of two components: the estimated mean; and the gap between the estimated mean and y (as of the forecast origin t) multiplied by the h^{th} power of $\hat{\rho}$

Note, $\hat{\rho}^h$ is a decreasing function of h ,

$$\hat{\rho} > \hat{\rho}^2 > \hat{\rho}^3 > \dots > \hat{\rho}^\infty$$

suggesting that as h increases the influence of the component $\hat{\mu}$ becomes more important.

In the limit, as $h \rightarrow \infty$, $\hat{\rho}^h \rightarrow 0$,

$$\text{and } \lim_{h \rightarrow \infty} \hat{y}_{t|t+h} \rightarrow \hat{\mu}$$

The smaller the value of $\hat{\rho}$ the faster $\hat{\rho}^h$ goes to zero and more rapid the convergence to model's implied steady-state.

Please refer to Clements and Hendry (1999) for further details.

Figure 1: Real-Time Long Run Forecasts

Notes for the figure: plots the real-time long-run forecasts from the Survey of Professional Forecasters (SPF), the 10-year-ahead average forecast. To facilitate comparison also plotted are the real-time estimates of the long-run forecasts implied from a fixed-parameter Small BVAR recursively estimated with data beginning in 1959:Q4, fixed-parameter Small BVAR estimated with data beginning in 1985, and time-varying parameter Small BVAR respectively. The short-term interest rate in the case of SPF refers to the 3-month T-bill rate, but in the case of BVARs, it refers to the federal funds rate.

Figure 2a: Cumulative Squared Errors, Small VAR est. 1960

Figure 2b: Cumulative CRPS, Small VAR est. 1960

Notes: Cumulative sum of squared errors (top); Cumulative sum of CRPS (bottom)

Figure 3a: Cumulative Squared Errors, Small VAR est. 1985

Figure 3b: Cumulative CRPS, Small VAR est. 1985

Notes: Cumulative sum of squared errors (top); Cumulative sum of CRPS (bottom)

Figure 4a: Cumulative Squared Errors, Small TVP VAR SV

Figure 4b: Cumulative CRPS, Small TVP-VAR SV

Notes: Cumulative sum of squared errors (top); Cumulative sum of CRPS (bottom)

Table 1: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	4.35	7.19	6.55	6.20	3.67	6.68	6.61	6.26
Relative MSE								
Baseline/Raw	0.62***	1.01	1.05**	1.02	0.74**	1.00	1.01	1.00
Hybrid/Raw	0.62***	0.94	0.91	1.00	0.74**	0.90**	0.86*	0.98
Hybrid/Baseline	1.00	0.93*	0.86*	0.98*	1.00	0.90**	0.85**	0.98
CPI								
Raw	3.11	5.61	6.70	8.52	2.94	5.26	5.50	6.50
Relative MSE								
Baseline/Raw	0.33***	0.90**	0.95	0.92*	0.35***	0.86***	0.99	0.94***
Hybrid/Raw	0.33***	0.83***	0.71***	0.61***	0.35***	0.85***	0.88*	0.79***
Hybrid/Baseline	1.00	0.92*	0.75***	0.67***	1.00	1.00	0.89**	0.83***
UR								
Raw	0.05	0.64	2.41	3.78	0.05	0.60	2.39	3.75
Relative MSE								
Baseline/Raw	0.29***	0.73*	0.92	0.95	0.32***	0.73**	0.90*	0.96
Hybrid/Raw	0.29***	0.75**	0.90	0.95	0.32***	0.74**	0.87	0.97
Hybrid/Baseline	1.00	1.03	0.98	0.99	1.00	1.01	0.97	1.00
FFR								
Raw	0.19	2.11	6.12	10.16	0.10	2.12	7.17	11.66
Relative MSE								
Baseline/Raw	0.03***	0.68***	0.87**	0.93***	0.05***	0.63***	0.85***	0.92**
Hybrid/Raw	0.03***	0.77**	0.77*	0.65***	0.05***	0.64***	0.68***	0.57***
Hybrid/Baseline	1.00	1.13*	0.88*	0.70***	1.00	1.02	0.80**	0.62***
Credit Spread								
Raw	0.09	0.67	1.14	1.28	0.08	0.73	1.30	1.53
Relative MSE								
Baseline/Raw	0.77*	0.95***	0.97***	0.99	0.89**	0.99	0.98**	0.99*
Hybrid/Raw	0.80	0.94*	0.85***	0.85***	0.89**	0.93***	0.80***	0.72***
Hybrid/Baseline	1.04	1.00	0.87***	0.86***	0.99	0.94**	0.81***	0.73***

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2: Out-of-Sample **Density** Forecasting Performance: **Small BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	1.21	1.48	1.43	1.41	1.08	1.39	1.42	1.42
Relative CRPS								
Baseline - Raw	-0.28***	0.01	0.05**	0.01	-0.15***	0.00	0.01	0.00
Hybrid - Raw	-0.29***	-0.04	-0.02	-0.01	-0.15***	-0.06**	-0.05	0.00
Hybrid - Baseline	0.00	-0.05*	-0.07	-0.02	0.00	-0.06**	-0.07	0.01
CPI								
Raw	0.95	1.22	1.42	1.57	0.95	1.20	1.34	1.46
Relative CRPS								
Baseline - Raw	-0.38***	-0.07***	-0.04	-0.02	-0.37***	-0.11***	-0.03	-0.05**
Hybrid - Raw	-0.38***	-0.12***	-0.22***	-0.26***	-0.37***	-0.11***	-0.09	-0.12***
Hybrid - Baseline	0.00	-0.05***	-0.18***	-0.24***	0.00	0.00	-0.06*	-0.07*
UR								
Raw	0.12	0.40	0.80	1.07	0.12	0.37	0.79	1.05
Relative CRPS								
Baseline - Raw	-0.05***	-0.06**	-0.04	-0.02	-0.05***	-0.06**	-0.05*	-0.02
Hybrid - Raw	-0.05***	-0.06**	-0.05	-0.02	-0.05***	-0.05**	-0.06	0.00
Hybrid - Baseline	0.00	0.00	-0.01	0.01	0.00	0.00	-0.01	0.02
FFR								
Raw	0.28	0.84	1.43	1.87	0.17	0.78	1.54	2.04
Relative CRPS								
Baseline - Raw	-0.24***	-0.13***	-0.08*	-0.06**	-0.13***	-0.16***	-0.13***	-0.06**
Hybrid - Raw	-0.24***	-0.10**	-0.15*	-0.33***	-0.13***	-0.15***	-0.27***	-0.43***
Hybrid - Baseline	0.00	0.03*	-0.07	-0.28***	0.00	0.01*	-0.14**	-0.37***
Credit Spread								
Raw	0.15	0.43	0.59	0.65	0.14	0.45	0.66	0.75
Relative CRPS								
Baseline - Raw	-0.01**	-0.01***	-0.01*	0.00	-0.01**	-0.01***	-0.01	0.00
Hybrid - Raw	-0.01*	-0.01	-0.06***	-0.07***	-0.01**	-0.02***	-0.08***	-0.11***
Hybrid - Baseline	0.00	0.00	-0.05***	-0.07***	0.00	-0.02***	-0.08***	-0.11***

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), the three rows immediately below report relative CRPS: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. A lower value for CRPS is preferable, as such a negative value for row labeled Hybrid Baseline suggests that on average the Hybrid forecast is more accurate compared to the Baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 3: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	3.43	6.23	6.65	6.46	3.42	6.07	6.51	6.62
Relative MSE								
Baseline/Raw	0.79**	0.93	1.00	1.03***	0.79**	0.96	1.01	1.00
Hybrid/Raw	0.79**	0.96	0.89*	0.96	0.79**	0.98	0.85***	0.92**
Hybrid/Baseline	1.00	1.02	0.89**	0.93*	1.00	1.02	0.84***	0.92*
CPI								
Raw	2.95	4.71	5.11	6.12	2.77	4.79	4.67	5.50
Relative MSE								
Baseline/Raw	0.35***	1.00	1.00	1.00	0.37***	0.95***	1.05	1.02
Hybrid/Raw	0.35***	0.95	0.88***	0.87***	0.37***	0.91***	0.97	0.89***
Hybrid/Baseline	1.00	0.95***	0.88***	0.87**	1.00	0.96*	0.93**	0.88***
UR								
Raw	0.04	0.48	2.22	3.94	0.04	0.46	2.09	3.84
Relative MSE								
Baseline/Raw	0.35***	0.75*	0.84	0.93	0.35***	0.76**	0.86	0.92*
Hybrid/Raw	0.35***	0.86	0.96	0.99	0.35***	0.87***	0.97	0.96
Hybrid/Baseline	1.00	1.14	1.14	1.06	1.00	1.15*	1.13	1.03
FFR								
Raw	0.08	1.95	6.79	10.58	0.08	1.95	6.85	11.33
Relative MSE								
Baseline/Raw	0.06***	0.57***	0.81**	0.94	0.06***	0.57***	0.81***	0.93*
Hybrid/Raw	0.06***	0.57***	0.71**	0.75	0.06***	0.57***	0.72***	0.76**
Hybrid/Baseline	1.00	1.00	0.87*	0.80*	1.00	1.00	0.89	0.82**
Credit Spread								
Raw	0.09	0.63	1.06	1.12	0.09	0.66	1.09	1.19
Relative MSE								
Baseline/Raw	0.72**	0.91**	1.00	1.07	0.84**	0.91**	0.99	1.00
Hybrid/Raw	0.73**	0.93	0.91	0.97	0.82*	0.90**	0.94	0.92*
Hybrid/Baseline	1.02*	1.02	0.91	0.91*	0.98	0.99	0.95	0.92*

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 4: Out-of-Sample **Density** Forecasting Performance: **Small BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	1.04	1.37	1.41	1.37	1.05	1.36	1.43	1.43
Relative CRPS								
Baseline - Raw	-0.11**	-0.03	0.01	0.04***	-0.13***	-0.03*	-0.01	0.00
Hybrid - Raw	-0.11**	-0.03	-0.10*	-0.02	-0.13***	-0.03	-0.10***	-0.03*
Hybrid - Baseline	0.00	0.00	-0.11**	-0.06**	0.00	0.01	-0.09***	0.03*
CPI								
Raw	0.94	1.14	1.20	1.33	0.92	1.14	1.16	1.26
Relative CRPS								
Baseline - Raw	-0.36***	-0.02	0.00	0.01	-0.35***	-0.04***	0.00	-0.01
Hybrid - Raw	-0.36***	-0.06**	-0.10***	-0.11***	-0.35***	-0.08***	-0.05**	-0.06***
Hybrid - Baseline	0.00	-0.04**	-0.11***	-0.13***	0.00	-0.04**	-0.04**	-0.05***
UR								
Raw	0.12	0.36	0.80	1.13	0.11	0.35	0.76	1.11
Relative CRPS								
Baseline - Raw	-0.05***	-0.05*	-0.04	-0.02	-0.04***	-0.05***	-0.05*	-0.04*
Hybrid - Raw	-0.05***	-0.03	0.00	0.00	-0.04***	-0.03**	-0.02	-0.04
Hybrid - Baseline	0.00	0.02*	0.04*	0.02	0.00	0.02**	0.04	0.00
FFR								
Raw	0.16	0.78	1.48	1.86	0.15	0.75	1.47	1.92
Relative CRPS								
Baseline - Raw	-0.12***	-0.19***	-0.12***	-0.05	-0.12***	-0.19***	-0.15***	-0.08*
Hybrid - Raw	-0.12***	-0.20***	-0.21**	-0.23	-0.12***	-0.19***	-0.22***	-0.27**
Hybrid - Baseline	0.00	0.00	-0.09	-0.19*	0.00	0.00	-0.07	-0.20**
Credit Spread								
Raw	0.15	0.43	0.57	0.62	0.15	0.43	0.57	0.62
Relative CRPS								
Baseline - Raw	-0.02**	-0.02*	0.00	0.02	-0.01**	-0.02***	-0.01**	-0.01***
Hybrid - Raw	-0.02**	-0.02	-0.03*	-0.03	-0.01**	-0.02**	-0.03**	-0.02*
Hybrid - Baseline	0.00	0.00	-0.03*	-0.04*	0.00	0.00	-0.02**	-0.02*

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), the three rows immediately below report relative CRPS: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. A lower value for CRPS is preferable, as such a negative value for row labeled Hybrid Baseline suggests that on average the Hybrid forecast is more accurate compared to the Baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 5: Out-of-Sample **Point** Forecasting Performance: **Small TVP-VAR**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small TVP VAR				Small TVP VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	4.65	6.91	6.33	6.00	3.99	6.55	6.24	6.08
Relative MSE								
Baseline/Raw	0.58***	0.95**	1.03	1.00	0.68***	0.92**	0.98	1.01*
Hybrid/Raw	0.58***	0.98	0.93	1.04*	0.68***	0.86**	0.90	1.01
Hybrid/Baseline	1.00	1.03	0.90	1.05	1.00	0.94	0.92	0.99
CPI								
Raw	3.41	6.01	6.65	7.82	3.10	5.21	5.45	6.25
Relative MSE								
Baseline/Raw	0.30***	0.91**	0.95	0.95***	0.33***	0.93**	0.94***	0.96**
Hybrid/Raw	0.30***	0.82***	0.69***	0.67***	0.33***	0.88***	0.82***	0.79***
Hybrid/Baseline	1.00	0.91***	0.73***	0.70**	1.00	0.94*	0.88**	0.82***
UR								
Raw	0.05	0.49	1.98	3.16	0.05	0.50	2.04	3.36
Relative MSE								
Baseline/Raw	0.32***	0.82***	0.93*	0.98	0.34***	0.82**	0.93	0.98
Hybrid/Raw	0.32***	0.91	1.04	1.13	0.34***	0.84***	1.00	1.09
Hybrid/Baseline	1.00	1.11	1.12	1.16	1.00	1.03	1.08	1.11

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 6: Out-of-Sample **Density** Forecasting Performance: **Small TVP-VAR**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small TVP VAR				Small TVP VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	1.27	1.49	1.44	1.42	1.14	1.42	1.35	1.33
Relative CRPS								
Baseline - Raw	-0.34***	-0.04**	0.01	-0.01	-0.21***	-0.07**	0.00	0.01
Hybrid - Raw	-0.34***	-0.01	-0.04	0.02**	-0.21***	-0.12**	-0.05	0.02
Hybrid - Baseline	0.00	0.02	-0.06	0.03*	0.00	-0.05*	-0.05	0.01
CPI								
Raw	1.03	1.30	1.45	1.54	1.01	1.18	1.27	1.33
Relative CRPS								
Baseline - Raw	-0.45***	-0.08***	-0.04**	-0.03***	-0.42***	-0.07***	-0.05***	-0.02***
Hybrid - Raw	-0.45***	-0.11**	-0.18**	-0.19**	-0.42***	-0.11***	-0.12**	-0.10**
Hybrid - Baseline	0.00	-0.03	-0.14**	-0.15*	0.00	-0.04***	-0.06*	-0.08**
UR								
Raw	0.12	0.38	0.76	0.99	0.12	0.38	0.80	1.08
Relative CRPS								
Baseline - Raw	-0.05***	-0.04***	-0.03**	-0.01	-0.05***	-0.04***	-0.02	0.00
Hybrid - Raw	-0.05***	-0.04**	-0.02	0.04	-0.05***	-0.05***	-0.02	0.01
Hybrid - Baseline	0.00	0.01	0.02	0.05	0.00	0.00	0.00	0.01

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), the three rows immediately below report relative CRPS: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. A lower value for CRPS is preferable, as such a negative value for row labeled Hybrid - Baseline suggests that on average the Hybrid forecast is more accurate compared to the Baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table 7: CPI Inflation Real-Time Out-of-Sample Point Forecasting Performance

Panel A: Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
Medium VAR est.1960					Medium VAR with SV est.1960			
CPI Inflation	h=2Q	h=4Q	h=8Q	h=12Q	h=2Q	h=4Q	h=8Q	h=12Q
MSE								
Hybrid	4.36	4.58	4.55	4.91	4.50	4.47	4.57	4.95
Relative MSE								
Hybrid/RW (AO)	0.86*	0.90*	0.84*	0.93**	0.88*	0.89**	0.84*	0.93**
Hybrid/UCSV (SW)	0.98	1.00	0.94	1.00	1.00	0.99	0.94	1.00
Hybrid/FW	0.98	1.00	0.93**	0.95**	1.00	0.99	0.93**	0.95**
Hybrid/SPF	1.08	1.05	N/A	N/A	1.10	1.04	N/A	N/A
Relative CRPS								
Baseline - UCSV	0.02	0.08***	0.12***	0.21**	0.04	0.00	0.01	0.06
Hybrid - UCSV	-0.01	0.03	0.02	0.07*	0.01	0.02	0.00	0.08

Panel B: Pre-Crisis Sample (Recursive evaluation: 1994.Q1-2006.Q4)								
Medium VAR est.1960					Medium VAR with SV est.1960			
CPI Inflation	h=2Q	h=4Q	h=8Q	h=12Q	h=2Q	h=4Q	h=8Q	h=12Q
MSE								
Hybrid	1.47	1.57	1.79	1.86	1.35	1.45	1.71	1.89
Relative MSE								
Hybrid/RW (AO)	0.96	0.91	0.79	0.88	0.89	0.84*	0.75*	0.89
Hybrid/UCSV (SW)	1.01	1.00	0.91	0.93	0.93	0.92	0.87	0.94
Hybrid/FW	1.01	0.98	0.94	0.87	0.92	0.90	0.89	0.88
Hybrid/SPF	0.99	0.94	N/A	N/A	0.91	0.87	N/A	N/A
Relative CRPS								
Baseline - UCSV	0.01	0.06	0.11	0.32***	0.00	-0.02	-0.01	0.08**
Hybrid - UCSV	0.01	0.04	0.06	0.11	-0.02	-0.03	-0.05	0.02

Notes for Table: The first row (in both panels A and B) report the mean squared error (MSE) corresponding to the hybrid forecast obtained from Medium BVAR. Rows 2 to 5 report relative MSE. So a ratio of less than 1 indicates that hybrid forecast is on average more accurate compared to the respective univariate forecast. Row 6 reports relative CRPS of baseline density forecast relative to UCSV, a negative value suggests that baseline density forecast is more accurate on average. Row 7 reports relative CRPS of hybrid density forecast relative to UCSV, a negative value suggests that hybrid density forecast is more accurate on average. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values. All models use the SPF nowcast for the one-step-ahead forecast; as a result, the relative MSE is equal to one and is not reported.