# Forecasting with Bayesian Vector Autoregressions estimated using Professional Forecasts*

Christoph Frey† Frieder Mokinski‡

University of Konstanz ZEW

May 27, 2014

**Abstract**

We propose a Bayesian shrinkage approach for vector autoregressions that uses survey forecasts as additional non-sample information. In particular, we augment the vector of dependent variables by their survey nowcasts, and claim that each variable of the VAR and its nowcast are likely to depend in a similar way on the lagged dependent variables. The idea is that this additional information will help us pin down the model coefficients. We find that the forecasts obtained from a VAR fitted by our new shrinkage approach typically yield smaller mean squared forecast errors than the forecasts obtained from a range of benchmark methods.

**Keywords:** Bayesian inference, survey forecasts, VAR, forecasting

**JEL classification:** C11, C32, C53, E17

*We would like to thank Luc Bauwens, Fabian Krüger and Winfried Pohlmeier for helpful comments. Financial support by the DFG (Project "Development and application of new methodologies of combining time series and expert survey data for economic forecasting.") is gratefully acknowledged. All remaining errors are ours.

†Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-5111, fax: +49-7531-88-4450, email: Christoph.Frey@uni-konstanz.de.

‡Centre for European Economic Research (ZEW), L7, 1, D-68161 Mannheim, Germany. Phone: +49-621-1235-143, fax: +49-621-1235-223, email: Mokinski@zew.de.

# 1 Introduction

Vector Autoregressions (VARs) are among the most popular tools in economic forecasting. However, as even medium-sized VARs (10-20 variables) have several hundred parameters to estimate, potential over-fitting is an immediate threat to forecast accuracy. The literature has therefore either used VARs with only a handful of variables (Chauvet and Potter, 2013; Faust and Wright, 2013), or it has resorted to Bayesian shrinkage methods (Banbura et al., 2010). Such methods include Doan et al. (1984)'s Minnesota prior, which assumes that each variable evolves according to a random walk, and Wright (2013)'s democratic steady-state prior, which uses long-run forecasts from an expert survey as prior information for the vector of unconditional means.

We build on Wright (2013)'s work and consider a *Bayesian shrinkage approach that additionally exploits the non-sample information in survey nowcasts*, i.e. forecasts for the current quarter or month. The idea of our approach is that the variables of the VAR and their corresponding survey nowcasts are likely to depend in a similar way on the lagged dependent variables. To exploit this conjecture, we first augment the vector of dependent variables of the VAR with survey nowcasts of the variables involved in the VAR and then express our belief of similar dependence on the lagged dependent variables through a Bayesian prior. The idea is best illustrated with a simple example: Consider a variable $y_t$, modeled as a univariate autoregression (AR) with a single lag $y_t = ay_{t-1} + \varepsilon_t$ and its nowcast $s_t$. The augmented model is

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a \\ a+\Delta \end{bmatrix} y_{t-1} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix},$$

and the prior distribution favoring pairwise identical coefficients can be stated as

$$p \begin{bmatrix} a \\ \Delta \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \underline{a} \\ 0 \end{bmatrix}, \begin{bmatrix} \underline{v}_a & 0 \\ 0 & \underline{v}_\Delta \end{bmatrix} \right).$$

Through $\underline{v}_\Delta$ we express our confidence in pairwise equality. If the dependence of the survey nowcasts on the lagged dependent variables is indeed not to dissimilar from the actuals, i.e. if $\Delta$ is small in the example, the extra information provided through the survey nowcasts will help us pin down the parameters of the original VAR. Put differently, the shrinkage method is likely to reduce the risk of over-fitting the model to the data and we therefore expect it to provide us with more accurate forecasts.[1]

The idea of similar dependence on the lagged dependent variables can be motivated in several ways: First, empirically, survey nowcasts have often been found to be very accurate predictions of the target variable (e.g. Faust and Wright, 2013). We would therefore expect that they exploit the available information in a way that resembles the true data generating process. Second, from an economic theory perspective, we show in Appendix A.1 that identical regressions coefficients arise if (i) expectations are formed in a fully rational manner based on an information set that includes the lagged dependent variables of the VAR, and (ii) the specification of the VAR is correct in an appropriate sense.

Similar approaches have been used in the frequentist estimation of a three-factor affine Gaussian model for U.S. Treasury yields by Kim and Orphanides (2012), and in the Bayesian estimation of a DSGE model by Del Negro and Schorfheide (2013). However, besides the different model class, a major difference to our method is that these authors have assumed that coefficients are *exactly equal* for each pair of actuals and nowcasts. By avoiding to impose equal coefficients deterministically, our Bayesian shrinkage method reduces the risk of deteriorating forecasts by imposing restrictions that may turn out to be severely erroneous.

We employ our new method to forecast with a ten-variable VAR featuring a range of U.S. macroeconomic and financial variables. We find that mean squared forecast errors (MSFEs) are typically lower with our method than with a univariate AR(1) estimated by OLS, uniformly lower than with the same VAR estimated using only the Minnesota prior, and comparable to those of survey forecasts.

The paper is structured as follows. Chapter 2 introduces the methodology and the underlying econometric ideas. Chapter 3 presents our empirical findings and chapter 4 summarizes our results.

[1]Taking a frequentist perspective, Ing and Wei (2003, Theorem 3) show that better coefficient estimates (in terms of mean squared error (MSE)) asymptotically translate into superior forecasts (in terms of MSFE).

## 2 VAR estimation using professional nowcasts

### 2.1 Augmenting a VAR with survey nowcasts

Our point of departure is a standard $M$-variate VAR model with $p$ lags

$$y_t = a_0 + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t, \tag{1}$$

where $y_t$ is the $M \times 1$ vector of dependent variables, $a_0$ is an $M \times 1$ vector of intercepts, $A_i$ is an $M \times M$ matrix of slope coefficients, and $\varepsilon_t$ is an $M \times 1$ vector of disturbances. We augment the VAR with:

$$s_t = b_0 + \sum_{i=1}^{p} B_i y_{t-i} + \eta_t, \tag{2}$$

where $s_t$ collects the survey nowcasts of the variables in $y_t$, $\eta_t$ is another $M \times 1$ vector of disturbances, and $\{b_0, B_1, \dots, B_p\}$ are used in the same way as in equation (1). The augmented VAR reads

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_i \\ B_i \end{bmatrix} y_{t-i} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}. \tag{3}$$

Equation (3) states that the survey nowcasts $s_t$ for the elements of $y_t$ depend on the same variables $\{y_{t-1}, \dots, y_{t-p}\}$ as $y_t$ itself, though they can have different coefficients. Estimating the augmented system (3) without imposing further restrictions on $\{b_0, B_1, \dots B_p\}$, we will hardly reduce the risk of over-fitting $\{a_0, A_1, \dots, A_p\}$ to the data. By contrast, if we impose $\{b_0 = a_0, B_1 = A_1, \dots, B_p = A_p\}$, provided that the restrictions are not *too incorrect*, this may help us to pin down the parameters of the VAR. To see that, it is convenient to take a frequentist perspective for a moment. To keep things simple, we consider the same AR(1) as in the introduction and impose equal coefficients:

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a \\ a \end{bmatrix} y_{t-1} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \qquad |a| < 1, \qquad \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,\eta} \\ \sigma_{\varepsilon,\eta} & \sigma_\eta^2 \end{bmatrix} \right). \tag{4}$$

By standard Maximum Likelihood theory, the asymptotic distribution of the parameter estimate from the augmented model is

$$\sqrt{T}(\hat{a}_{aug} - a) \xrightarrow{d} \mathcal{N}\left(0\,,\,(1-a^2)\frac{(\sigma_\eta^2\sigma_\varepsilon^2 - \sigma_{\varepsilon,\eta}^2)}{\sigma_\varepsilon^2(\sigma_\eta^2 - 2\sigma_{\varepsilon,\eta} + \sigma_\varepsilon^2)}\right). \tag{5}$$

By contrast, the standard OLS estimation approach for the AR ($y_t = ay_{t-1} + \varepsilon_t$), which makes no use of survey nowcasts, is asymptotically distributed as

$$\sqrt{T}(\hat{a}_{std} - a) \xrightarrow{d} \mathcal{N}\left(0\,,\,1-a^2\right). \tag{6}$$

Thus, the ratio of the two asymptotic variances is

$$\mathrm{VR} := \frac{\mathrm{V}_a[\hat{a}_{aug}]}{\mathrm{V}_a[\hat{a}_{std}]} = \frac{(\sigma_\eta^2\sigma_\varepsilon^2 - \sigma_{\varepsilon,\eta}^2)}{\sigma_\varepsilon^2(\sigma_\eta^2 - 2\sigma_{\varepsilon,\eta} + \sigma_\varepsilon^2)} = \frac{r^2(1-\rho^2)}{r^2 - 2\rho r + 1}, \tag{7}$$

where $r = \sigma_\eta/\sigma_\varepsilon$ measures the imprecision of the survey nowcast as a signal about the conditional mean $\mathrm{E}[y_t|y_{t-1},\ldots,y_{t-p}]$, and $\rho$ is the correlation between the two disturbances $\varepsilon_t$ and $\eta_t$. A value of VR below one means that the parameter estimate from the augmented model is asymptotically more precise than the standard OLS estimate. It is easy to show that VR can never exceed one, meaning that the estimator based on the augmented model never produces asymptotically less efficient parameter estimates. Figure 1 depicts VR as a function of $\rho$ and $r$. It shows that gains are particularly high when $r$ is small, i.e. if survey nowcasts tend to be relatively close to the true conditional mean, and if the correlation $\rho$ among the two disturbances is either negative or close to one.

## 2.2 How to address that survey nowcasts may not be "correctly specified"?

In the previous section, we have derived the efficiency gain implied by the augmented model conditional on the assumption of equal coefficient matrices for the actuals $y_t$ and survey nowcasts $s_t$, i.e. $b_0 = a_0, B_1 = A_1, \ldots, B_p = A_p$. This is arguably a demanding assumption that is not likely to be exactly met in practice. Indeed, in Appendix A.1, we show that sufficient conditions for it to hold are that expectations are formed in a fully rational manner based on an information set that includes the conditioning information of the correctly specified VAR.

In this section, we propose a Bayesian estimation approach, that uses equal coefficients as a shrinkage target, but does not impose them deterministically. We thus conserve some of the potential gains sketched in the previous section without running into the risk of deteriorating forecasts by imposing severely erroneous restrictions.

To express the belief that coefficients are equal, it is helpful to adjust the parametrization of the augmented VAR in equation (3). Specifically, we replace $b_0$ with $a_0 + \Delta_0$, $B_1$ with $A_1 + \Delta_1$, $B_2$ with $A_2 + \Delta_2$, etc. such that (3) becomes

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a_0 \\ a_0 + \Delta_0 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} y_{t-i} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \tag{8}$$

and, for convenience, we assume $\begin{bmatrix} \varepsilon_t' & \eta_t' \end{bmatrix}' \sim \mathcal{N}\left(0\,,\, \Sigma\right)$. Using the new parameterization, we specify a multivariate normal prior distribution for $\{a_0, A_1, \ldots, A_p, \Delta_0, \Delta_1, \ldots, \Delta_p\}$. Given that we assume that all prior covariances are zero, it suffices to define the marginal prior distribution for each element of the aforementioned matrices and vectors. Denoting by $A_i^{k,l}$ and $\Delta_i^{k,l}$ the (k,l)-cell of $A_i$ and $\Delta_i$ respectively, the marginal priors are

$$p(A_i^{k,l}) \sim \mathcal{N}\left(\underline{A}_i^{k,l}\,,\, \lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right),$$
$$\text{with } \underline{A}_i^{k,l} = d_k \text{ if } k = l \wedge i = 1, \text{ and } \underline{A}_i^{k,l} = 0 \text{ otherwise,} \tag{9}$$
$$p(\Delta_i^{k,l}) \sim \mathcal{N}\left(0\,,\, \zeta^2\left(\lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right)\right), \tag{10}$$
$$p(a_0) \sim \mathcal{N}\left(0\,,\, \kappa \cdot I_M\right), \tag{11}$$
$$p(\Delta_0) \sim \mathcal{N}\left(0\,,\, \kappa \cdot I_M\right), \tag{12}$$

where $\kappa \to \infty$. The joint prior distribution is the product of the independent marginals. We complete the specification by assuming a diffuse prior distribution for $\Sigma$ that is independent from the prior distribution of the remaining model parameters: $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$.

Next, we discuss the prior element-by-element. The prior for $\{a_0, A_1, \ldots, A_p\}$ is a variant of the Minnesota prior (Doan et al., 1984) that has been used by Wright (2013). While being diffuse about the vector of intercepts $a_0$, it is informative about the matrices of slope parameters $\{A_1, \ldots, A_p\}$. By setting all prior means except for the first lag of the dependent variable to zero, it expresses the belief that the variables are generated from univariate AR(1) processes.[2]

[2] This contrasts with Doan et al. (1984), who have suggested a random walk prior with $d_1 = \ldots = d_M = 1$.

In the specification of the prior variances in equation (9), the hyperparameter $\lambda$ governs the overall tightness of the prior for $A_1, \ldots, A_p$: If $\lambda = 0$, the prior expresses that we are absolutely certain about the prior means. If, by contrast, $\lambda \to \infty$, the prior becomes diffuse. The factor $1/i^2$ implies that the prior gets tighter, the higher the lag we consider. It thus reflects the belief that more distant lags play a minor role. Finally, the ratio $\sigma_k^2/\sigma_l^2$ accommodates differences in the scale and variability of the different variables. As we do not have a good prior guess about the term, we follow common practice and proxy $\sigma_k^2$ by the residual variance of an AR(1) regression for the $k$-th variable.

The prior for $\{\Delta_0, \Delta_1, \ldots, \Delta_p\}$ is centered at zero, reflecting that we expect the coefficients to be equal for the actuals $y_t$ and their survey nowcasts $s_t$. By specifying the prior variances of the $\Delta_i$'s relative to the corresponding elements of $\{a_0, A_1, \ldots, A_p\}$, we obtain a parsimonious way to express our confidence in equal coefficients. Details about the posterior distribution are given in Appendix A.2.

## 2.3 Adding Wright's democratic steady-state prior

Wright (2013) suggests using long-term survey forecasts to form a prior for the unconditional mean of the variables involved in a VAR. The underlying idea is that professional forecasters should realize shifts in time series endpoints well before they can be inferred from realizations of the process. Villani (2009) outlines the Bayesian estimation of a VAR where a prior is specified for the unconditional mean instead of the vector of intercepts as in section 2.2. We extend his approach to the augmented VAR. To implement a prior for the unconditional mean, we set up the following steady-state representation of the augmented VAR:

$$\begin{bmatrix} y_t - \psi \\ s_t - \psi^+ \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} (y_{t-i} - \psi) + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \tag{13}$$

Their specification makes perfect sense when time series are modeled in levels, but it is inappropriate for the stationary variables we consider (see e.g. Banbura et al., 2010).

where $\psi = \mathrm{E}[y_t]$ and $\psi^+ = \mathrm{E}[s_t]$. Equation (13) has been obtained by subtracting

$$\underbrace{\begin{bmatrix} \mathrm{E}[y_t] \\ \mathrm{E}[s_t] \end{bmatrix}}_{=[\psi' \ \psi^{+\prime}]'} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} \underbrace{\mathrm{E}[y_{t-i}]}_{=\psi}$$

from equation (8).

Parameterizing $\psi^+ = \psi + \Delta_\psi$, we specify a multivariate normal prior for $\{\psi, A_1, \ldots, A_p, \Delta_\psi, \Delta_1, \ldots, \Delta_p\}$. Denoting by $A_i^{k,l}$ and $\Delta_i^{k,l}$ the (k,l)-cell of $A_i$ and $\Delta_i$ respectively and by $\psi_k$ and $\Delta_\psi^k$ the k-th entry of $\psi$ and $\Delta_\psi$, we set

$$p\left(A_i^{k,l}\right) \sim \mathcal{N}\left(\underline{A}_i^{k,l}, \lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right)$$
$$\text{with } \underline{A}_i^{k,l} = d_k \text{ if } k = l \wedge i = 1, \text{ and } \underline{A}_i^{k,l} = 0 \text{ otherwise,} \tag{14}$$

$$p\left(\Delta_i^{k,l}\right) \sim \mathcal{N}\left(0, \zeta^2\left(\lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right)\right), \tag{15}$$

$$p\left(\psi_j\right) \sim \mathcal{N}\left(\underline{\psi}_j, \lambda_0^2\right), \tag{16}$$

$$p\left(\Delta_{\psi_j}\right) \sim \mathcal{N}\left(0, \zeta_0^2 \cdot \lambda_0^2\right). \tag{17}$$

Once again, provided that we assume that the prior covariances are zero, the joint prior can be obtained by multiplying the marginals. With regard to the elements of $A_i$ and $\Delta_i$, the prior is identical to section 2.2, but instead of being diffuse about the vector of intercepts, it uses an informative prior for the vector of unconditional means $\psi$ and for the difference vector $\Delta_\psi$. Following Wright (2013), we set the elements of $\underline{\psi}_j$ to the most recent average long-term survey forecasts.[3] The hyperparameter $\lambda_0$ governs the tightness of the prior for $\psi$, and thus reflects how optimistic we are about the informativeness of the long-term forecasts. Eventually, $\zeta_0$ expresses our confidence in the equality of $\psi$ and $\psi^+$, where $\psi^+$ is the unconditional mean implied by the survey nowcasts. The specification is completed by assuming an independent diffuse prior for $\Sigma$, $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$. Details about the posterior distribution can be found in Appendix A.3.

[3] For example, for the CPI inflation rate we use the forecasts with a ten year horizon collected by the Philadelphia Federal Reserve's Survey of Professional Forecasters.

## 3 Empirical application

In this section, we evaluate the forecasts of a ten-variable quarterly VAR(4) that is estimated using our novel approach. As in Wright (2013), our model features eight U.S. macroeconomic variables, a short-term and a long-term yield. To produce the forecasts, we use real-time data from the *Philadelphia Federal Reserve Bank's Real-Time Data Set for Macroeconomists* and *average* survey forecasts from its quarterly *Survey of Professional Forecasters* (SPF). Table 1 gives details about the data and how we have processed it.

We conduct the following forecasting experiment: Each period from the second quarter of 1984 through the second quarter of 2011, we re-estimate the VAR on an expanding real-time data window, and use the estimated model to produce forecasts at horizons of one, four, eight and twelve quarters. Thereby, the estimation window has an atypical design: Whereas the time series of actuals ($y_t$) starts in the second quarter of 1962, survey nowcasts ($s_t$) are first observed in the fourth quarter of 1968. In Appendix A.4, we modify our approach to this setting, additionally acknowledging that we observe survey nowcasts only for a subset of the variables in the VAR. The reason for adjusting our approach is that we want to avoid wasting time-series information.

In what follows, we try to discern the impact of the different sets of non-sample information by considering alternative specifications of the prior specified in equations (14 - 17) of section 2.3. Table 2 gives the details: Specification $M$ has the structure of Doan et al. (1984)'s Minnesota prior and ignores all survey information. $W$ adds Wright (2013)'s democratic steady-state prior and thus additionally exploits the long-run survey forecasts. $S$ extends $W$ by using the non-sample information provided through the survey nowcasts. Finally, $S2$ sets the prior variances of the difference parameters to very low values and thus virtually imposes that the slope and unconditional mean parameters are exactly identical for the elements of $y_t$ and of $s_t$.

Below, we study the forecasts for real GDP growth, GDP deflator inflation, CPI inflation, industrial production growth, the three-month Treasury bill rate and the unemployment rate. We evaluate the forecasts by their MSFE, specifying as the forecast target the value recorded in the second vintage following the quarter, to which the prediction refers. Benchmark forecasts are generated from an AR(1) model, which is estimated by OLS. The AR(1) is often found to be a tough competitor to more complex forecasting models (Chauvet and Potter, 2013; Del Negro and Schorfheide, 2013).[4]

Table 3 reports the results of the forecasting experiment. Its key message is that our shrinkage method that uses the full set of non-sample information (specifications $S$ and $S2$) produces better forecasts for most variables and horizons than all the benchmarks we consider. The result highlights that it pays off in terms of forecast accuracy to exploit the non-sample information provided through the survey nowcasts.

More specifically, our main findings from Table 3 are: First, in terms of its MSFE, the OLS-VAR(4) is typically inferior to the OLS-AR(1). As the AR(1) model is nested in the VAR(4), this deterioration is likely to reflect over-fitting. Second, the Minnesota prior ($M$) turns out to improve the VAR forecasts, yet only to a level that is comparable to that of the OLS-AR(1). Third, adding the democratic steady-state prior as in specification $W$ increases the forecast precision (relative to $M$) for the long-run inflation forecasts, but turns out to make little difference for the remaining variables and horizons. To understand the differences between our results and those of Wright (2013), it is important to note that we use different long-term survey forecasts. Whereas he uses data from the Blue Chip Survey that has collected long-term forecasts of all the ten variables twice a year since 1984, the SPF's ten-year forecasts are available for only four variables and start in 1991:Q4 earliest. It is therefore not surprising that he finds a much larger improvement in predictive ability from the steady-state prior than we do. Fourth, in most cases, augmenting the VAR with survey nowcasts as in specification $S$ gives superior forecasts. The strongest improvements are obtained for the two inflation series (with a relative gain above 50 percent for GDP deflator inflation on the longest horizon) and for the unemployment rate. For real GDP growth, industrial production growth and the Treasury bill yield, the improvements are less profound but still visible. Fifth, adjusting the prior to rely even more on the survey nowcasts, specification $S2$ gives an additional improvement in predictive ability. This is indirect evidence for our initial guess that survey nowcasts and actuals depend in a very similar way on the lagged dependent variables.

To test if a method improves significantly over the OLS-AR(1), we apply the test for equal finite sample predictive ability proposed by Giacomini and White (2006).[5] While the test results

[4] As an alternative, following Wright (2013), we have considered the forecasts of an AR(p) model with the lag length selected by the BIC. We found that, on average, the AR(1) was harder to beat.

[5] Due to the expanding estimation window, the asymptotics presented in Giacomini and White (2006) are not valid in our context. In favor of using the method with expanding estimation windows anyway, Clark and McCracken (2009) show in a simulation study that the test has reasonable size properties.

support that the OLS-VAR(4) tends to produce inferior forecasts, the predictive ability of the specifications $M$ and $W$ is rarely significantly different from the OLS-AR(1). By contrast, the forecasts of specifications $S$ and $S2$ are significantly superior at all horizons for the two inflation rates and the unemployment rate. Moreover, they significantly improve over the AR(1) at longer forecast horizons for industrial production growth and the 3-month T-Bill yield.

### 3.1 Trained hyperparameters

So far, we have considered four alternative specifications of the set of prior hyperparameters $\{\lambda, \zeta, \lambda_0, \zeta_0\}$, finding that their choice strongly affects forecasting performance: On an evaluation sample spanning from the second quarter of 1984 through the second quarter of 2011, we found that stronger shrinkage, i.e. lower parameter values, typically implies better forecasting performance. Despite the promising result, a valid criticism is the arbitrary choice of the hyperparameters values. To address this concern, we choose the hyperparameters based on a training sample and evaluate the performance of the specification on a subsequent evaluation sample. Specifically, using the real-time data as of the fourth quarter of 1990, we evaluate each combination of the following values for the hyperparameters: $\lambda = \{.01, .05, .1, .15, .2\}$, $\zeta = \{.01, .1, .5, 1, 2, 10\}$, $\lambda_0 = .5$, $\zeta_0 = \{.01, .1, .5, 1, 2, 10\}$. On this occasion, we need to fix $\lambda_0$ because our data set features no long-term survey forecasts in the training sample.[6] To choose a single best specification, we use a criterion that aggregates the forecast performance across several variables and horizons. In the spirit of Wright (2013), we compute for each variable-horizon combination the relative MSFE versus the $AR(1)$ model, and aggregate by averaging across variables and forecast horizons (considering only the six variables and four horizons evaluated in Table 3). We find that the criterion prefers the following specification, which we subsequently denote by $T$: $\lambda^* = 0.1$, $\zeta^* = 0.01$, $\lambda_0^* = 0.5$ and $\zeta_0^* = 0.01$. This is the tightest specification available with respect to $\zeta$ and $\zeta_0$, the two hyperparameters that relate to the survey nowcasts, but not with respect to $\lambda$, the hyperparameter that governs the tightness of the Minnesota prior. We use $T$ to produce pseudo real-time out-of-sample forecasts starting with the 1990:Q4 real-time data vintage.

Table 4 summarizes the results of the forecasting experiment: The four specifications $M$, $W$,

[6] The value of 0.5 roughly coincides with the specification that Wright (2013) infers from his training sample (0.557 in terms of our specification of the prior) using a richer data set of survey long-run forecasts.

$S$, and $S2$ perform similarly on this shorter evaluation sample as on the full sample considered in the previous section: The tightest variant $S2$ typically provides the best forecasts. Interestingly, the trained specification $T$ roughly performs at eye-level with the best specification ($S2$), indicating that the real-time choice of hyperparameters works pretty well.

### 3.2 Survey forecasts and forecast combination

In this section, we compare the forecasts from our method, using prior specification $T$, to two additional benchmarks: The SPF survey forecasts themselves, and different linear combinations of the survey forecasts and the Bayesian VAR forecasts. Contrary to the previous evaluations, due to the limited availability of the survey data, we can only consider forecasts at horizons of one, two, three, and four quarters.

The comparison of the model-based forecasts with survey-forecasts raises some intricate timing issues: For a fair comparison, the two methods should have similar information sets available. To illustrate the difficulty, we consider the one-quarter ahead forecast for the growth of real GDP in 1990:Q4: The latest information used by the VAR refers to 1990:Q3, whereas (i) the one-quarter ahead survey forecast produced by the quarter-mid of 1990:Q3 has only limited information about the 1990:Q3 data and (ii) the survey nowcast made in 1990:Q4 has extra-information (relative to the VAR) about the ongoing quarter, such as the industrial production growth in 1990:M10. Here, we follow Wright (2013) and use the one-quarter ahead survey forecast, thus putting the survey forecasts at a slight information disadvantage relative to the VAR.

Despite this disadvantage, Table 5 shows that survey forecasts are a tough competitor to our method. Considering the two inflation series, the gain from using the survey forecast is considerable with respect to GDP deflator inflation and moderate for CPI inflation. Considering the remaining four series, the table suggests that the two methods roughly perform at eye-level with a slight edge for our method. It should be kept in mind that even though our method cannot clearly beat survey forecasts, it has the advantage of providing forecasts at any horizon and any point in time.

The head-to-head race among our method and the survey forecasts suggests that we may benefit from forecast combinations. We consider three approaches with pseudo real-time updates of the forecast weights:

1. The **MSFE approach** weighs the forecasts according to the inverse of their MSFE.

2. The **Granger and Ramanathan (1984) approach** obtains weights by regressing the realization on the two forecasts, subject to the restriction the regression coefficient sum to unity.

3. The $\mathbf{\frac{1}{N}}$ **approach** weighs each forecast by 0.5.

The results are also found in Table 5: The first insight is that the different weighting approaches perform similarly, allowing no uniform ranking across the variables and horizons. Moreover, the MSFE of the combined forecast is typically marginally higher than the MSFE of the better individual forecast. This is a typical result in forecast combination experiments (e.g. Krüger, 2014) and suggests that without reliable ex-ante knowledge of the relative performance of the two forecast methods, combination is an advisable strategy.

# 4 Concluding Remarks

In this paper, we have proposed a Bayesian shrinkage method for VARs that uses both long- and short-run survey forecasts as non-sample information. Our empirical application has shown that the method typically improves forecast accuracy relative to approaches that do not use such (non-sample) information. The shrinkage approach is easy to implement and it can be transfered to other types of time-series models, such as the non-linear class of vector STAR models (e.g. Schleer, 2013).

# References

Banbura M, Giannone D, Reichlin L. 2010. Large bayesian vector auto regressions. *Journal of Applied Econometrics* **25**: 71–92.

Bauwens L, Lubrano M, Richard J. 1999. *Bayesian Inference in Dynamic Econometric Models*. Advanced Texts in Econometrics. Oxford University Press.

Chauvet M, Potter S. 2013. Forecasting output. In *Handbook of Economic Forecasting, Vol. 2*. Elsevier, 141–194.

Clark TE, McCracken MW. 2009. Nested forecast model comparisons: A new approach to testing equal accuracy. Research Working Paper RWP 09-11, Federal Reserve Bank of Kansas City.

Croushore D. 1993. Introducing: The Survey of Professional Forecasters. *Business Review* **6**: 3–15.

Croushore D, Stark T. 2001. A Real-Time Data Set for Macroeconomists. *Journal of Econometrics* **105**: 111–130.

Del Negro M, Schorfheide F. 2013. DSGE model-based forecasting. In *Handbook of Economic Forecasting, Vol. 2*. Elsevier, 57–140.

Doan T, Litterman R, Sims C. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3**: 1–100.

Faust J, Wright JH. 2013. Forecasting inflation. In *Handbook of Economic Forecasting, Vol. 2*. Elsevier, 2 – 56.

Geweke J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics. Wiley.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.

Granger CWJ, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3**: 197–204.

Ing CK, Wei CZ. 2003. On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* **85**: 130–155.

Kim D, Orphanides A. 2012. Term structure estimation with survey data on interest rate forecasts. *Journal of Financial & Quantitative Analysis* **47**: 241–272.

Krüger F. 2014. Combining density forecasts under various scoring rules: An analysis of uk inflation. Working Paper, Heidelberg Institute for Theoretical Studies.

Pesaran MH, Weale M. 2006. Survey Expectations. In *Handbook of Economic Forecasting*. Amsterdam: Elsevier, 715–776.

Schleer F. 2013. Finding starting-values for maximum likelihood estimation of vector star models. ZEW Discussion Papers 13-076.

Schmidt P. 1977. Estimation of seemingly unrelated regressions with unequal numbers of observations. *Journal of Econometrics* **5**: 365–377.

Swamy P, Mehta J. 1975. On bayesian estimation of seemingly unrelated regressions when some observations are missing. *Journal of Econometrics* **3**: 157–169.

Villani M. 2009. Steady-state priors for vector autoregressions. *Journal of Applied Econometrics* **24**: 630–650.

Wright JH. 2013. Evaluating real-time var forecasts with an informative democratic prior. *Journal of Applied Econometrics* **28**: 762–776.

# A Appendix

## A.1 A set of assumptions that ensures equal coefficients

Considering the augmented VAR of equation (3), in this Appendix we demonstrate that equal coefficient matrices, i.e. $b_0 = a_0, B_1 = A_1, \ldots, B_p = A_p$, arise under a set of assumptions about the specification of the VAR (assumption i) and the way survey nowcasts are produced (assumptions ii and iii). These assumptions are:

(i) The VAR model is correctly specified, i.e. $\mathrm{E}[y_t | y_{t-1}, \ldots, y_{t-p}] = a_0 + \sum_{i=1}^{p} A_i y_{t-i}$, implying that the disturbance vector $\varepsilon_t$ has a mean of zero and is orthogonal to the lagged dependent variables.

(ii) Survey nowcasts do not systematically deviate from the expectations of the corresponding variables conditional on some public information set $\Omega_t$, i.e. $s_t = \mathrm{E}[y_t | \Omega_t] + u_t^*$ with $u_t^*$, where $u_t^*$ is a mean zero error independent of $\Omega_t$.

(iii) The public information set includes the lagged dependent variables of the VAR system, i.e. $\{Y_{t-1}, \ldots, Y_{t-p}\} \subseteq \Omega_t$.

Assumption (i) is standard for any frequentist estimation approach. Assumption (ii) requires a weak form of rationality - the so-called "average form of the rational expectations hypothesis" (cp. Pesaran and Weale, 2006, p.722ff): Forecasters must on average make efficient use of the information they have. Eventually, assumption (iii) demands that forecasters make their predictions employing at least the conditional information set of the VAR. We would argue that it is likely that professional forecasters make their forecasts considering additional variables. For example, we think of factors that cannot easily be quantified, such as policy changes, or factors that have a highly non-linear impact. Note that assumption (iii) implies that $s_t$ must be a nowcast, i.e. a prediction referring to the current quarter or month. If $s_t$ were a one-quarter ahead forecast of $y_t$ formed in quarter $t-1$, then $y_{t-1}$ could not be part of the information set used to make the forecast because it is first observed in quarter $t$. Assumption (iii) would thus be violated.

**Lemma A.1**

Let assumptions (i) - (iii) hold. Then, $s_t = a_0 + \sum_{i=1}^{p} A_i y_{t-i} + \eta_t$ with $\mathrm{E}[\eta_t | y_{t-1}, \ldots, y_{t-p}] = 0$.

**Proof A.1**

We need to show that $s_t = a_0 + \sum_{i=1}^{p} A_i y_{t-i} + \eta_t$ with $\mathrm{E}[\eta_t | y_{t-1}, \ldots, y_{t-p}] = 0$.

First, we show that the representation is valid:

$$\begin{aligned} s_t &\overset{(ii)}{=} \mathrm{E}[y_t|\Omega_{t-1}] + u_t^* \\ &\overset{(i),(iii)}{=} a_0 + \sum\nolimits_{i=1}^{p} A_i y_{t-i} + \mathrm{E}[\varepsilon_t|\Omega_{t-1}] + u_t^* \\ &= a_0 + \sum\nolimits_{i=1}^{p} A_i y_{t-i} + \eta_t, \quad \text{if we define } \eta_t := \mathrm{E}[\varepsilon_t|\Omega_{t-1}] + u_t^*. \end{aligned}$$

Next, we show that the disturbance has a conditional mean of zero:

$$\begin{aligned} \mathrm{E}[\eta_t|y_{t-1}, \ldots, y_{t-p}] &= E\left[\mathrm{E}[\varepsilon_t|\Omega_{t-1}]|y_{t-1}, \ldots, y_{t-p}\right] + E\left[u_t^*|y_{t-1}, \ldots, y_{t-p}\right] \\ &\overset{(iii)}{=} \underbrace{E\left[\varepsilon_t|y_{t-1}, \ldots, y_{t-p}\right]}_{\overset{(i)}{=}0} + \underbrace{E\left[u_t^*|y_{t-1}, \ldots, y_{t-p}\right]}_{\overset{(ii),(iii)}{=}0} \\ &= 0. \end{aligned}$$

□

## A.2 Posterior distribution for the augmented VAR

To derive the posterior distribution of the model of section 2.2, we put (8) in matrix notation:

$$\begin{bmatrix} Y & S \end{bmatrix} = X \begin{bmatrix} A & A + \Delta \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix}, \tag{18}$$

where $Y = \begin{bmatrix} y_1 & \ldots & y_T \end{bmatrix}'$, $S = \begin{bmatrix} s_1 & \ldots & s_T \end{bmatrix}'$, $X = \begin{bmatrix} x_1 & \ldots & x_T \end{bmatrix}'$, $x_t = \begin{bmatrix} 1 & y_{t-1}' & \ldots & y_{t-p}' \end{bmatrix}'$, $A' = \begin{bmatrix} a_0 & A_1 & \ldots & A_p \end{bmatrix}$, $\Delta' = \begin{bmatrix} \Delta_0 & \Delta_1 & \ldots & \Delta_p \end{bmatrix}$, $E = \begin{bmatrix} \varepsilon_1 & \ldots & \varepsilon_T \end{bmatrix}'$, and $H = \begin{bmatrix} \eta_1 & \ldots & \eta_T \end{bmatrix}'$.

Vectorizing the matrix representation column-by-column, we obtain

$$\underbrace{\begin{bmatrix} \text{vec}(Y) \\ \text{vec}(S) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X & O \\ I_M \otimes X & I_M \otimes X \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \text{vec}(A) \\ \text{vec}(\Delta) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H) \end{bmatrix}}_{\epsilon}, \tag{19}$$

where $\text{V}[\epsilon] = \Sigma \otimes I_T$. The vectorized representation has the structure of a multivariate linear seemingly unrelated regression (SUR) model, such that we can use standard results outlined e.g. in Geweke (2005, p.162ff) for its Bayesian estimation. Given our normal-diffuse prior of the form $p(\beta) \sim \mathcal{N}\left(\underline{\beta}\,, \underline{V}_\beta\right)$, $\Sigma \propto |\Sigma|^{-2(2M+1)/2}$, the full conditional posterior of $\beta$ is

$$p(\beta|Y_T, \Sigma) \sim \mathcal{N}\left(\bar{\beta}\,, \bar{V}_\beta\right), \tag{20}$$

with

$$\bar{V}_\beta = \left(\underline{V}_\beta^{-1} + Z'(\Sigma^{-1} \otimes I_T)Z\right)^{-1}$$
$$\bar{\beta} = \bar{V}_\beta\left(\underline{V}_\beta^{-1}\underline{\beta} + Z'(\Sigma^{-1} \otimes I_T)y\right).$$

And the full conditional posterior of $\Sigma$ is

$$p(\Sigma|Y_T, \beta) \sim IW(T, U) \tag{21}$$

where $U = (\begin{bmatrix} Y & S \end{bmatrix} - X\begin{bmatrix} A & A+\Delta \end{bmatrix})'(\begin{bmatrix} Y & S \end{bmatrix} - X\begin{bmatrix} A & A+\Delta \end{bmatrix})$, and IW is the inverted Wishart distribution (see Bauwens et al., 1999, section A.2.6). Thus, we can use the Gibbs sampler to obtain draws from the posterior distribution, iterating between (20) and (21).

### A.3 Posterior distribution for augmented VAR with a democratic steady-state prior

Below we discuss the posterior distribution of the model of section 2.3. First, it proves helpful to put the model in matrix notation as

$$\begin{bmatrix} Y_\psi & S_{\psi^+} \end{bmatrix} = X_\psi \begin{bmatrix} \Lambda & \Lambda + \Delta_\Lambda \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix} \tag{22}$$

where $Y_\psi = \begin{bmatrix} y_1 - \psi & \dots & y_T - \psi \end{bmatrix}'$, $S_{\psi^+} = \begin{bmatrix} s_1 - \psi^+ & \dots & s_T - \psi^+ \end{bmatrix}'$, $X_\psi = \begin{bmatrix} x_{\psi 1} & \dots & x_{\psi T} \end{bmatrix}'$, $x_{\psi t} = \begin{bmatrix} y'_{t-1} - \psi' & \dots & y'_{t-p} - \psi' \end{bmatrix}'$, $\Lambda' = \begin{bmatrix} A_1 & \dots & A_p \end{bmatrix}$, $\Delta'_\Lambda = \begin{bmatrix} \Delta_1 & \dots & \Delta_p \end{bmatrix}$, $E = \begin{bmatrix} \varepsilon_1 & \dots & \varepsilon_T \end{bmatrix}'$, and $H = \begin{bmatrix} \eta_1 & \dots & \eta_T \end{bmatrix}'$. Vectorizing the matrix representation column-by-column, we obtain

$$\underbrace{\begin{bmatrix} \text{vec}(Y_\psi) \\ \text{vec}(S_{\psi^+}) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X_\psi & 0 \\ I_M \otimes X_\psi & I_M \otimes X_\psi \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \text{vec}(\Lambda) \\ \text{vec}(\Delta_\Lambda) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H) \end{bmatrix}}_{\epsilon}, \tag{23}$$

where as before $\text{V}[\epsilon] = \Sigma \otimes I_T$. Noting that conditional on $\psi$ and $\Delta_\psi$, the model has the structure of a multivariate linear seemingly unrelated regression (SUR) model, we can use standard results for its Bayesian estimation. The full conditional posterior of $\beta$ is

$$p(\beta | Y_T, \Sigma, \psi, \Delta_\psi) \sim \mathcal{N}\left(\bar{\beta}, \bar{V}_\beta\right), \tag{24}$$

with $\bar{V}_\beta = \underline{V}_\beta^{-1} + Z'(\Sigma^{-1} \otimes I_T)Z$ and $\bar{\beta} = \bar{V}_\beta \left(\underline{V}_\beta^{-1}\underline{\beta} + Z'(\Sigma^{-1} \otimes I_T)y\right)$. And the full conditional posterior of $\Sigma$ is

$$p(\Sigma | Y_T, \Lambda, \Delta_\Lambda, \psi, \Delta_\psi) \sim IW(T, U) \tag{25}$$

where $U = \left(\begin{bmatrix} Y_\psi & S_{\psi^+} \end{bmatrix} - X_\psi \begin{bmatrix} \Lambda & \Lambda + \Delta_\Lambda \end{bmatrix}\right)' \left(\begin{bmatrix} Y_\psi & S_{\psi^+} \end{bmatrix} - X_\psi \begin{bmatrix} \Lambda & \Lambda + \Delta_\Lambda \end{bmatrix}\right)$.

To obtain the full conditional posterior of $\Psi = \begin{bmatrix} \psi' & \Delta'_\psi \end{bmatrix}'$, we rewrite the steady-state representation of the VAR (13) as

$$\overbrace{y_t - \sum_{i=1}^p A_i y_{t-i}}^{y_t^d} = \overbrace{\left(I_M - \sum_{i=1}^p A_i\right)}^{\Pi_1} \psi + \varepsilon_t, \tag{26}$$

$$\underbrace{s_t - \sum_{i=1}^p (A_i + \Delta_i)\, y_{t-i}}_{s_t^d} = \underbrace{\left(I_M - \sum_{i=1}^p (A_i + \Delta_i)\right)}_{\Pi_2} \psi + I_M \Delta_\psi + \eta_t, \tag{27}$$

and put it into matrix notation as

$$\underbrace{\begin{bmatrix} y_1^d \\ s_1^d \\ \vdots \\ y_T^d \\ s_T^d \end{bmatrix}}_{y_d} = \underbrace{\begin{bmatrix} \Pi_1 & 0 \\ \Pi_2 & I_M \\ \vdots & \\ \Pi_1 & 0 \\ \Pi_2 & I_M \end{bmatrix}}_{Z_d} \Psi + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \eta_1 \\ \vdots \\ \varepsilon_1 \\ \eta_1 \end{bmatrix}}_{\epsilon_d}, \text{ where } \mathrm{V}[\epsilon_d] = I_T \otimes \Sigma. \tag{28}$$

Conditional on $\Lambda,\ \Delta_\Lambda,$ and $\Sigma$, equation (28) is a multivariate linear regression of $y_d$ on $Z_d$, such that the posterior distribution is obtained from standard SUR results as

$$p(\Psi|Y_T, \Sigma, \Lambda, \Delta_\Lambda) \sim N(\bar{\Psi}, \bar{V}_\Psi), \tag{29}$$

where $\bar{V}_\Psi = \left(\underline{V}_\Psi^{-1} + Z_d'(I_T \otimes \Sigma^{-1})Z_d\right)^{-1}$ and $\bar{\Psi} = \bar{V}_\Psi \left(\underline{V}_\Psi^{-1}\underline{\Psi} + Z_d'(I_T \otimes \Sigma^{-1})y_d\right)$.

To obtain draws from the posterior distribution, Villani (2009) suggests using a three-block Gibbs sampler that iterates between (24), (25) and (29).

### A.4 Estimation under limited availability of the survey nowcasts

Below, we outline the Bayesian estimation of the augmented VAR when survey nowcasts are available only for a subset of the variables in $y_t$ and when the time series for $y_t$ starts earlier than the nowcasts. We distinguish the cases of section 2.2 and 2.3 respectively.

#### A.4.1 Model without Wright's democratic steady-state prior (section 2.2)

In what follows, we derive a SUR representation of the model that acknowledges the two particularities of the data: First, we show how equation (8) changes when survey nowcasts are observed for a subset of the variables in $y_t$. For that purpose, we make the following definitions

$s_t :=$ partly observed $M \times 1$ vector of survey nowcasts in equation (8)

$s_t^* :=$ observed $N \times 1$ vector of survey nowcasts with $N < M$

$R :=$ $N \times M$ selection matrix obtained from $I_M$ by deleting the rows that refer to the elements of $y_t$, for which we do not observe nowcasts, such that $s_t^* = Rs_t$

Pre-multiplying the part of equation (8) that refers to $s_t$ by $R$, we obtain

$$\begin{bmatrix} y_t \\ s_t^* \end{bmatrix} = \begin{bmatrix} a_0 \\ Ra_0 + \Delta_0^* \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_i \\ RA_i + \Delta_i^* \end{bmatrix} y_{t-i} + \begin{bmatrix} \varepsilon_t \\ \eta_t^* \end{bmatrix}, \tag{30}$$

where we have replaced $\{R\Delta_0, R\Delta_1, \ldots, R\Delta_p\}$ with $\left\{\Delta_0^*, \Delta_1^*, \ldots, \Delta_p^*\right\}$ to highlight that we can only identify these reduced form parameters.

Next, we show how the SUR representation of equation (19) changes when the time series of survey nowcasts $s_t$ starts later. For that purpose, we denote the periods when only $y_t$ is observed by $t = 1, \ldots, t_1$. Accordingly, the periods when both $y_t$ and $s_t$ are observed are $t = t_1 + 1, \ldots, T$. As an intermediate step to the SUR representation, we consider the matrix representation of equation (18). Specifically, the part referring to the survey nowcasts becomes

$$S_{II}^* = X_{II} \left[ AR' + \Delta^* \right] + H_{II}^*, \tag{31}$$

where $S_{II}^* = \begin{bmatrix} s_{t_1+1}^* & \ldots & s_T^* \end{bmatrix}'$, $X_{II} = \begin{bmatrix} x_{t_1+1} & \ldots & x_T \end{bmatrix}'$, $\Delta^{*\prime} = \begin{bmatrix} \Delta_0^* & \Delta_1^* & \ldots & \Delta_p^* \end{bmatrix}$, and $H_{II}^* = \begin{bmatrix} \eta_{t_1+1}^* & \ldots & \eta_T^* \end{bmatrix}'$ (other elements are defined as before). By vectorizing, we obtain the SUR representation:

$$\underbrace{\begin{bmatrix} \text{vec}(Y) \\ \text{vec}(S_{II}^*) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X & O \\ R \otimes X_{II} & I_N \otimes X_{II} \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \text{vec}(A) \\ \text{vec}(\Delta^*) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H_{II}^*) \end{bmatrix}}_{\epsilon}, \tag{32}$$

with

$$\mathrm{V}[\epsilon] = \begin{bmatrix} \Sigma_Y \otimes I_{t_1} & 0 \\ 0 & \Sigma^* \otimes I_{T-t_1} \end{bmatrix} \tag{33}$$

where $\Sigma^* = V([\varepsilon_t' \quad \eta_t^{*\prime}]')$, and $\Sigma_Y = V(\varepsilon_t)$. Eventually, as the final step to the posterior, we adjust the prior distribution specified in section 2.3 by dropping the elements that refer to the unobserved parts of $s_t$.

The model specified through (32-33) has the structure of a multivariate linear seemingly unrelated regression (SUR) model, such that we can use (once again) standard results for its Bayesian estimation (see Geweke, 2005, p.162ff). Specifically, the full conditional posterior of $\beta$ is

$$p(\beta|Y_T, \Sigma) \sim \mathcal{N}\left(\bar{\beta}\,,\,\bar{V}_\beta\right), \tag{34}$$

with $\bar{V}_\beta = \left(\underline{V}_\beta^{-1} + Z'\mathrm{V}[\epsilon]^{-1}Z\right)^{-1}$ and $\bar{\beta} = \bar{V}_\beta\left(\underline{V}_\beta^{-1}\underline{\beta} + Z'\mathrm{V}[\epsilon]^{-1}y\right)$.

One complication arises with the full conditional posterior of the variance-covariance matrix (see Swamy and Mehta, 1975): It is not completely clear, how to use the time series of different length in computing the sample estimate of $\Sigma^*$. If we use for each of its elements the maximum number of observations available, $\Sigma^*$ is not guaranteed to be invertible (see Schmidt, 1977). We avoid the problem by estimating the parameter of the full conditional posterior of $\Sigma^*$ on the overlapping sample ($t = t_1 + 1, \ldots, T$), thus throwing away the observations of the earlier sub-sample in its computation. Specifically, as an approximation to the full conditional posterior of $\Sigma^*$ we use

$$p(\Sigma^*|Y_T, \beta) \sim IW(T - t_1, U_{T-t_1}) \tag{35}$$

where $U_{T-t_1} = \left(\begin{bmatrix} Y_{II} & S_{II}^* \end{bmatrix} - X_{II}\begin{bmatrix} A & AR' + \Delta^* \end{bmatrix}\right)' \left(\begin{bmatrix} Y_{II} & S_{II}^* \end{bmatrix} - X_{II}\begin{bmatrix} A & AR' + \Delta^* \end{bmatrix}\right)$, and $Y_{II}$ is the part of $Y$ referring to $t = t_1 + 1, \ldots, T$.

#### A.4.2 Model with Wright's democratic steady-state prior (section 2.3)

In appendix A.3, we have derived the full conditional posterior distributions of $\{\Lambda, \Delta_\Lambda\}$, $\{\psi, \Delta_\psi\}$ and $\Sigma$ respectively. Here, we adjust the SUR representations leading to the full conditional posteriors of $\{\Lambda, \Delta_\Lambda\}$ (eq. 23) and $\{\psi, \Delta_\psi\}$ (eq. 29) to the case, when survey nowcasts are observed for a subset of the variables of the VAR, and their observations start later. Given the SUR representations and an appropriately adjusted prior distribution (see Appendix A.4.1), the posterior is obtained from standard results (see Geweke, 2005, p.162ff). The same complications as in Appendix A.4.1 arise with respect to the posterior of $\Sigma^*$.

*I.* Full conditional posterior of $\{\Lambda, \Delta_\Lambda\}$:

*I.a.* Pre-multiplying the part of (13) that refers to $s_t$ by $R$, we obtain

$$\begin{bmatrix} y_t - \psi \\ s_t^* - \psi^{+*} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} A_i \\ RA_i + \Delta_i^* \end{bmatrix} (y_{t-i} - \psi) + \begin{bmatrix} \varepsilon_t \\ \eta_t^* \end{bmatrix}, \tag{36}$$

where $\psi^{+*}$ is the reduced form of $R\psi_+$ (other elements are defined as before).

*I.b.* As an intermediate step to the SUR representation, we consider the matrix representation of equation (22). Specifically, the part referring to the survey nowcasts becomes

$$S_{II}^{\psi^{+*}} = X_{II\psi} \left( \Lambda R' + \Delta_\Lambda^* \right) + H_{II}^*, \tag{37}$$

where $S_{II}^{\psi^{+*}} = \begin{bmatrix} s_{t_1+1}^* - \psi^{+*} & \dots & s_T^* - \psi^{+*} \end{bmatrix}'$, $X_{II\psi} = \begin{bmatrix} x_{\psi t_1} & \dots & x_{\psi T} \end{bmatrix}'$, $\Delta_\Lambda^{*\prime} = \begin{bmatrix} \Delta_1^* & \dots & \Delta_p^* \end{bmatrix}$, and $H_{II}^* = \begin{bmatrix} \eta_{t_1+1}^* & \dots & \eta_T^* \end{bmatrix}'$. By vectorizing, we see that the SUR representation (eq. 23) becomes

$$\underbrace{\begin{bmatrix} \text{vec}(Y_\psi) \\ \text{vec}(S_{II\psi^+}^*) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X_\psi & 0 \\ R \otimes X_{II\psi} & I_N \otimes X_{II\psi} \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \text{vec}(\Lambda) \\ \text{vec}(\Delta_\Lambda^*) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H_{II}^*) \end{bmatrix}}_{\epsilon}, \tag{38}$$

with

$$\text{V}[\epsilon] = \begin{bmatrix} \Sigma_Y \otimes I_{t_1} & 0 \\ 0 & \Sigma^* \otimes I_{T-t_1} \end{bmatrix}. \tag{39}$$

The full conditional posterior is obtained as in (34).

*II.* Full conditional posterior of $\left\{\psi, \Delta_\psi^*\right\}$:

*II.a.* Pre-multiplying (eq. 27) by $R$, we obtain

$$\underbrace{s_t^* - \sum (RA_i + \Delta_i^*)\, y_{t-i}}_{s_t^{d*}} = \underbrace{\Big(R - \sum (RA_i + \Delta_i^*)\Big)}_{\Pi_2} \psi + I_M \Delta_\psi^* + \eta_t^*, \tag{40}$$

where $\Delta_\psi^*$ is the reduced form of $R\Delta_\psi$.

*II.b.* Accordingly, the SUR representation in the spirit of (28) - that allows for a standard Bayesian treatment as in (34) - becomes

$$\underbrace{\begin{bmatrix} y_1^d \\ \vdots \\ y_{t_1}^d \\ y_{t_1}^d \\ s_{t_1}^d \\ \vdots \\ y_T^d \\ s_T^d \end{bmatrix}}_{y_d} = \underbrace{\begin{bmatrix} \Pi_1 & 0 \\ \vdots & \vdots \\ \Pi_1 & 0 \\ \Pi_1 & 0 \\ \Pi_2 & I_N \\ \vdots & \vdots \\ \Pi_1 & 0 \\ \Pi_2 & I_N \end{bmatrix}}_{Z_d} \underbrace{\begin{bmatrix} \psi \\ \Delta_\psi^* \end{bmatrix}}_{\Psi^*} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{t_1} \\ \varepsilon_{t_1+1} \\ \eta_{t_1+1}^* \\ \vdots \\ \varepsilon_T \\ \eta_T^* \end{bmatrix}}_{\epsilon_d}, \tag{41}$$

with

$$\mathrm{V}[\epsilon_d] = \begin{bmatrix} I_{t_1} \otimes \Sigma_Y & 0 \\ 0 & I_{T-t_1} \otimes \Sigma^* \end{bmatrix}. \tag{42}$$

For the posterior of $\Sigma^*$, essentially, we use the same approximation as in (35).

# B Figures & Tables

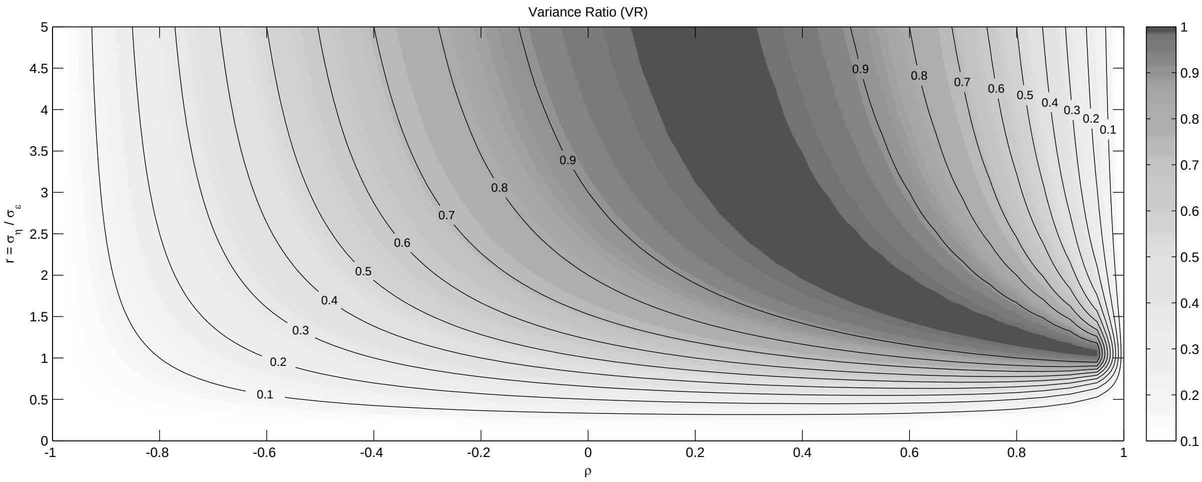


Figure 1: Variance Ratio VR$(r, \rho)$

Table 1: Data description and variable transformation

| | Variable ($x_t$) | Transformation used in VAR ($y_t$)[1] | Real-time data[2] (earliest vintage used) | Original data frequency[3] | Survey nowcasts starting in 1968:Q4[4] | Long-term forecasts w. 10y horizon |
|---|---|---|---|---|---|---|
| 1 | Real GDP | $100\left((x_t/x_{t-1})^4 - 1\right)$ | yes (1984:Q2) | Quarterly | yes | yearly fr. 1992:Q1 |
| 2 | GDP deflator | $400\,(\ln x_t - \ln x_{t-1})$ | yes (1984:Q2) | Quarterly | yes | - |
| 3 | CPI | $100\left((x_t/x_{t-1})^4 - 1\right)$ | yes (1994:Q3[5]) | Monthly | yes[6] | quarterly fr. 1991:Q4 |
| 4 | Industrial production | $400\,(\ln x_t - \ln x_{t-1})$ | yes (1984:Q2) | Monthly[7] | yes | - |
| 5 | Nonresidential fixed investment | $400\,(\ln x_t - \ln x_{t-1})$ | yes (1984:Q2) | Quarterly | - | - |
| 6 | Real personal consumption expenditures | $400\,(\ln x_t - \ln x_{t-1})$ | yes (1984:Q2) | Quarterly | - | - |
| 7 | Housing starts | $x_t$ | yes (1984:Q2) | Monthly[6] | yes | - |
| 8 | Unemployment rate | $x_t$ | yes (1984:Q2) | Monthly | yes | - |
| 9 | 10y Treasury bond yield[8] | $x_t$ | - | Daily | - | yearly fr. 1992:Q1 |
| 10 | 3m Treasury bill yield[8] | $x_t$ | - | Daily | - | yearly fr. 1992:Q1 |

*Notes:*
[1] We transform real GDP and CPI to *geometric* growth rates, because in the two cases the survey forecasts refer to growth rates instead of index levels. Thus, we make sure that the variables in the VAR and the survey nowcasts use identical definitions.
[2] Source of the real-time data: Philadelphia Federal Reserve Bank's Real-Time Data Set for Macroeconomists (http://www.phil.frb.org/research-and-data/real-time-center/real-time-data/, see Croushore and Stark, 2001).
[3] We obtain quarterly time series by averaging the monthly or daily observations. Note that growth rates are computed after averaging across the high-frequency observations in levels.
[4] Source of the Survey forecasts: Philadelphia Federal Reserve Bank's quarterly Survey of Professional Forecasters (SPF, http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/, see Croushore, 1993).
[5] We extract the appropriate data for earlier pseudo real-time periods from the 1994:Q3 vintage.
[6] CPI has only been part of the survey since 1981:Q3. We impute values based on a regression of the average CPI inflation nowcast on the average GDP deflator inflation nowcast and an intercept based on post 1981:Q2 data.
[7] Vintages are monthly, we extract the quarter-middle vintage.
[8] Downloaded from Thomson Reuters Datastream.

Table 2: Prior Specifications

| | $\boldsymbol{A_1, \ldots, A_p}$ $\left[p(A_i^{k,l}) \sim \mathcal{N}\left(\underline{A}_i^{k,l}, \frac{\lambda^2}{i^2}\frac{\sigma_k^2}{\sigma_l^2}\right)\right]$ | | | | $\boldsymbol{\Delta_1, \ldots, \Delta_p}$ $\left[p(\Delta_i^{k,l}) \sim \mathcal{N}\left(0, \zeta^2 \frac{\lambda^2}{i^2}\frac{\sigma_k^2}{\sigma_l^2}\right)\right]$ | |
|---|---|---|---|---|---|---|
| *Spec.* | $\underline{A}_i^{k,l}\vert_{i\neq 1}^{k\neq l\vee}$ | $\underline{A}_1^{k,k}$ | $\lambda$ | $Pr(\cdot) \approx 0.95$ | $\zeta$ | $Pr(\cdot) \approx 0.95$ |
| $M$ | 0 | $d_k{}^{\dagger}$ | 0.2 | $A_1^{k,k} \in [d_k \pm 0.4]$ | 1000 | $\Delta_1^{k,k} \in [\pm 400]$ |
| $W$ | ″ | ″ | ″ | ″ | ″ | ″ |
| $S$ | ″ | ″ | ″ | ″ | 0.100 | $\Delta_1^{k,k} \in [\pm 0.04]$ |
| $S2$ | ″ | ″ | ″ | ″ | 0.001 | $\Delta_1^{k,k} \in [\pm 0.0004]$ |

| | $\boldsymbol{\psi}$ $\left[p(\psi_j) \sim \mathcal{N}\left(\underline{\psi}_j, \lambda_0^2\right)\right]$ | | | $\boldsymbol{\Delta_\psi}$ $\left[p(\Delta_{\psi_j}) \sim \mathcal{N}\left(0, \zeta_0^2 \cdot \lambda_0^2\right)\right]$ | |
|---|---|---|---|---|---|
| *Spec.* | $\underline{\psi}_j$ | $\lambda_0$ | $Pr(\cdot) \approx 0.95$ | $\zeta_0$ | $Pr(\cdot) \approx 0.95$ |
| $M$ | 0 | $10^5$ | $\psi_k \in [\pm 2 \times 10^5]$ | 1 | $\Delta_{\psi_k} \in [\pm 2 \times 10^5]$ |
| $W$ | $l_k \mathbin{/\!/} 0^*$ | $0.5 \mathbin{/\!/} 10^{5*}$ | $\psi_k \in [l_k \pm 1] \mathbin{/\!/} [\pm 2 \times 10^5]^*$ | $2 \times 10^5$ | ″ |
| $S$ | ″ | ″ | ″ | 0.2 | $\Delta_{\psi_k} \in [\pm 0.2]$ |
| $S2$ | ″ | ″ | ″ | 0.001 | $\Delta_{\psi_k} \in [\pm 0.001]$ |

*Note:* The table presents the different prior specifications employed in the forecasting experiment of section 3. $l_k$ is the mean long-term forecast of variable $k$. The * indicates that for the variables and points in time, for which no long-term forecasts $l_k$ are available, we use the value after the // to specify the prior distribution.
† Following Wright (2013), we set $d_k = 0$ for each real variable (real GDP , nonresidential fixed investment, and real personal consumption expenditures), and $d_k = 0.8$ for the nominal variables.

Table 3: Forecasts with Different Prior Specifications:
Relative Mean Squared Forecast Errors (Evaluation Sample: 1984:Q2 - 2011:Q2)

| | MSFE | Relative MSFE to the OLS-AR1 | | | | |
|---|---|---|---|---|---|---|
| Horizon | OLS-AR1 | OLS-VAR | *M* | *W* | *S* | *S2* |
| | | | Real GDP growth | | | |
| $h = 1$ | 4.615 | 1.599** | 1.188 | 1.158 | 0.910 | 0.887 |
| $h = 4$ | 5.732 | 1.338** | 1.190 | 1.148 | 0.932 | 0.882 |
| $h = 8$ | 5.807 | 1.339** | 1.101 | 1.082 | 0.942 | 0.922 |
| $h = 12$ | 5.855 | 0.972 | 0.920 | 0.921 | 0.937 | 1.011 |
| | | | GDP deflator inflation | | | |
| $h = 1$ | 1.443 | 1.164 | 0.928 | 0.933 | 0.940 | 0.846* |
| $h = 4$ | 2.359 | 1.196 | 0.948 | 0.911 | 0.781*** | 0.625*** |
| $h = 8$ | 4.091 | 1.385 | 0.965 | 0.840 | 0.610*** | 0.438*** |
| $h = 12$ | 5.084 | 1.204 | 0.954 | 0.759 | 0.498*** | 0.332*** |
| | | | CPI inflation | | | |
| $h = 1$ | 5.021 | 1.095 | 0.919 | 0.919 | 0.835 | 0.772* |
| $h = 4$ | 6.218 | 1.388 | 1.059 | 1.055 | 0.818* | 0.734** |
| $h = 8$ | 6.893 | 1.586 | 1.128 | 1.039 | 0.716*** | 0.656*** |
| $h = 12$ | 7.411 | 1.295 | 1.037 | 0.895 | 0.604*** | 0.592*** |
| | | | IP growth | | | |
| $h = 1$ | 14.772 | 1.697** | 1.266* | 1.264* | 1.153 | 1.159 |
| $h = 4$ | 25.180 | 1.301** | 1.163 | 1.149 | 0.997 | 0.933 |
| $h = 8$ | 25.825 | 1.140 | 0.969 | 0.957 | 0.906* | 0.929** |
| $h = 12$ | 26.250 | 0.968 | 0.899 | 0.886 | 0.941* | 0.983 |
| | | | Three-month Treasury bill yield | | | |
| $h = 1$ | 0.252 | 2.091** | 1.072 | 1.092 | 1.150 | 1.128 |
| $h = 4$ | 2.301 | 1.345 | 0.994 | 1.014 | 0.990 | 0.901 |
| $h = 8$ | 5.744 | 1.118 | 0.984 | 0.987 | 0.850 | 0.742 |
| $h = 12$ | 7.942 | 1.071 | 1.022 | 1.004 | 0.756* | 0.568** |
| | | | Unemployment rate | | | |
| $h = 1$ | 0.092 | 0.819 | 0.606* | 0.612* | 0.598* | 0.644** |
| $h = 4$ | 0.885 | 0.930 | 0.834 | 0.850 | 0.766** | 0.774** |
| $h = 8$ | 2.433 | 1.023 | 0.951 | 0.969 | 0.818* | 0.747*** |
| $h = 12$ | 3.444 | 0.883 | 0.780*** | 0.794*** | 0.715*** | 0.685*** |

*Note:* The table reports results from a pseudo real-time out-of-sample forecasting experiment. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns titled 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the one/five/ten percent test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the ten-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications *M, W, S, S2* are outlined in table 2 and on page 8.

Table 4: Forecasting with Trained Hyperparameters:
Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)

| | MSFE | Relative MSFE to the OLS-AR1 | | | | | |
|---|---|---|---|---|---|---|---|
| Horizon | OLS-AR1 | OLS-VAR | $M$ | $W$ | $S$ | $S2$ | $T$ |
| | | | Real GDP growth | | | | |
| $h = 1$ | 5.249 | 1.458* | 1.186 | 1.154 | 0.873 | 0.835 | 0.830 |
| $h = 4$ | 5.988 | 1.446** | 1.291** | 1.243* | 1.014 | 0.924 | 0.890 |
| $h = 8$ | 6.187 | 1.331 | 1.161 | 1.140 | 1.024 | 0.946 | 0.930 |
| $h = 12$ | 6.367 | 0.904 | 0.895 | 0.894 | 0.985 | 0.961 | 0.957 |
| | | | GDP deflator inflation | | | | |
| $h = 1$ | 1.314 | 1.013 | 0.891 | 0.891 | 0.899 | 0.816* | 0.841* |
| $h = 4$ | 2.313 | 1.029 | 0.907 | 0.857 | 0.736** | 0.612*** | 0.690*** |
| $h = 8$ | 4.088 | 0.968 | 0.836 | 0.673 | 0.524*** | 0.413*** | 0.458*** |
| $h = 12$ | 5.422 | 0.940 | 0.875 | 0.621** | 0.446*** | 0.325*** | 0.372*** |
| | | | CPI inflation | | | | |
| $h = 1$ | 5.540 | 1.060 | 0.895 | 0.892 | 0.814 | 0.736* | 0.739* |
| $h = 4$ | 6.586 | 1.168 | 1.025 | 1.023 | 0.781 | 0.674*** | 0.685*** |
| $h = 8$ | 7.307 | 1.147 | 0.983 | 0.883 | 0.684** | 0.612*** | 0.634*** |
| $h = 12$ | 8.797 | 1.018 | 0.966 | 0.803 | 0.593*** | 0.529*** | 0.563*** |
| | | | IP growth | | | | |
| $h = 1$ | 17.752 | 1.167 | 1.129 | 1.129 | 1.029 | 1.025 | 1.080 |
| $h = 4$ | 27.972 | 1.243 | 1.186 | 1.166 | 1.022 | 0.937 | 0.932 |
| $h = 8$ | 27.980 | 1.163 | 1.026 | 1.011 | 0.965 | 0.951* | 0.942** |
| $h = 12$ | 29.185 | 0.911 | 0.866 | 0.853* | 0.956* | 0.971* | 0.965* |
| | | | Three-month Treasury bill yield | | | | |
| $h = 1$ | 0.242 | 1.641 | 0.992 | 1.007 | 1.026 | 0.951 | 0.847 |
| $h = 4$ | 2.465 | 1.103 | 0.851 | 0.875 | 0.855 | 0.743* | 0.736** |
| $h = 8$ | 6.099 | 0.948 | 0.824 | 0.829 | 0.694** | 0.585** | 0.627*** |
| $h = 12$ | 8.330 | 1.009 | 0.896 | 0.875 | 0.582*** | 0.434*** | 0.509*** |
| | | | Unemployment rate | | | | |
| $h = 1$ | 0.110 | 0.651 | 0.586* | 0.589* | 0.580* | 0.622** | 0.641** |
| $h = 4$ | 1.052 | 0.940 | 0.888 | 0.907 | 0.814* | 0.817** | 0.821** |
| $h = 8$ | 2.821 | 1.123 | 1.065 | 1.085 | 0.909 | 0.824** | 0.822** |
| $h = 12$ | 3.889 | 0.839** | 0.822*** | 0.838** | 0.769*** | 0.729*** | 0.760*** |

*Note:* The table reports results of a pseudo real-time out-of-sample forecasting experiment. Relative to table 3, to evaluate the different forecast methods, this experiment evaluates a smaller sample of forecasts produced from 1990:Q4 through 2011:Q2 in pseudo real-time. The reason is that the prior specification $T$ uses hyperparameters that have been trained on a sample extending from 1984:Q2 through 1990:Q3. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns titled 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the one/five/ten percent test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the ten-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications $M$, $W$, $S$, $S2$ are outlined in table 2 and on page 8.

Table 5: Comparing the Forecasts with both Survey Expectations and Combined Forecasts: Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)

| | MSFE | Relative MSFE to the OLS-AR1 | | | | |
|---|---|---|---|---|---|---|
| | | | | Combined Forecasts (VAR+Surveys) | | |
| Horizon | OLS-AR1 | VAR, specification $T$ | Survey Forecasts | Inv. MSFE | G.R. | $\frac{1}{N}$ |
| | | | Real GDP growth | | | |
| $h=1$ | 5.249 | 0.830 | 0.831 | 0.785 | 0.810 | 0.783 |
| $h=2$ | 6.148 | 0.849 | 0.848 | 0.814 | 0.847 | 0.813 |
| $h=3$ | 6.019 | 0.859 | 0.950 | 0.881 | 0.923 | 0.881 |
| $h=4$ | 5.988 | 0.890 | 1.006 | 0.925 | 0.937 | 0.929 |
| | | | GDP deflator inflation | | | |
| $h=1$ | 1.314 | $0.841^{*}$ | $0.592^{**}$ | $0.625^{**}$ | $0.602^{**}$ | $0.636^{**}$ |
| $h=2$ | 1.555 | $0.662^{***}$ | $0.533^{***}$ | $0.534^{***}$ | $0.534^{***}$ | $0.541^{***}$ |
| $h=3$ | 1.810 | $0.681^{***}$ | $0.525^{***}$ | $0.550^{***}$ | $0.506^{***}$ | $0.556^{***}$ |
| $h=4$ | 2.313 | $0.690^{***}$ | $0.495^{***}$ | $0.549^{***}$ | $0.483^{***}$ | $0.555^{***}$ |
| | | | CPI inflation | | | |
| $h=1$ | 5.540 | $0.739^{*}$ | 0.695 | $0.690^{*}$ | $0.710^{*}$ | $0.688^{*}$ |
| $h=2$ | 7.092 | $0.618^{**}$ | $0.567^{**}$ | $0.580^{**}$ | $0.579^{**}$ | $0.581^{**}$ |
| $h=3$ | 6.505 | $0.688^{**}$ | $0.625^{*}$ | $0.641^{**}$ | $0.636^{**}$ | $0.642^{**}$ |
| $h=4$ | 6.586 | $0.685^{***}$ | $0.622^{**}$ | $0.638^{***}$ | $0.626^{**}$ | $0.640^{***}$ |
| | | | IP growth | | | |
| $h=1$ | 17.752 | 1.080 | 1.161 | 1.079 | 1.073 | 1.081 |
| $h=2$ | 26.781 | 1.004 | 0.966 | 0.953 | 0.966 | 0.953 |
| $h=3$ | 27.284 | 0.931 | 1.009 | 0.948 | 0.963 | 0.948 |
| $h=4$ | 27.972 | 0.932 | 1.002 | 0.951 | 0.971 | 0.952 |
| | | | Three-month Treasury bill yield | | | |
| $h=1$ | 0.242 | 0.847 | 1.086 | $0.869^{*}$ | 0.897 | $0.873^{*}$ |
| $h=2$ | 0.796 | $0.814^{*}$ | 0.949 | $0.824^{***}$ | $0.868^{*}$ | $0.825^{***}$ |
| $h=3$ | 1.550 | $0.775^{**}$ | 0.956 | $0.820^{***}$ | $0.841^{**}$ | $0.825^{***}$ |
| $h=4$ | 2.465 | $0.736^{**}$ | 0.987 | $0.819^{**}$ | $0.793^{*}$ | $0.829^{**}$ |
| | | | Unemployment rate | | | |
| $h=1$ | 0.110 | $0.641^{**}$ | 0.897 | $0.664^{**}$ | $0.667^{*}$ | $0.702^{**}$ |
| $h=2$ | 0.336 | $0.703^{**}$ | 0.738 | $0.675^{**}$ | $0.728^{**}$ | $0.681^{**}$ |
| $h=3$ | 0.668 | $0.771^{**}$ | $0.768^{*}$ | $0.734^{**}$ | $0.815^{*}$ | $0.738^{**}$ |
| $h=4$ | 1.052 | $0.821^{**}$ | $0.824^{*}$ | $0.791^{***}$ | 0.898 | $0.795^{***}$ |

*Note:* The table reports results of a pseudo real-time out-of-sample forecasting experiment. Relative to table 3, this experiment uses a smaller sub-sample of realizations, which spans from 1990:Q4 through 2011:Q2, to evaluate the different methods. The reason is that the forecast combination methods and the trained VAR prior ('VAR, specification $T$') require a training sample. Note that the VAR prior is trained only once, whereas forecast combination weights are re-estimated recursively. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns title 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the one/five/ten percent test level.