JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **31**: 1083–1099 (2016)
Published online 26 August 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2483

FORECASTING WITH BAYESIAN VECTOR AUTOREGRESSIONS ESTIMATED USING PROFESSIONAL FORECASTS

CHRISTOPH FREY[a] AND FRIEDER MOKINSKI[b]*
[a] *Department of Economics, University of Konstanz, Germany*
[b] *Deutsche Bundesbank, Frankfurt am Main, Germany*

SUMMARY

We propose a Bayesian shrinkage approach for vector autoregressions (VARs) that uses short-term survey forecasts as an additional source of information about model parameters. In particular, we augment the vector of dependent variables by their survey nowcasts, and claim that each variable modelled in the VAR and its nowcast are likely to depend in a similar way on the lagged dependent variables. In an application to macroeconomic data, 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. Copyright © 2015 John Wiley & Sons, Ltd.

Received 20 February 2014; Revised 6 June 2015

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Vector autoregressions (VARs) are among the most popular tools in economic forecasting. VARs offer great flexibility in modelling the complex dynamic relations among macroeconomic variables, they are easy to estimate and can be used to generate forecasts at multiple horizons (see, for example, Stock and Watson, 2001). However, as even medium-sized VARs (10–20 variables) have several hundred parameters to estimate, potential overfitting 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.*'s (1984) Minnesota prior, which assumes that each variable evolves according to a random walk, and Wright's (2013) 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's (2013) 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 modelled in 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 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, modelled as a univariate autoregression (AR) with a single lag, i.e. $y_t = ay_{t-1} + \varepsilon_t$, and its nowcast s_t. The augmented model is

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

* Correspondence to: Frieder Mokinski, Deutsche Bundesbank, Wilhelm-Epstein-Straβe 14, 60431 Frankfurt am Main, Germany. E-mail: frieder.mokinski@googlemail.com

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

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

This prior implies $E[b] = \underline{a}$, i.e. we expect that y_t and s_t depend on y_{t-1} in the same way. Through $\underline{v}_\Delta$ we express our confidence in this conjecture. If the dependence of the survey nowcasts (s_t) on the lagged dependent variables is indeed not too dissimilar from the actuals (y_t), i.e. if $(b-a)$ is small, then 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 overfitting the model to the data and we therefore expect it to provide us with more accurate forecasts.[1] Note that our interpretation of shrinkage differs somewhat from the above-mentioned approaches: the Minnesota prior, for example, shrinks the coefficients of a VAR towards a system of univariate random walks. Thus shrinkage is directly provided through the prior. In our case, instead, shrinkage comes both from the prior and from the data: we shrink one set of unknown parameters (regression of survey nowcasts on lagged dependent variables) towards a second set of unknown parameters (regression coefficients of the VAR). Thus the method relies on 'learning' about the parameters of the original VAR from survey nowcasts.

In a forecasting application with US macroeconomic and macro-financial data, we find that a 10-variable VAR estimated using our novel shrinkage approach produces forecasts that are superior to a range of benchmark methods. Specifically, we find that MSFEs are typically lower with our method than with a univariate AR(1) estimated by ordinary least squares (OLS), uniformly lower than with the same VAR estimated using only the Minnesota prior, and comparable to those of survey forecasts.

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, online Appendix B.1 (supporting information) shows that the shrinkage target $\Delta = 0$ can alternatively be motivated from assumptions about the expectations formation process and about the time series model specification. Specifically, if (i) average 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 VAR is correctly specified, then the true value of Δ is zero. The fact that these ideal conditions are not likely to be fully satisfied in practice is one motivation to use $\Delta = 0$ as a shrinkage target instead of imposing it deterministically.

Similar approaches to incorporate information from survey nowcasts have been used in the frequentist estimation of a three-factor affine Gaussian model for US Treasury yields by Kim and Orphanides (2012), and in the Bayesian estimation of a dynamic stochastic general equilibrium (DSGE) model by Del Negro and Schorfheide (2013). However, besides the different model class, a major difference is that these studies 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.

Recently, a number of studies have used exponential tilting (Robertson *et al.*, 2005) to incorporate moment restrictions—e.g. from survey forecasts—into predictive densities obtained from macroeconomic time series models. Exponential tilting proceeds in the following way: from the universe of densities fulfilling the moment restrictions, it chooses the one closest in terms of relative entropy to the predictive density obtained from the time series model. Using this method, Cogley *et al.* (2005) have considered forecasting UK inflation with moment restrictions for the mean and variance taken from fan charts of the Bank of England. Alternatively, Altavilla *et al.* (2014) have used survey point

[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 mean squared forecast error (MSFE)).

DOI: 10.1002/jae

forecasts of short-term interest rates to adjust the forecasts of a dynamic Nelson–Siegel model of the US yield curve. Lately, Krüger *et al.* (2015) have employed moment restrictions which represent the mean and variance of survey nowcasts in order to modify the forecast density of a Bayesian VAR. Incorporating survey-based information through exponential tilting differs in a number of ways from our approach. First, it only exploits the survey data after model estimation. Thus, although such information is deemed informative, it is not used to learn about the data-generating process but only to modify the forecast density. Second, the method makes no attempt to evaluate empirically whether the moment restrictions (obtained from survey forecasts) it imposes are likely to hold in the data. Exponential tilting therefore relies heavily on carefully selecting the 'right' moment restrictions. Our method instead lets the data decide how informative survey forecasts are about the data-generating process. Eventually, exponential tilting is forecast-horizon specific, i.e. it can only be used to adjust forecasts at horizons for which moment restrictions are available. By contrast, in our method survey nowcasts are used to shrink coefficients of a time series model that can provide forecasts at any horizon.

Baştürk *et al.* (2014) present another approach for incorporating survey data into a forecasting model. Specifically, they estimate a New Keynesian Phillips Curve model, using inflation expectations to facilitate estimation of the expectations mechanism. A major difference from our approach is that they effectively include survey forecasts as a regressor, whereas we model survey nowcasts as a by-product of the data-generating process. Additionally, whereas their method is tailor made for inflation forecasting, ours can in principle be applied to any macroeconomic variable.

The paper proceeds as follows. Section 2 introduces the methodology and the underlying econometric ideas. Section 3 presents our empirical findings and Section 4 summarizes our results.

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 ε_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, η_t is another $M \times 1$ vector of disturbances and $\{b_0, B_1, \ldots, B_p\}$ 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 $\{y_{t-1}, \ldots, y_{t-p}\}$ as y_t itself, though they can have different coefficients. Estimating the augmented system (3) without imposing further restrictions on $\{b_0, B_1, \ldots B_p\}$, we will hardly reduce the risk of overfitting $\{a_0, A_1, \ldots, A_p\}$ to the data. By contrast, if we impose $\{b_0 = a_0, B_1 = A_1, \ldots, B_p = A_p\}$, provided that the restrictions are not *too incorrect*, this may help us to pin down the parameters

DOI: 10.1002/jae

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:

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

By standard maximum likelihood theory, the asymptotic distribution of the parameter estimate from the augmented model is

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

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

$$\sqrt{T}(\hat{a}_{\text{std}} - a) \overset{d}{\to} N(0, 1 - a^2) \tag{6}$$

Thus the ratio of the two asymptotic variances is

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

where $r = \sigma_\eta / \sigma_\varepsilon$ measures the imprecision of the survey nowcast as a signal about the conditional mean $\text{E}[y_t | y_{t-1}, \ldots, y_{t-p}]$, and ρ is the correlation between the two disturbances ε_t and η_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 ρ 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 ρ 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, online Appendix B.1 shows 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 equation (3) becomes

DOI: 10.1002/jae

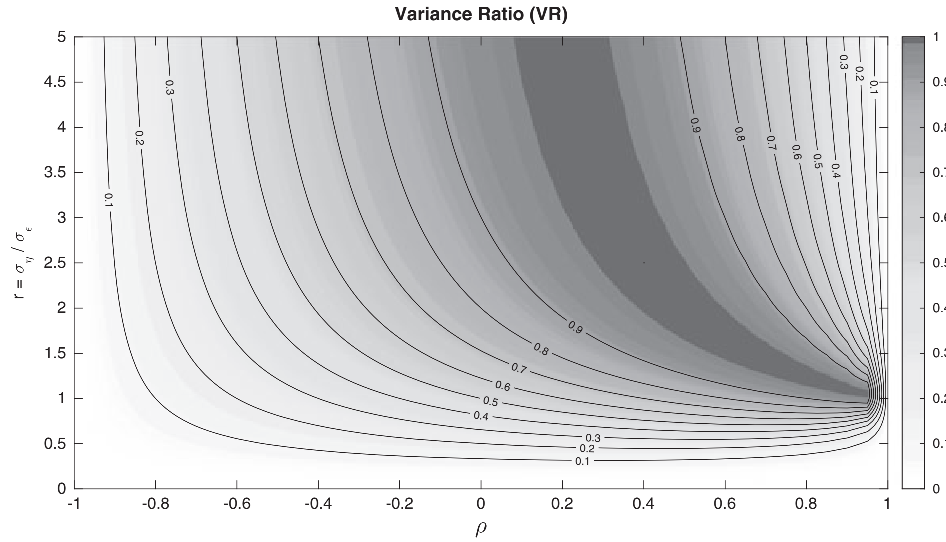

Figure 1. Variance ratio $\mathrm{VR}(r, \rho)$

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

and, for convenience, we assume $\begin{bmatrix} \varepsilon_t' & \eta_t' \end{bmatrix}' \sim N(0, \Sigma)$. Using the new parametrization, 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 Δ_i respectively, the marginal priors are

$$\begin{gathered} p(A_i^{k,l}) \sim N\left(\underline{A}_i^{k,l}, \lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right), \\ \text{with } \underline{A}_i^{k,l} = d_k \text{ if } k = l \wedge i = 1, \text{and } \underline{A}_i^{k,l} = 0 \text{ otherwise} \end{gathered} \tag{9}$$

$$p\left(\Delta_i^{k,l}\right) \sim N\left(0, \zeta^2\left(\lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right)\right) \tag{10}$$

$$p(a_0) \sim N(0, \kappa \cdot I_M) \tag{11}$$

$$p(\Delta_0) \sim N(0, \kappa \cdot I_M) \tag{12}$$

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 Σ 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 $\{A_1, \ldots, A_p\}$.

DOI: 10.1002/jae

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] In the specification of the prior variances in equation (9), the hyperparameter λ 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 σ_k^2/σ_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 σ_k^2 by the residual variance of an AR(1) regression for the kth variable.

The prior for $\{\Delta_0, \Delta_1, \ldots, \Delta_p\}$ is centred 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 Δ_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.[3] Details of the posterior distribution are given in Appendix A.1.

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:

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

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).

Parametrizing $\psi^+ = \psi + \Delta_\psi$, 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 Δ_i respectively and by ψ_k and Δ_ψ^k the kth entry of ψ and Δ_ψ, we set

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

[2]This contrasts with Doan *et al.* (1984), who have suggested a random walk prior with $d_1 = \ldots = d_M = 1$. Their specification makes perfect sense when time series are modelled in levels, but it is inappropriate for the stationary variables we consider (see, for example, Banbura *et al.*, 2010).

[3]Similar to the specification of the prior variance in the Minnesota prior, the $1/i^2$ term implies that survey respondents consider higher lags as less important. Following the advice of an anonymous referee, we have alternatively considered dropping the $1/i^2$ term, finding very similar results in terms of forecast accuracy in our empirical application.

DOI: 10.1002/jae

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

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

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

Once again, provided 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 Δ_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 ψ and for the difference vector Δ_ψ. Following Wright (2013), we set the elements of $\underline{\psi}_j$ to the most recent average long-term survey forecasts.[4] The hyperparameter λ_0 governs the tightness of the prior for ψ and thus reflects how optimistic we are about the informativeness of the long-term forecasts. Eventually, ζ_0 expresses our confidence in the equality of ψ and ψ^+, where ψ^+ is the unconditional mean implied by the survey nowcasts. The specification is completed by assuming an independent diffuse prior for Σ, $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$. Details of the posterior distribution can be found in Appendix A.2.

3. EMPIRICAL APPLICATION

In this section, we evaluate the forecasts of a 10-variable quarterly VAR(4) that is estimated using our novel approach. As in Wright (2013), our model features eight US macroeconomic variables, and 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 I 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 produce point forecasts at horizons of one, four, eight and twelve quarters using the iterated approach outlined in Carriero *et al.* (2015). This approach implies drawing a simulated sample from the posterior predictive density and using the sample mean as a point forecast. To avoid throwing away time series information, we use an estimation window with atypical design: Whereas the time series of actuals (y_t) starts in the second quarter of 1962, the time series of survey nowcasts (s_t) only begins in the fourth quarter of 1968. An additional complication arises from the fact that our vector of survey nowcasts (s_t) comprises only six of the ten variables included in the VAR.[5] Online Appendix B.2 modifies our approach to this setting.

In what follows, we try to discern the impact of the different sets of non-sample information by considering alternative specifications of the prior given in equations (14–17) of Section 2.3. Table II shows the details: specification *M* has the structure of Doan *et al.*'s (1984) Minnesota prior and ignores all survey information. *W* adds Wright's (2013) democratic steady-state prior and thus additionally exploits the long-run survey forecasts. *S* extends *W* by using the non-sample information provided

---

[4]For example, for the consumer price index (CPI) inflation rate we use the forecasts with a 10-year horizon collected by the Philadelphia Federal Reserve's Survey of Professional Forecasters.

[5]Note that the time series of survey nowcasts for CPI inflation only starts in 1981:Q3. To obtain pre-1981:Q3 survey nowcasts of CPI inflation, we use an imputation regression based on survey nowcasts of gross domestic product (GDP) deflator inflation. For details see Table I.

DOI: 10.1002/jae

Table I. Data description and variable transformation

	Variable (x_t)	Transformation used in VAR (y_t)[a]	Real-time data[b] (earliest vintage used)	Original data frequency[c]	Survey nowcasts starting in 1968:Q4[d]	Long-term forecasts with 10y horizon
1	Real GDP	$100\left((x_t/x_{t-1})^4 - 1\right)$	Yes (1984:Q2)	Quarterly	yes	yearly from 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[e])	Monthly	Yes[f]	quarterly from 1991:Q4
4	Industrial production	$400\,(\ln x_t - \ln x_{t-1})$	Yes (1984:Q2)	Monthly[g]	Yes	—
5	Non-residential 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[g]	Yes	—
8	Unemployment rate	x_t	Yes (1984:Q2)	Monthly	Yes	—
9	10y Treasury bond yield[h]	x_t	—	Daily	—	yearly from 1992:Q1
10	3m Treasury bill yield[h]	x_t	—	Daily	—	yearly from 1992:Q1

[a]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.

[b]Source of 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).

[c]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.

[d]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).

[e]We extract the appropriate data for earlier pseudo real-time periods from the 1994:Q3 vintage.

[f]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. Alternatively, acknowledging the real-time nature of our forecasting experiment, we have considered imputing the average GDP deflator nowcast without making an adjustment to it. The empirical results show only a very slight deterioration of the forecasts, suggesting that the imputation method plays only a minor role for forecast accuracy.

[g]Vintages are monthly; we extract the quarter-middle vintage.

[h]Downloaded from Thomson Reuters Datastream.

Table II. Prior specifications

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

	ψ $\left[p(\psi_j) \sim N\left(\underline{\psi}_j, \lambda_0^2\right)\right]$			Δ_ψ $\left[p(\Delta_{\psi_j}) \sim N\left(0, \zeta_0^2 \cdot \lambda_0^2\right)\right]$	
Spec.	$\underline{\psi}_j$	λ_0	$\Pr(\cdot) \approx 0.95$	ζ_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	$l_k // 0$*	$0.5 // 10^5$*	$\psi_k \in [l_k \pm 1] // [\pm 2 \times 10^5]$*	2×10^5	$\Delta_{\psi_k} \in [\pm 2 \times 10^5]$
S	$l_k // 0$*	$0.5 // 10^5$*	$\psi_k \in [l_k \pm 1] // [\pm 2 \times 10^5]$*	0.2	$\Delta_{\psi_k} \in [\pm 0.2]$
$S2$	$l_k // 0$*	$0.5 // 10^5$*	$\psi_k \in [l_k \pm 1] // [\pm 2 \times 10^5]$*	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. * 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.
[a]Following Wright (2013), we set $d_k = 0$ for each real variable (real GDP, non-residential fixed investment and real personal consumption expenditures), and $d_k = 0.8$ for the nominal variables.
[b]According to Carriero *et al.* (2015), this choice for the tightness parameter of the Minnesota prior is common in the Bayesian VAR forecasting literature. In a forecast experiment that uses macroeconomic data, the authors find that the optimal value of λ is close to 0.2.

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 3-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).[6]

Table III reports the results of the forecasting experiment. Its key message is that specifications S and $S2$ produce better forecasts for most variables and horizons than all the benchmarks we consider. This result highlights that it pays off in terms of forecast accuracy to exploit the additional information provided through the survey nowcasts. A few more points are notable in Table III. 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 overfitting. 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

[6]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.

Table III. 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
3-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). Asterisks denote rejection of the null hypothesis of equal predictive ability at the *10%/**5%/***1% test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the 10-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications *M, W, S, S2* are outlined in Table II and in the beginning of Section 3.

(relative to M) for the long-run inflation forecasts, but turns out to make little difference for the remaining variables and horizons.[7] 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% for GDP deflator inflation on the longest horizon) and for the unemployment rate. For real GDP growth, industrial production growth and the Treasury bill

[7] 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, which has collected long-term forecasts of all 10 variables twice a year since 1984, the SPF's 10-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 democratic steady-state prior than we do.

DOI: 10.1002/jae

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 whether 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).[8] While the test results support that the OLS-VAR(4) tends to produce inferior forecasts, the predictive ability of 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, the two specifications significantly improve over the AR(1) at longer forecast horizons for industrial production growth and the 3-month Treasury 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. smaller parameter values, typically implied better forecasting performance. Despite the promising result, a valid criticism is the arbitrary choice of hyperparameter values.

To address this concern, we have considered choosing the hyperparameters based on a training sample and evaluating the performance of this prior specification on a subsequent evaluation sample. Specifically, we think ourselves back to 1990:Q4. Using all data available at that time, we evaluate the pseudo out-of-sample forecasts obtained from each possible combination of the following hyperparameters values: $\lambda = \{0.01, 0.05, 0.1, 0.15, 0.2\}$, $\zeta = \{0.01, 0.1, 0.5, 1, 2, 10\}$, $\lambda_0 = 0.5$, $\zeta_0 = \{0.01, 0.1, 0.5, 1, 2, 10\}$, where we need to fix λ_0 because the data on long-term survey forecasts only start in 1991:Q4.[9] 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 III). 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 ζ and ζ_0, the two hyperparameters that relate to the survey nowcasts, but not with respect to λ, the hyperparameter that governs the tightness of the Minnesota prior. Based on prior specification T, we start generating real-time out-of-sample forecasts with the 1990:Q4 real-time data vintage. Table IV summarizes the results of the forecasting experiment: the four specifications M,

[8]Note that the test we use differs from the one employed by Wright (2013). Whereas we use a test for finite-sample predictive ability of alternative forecast methods (Giacomini and White, 2006), he uses a test for equal population level predictive ability that is suitable for nested forecast models (Clark and West, 2007). We prefer our test for two reasons. First, our test is far more demanding with respect to the extent of forecast improvement. The test used by Wright (2013) only requires the richer model to produce better forecasts at population level. It therefore ignores estimation uncertainty that is likely to deteriorate the forecasts of the richer model relative to the nested model in a finite-sample context. As the samples encountered in real-world macroeconomic forecasting applications are not even close to the population level, we consider the finite-sample context as more relevant. Second, the finite-sample test allows for comparisons among different forecasting methods, i.e. combinations of a forecasting model and an estimation strategy, whereas the population-level test can only distinguish models (because estimation is irrelevant at the population level).
Owing to the expanding estimation window, the asymptotics presented in Giacomini and White (2006) are not valid in our context. In favour of using the method with expanding estimation windows anyway, Clark and McCracken (2015) show in a simulation study that the test has reasonable size properties.
Note also that the use of real-time data may invalidate the asymptotics of tests for equal predictive ability such as the one we use; for details see Clark and McCracken (2009).

[9]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 dataset of survey long-run forecasts.

Table IV. 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 III, 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). Asterisks denote rejection of the null hypothesis of equal predictive ability at the *10%/**5%/***1% test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the 10-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications *M, W, S, S2* are outlined in Table II and in the beginning of Section 3.

W, 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

DOI: 10.1002/jae

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

Table V. 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	GR	$\frac{1}{N}$
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
3-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 III, this experiment uses a smaller subsample 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). Asterisks denote rejection of the null hypothesis of equal predictive ability at the *10%/**5%/***1% test level.

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 V 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 two 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* weights each forecast by 0.5.

The results are also found in Table V. 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 can be transferred to other types of time series models, such as the nonlinear class of vector STAR models (e.g. Schleer, 2015).

ACKNOWLEDGEMENTS

We would like to thank Luc Bauwens, Fabian Krüger, François Laisney 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. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

REFERENCES

Altavilla C, Giacomini R, Ragusa G. 2014. Anchoring the yield curve using survey expectations. *ECB working paper 1632*.

Baştürk N, Çakmakli C, Ceyhan SP, Van Dijk HK. 2014. Posterior-predictive evidence on us inflation using extended new Keynesian Phillips curve models with non-filtered data. *Journal of Applied Econometrics* **29**: 1164 –1182.

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: Oxford UK.
Carriero A, Clark TE, Marcellino M. 2015. Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics* **30**: 46–73.
Chauvet M, Potter S. 2013. Forecasting output. In *Handbook of Economic Forecasting*, Elliott G, Timmermann A (eds), Vol. 2. Elsevier: Amsterdam; 141–194.
Clark TE, McCracken MW. 2009. Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics* **27**: 441–454.
Clark TE, McCracken MW. 2015. Nested forecast model comparisons: a new approach to testing equal accuracy. *Journal of Econometrics* **186**: 160–177.
Clark TE, West KD. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* **138**: 291–311.
Cogley T, Morozov S, Sargent TJ. 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.
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*, Elliott G, Timmermann A (eds), Vol. 2. Elsevier: Amsterdam; 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*, Elliott G, Timmermann A (eds), Vol. 2. Elsevier: Amsterdam; 2–56.
Geweke J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley: Hoboken, NJ.
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 and 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.
Krüger F, Clark TE, Ravazzolo F. 2015. Using entropic tilting to combine BVAR forecasts with external nowcasts. Working paper 1439. Federal Reserve Bank of Cleveland.
Robertson JC, Tallman EW, Whiteman CH. 2005. Forecasting using relative entropy. *Journal of Money, Credit and Banking* **37**: 383–401.
Schleer F. 2015. Finding starting-values for the estimation of vector STAR models. *Econometrics* **3**: 65–90.
Stock JH, Watson MW. 2001. Vector autoregressions. *Journal of Economic Perspectives* **15**: 101–115.
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.

APPENDIX A.1: POSTERIOR DISTRIBUTION FOR THE AUGMENTED VAR

To derive the posterior distribution of the model of Section 2.2, we put equation (8) into matrix notation:

$$\begin{bmatrix} Y & S \end{bmatrix} = X \begin{bmatrix} A & A + \Delta \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix} \tag{A.1}$$

where $Y = \begin{bmatrix} y_1 & \dots & y_T \end{bmatrix}'$, $S = \begin{bmatrix} s_1 & \dots & s_T \end{bmatrix}'$, $X = \begin{bmatrix} x_1 & \dots & x_T \end{bmatrix}'$, $x_t = \begin{bmatrix} 1 & y'_{t-1} & \dots & y'_{t-p} \end{bmatrix}'$, $A' = \begin{bmatrix} a_0 & A_1 & \dots & A_p \end{bmatrix}$, $\Delta' = \begin{bmatrix} \Delta_0 & \Delta_1 & \dots & \Delta_p \end{bmatrix}$, $E = \begin{bmatrix} \varepsilon_1 & \dots & \varepsilon_T \end{bmatrix}'$, and $H = \begin{bmatrix} \eta_1 & \dots & \eta_T \end{bmatrix}'$. Vectorizing the matrix representation column by column, we obtain

$$\underbrace{\begin{bmatrix} \mathrm{vec}(Y) \\ \mathrm{vec}(S) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X & 0 \\ I_M \otimes X & I_M \otimes X \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \mathrm{vec}(A) \\ \mathrm{vec}(\Delta) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \mathrm{vec}(E) \\ \mathrm{vec}(H) \end{bmatrix}}_{\epsilon} \tag{A.2}$$

DOI: 10.1002/jae

where $\mathrm{V}[\epsilon] = \Sigma \otimes I_T$. The vectorized representation has the structure of a multivariate linear seemingly unrelated regression (SUR) model, such that we can use standard results outlined, for example, in Geweke (2005, p. 162ff.) for its Bayesian estimation.

Given our normal-diffuse prior of the form $p(\beta) \sim N(\underline{\beta}, \underline{V}_\beta)$, $\Sigma \propto |\Sigma|^{-2(2M+1)/2}$, the full conditional posterior of β is

$$p(\beta|Y_T, \Sigma) \sim N(\bar{\beta}, \bar{V}_\beta) \tag{A.3}$$

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

The full conditional posterior of Σ is

$$p(\Sigma|Y_T, \beta) \sim \mathcal{IW}(T, U) \tag{A.4}$$

where $U = \left(\begin{bmatrix} Y & S \end{bmatrix} - X\begin{bmatrix} A & A + \Delta \end{bmatrix}\right)' \left(\begin{bmatrix} Y & S \end{bmatrix} - X\begin{bmatrix} A & A + \Delta \end{bmatrix}\right)$ and $\mathcal{IW}$ 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 equations (A.3) and (A.4).

APPENDIX A.2: 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

$$\begin{bmatrix} Y_\psi & S_{\psi+} \end{bmatrix} = X_\psi \begin{bmatrix} \Lambda & \Lambda + \Delta_\Lambda \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix} \tag{A.5}$$

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

$$\underbrace{\begin{bmatrix} \mathrm{vec}(Y_\psi) \\ \mathrm{vec}(S_{\psi+}) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} I_M \otimes X_\psi & 0 \\ I_M \otimes X_\psi & I_M \otimes X_\psi \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \mathrm{vec}(\Lambda) \\ \mathrm{vec}(\Delta_\Lambda) \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \mathrm{vec}(E) \\ \mathrm{vec}(H) \end{bmatrix}}_{\epsilon} \tag{A.6}$$

where, as before, $\mathrm{V}[\epsilon] = \Sigma \otimes I_T$. Noting that, conditional on ψ and Δ_ψ, 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 β is

$$p(\beta|Y_T, \Sigma, \psi, \Delta_\psi) \sim N(\bar{\beta}, \bar{V}_\beta) \tag{A.7}$$

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

$$p(\Sigma|Y_T, \Lambda, \Delta_\Lambda, \psi, \Delta_\psi) \sim \mathcal{IW}(T, U) \tag{A.8}$$

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

DOI: 10.1002/jae

To obtain the full conditional posterior of $\Psi = \left[\psi' \;\; \Delta'_{\psi}\right]'$, we rewrite the steady-state representation of the VAR (13) as

$$\overbrace{y_t - \sum_{i=1}^{p} A_i y_{t-i}}^{y_t^d} = \overbrace{\left(I_M - \sum_{i=1}^{p} A_i\right)}^{\Pi_1} \psi + \varepsilon_t, \tag{A.9}$$

$$\underbrace{s_t - \sum_{i=1}^{p} (A_i + \Delta_i)\, y_{t-i}}_{s_t^d} = \underbrace{\left(I_M - \sum_{i=1}^{p} (A_i + \Delta_i)\right)}_{\Pi_2} \psi + I_M \Delta_{\psi} + \eta_t \tag{A.10}$$

and put it into matrix notation as

$$\underbrace{\begin{bmatrix} y_1^d \\ s_1^d \\ \vdots \\ y_T^d \\ s_T^d \end{bmatrix}}_{y_d} = \underbrace{\begin{bmatrix} \Pi_1 & 0 \\ \Pi_2 & I_M \\ \vdots & \\ \Pi_1 & 0 \\ \Pi_2 & I_M \end{bmatrix}}_{Z_d} \Psi + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \eta_1 \\ \vdots \\ \varepsilon_1 \\ \eta_1 \end{bmatrix}}_{\epsilon_d}, \text{ where } \mathrm{V}[\epsilon_d] = I_T \otimes \Sigma \tag{A.11}$$

Conditional on Λ, Δ_{Λ}, and Σ, equation (A.11) is a multivariate linear regression of y_d on Z_d, such that the posterior distribution is obtained from standard SUR results as

$$p(\Psi | Y_T, \Sigma, \Lambda, \Delta_{\Lambda}) \sim N(\bar{\Psi}, \bar{V}_{\Psi}) \tag{A.12}$$

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 equations (A.7), (A.8) and (A.12).

DOI: 10.1002/jae