

Weigt, Till; Wilfling, Bernd

Article — Published Version

An approach to increasing forecast-combination accuracy through VAR error modeling

Journal of Forecasting

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Weigt, Till; Wilfling, Bernd (2021) : An approach to increasing forecast-combination accuracy through VAR error modeling, Journal of Forecasting, ISSN 1099-131X, Wiley, Hoboken, NJ, Vol. 40, Iss. 4, pp. 686-699, <https://doi.org/10.1002/for.2733>

This Version is available at:

<https://hdl.handle.net/10419/233668>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

RESEARCH ARTICLE

WILEY

An approach to increasing forecast-combination accuracy through VAR error modeling

Till Weigt  | Bernd Wilfling 

Department of Economics (CQE),
Westfälische Wilhelms-Universität, Am
Stadtgraben 9, 48143 Münster, Germany

Correspondence

Bernd Wilfling, Westfälische Wilhelms-
Universität, Department of Economics
(CQE), Am Stadtgraben 9, 48143 Münster,
Germany.

Email: bernd.wilfling@wiwi.uni-
muenster.de

Abstract

We consider a situation in which the forecaster has available M individual forecasts of a univariate target variable. We propose a 3-step procedure designed to exploit the interrelationships among the M forecast-error series (estimated from a large time-varying parameter VAR model of the errors, using past observations) with the aim of obtaining more accurate predictions of future forecast errors. The refined future forecast-error predictions are then used to obtain M new individual forecasts that are adapted to the information from the estimated VAR. The adapted M individual forecasts are ultimately combined and any potential accuracy gains from the adapted combination forecasts analyzed. We evaluate our approach in an out-of-sample forecasting analysis, using a well-established 7-country data set on output growth. Our 3-step procedure yields substantial accuracy gains (in terms of loss reductions of up to 18%) for the simple average and three time-varying-parameter combination forecasts.

KEYWORDS

Bayesian VAR estimation, dynamic model averaging, forecast combinations, forgetting factors, large time-varying parameter VARs, state-space model

JEL CLASSIFICATION

C53; C32; C11

1 | INTRODUCTION

Starting with the seminal paper of Bates and Granger (1969), the combining of M individual forecasts to produce a pooled univariate forecast has become an established field of research. Hsiao and Wan (2014), *inter alia*, summarize the main arguments in favor of combining individual forecasts, and note that forecast combinations'... may be viewed as a way to make the forecast more robust against misspecification biases and measurement errors in the data set.'. Theoretical justification and

empirical evidence indicating a superior performance of forecast combinations to individual predictor and other forecasts are well documented in the literature (see Aiolfi & Timmermann, 2006; Baumeister & Kilian, 2015; Diebold & Pauly, 1987; Pesaran & Pick, 2011; Stock & Watson, 2004; Timmermann, 2006; and the literature cited therein).

From a statistical perspective, a plausible conjecture is that accuracy gains in the M individual forecasts should also improve the performance of forecast combinations obtained from them. Any feasible accuracy-

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Journal of Forecasting published by John Wiley & Sons Ltd

increasing methodology should be applied to the M individual forecasts prior to merging them into combination forecasts. In this paper, we propose such an approach, which is designed to exploit the probabilistic structure among the M individual forecast-error series (observed from past observations) with the objective of improving the forecast-accuracy of the M individual forecasts for future realizations. Our procedure consists of the following three steps (to be executed at each point in time). (i) We interrelate the forecast-error series of the M individual forecasts within a vector autoregressive (VAR) model and estimate these interrelationships from past observations. (ii) We use the information contained in the estimates to obtain more accurate predictions of the M future forecast errors. (iii) We adapt the original M individual forecasts to the refined error predictions from Step (ii), thus striving for a reduction in the future mean-squared error losses of the M individual forecasts. After executing this 3-step procedure, the adapted M individual forecasts can be combined, the forecast performance of the adapted combination(s) evaluated and compared with the losses of the corresponding combinations obtained from the original M individual forecasts.

The contributions of this paper are twofold. (i) We present the econometric concept behind our VAR forecast-error-modeling 3-step procedure. At this stage, we consider VAR modeling on the basis of the classic (covariance-stationary and stable) VAR(p) process. (ii) For selected forecast combinations, we apply our 3-step procedure to the established and standard 7-country data set on output growth, as introduced by Stock and Watson (2004). This G7 data set provides a benchmark setting, in which we benefit substantially from the authors' extensive data handling and preliminary work (detection of outliers, data transformations, construction of individual predictors, and so forth). This allows us to focus exclusively on analyzing the potential out-of-sample forecasting improvements associated with our approach.

Owing to the many individual forecast models provided by the G7 data set, our 3-step procedure involves the estimation of high-dimensional VARs. This suggests using the Bayesian methods of Koop and Korobilis (2013), designed to handle large time-varying parameter VARs.¹ At this point, we extend the classic VAR framework and include heteroscedastic, time-varying parameter VAR specifications in our forecast-error-modeling approach.

In the ensuing out-of-sample forecasting analysis, we consider two simple combination schemes: (i) the simple average (mean) combination forecast, and (ii) 3 distinct time-varying-parameter forecast combinations. This selection has two rationales. First, the latter combination schemes perform best in the Stock and Watson (2004) benchmark analysis. Second, there is some consensus that simple combination schemes (such as the mean with given, equal weights) are often hard to beat in practice (Palm & Zellner, 1992, p. 699; Rossi, 2013, p. 1213; Timmermann, 2006, pp. 181-182).

Our out-of-sample forecasting analysis yields two major findings. (i) In an idealized setting, our VAR-forecast-error-modeling procedure is able to produce an accuracy gain (aggregated over the entire data set) of 18% for the adapted mean combination over the original mean combination forecast (in terms of an 18% mean-squared-forecast-error reduction). However, since the improvement stems crucially from an *ex-post* perspective on the data, this sizable gain does not necessarily reflect realistic potential that is generally available to the forecaster. We rather interpret these 18% as an upper benchmark for the accuracy gain achievable for the G7 data set via our approach, when applied to the mean combination forecast. (ii) If we adopt a realistic *ex-ante* stance on the VAR specification in our 3-step procedure, we still obtain robust and substantial accuracy gains. For example, using a specific VAR model-selection criterion and/or rigidly standardized VAR specifications (with fixed lag-length), we still find a notable accuracy gain of 17.1% for the mean combination, and gains ranging between 5.7 and 18.0% for distinct time-varying-parameter combinations.

Our paper is organized as follows: Section 2 establishes our accuracy-increasing 3-step procedure. In Section 3, we briefly review the G7 data set, the relevant cornerstones of the Stock and Watson (2004) analysis, and the techniques of Koop and Korobilis (2013) for handling large time-varying parameter VARs. Section 4 contains the out-of-sample forecasting analysis. Section 5 concludes.

2 | VAR FORECAST ERROR MODELING (VAFEM)

For $t = 0, \pm 1, \pm 2, \dots$ we consider the univariate target variable y_t and for $h > 0$, we denote a forecast of y_{t+h} , based on information available at date t , by $\hat{y}_{t+h|t}$. We assume M given alternative forecast models and, associated with each model, the corresponding individual forecasts $\hat{y}_{t+h|t,1}, \dots, \hat{y}_{t+h|t,M}$, which we collect in the $M \times 1$ vector $\hat{\mathbf{y}}_{t+h|t} = (\hat{y}_{t+h|t,1}, \dots, \hat{y}_{t+h|t,M})'$. The information set at date t , \mathcal{F}_t , consists of (i) these M forecasts, (ii) their

¹Two alternative strategies for interrelating the forecast-error series are conceivable, but not discussed further in this paper: (i) via large VARs with shrinkage (e.g. George, Sun, & Ni, 2008; Korobilis, 2013), and (ii) via dynamic factor models (e.g. De Mol, Giannone, & Reichlin, 2008; Stock & Watson, 2011).

entire history, and (iii) the entire history of our target variable, i.e. $\mathcal{F}_t = \{\dots, \hat{y}_{t-1+h|t-1}, \hat{y}_{t+h|t}, \dots, y_{t-1}, y_t\}$. We collect the forecast errors $e_{t+h|i} = y_{t+h} - \hat{y}_{t+h|i}$ for $i = 1, \dots, M$, in the $M \times 1$ vector $\mathbf{e}_{t+h|t} = (e_{t+h|1}, \dots, e_{t+h|M})'$.

Our 3-step procedure starts with the assumption that the forecast-error vector $\mathbf{e}_{t+h|t}$ is governed by a covariance-stationary, stable VAR(p) process. In Step 1, we model the dynamics of the forecast errors as

$$\mathbf{e}_{t+h|t} = \boldsymbol{\nu} + \mathbf{A}_1 \mathbf{e}_{t+h-1|t-1} + \dots + \mathbf{A}_p \mathbf{e}_{t+h-p|t-p} + \epsilon_{t+h}, \quad (1)$$

where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)'$ is a vector of intercept terms, $\mathbf{A}_1, \dots, \mathbf{A}_p$ denote $M \times M$ parameter matrices, and $\epsilon_{t+h} = (\epsilon_{t+h,1}, \dots, \epsilon_{t+h,M})'$ represents an i.i.d. white noise process with a non-singular covariance matrix. We denote the i th $1 \times M$ row vector of the matrix \mathbf{A}_k (for $k = 1, \dots, p$) by $\mathbf{A}_{k,i}$ and the conditional expectation operator by $\mathbb{E}(\cdot | \cdot)$. Our goal is to find (consistently estimate) the optimal weights in Equation (1), which—under the mean-squared-error (MSE) loss function—provide solutions to the M separate minimization problems

$$\begin{aligned} & \left(\nu_i^*, \mathbf{A}_{1,i}^*, \dots, \mathbf{A}_{p,i}^* \right) = \\ & \arg \min_{\nu_i, \mathbf{A}_{1,i}, \dots, \mathbf{A}_{p,i}} \mathbb{E} \left\{ \left[\epsilon_{t+h,i}(\nu_i, \mathbf{A}_{1,i}, \dots, \mathbf{A}_{p,i}) \right]^2 | \mathcal{F}_t \right\} \end{aligned} \quad (2)$$

for $i = 1, \dots, M$.² From Equation (1), it follows directly that

$$\begin{aligned} & \mathbb{E} \left\{ \left[\epsilon_{t+h,i}(\nu_i^*, \mathbf{A}_{1,i}^*, \dots, \mathbf{A}_{p,i}^*) \right]^2 | \mathcal{F}_t \right\} \\ & \leq \mathbb{E} \left\{ e_{t+h|i}^2 | \mathcal{F}_t \right\} \text{ for } i = 1, \dots, M. \end{aligned} \quad (3)$$

Subsequently, we refer to our approach as Vector Autoregressive Forecast Error Modeling of order p [in symbols: VAFEM(p)].

Step 2 of our procedure consists of predicting the M VAFEM forecast errors in $\mathbf{e}_{t+h|t}$. Ideally, we would base these forecast-error predictions, which we collect in the $M \times 1$ vector $\hat{\mathbf{e}}_{t+h|t}$, on the optimal weights $\nu_i^*, \mathbf{A}_{1,i}^*, \dots, \mathbf{A}_{p,i}^*$ from the M minimization problems in Equation (2). However, since the optimal weights are unknown in practice, we replace them with their corresponding estimates from the VAR specification (1), and then use conventional VAR forecasting methods (recursive or direct forecasts) to obtain our predicted VAFEM forecast errors in $\hat{\mathbf{e}}_{t+h|t}$ (see our formal

description below). We note that the above-stated classic assumptions for $\mathbf{e}_{t+h|t}$ are simplifying at this stage. In Sections 3 and 4, we abandon these assumptions and consider time-varying parameter VARs.

In the final Step 3, we adapt the initial M individual forecasts in $\hat{y}_{t+h|t}$ by adding to them the predicted VAFEM errors $\hat{\mathbf{e}}_{t+h|t}$ from Step 2, i.e. we compute the adapted M individual forecasts as $\tilde{y}_{t+h|t} = \hat{y}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t}$. Under the classic assumptions and according to Equation (3), the MSEs of our adapted M individual forecasts in $\tilde{y}_{t+h|t}$ should not, on average, exceed the MSEs of their M initial counterparts in $\hat{y}_{t+h|t}$. Intuitively, these potential forecast accuracy gains (in terms of MSE reductions) may result from exploiting (i) systematic components in the history of the original M forecast errors in $\mathbf{e}_{t+h|t}$, and (ii) any correlation among these forecast errors, as modeled in the VAR system from Equation (1).

Ultimately, the adapted M individual forecasts in $\tilde{y}_{t+h|t}$ may be combined in the hope of achieving additional forecast accuracy. To this end, any conceivable forecast combination may be applied. (We refer to an overview article describing alternative combination schemes in Section 3.) However, in our application in Section 4, we confine our analysis to (i) the arithmetic mean combination, and (ii) three distinct time-varying-parameter combination forecasts.

To establish analytical expressions, we adopt the notation of Koop and Korobilis (2009, 2013) and define (i) the $M \times M(1 + p \cdot M)$ matrix

$$\mathbf{z}_t^{(h)} \equiv \begin{pmatrix} \left(\mathbf{z}_{1t}^{(h)} \right)' & 0 & \dots & 0 \\ 0 & \left(\mathbf{z}_{2t}^{(h)} \right)' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \left(\mathbf{z}_{Mt}^{(h)} \right)' \end{pmatrix}$$

with entries $\mathbf{z}_{it}^{(h)} \equiv (1, \mathbf{e}'_{t-1|t-h-1}, \dots, \mathbf{e}'_{t-p|t-h-p})'$ for $i = 1, \dots, M$, and (ii) the $M(1 + p \cdot M) \times 1$ vector $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_M)'$, where $\boldsymbol{\beta}_i \equiv (\nu_i, \mathbf{A}_{1,i}, \dots, \mathbf{A}_{p,i})$ for $i = 1, \dots, M$. With this notation, we rewrite Equation (1) in Step 1 of our VAFEM procedure as

$$\mathbf{e}_{t+h|t} = \mathbf{z}_{t+h}^{(h)} \boldsymbol{\beta} + \epsilon_{t+h}. \quad (4)$$

To formalize Step 2, we first consider the (theoretical) 1-step-ahead forecast, given by the conditional expectation

$$\mathbb{E} \left\{ \mathbf{e}_{t+1|t} | \mathcal{F}_t \right\} = \mathbf{z}_{t+1}^{(1)} \boldsymbol{\beta}. \quad (5)$$

²Under our assumptions, the optimal weights can be estimated consistently by the multivariate least squares estimator (Lütkepohl, 2006, pp. 69-72).

In principle, the (theoretical) h -step-ahead forecasts can be obtained via two alternative routes: either (i) by recursively applying Equation (5) h times, or (ii) directly, by regressing $\mathbf{e}_{t+h|t}$ on (measurable) variables at date t (and earlier) contained in $\mathbf{Z}_{t+1}^{(h)}$. In our empirical application below, we follow the second route, i.e. we consider the regression

$$\mathbf{e}_{t+h|t} = \mathbf{Z}_{t+1}^{(h)}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{t+h}. \quad (6)$$

Estimating $\boldsymbol{\beta}$ in Equation (6) with the techniques described in Section 3, we predict the VAFEM forecast errors by

$$\hat{\mathbf{e}}_{t+h|t} = \mathbf{Z}_{t+1}^{(h)}\hat{\boldsymbol{\beta}}. \quad (7)$$

In Step 3, we use the VAFEM error predictions from Equation (7) and compute our adapted M individual VAFEM forecasts as

$$\tilde{\mathbf{y}}_{t+h|t} = \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t} = \hat{\mathbf{y}}_{t+h|t} + \mathbf{Z}_{t+1}^{(h)}\hat{\boldsymbol{\beta}}, \quad (8)$$

which we then use as inputs in our forecast combinations.

3 | DATA AND LARGE VARS

3.1 | G7 data set, individual forecasts, combinations

Our aim is to apply the VAFEM procedure to the G7 data set provided by Stock and Watson (2004). The set covers quarterly data between 1959:I and 1999:IV for up to 43 time series, for each of the G7 countries Canada, France, Germany, Italy, Japan, the UK and the USA. A detailed list of the time series involved (including various asset prices, wages and prices, selected measures of real economic activity and the money stock) is compiled in Stock and Watson (2004, Table Ia). To cope with specific data characteristics (seasonality, outliers, stochastic trends), the authors provide adequate data transformations.

Besides the raw data, we also borrow individual forecast models and some combination forecasts from the original article. The target variable y_{t+h} represents output growth over the next h quarters (expressed at an annual rate) and is measured either in terms of real GDP or Industrial Production (IP). For each country, the i th individual forecast, $\hat{y}_{t+h|t,i}$, is based on a single country-specific predictor variable, say x_t , and is obtained using h -

step-ahead projections for the quarterly horizons $h = 1, 2, 4, 8$.³ Formally, the country's i th forecast is made by using the h -step-ahead regression model

$$y_{t+h} = \beta_0 + \beta_1(L)x_t + \beta_2(L)y_t + u_{t+h}, \quad (9)$$

where u_{t+h} represents the error term and $\beta_1(L), \beta_2(L)$ appropriately specified lag polynomials. Apart from these predictor-based individual forecasts, we additionally use a multistep autoregressive (AR) forecast, which—as in the original article—serves as the benchmark forecast in our application below.

Stock and Watson (2004) present a comprehensive collection of forecast combination schemes, which they subsume under the categories 'simple combination forecasts', 'discounted MSE forecasts', 'shrinkage forecasts', 'factor model forecasts', and 'time-varying-parameter (tvp) combination forecasts'. In our out-of-sample analysis, we consider only two types of time- t combination forecasts: (i) the mean $1/M \sum_{i=1}^M \hat{y}_{t+h|t,i}$ of the M individual forecasts, and (ii) a selection of time-varying-parameter combination forecasts. In order to compute the latter, the authors essentially apply a methodology suggested by Sessions and Chatterjee (1989) and LeSage and Magura (1992), but introduce the parameter $\phi \in \{0.1, 0.2, 0.4\}$ to control for the degree of time variation. We refer to these time-varying-parameter combination forecasts as tvp(0.1), tvp(0.2) and tvp(0.4).

3.2 | Large VARS

The inclusion of a large number of individual forecast models in our VAFEM approach necessarily involves the handling of high-dimensional VARs. Recently, large VARs containing more than 100 dependent variables have been analyzed with respect to estimation and forecasting issues (Banbura, Giannone, & Reichlin, 2010; Canova & Ciccarelli, 2009; Koop, 2013; Koop & Korobilis, 2016). Koop and Korobilis (2013) establish a computationally feasible Bayesian framework for large time-varying parameter VARs based on forgetting factors and Dynamic Model Averaging (DMA). We briefly review their methodology, in which the DMA part draws on technical details from Raftery, Kárný, and Ettlér (2010).

We consider the following state-space generalization of our VAFEM Equation (4),

$$\mathbf{e}_{t+h|t} = \mathbf{Z}_{t+h}^{(h)}\boldsymbol{\beta}_{t+h} + \boldsymbol{\epsilon}_{t+h}, \quad (10)$$

³Stock and Watson (2004) focus on the three forecast horizons $h = 2, 4, 8$. In order to include the short-term forecasting effects of our approach, we additionally consider $h = 1$.

$$\beta_{t+h+1} = \beta_{t+h} + \mathbf{u}_{t+h+1}, \quad (11)$$

where (i) ϵ_{t+h} and \mathbf{u}_{t+h} are i.i.d. $N(\mathbf{0}, \Sigma_{t+h})$ and $N(\mathbf{0}, \mathbf{Q}_{t+h})$, respectively, and (ii) ϵ_t and \mathbf{u}_s are independent of one another for all s and t . Obviously, setting the $M(1+pM) \times M(1+pM)$ matrix $\mathbf{Q}_{t+h} \equiv \mathbf{0}$ in transition Equation (11), and $\Sigma_{t+h} \equiv \Sigma$ in observation Equation (10) renders the state-space representation equivalent to VAFEM Equation (4), in which the parameter vector β is constant. We refer to (i) the generalized state-space model as the heteroscedastic VAR with time-varying parameters, and (ii) the special case in VAFEM Equation (4) as the homoscedastic VAR with constant parameters.

The key idea is to estimate the VAR in a Bayesian framework involving the Kalman filter, but assuming simplifying covariance structures in Equations (10) and (11). The prior on the parameter vector β_{t+h} is a Minnesota prior with a large variance for the intercepts in ν , and only one fixed scalar $\gamma \in [0, 1]$ representing the precision of the coefficients in β_{t+h} . The Minnesota prior is defined as γ/r^2 , where r is the lag of the parameter. Given the information set $\mathbf{e}^{t+h-1|t-1} \equiv \{\mathbf{e}'_{t+h-1|t-1}, \dots, \mathbf{e}'_{t+h-p|t-p}\}$, the update and prediction steps in Kalman filtering are based on $\beta_{t+h-1} | \mathbf{e}^{t+h-1|t-1} \sim N(\beta_{t+h-1|t+h-1}, \mathbf{V}_{t+h-1|t+h-1})$ and $\beta_{t+h} | \mathbf{e}^{t+h-1|t-1} \sim N(\beta_{t+h|t+h-1}, \mathbf{V}_{t+h|t+h-1})$, respectively. The estimated covariance matrix in Equation (10) at date t is the lagged covariance matrix at time $t-1$, multiplied by a fixed scalar $\kappa \in [0, 1]$, plus the residual covariance matrix multiplied by $1-\kappa$, i.e. $\hat{\Sigma}_{t+h} = \kappa \hat{\Sigma}_{t+h-1} + (1-\kappa) \hat{\epsilon}_{t+h} \hat{\epsilon}'_{t+h}$. The covariance matrix in the Kalman filtering formulae is reduced to $\mathbf{V}_{t+h|t+h-1} = \frac{1}{\lambda} \mathbf{V}_{t+h-1|t+h-1}$, where the fixed scalar $\lambda \in (0, 1]$ is a so-called forgetting factor.

Theoretically, we could estimate β_{t+h} involving the Kalman filter by Markov Chain Monte Carlo (MCMC) methods (e.g. Primiceri, 2005). In that case, we would have to set priors on Σ_{t+h} and \mathbf{Q}_{t+h} , rather than using the simplifying covariance structures described above. Owing to the normal distributions specified in Equations (10) and (11), this would produce MSE-minimizing parameter estimates, and would thus correspond closely to our motivation of the VAFEM approach in Section 2. However, due to computational burdens arising in the MCMC estimation of high-dimensional VARS, we resort to the approximate (DMA) grid-approach from Koop and Korobilis (2013), which consists of repeating the Kalman filter for different values of the scalars γ , λ and κ . Each Kalman-filter repetition is viewed as a model on its own, and ultimately, all models thus

created are combined by DMA with model-specific weights that basically correspond to their predictive likelihoods from the Kalman filter. More precisely, the DMA procedure uses the predictive likelihoods raised to the power of a second forgetting factor, $\alpha \in [0, 1]$. Thus, for $\alpha = 1$, this reduces to (recursively estimated) Bayesian model averaging, whereas smaller α -values assign less weight to past predictive likelihoods.

4 | VAFEM COMBINATION FORECASTING RESULTS

In this section, we analyze forecasting accuracy gains, obtainable by our VAFEM procedure, for mean and tvp(\cdot)-combination forecasts when used to predict output growth in the G7 data set. We implemented the entire VAFEM procedure, including the Bayesian large-VAR estimation methodology, in R.

4.1 | Timing and in-sample VAR estimation

We adopt the in-sample/out-of-sample timing from Stock and Watson (2004), including their Formula (3) for computing recursive mean-squared-forecast errors (MSFEs). All observations prior to 1973:I are used for estimating the individual forecasting regressions in Equation (9). The computation of the original (non-VAFEM) individual forecasts starts in 1973:I, while the (pseudo) out-of-sample forecasts of the non-VAFEM combination forecasts are calculated from 1981:I+ h onwards. For our in-sample VAR estimation, we use a recursively expanding estimation window, which we initialize for 1973:I to 1981:I (shortest in-sample estimation window). Our adapted individual VAFEM forecasts, plus the VAFEM combination forecasts obtained from them, are then available from 1981:I+ h onwards.

For our (recursively expanding) Bayesian in-sample VAR estimation, we adopt the grid values from Koop and Korobilis (2013): $\alpha = 0.99, \gamma \in \{0.0001, 0.001, 0.005, 0.01, 0.05\}, \kappa \in \{0.94, 0.96, 0.98\}, \lambda \in \{0.97, 0.98, 0.99, 1\}$.⁴ In our estimation procedure, we encounter a technical problem, stemming from high forecast-error correlation among the individual forecast models. Typically, an 'excessively large' forecast-error correlation renders the inversion of the covariance matrix $\mathbf{V}_{t+h|t+h-1}$ in the Kalman filter infeasible. To enhance numerical stability, we reduce the

⁴The G7 data set and the data used in the Koop and Korobilis (2013) analysis are of a similar type and have the same data frequency (quarterly observations of macroeconomic variables).

number of individual forecast models at the outset of the VAFEM procedure via the following mechanism. (i) We compute pairwise forecast-error correlation coefficients among all original individual forecast models. (ii) We randomly eliminate from the analysis,

one (of the two) individual forecast models with a forecast-error correlation coefficient exceeding a certain threshold value. For the four distinct forecast-error correlation thresholds $CT \in \{0.85, 0.90, 0.95, 0.99\}$, Table 1 displays the number of individual forecast models that

TABLE 1 Number of individual forecast models

Correlation Threshold (CT)	Canada	France	Germany	Italy	Japan	UK	USA
Forecast horizon: $h = 2$, target variable: real GDP growth							
0.85	8	–	6	6	8	3	8
0.90	12	–	14	11	8	6	10
0.95	21	–	25	21	17	17	22
0.99	32	–	32	24	26	29	27
Total	41	–	41	41	33	31	63
Forecast horizon: $h = 4$, target variable: real GDP growth							
0.85	10	–	13	9	8	14	12
0.90	11	–	16	11	11	17	19
0.95	16	–	25	16	20	24	31
0.99	32	–	32	27	29	31	32
Total	40	–	39	31	33	31	63
Forecast horizon: $h = 8$, target variable: real GDP growth							
0.85	8	–	11	7	7	10	9
0.90	10	–	15	15	12	13	16
0.95	13	–	23	20	23	20	30
0.99	27	–	31	29	32	30	30
Total	40	–	39	31	33	31	63
Forecast horizon: $h = 2$, target variable: Industrial Production (IP) growth							
0.85	7	9	9	4	9	4	7
0.90	12	14	11	10	15	7	16
0.95	19	20	22	13	25	18	29
0.99	32	22	30	28	31	30	32
Total	41	41	41	31	33	31	63
Forecast horizon: $h = 4$, target variable: Industrial Production (IP) growth							
0.85	7	6	11	6	11	13	11
0.90	12	9	16	10	20	18	20
0.95	21	12	20	17	23	24	38
0.99	31	21	29	29	30	30	40
Total	40	25	39	31	33	31	63
Forecast horizon: $h = 8$, target variable: Industrial Production (IP) growth							
0.85	8	8	13	9	11	9	13
0.90	13	10	18	13	11	14	23
0.95	23	17	22	20	18	19	36
0.99	30	19	28	24	31	29	39
Total	40	25	39	31	33	31	63

Note: The first four rows of each block indicate the numbers of individual forecast models used in computing the VAFEM combination forecasts, after randomly eliminating one of the two individual forecast models with pairwise forecast-error correlation coefficients exceeding the thresholds 0.85, 0.90, 0.95, and 0.99, respectively. The row 'Total' indicates the number of all individual forecast models provided by the G7 data set.

remain in our VAFEM analysis after executing this model-reducing mechanism.

4.2 | Out-of-sample evaluation of VAFEM mean combinations

Tables 2–5 report the country-specific out-of-sample forecasting results grouped by the four forecast horizons $h = 1, 2, 4, 8$ quarters, and the two target variables (real GDP, Industrial Production [IP]). The four tables present results for 52 cases: four forecast horizons, seven countries, two target variables, except for France for which the real GDP series is too short. The country-specific out-of-sample forecast periods are given in the table headings. In Block (1), the tables display the root-mean-squared forecast errors of the benchmark autoregressive model (AR RMSFE, in decimal values of the h -period growth). The MSFEs of all other forecasts in Blocks (2)–(4) are expressed relative to their corresponding AR MSFE (the squared AR RMSFE). For $h = 2, 4, 8$ the entries in Blocks (1) and (2) were compiled from Stock and Watson ((2004), Tables II–VII). In Blocks (3) and (4), we report the MSFEs of distinct VAFEM mean combination forecasts. These differ in their in-sample VAR specifications, which were used to compute the adapted individual VAFEM forecasts in the vector $\tilde{\mathbf{y}}_{t+h|t}$ according to Equation (8).

In Block (3), we report the MSFEs of two VAFEM mean combination forecasts obtained from a rigidly standardized in-sample VAR specification with correlation threshold $CT = 0.85$ and VAR lag-length $p = 4$ for each of the 52 cases. We estimated two in-sample VAR variants: (i) a homoscedastic VAR with constant parameter vector β , and (ii) a heteroscedastic VAR with time-varying parameter vector β_t . We label the MSFE of a VAFEM mean combination forecast with \bullet -superscripts, whenever it outperforms its corresponding original, non-VAFEM mean combination forecast from Block (2) in terms of a percentage MSFE reduction exceeding 0% (*), 5% (**), or 10% (***). We also report statistical significance for the improvements of the VAFEM mean combinations on the non-VAFEM mean combination forecasts, on the basis of the modified Diebold-Mariano test (Harvey et al. 1997). Here, we use the subscripts *, **, *** to denote significant improvements at the 10, 5, and 1% levels, respectively.

Throughout the 52 cases analyzed in Block (3) of Tables 2–5, we find that under the constant-parameter in-sample VAR specification (lines ' β (homoscedastic VAR)'), the VAFEM mean combination outperforms its non-VAFEM counterpart from Block (2) in 21 out of 52 cases (40.4%). Among these 21 improvements, we

observe 4 '***', 2 '**', and 15 '***' MSFE reductions, where 5 improvements are significant (3 '***', 2 '**'). *Prima facie*, this first result appears modest. An initial and substantial improvement can be achieved when using a time-varying parameter in-sample VAR specification (lines ' β_t (heteroscedastic VAR)'). Here, the VAFEM mean combinations outperform their non-VAFEM counterparts in 26 out of 52 cases (50.0%), with 5 '***', 2 '**', and 19 '***' MSFE reductions. Among this total of 26 MSFE reductions, 9 improvements are significant (4 '***', 4 '**', 1 '*').

Table 6 provides an overview of VAFEM accuracy gains by means of an overall performance ranking involving all combination forecasts from Tables 2–5. Table 6 ranks the alternative (non-VAFEM and VAFEM) mean combination forecasts according to their overall average losses aggregated over the 52 cases.⁵ The reference combination forecast is the non-VAFEM mean (Rank 8) with average loss 0.607. The VAFEM mean combination using the rigid ($CT = 0.85, p = 4$)- β (homoscedastic, constant-parameter) in-sample VAR specification performs worse than the non-VAFEM mean, in terms of a 5.1% higher average loss (Rank 10). By contrast, the VAFEM mean combination induced by the ($CT = 0.85, p = 4$)- β_t (heteroscedastic, time-varying parameter) in-sample VAR specification outperforms the non-VAFEM mean in terms of a 2.5% accuracy gain (Rank 7).⁶

Table 6 addresses additional VAFEM mean combinations exhibiting intermediate and large accuracy improvements to the non-VAFEM mean (Ranks 1–6). We first focus on the VAFEM specifications given for the Ranks 1, 3, 5. The associated VAFEM mean combinations refer to Block (4) of Tables 2–5, where we report out-of-sample MSFEs of VAFEM means obtained from (CT, p)-flexibilized in-sample VAR specifications, each estimated as constant-parameter (β (homoscedastic VAR)) and time-varying parameter (β_t (heteroscedastic VAR)) variants. More precisely, for all 52 cases and both VAR-parameter variants, we disclose the MSFEs of the best VAFEM mean combinations, obtained from *ex-post* searching for that tuple $(CT, p) \in \{0.85, 0.90, 0.95, 0.99\} \times \{1, 2, 3, 4\}$ producing the (case-specific) minimal out-of-sample MSFE. (The optimizing tuple—which is kept constant during the sample period—is displayed in the lines 'Specification (CT, p)' in Block (4) of Tables 2–5.) In Table 6, we denote these two in-sample VAR

⁵The concept of this ranking is from Stock and Watson (2004). The authors suggest computing the aggregated losses over all 52 cases as weighted averages of the single combination forecast losses, with weights equal to the inverse of the full-sample standard deviation of the two target variables (real GDP and Industrial Production).

⁶We note that no statistical tests for comparing the average losses from Table 6 are currently available in the literature.

TABLE 2 MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of one-quarter growth of real GDP and Industrial Production (IP) ($h = 1$)

Forecast Period	Canada 81:III – 98:IV	France 81:III – 98:IV	Germany 81:III – 98:IV	Italy 81:III – 98:IV	Japan 81:III – 98:IV	UK 81:III – 98:IV	USA 81:III – 98:IV
(1) Benchmark forecasts (Stock & Watson, 2004)							
AR RMSFE, real GDP	0.012	–	0.009	0.008	0.010	0.007	0.007
AR RMSFE, IP	0.022	0.013	0.017	0.018	0.016	0.012	0.013
(2) Non-VAFEM mean combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean, real GDP	0.94	–	0.96	1.01	0.98	0.96	0.92
mean, IP	0.98	1.06	0.94	0.98	1.01	0.99	0.86
(3) VAFEM mean combination forecasts: CT = 0.85, $p = 4$							
Real GDP:							
β (homoscedastic VAR)	0.93*	–	0.98	0.72***	0.85***	1.13	1.07
β_t (heteroscedastic VAR)	0.75***	–	1.01	0.52***	0.88***	1.07	0.96
IP:							
β (homoscedastic VAR)	1.07	1.18	1.06	0.87***	0.93**	1.22	1.09
β_t (heteroscedastic VAR)	0.98	1.22	1.02	0.87***	0.76***	1.17	1.04
(4) VAFEM mean combination forecasts: best (<i>ex post</i>) in-sample VAR specification							
Real GDP:							
Specification (CT, p)	(0.90, 2)	–	(0.99, 1)	(0.90, 1)	(0.99, 1)	(0.95, 3)	(0.95, 3)
β (homoscedastic VAR)	0.79***	–	0.81***	0.41***	0.79***	1.09	0.98
Specification (CT, p)	(0.95, 2)	–	(0.90, 3)	(0.95, 3)	(0.90, 3)	(0.85, 2)	(0.95, 3)
β_t (heteroscedastic VAR)	0.67***	–	0.91**	0.49***	0.78***	1.04	0.92
IP:							
Specification (CT, p)	(0.90, 1)	(0.99, 1)	(0.99, 2)	(0.90, 1)	(0.99, 2)	(0.99, 3)	(0.90, 2)
β (homoscedastic VAR)	1.03	1.00**	0.96	0.75***	0.60***	1.09	0.75***
Specification (CT, p)	(0.95, 2)	(0.90, 2)	(0.90, 3)	(0.95, 1)	(0.95, 2)	(0.85, 3)	(0.90, 4)
β_t (heteroscedastic VAR)	0.94*	1.04*	0.92*	0.70***	0.71***	1.05	0.90

Notes: (i) In Block (1), AR RMSFE denotes the root-mean-squared forecast error of the benchmark autoregressive model computed over the out-of-sample forecast periods, as indicated in the table heading. All other MSFEs in the table are expressed relative to the AR MSFEs. (ii) In Block (2), 'mean' denotes the simple average combination forecasts of the original M individual forecasts. These MSFEs were compiled from Stock and Watson ((2004), Table 2–5). (iii) The Blocks (3) and (4) contain the MSFEs of the VAFEM mean combination forecasts computed on the basis of Eq. (8). In Block (3), both in-sample VARs [constant- (homoscedastic) and time-varying-parameter (heteroscedastic) specifications] use the fixed correlation threshold $CT = 0.85$ and the VAR lag-length $p = 4$. Block (4) displays the MSFEs of the (*ex post*) best VAFEM mean combination forecasts. These are obtained by searching (*ex post*) for that particular tuple $(CT, p) \in \{0.85, 0.90, 0.95, 0.99\} \times \{1, 2, 3, 4\}$ in the constant- and time-varying-parameter in-sample VAR specifications, which produces minimal MSFEs. (iv) *, **, *** indicate that the VAFEM mean combination forecast outperforms its corresponding non-VAFEM mean combination forecast from Block (2) in terms of a percentage MSFE reduction of more than 0%, 5%, and 10%, respectively. We also test for significant improvements via the (modified) Diebold-Mariano test. *, **, *** denote significantly different MSFEs at the 10, 5, and 1% levels, respectively.

specifications as '(CT, p)- β (homoscedastic VAR)' and '(CT, p)- β_t (heteroscedastic VAR)'. The table shows that the (CT, p)- β VAFEM mean combination yields a 5.6% accuracy gain (Rank 5), compared to the non-VAFEM mean. The (CT, p)- β_t VAFEM mean combination outperforms the non-VAFEM mean combination forecast in terms of a 13.5% loss reduction (Rank 3). Additionally, we consider the VAFEM mean combination forecasts under those scenarios, in which—for each of the 52 cases—we select *ex post* the better of the two in-

sample VAR specifications (CT, p)- β and (CT, p)- β_t . Here, the VAFEM mean combinations outperform the non-VAFEM means by 18% (Rank 1).

We emphasize that the three VAFEM specifications given for the Ranks 1, 3, 5 of Table 6 are susceptible to hindsight-criticism. In practice, it appears infeasible *ex ante* to select that in-sample VAR specification ultimately producing the best VAFEM mean combination forecast. Therefore, we interpret the substantial accuracy gain of 18% (Rank 1) as an upper benchmark for the accuracy

TABLE 3 MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of two-quarter growth of real GDP and Industrial Production (IP) ($h = 2$)

Forecast Period	Canada 81:III – 98:IV	France 81:III – 98:IV	Germany 81:III – 98:IV	Italy 81:III – 98:IV	Japan 81:III – 98:IV	UK 81:III – 98:IV	USA 81:III – 98:IV
(1) Benchmark forecasts (Stock & Watson, 2004)							
AR RMSFE, real GDP	0.016	–	0.013	0.011	0.013	0.010	0.011
AR RMSFE, IP	0.031	0.018	0.026	0.028	0.026	0.018	0.019
(2) Non-VAFEM mean combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean, real GDP	0.90	–	0.92	0.99	0.99	0.95	0.95
mean, IP	0.93	1.08	0.90	1.00	1.02	0.98	0.89
(3) VAFEM mean combination forecasts: CT = 0.85, $p = 4$							
Real GDP:							
β (homoscedastic VAR)	0.76*** **	–	0.95	0.75*** ***	0.96*	1.07	1.05
β_t (heteroscedastic VAR)	0.75*** **	–	1.02	0.76*** ***	1.01	0.93*	0.99
IP:							
β (homoscedastic VAR)	0.90*	1.18	0.98	0.95**	0.88***	1.16	0.99
β_t (heteroscedastic VAR)	0.91*	1.24	1.03	0.96*	0.89***	1.20	0.93
(4) VAFEM mean combination forecasts: best (<i>ex post</i>) in-sample VAR specification							
Real GDP:							
Specification (CT, p)	(0.90, 1)	–	(0.99, 1)	(0.90, 1)	(0.99, 3)	(0.99, 4)	(0.99, 3)
β (homoscedastic VAR)	0.73*** **	–	0.89*	0.67*** ***	0.87***	0.95	0.98
Specification (CT, p)	(0.95, 2)	–	(0.85, 4)	(0.99, 1)	(0.95, 3)	(0.85, 2)	(0.95, 3)
β_t (heteroscedastic VAR)	0.73*** **	–	1.02	0.68*** ***	0.87***	0.86**	0.91*
IP:							
Specification (CT, p)	(0.85, 4)	(0.99, 1)	(0.99, 4)	(0.99, 2)	(0.99, 1)	(0.95, 3)	(0.99, 4)
β (homoscedastic VAR)	0.90*	0.98**	0.91	0.81*** *	0.72*** **	1.01	0.96
Specification (CT, p)	(0.99, 1)	(0.85, 1)	(0.90, 1)	(0.95, 1)	(0.99, 1)	(0.95, 2)	(0.85, 4)
β_t (heteroscedastic VAR)	0.85** *	1.09	1.00	0.85*** *	0.87*** *	0.95*	0.93

Notes: Analogous to the notes to Table 2.

gain in the mean combination forecast, achievable via our VAFEM approach for the G7 data set. In the next section, we reconsider this point and discuss in-sample VAR model selection issues within our VAFEM framework.

4.3 | VAR selection issues and $\text{tvp}(\cdot)$ -combination forecasts

The question of whether the benchmark VAFEM accuracy gain, obtained under hindsight, is realistically exploitable by the econometrician is closely related to providing practical criteria for in-sample VAR model selection. Within the VAFEM approach, the two crucial quantities are the correlation threshold CT and the VAR lag-length p , where the latter can be integrated naturally into the Bayesian estimation framework from Section 3.2.

More precisely, the DMA approach was represented as a grid over the parameters γ, λ, κ , resulting in a total of $\gamma \cdot \lambda \cdot \kappa$ models. Adding the lag-length to this grid increases the number of models in the DMA procedure by the factor p , which does not induce serious computational problems.

By contrast, the appropriate selection of the correlation threshold CT turns out to be far more problematic. According to Table 1, distinct CT-values generally lead to substantially differing numbers of individual forecast models to be included in the in-sample VAR. Koop and Korobilis (2013) execute DMA over different VAR dimensions, for which model averaging is based on the predictive likelihood of a set of variables included in a baseline VAR of minimal dimension. The forecasts of the baseline-set variables obtained from this VAR of minimal dimension are then averaged with their own forecasts

TABLE 4 MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of four-quarter growth of real GDP and Industrial Production (IP) ($h = 4$)

	Canada	France	Germany	Italy	Japan	UK	USA
Forecast Period	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV	82:I – 98:IV
(1) Benchmark forecasts (Stock & Watson, 2004)							
AR RMSFE, real GDP	0.025	–	0.018	0.019	0.023	0.018	0.016
AR RMSFE, IP	0.047	0.031	0.037	0.041	0.052	0.026	0.029
(2) Non-VAFEM mean combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean, real GDP	0.96	–	1.05	1.06	0.98	0.94	0.89
mean, IP	0.96	1.16	0.98	1.03	1.02	0.95	0.86
(3) VAFEM mean combination forecasts: CT = 0.85, $p = 4$							
Real GDP:							
β (homoscedastic VAR)	0.90***	–	1.19	0.52***	0.80***	1.17	1.03
β_t (heteroscedastic VAR)	0.73**	–	1.07	0.51***	0.83***	1.04	1.03
IP:							
β (homoscedastic VAR)	1.03	1.28	1.10	0.77***	0.83***	1.06	1.17
β_t (heteroscedastic VAR)	1.05	1.40	1.06	0.82***	0.87***	1.08	1.41
(4) VAFEM mean combination forecasts: best (<i>ex post</i>) in-sample VAR specification							
Real GDP:							
Specification (CT, p)	(0.90, 1)	–	(0.99, 1)	(0.90, 1)	(0.85, 3)	(0.90, 2)	(0.90, 3)
β (homoscedastic VAR)	0.83***	–	1.03*	0.46***	0.79***	1.08	0.99
Specification (CT, p)	(0.95, 3)	–	(0.85, 4)	(0.95, 3)	(0.85, 3)	(0.85, 4)	(0.90, 3)
β_t (heteroscedastic VAR)	0.72**	–	1.07	0.40***	0.82**	1.04	0.90
IP:							
Specification (CT, p)	(0.85, 4)	(0.99, 4)	(0.99, 1)	(0.95, 1)	(0.99, 1)	(0.85, 4)	(0.95, 2)
β (homoscedastic VAR)	1.03	1.01**	0.97*	0.74***	0.78**	1.06	0.96
Specification (CT, p)	(0.85, 4)	(0.95, 4)	(0.90, 4)	(0.90, 4)	(0.85, 2)	(0.95, 1)	(0.95, 4)
β_t (heteroscedastic VAR)	1.05	1.17	0.94*	0.77***	0.85**	0.90**	0.93

Notes: Analogous to the notes to Table 2.

obtained from VARs of higher dimensions. Ultimately, however, their procedure rests on several econometric technicalities (e.g. on the variable ordering) that cannot easily be applied to our setting. A straightforward alternative could consist of directly integrating our VAR variables (the forecast error series) into a DMA procedure. However, this leads rapidly to substantial computational challenges, since a number of M individual forecast series would imply 2^M model constellations to consider in DMA.⁷

Instead of directly integrating the parameters p and CT into our Bayesian estimation framework, a feasible alternative consists of using appropriate in-sample model-selection techniques. Subsequently, we apply two distinct selection procedures, namely (i) the Akaike information criterion (AIC), and (ii) the discounted mean-squared-forecast-error procedure (MSFE).⁸ To formalize,

consider the N distinct in-sample VAR specifications S_1, \dots, S_N (each representing a specific (CT, p)-tuple). (i) Let $\tilde{\mathbf{y}}_{S_j:t+h|t}$ be the vector of the adapted individual VAFEM forecasts from Equation (8), computed under the in-sample VAR specification S_j , and (ii) let $\text{mean}(\tilde{\mathbf{y}}_{S_j:t+h|t})$ denote the associated mean combination forecast. Furthermore, let (iii) T_0 be the date of the first available in-sample VAFEM forecast, (iv) $\delta \in (0, 1)$ denote a discount factor, and (v) y_t be—as before—the target variable (real GDP, IP).

⁷Recently, Onorante and Raftery (2016) proposed a heuristic approach to (practically) handling a large number of model combinations in DMA, which could be a promising avenue for future research.

⁸We also considered the Bayesian information criterion (BIC), which virtually produces identical results as AIC.

TABLE 5 MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of eight-quarter growth of real GDP and Industrial Production (IP) ($h = 8$)

Forecast Period	Canada 83:I – 97:IV	France 83:I – 97:IV	Germany 83:I – 97:IV	Italy 83:I – 97:IV	Japan 83:I – 97:IV	UK 83:I – 97:IV	USA 83:I – 97:IV
(1) Benchmark forecasts (Stock & Watson, 2004)							
AR RMSFE, real GDP	0.046	–	0.030	0.038	0.046	0.034	0.025
AR RMSFE, IP	0.070	0.050	0.054	0.059	0.111	0.041	0.042
(2) Non-VAFEM mean combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean, real GDP	1.00	–	1.05	0.96	0.99	1.06	0.98
mean, IP	1.00	1.04	0.99	0.98	0.99	1.03	0.89
(3) VAFEM mean combination forecasts: $CT = 0.85, p = 4$							
Real GDP:							
β (homoscedastic VAR)	1.12	–	1.13	0.30***	0.79***	1.66	1.55
β_t (heteroscedastic VAR)	0.71**	–	0.90***	0.32***	0.82***	1.38	0.92**
IP:							
β (homoscedastic VAR)	1.32	1.00*	1.07	0.87***	0.45***	1.89	1.18
β_t (heteroscedastic VAR)	0.99*	0.98**	0.97*	0.87***	0.45***	1.41	1.02
(4) VAFEM mean combination forecasts: best (<i>ex post</i>) in-sample VAR specification							
Real GDP:							
Specification (CT, p)	(0.85, 1)	–	(0.99, 1)	(0.85, 1)	(0.99, 1)	(0.85, 4)	(0.85, 1)
β (homoscedastic VAR)	0.94**	–	0.44***	0.30***	0.74***	1.66	1.12
Specification (CT, p)	(0.95, 4)	–	(0.90, 4)	(0.90, 4)	(0.85, 4)	(0.85, 4)	(0.90, 2)
β_t (heteroscedastic VAR)	0.65***	–	0.87***	0.23***	0.82***	1.38	0.90**
IP:							
Specification (CT, p)	(0.90, 2)	(0.85, 2)	(0.95, 4)	(0.99, 4)	(0.85, 4)	(0.85, 1)	(0.95, 1)
β (homoscedastic VAR)	1.25	0.99*	1.06	0.76**	0.45***	1.67	0.33***
Specification (CT, p)	(0.90, 4)	(0.95, 3)	(0.99, 1)	(0.95, 2)	(0.85, 4)	(0.99, 1)	(0.90, 2)
β_t (heteroscedastic VAR)	0.86***	0.81***	0.85**	0.59***	0.45***	0.96**	0.93

Notes: Analogous to the notes to Table 2.

For VAR specification S_j , we compute the AIC on the basis of the 'residuals' $\left\{ y_{s+h} - \text{mean}(\tilde{y}_{S_j:s+h|s}) \right\}_{s=T_0}^{t-h}$ up to date t , and denote this value by $\text{AIC}_{S_j:t}$. At date t , we then use the VAR specification $\text{argmin}_{S_j, j=1, \dots, N} \text{AIC}_{S_j:t}$ in computing future VAFEM forecasts for date $t+h$. In Table 6, we refer to this in-sample VAR selection method as 'AIC: $\{S_1, \dots, S_N\}$ '. Similarly, for VAR specification S_j , we may consider the discounted MSFE at date t (cf. Stock & Watson, 2004, Equation (4)),

$$\text{MSFE}_{S_j:t}(\delta) = \sum_{s=T_0}^{t-h} \delta^{t-h-s} \left[y_{s+h} - \text{mean}(\tilde{y}_{S_j:s+h|s}) \right]^2,$$

and then use the VAR specification $\text{argmin}_{S_j, j=1, \dots, N} \text{MSFE}_{S_j:t}(\delta)$ in computing future VAFEM forecasts for date $t+h$. In Table 6, we refer to this VAR selection method as 'MSFE(δ): $\{S_1, \dots, S_N\}$ ', where—in line with the

values suggested by Stock and Watson (2004)—we use the discount factor $\delta = 0.9$.

Prior to interpreting the results, we note the crucial difference between (i) the in-sample specifications selected via AIC/MSFE(δ), and (ii) all other in-sample VAFEM specifications from Table 6. Under the two selection criteria, the VAR specification at any in-sample date t is chosen among $\{S_1, \dots, S_N\}$. By contrast, all other VAFEM specifications are based on (CT, p)-tupels that are kept constant during the entire in-sample period. Thus, theoretically, both in-sample selection criteria could produce larger accuracy gains than the *ex-post* VAFEM specification 'Best of (CT, p)- β_t and (CT, p)- β ' with the 18% accuracy gain (Rank 1, Table 6). However, Table 6 shows that AIC produces forecasting losses of 2.6, 5.9, and 6.9% (Ranks 9, 11, 12), compared with the non-VAFEM mean combination. By contrast, MSFE(0.9) creates substantial accuracy gains (Ranks

TABLE 6 Mean combination forecasts ranked by average losses: both output measures (real GDP, IP), all horizons ($h = 1, 2, 4, 8$)

Rank	VAFEM specification (of mean combination forecast)	VAFEM / Non-VAFEM	Average loss	Deviation from loss of non-VAFEM mean combination (in %)
1	Best of $(CT, p)\text{-}\beta_t$ and $(CT, p)\text{-}\beta$	VAFEM (<i>ex post</i>)	0.498	-18.0
2	MSFE(0.9): $\{(CT, p)\text{-}\beta_i\}$ (heteroscedastic VAR)	VAFEM	0.503	-17.1
3	$(CT, p)\text{-}\beta_t$ (heteroscedastic VAR)	VAFEM (<i>ex post</i>)	0.525	-13.5
4	MSFE(0.9): $\{(CT, p)\text{-}\beta_i\} \cup \{(CT, p)\text{-}\beta\}$	VAFEM	0.539	-11.2
5	$(CT, p)\text{-}\beta$ (homoscedastic VAR)	VAFEM (<i>ex post</i>)	0.573	-5.6
6	MSFE(0.9): $\{(CT, p)\text{-}\beta\}$ (homoscedastic VAR)	VAFEM	0.582	-4.1
7	$(CT = 0.85, p = 4)\text{-}\beta_t$ (heteroscedastic VAR)	VAFEM	0.592	-2.5
8	mean	Non-VAFEM	0.607	
9	AIC: $\{(CT, p)\text{-}\beta_i\}$ (homoscedastic VAR)	VAFEM	0.623	2.6
10	$(CT = 0.85, p = 4)\text{-}\beta$ (homoscedastic VAR)	VAFEM	0.638	5.1
11	AIC: $\{(CT, p)\text{-}\beta_i\}$ (heteroscedastic VAR)	VAFEM	0.643	5.9
12	AIC: $\{(CT, p)\text{-}\beta_i\} \cup \{(CT, p)\text{-}\beta\}$	VAFEM	0.649	6.9

Notes: (i) The average losses are weighted averages of the losses of the VAFEM and non-VAFEM combination forecasts across all countries, horizons and target variables (a total of 52 cases). The weighting is obtained from the inverse of the full-sample standard deviation of the target variable (real GDP, IP) being forecasted. (ii) The VAFEM mean combination forecasts are denoted according to their in-sample VAR specifications from Tables 2–5. $(CT = 0.85, p = 4)\text{-}\beta_t$ and $(CT = 0.85, p = 4)\text{-}\beta$ refer to the in-sample VAR specifications from Block (3) of Tables 2–5, $(CT, p)\text{-}\beta_t$ and $(CT, p)\text{-}\beta$ refer to those from Block (4) of Tables 2–5. 'mean' denotes the non-VAFEM mean combination forecasts from Block (2) of Tables 2–5. (iii) 'Best of $(CT, p)\text{-}\beta_t$ and $(CT, p)\text{-}\beta$ ' represents the VAFEM mean combination forecasts with the (*ex post*) smaller MSFE, when specifying the in-sample VAR either as $(CT, p)\text{-}\beta_t$ or $(CT, p)\text{-}\beta$. (iv) 'MSFE(0.9):' and 'AIC:' denote the Mean-Squared-Forecast-Error and the Akaike VAR selection criteria applied to the set of VAR models labeled after the ':', where $(CT, p) \in \{0.85, 0.90, 0.95, 0.99\} \times \{1, 2, 3, 4\}$.

2, 4, 6). In particular, when applied to the set of heteroscedastic VARs, MSFE(0.9) produces a 17.1% accuracy gain (Rank 2), and thus only falls slightly below the upper benchmark gain induced by the (infeasible) *ex post* 'Best of $(CT, p)\text{-}\beta_t$ and $(CT, p)\text{-}\beta$ ' VAFEM specification.

We report a final empirical result, demonstrating that our VAFEM approach has the potential to produce large accuracy-gains for combinations other than the mean forecast, even without 'optimizing' the in-sample VAR specification. Table 7 displays the average MSFEs of several non-VAFEM and VAFEM time-varying parameter forecast combinations (relative to autoregression) aggregated over all 52 cases of the G7 data set. Block (1) reports the results for the non-VAFEM $\text{tvp}(0.1)\text{-}$, $\text{tvp}(0.2)\text{-}$, and $\text{tvp}(0.4)\text{-}$ combination forecasts, which deviate slightly from the

values in Table XI from Stock and Watson (2004), due to our inclusion of $h = 1$. Blocks (2) and (3) display the average (aggregated) MSFEs of the corresponding VAFEM $\text{tvp}(\cdot)\text{-}$ forecast combinations under the rigid in-sample VAR specifications $(CT = 0.85, p = 4)\text{-}\beta$ and $(CT = 0.85, p = 4)\text{-}\beta_t$. Obviously, our VAFEM approach provides substantially larger aggregated accuracy gains for $\text{tvp}(\cdot)\text{-}$ combination forecasts than for the mean combinations under the same VAFEM specifications (cf. Ranks 7, 10 in Table 6). In particular, the homoscedastic $(CT = 0.85, p = 4)\text{-}\beta$ specification produces aggregated accuracy gains between 9.1% [$\text{tvp}(0.1)$] and 18.0% [$\text{tvp}(0.4)$]. Interestingly, the accuracy gains under the heteroscedastic $(CT = 0.85, p = 4)\text{-}\beta_t$ specification are lower, but—ranging between 5.7% [$\text{tvp}(0.1)$] and 16.2% [$\text{tvp}(0.4)$]]—remain substantial.

TABLE 7 Average (equally weighted) MSFEs of tvp(\cdot)-combination forecasts relative to autoregression over all 52 cases (forecast period: 90:III – 99:IV)

tvp(\cdot)-combination	Average MSFE	Deviation from MSFE of non-VAFEM tvp(\cdot)-combination (in %)
(1) Non-VAFEM tvp(\cdot)-combination forecasts, balanced panel (Stock & Watson, 2004)		
tvp(0.1)	0.88	
tvp(0.2)	0.98	
tvp(0.4)	1.11	
(2) VAFEM tvp(\cdot)-combination forecasts: In-sample VAR specification: (CT = 0.85, $p = 4$)- β (homoscedastic VAR)		
tvp(0.1)	0.80	-9.1
tvp(0.2)	0.81	-17.3
tvp(0.4)	0.91	-18.0
(3) VAFEM tvp(\cdot)-combination forecasts: In-sample VAR specification: (CT = 0.85, $p = 4$)- β_t (heteroscedastic VAR)		
tvp(0.1)	0.83	-5.7
tvp(0.2)	0.87	-11.2
tvp(0.4)	0.93	-16.2

Notes: (i) In Block (1), the MSFEs of the non-VAFEM time-varying-parameter combination forecasts of the original individual forecasts deviate slightly from the values in Stock and Watson ((2004), Table XI), due to our inclusion of $h = 1$. (ii) In Blocks (2) and (3), the in-sample VAR specifications are denoted as in Tables 2–6. (iii) The percentage deviations in Column 3 are computed as pairwise deviations between the MSFEs of the VAFEM tvp(ϕ)-combination forecasts from Blocks (2) and (3), and the MSFEs of the corresponding non-VAFEM tvp(ϕ)-combination forecasts from Block (1) with the same degree-of-time-variation $\phi \in \{0.1, 0.2, 0.4\}$.

5 | CONCLUDING REMARKS

We establish a new forecasting approach (VAFEM) aimed at reducing the future forecast errors of M individual forecast models by exploiting structural interrelationships among the M individual forecast-error series, estimated from past observations. We formally motivate and empirically implement a 3-step procedure that estimates the interrelationships among the M individual forecast-error processes as high-dimensional VARs (within a Bayesian framework using forgetting factors and DMA). The objective is to exploit these estimates to obtain refined predictions on future forecast errors, which we then use to obtain new individual forecasts, adapted to the information on the forecast-error interrelationships. These M adapted individual

forecasts can subsequently be merged into combination forecasts, and potential accuracy gains analyzed.

In the empirical part, we evaluate our VAFEM approach in an out-of-sample forecasting analysis using the G7 data set on output growth, as introduced to the forecast combination literature by Stock and Watson (2004). Focusing on two types of combination forecasts, the simple average and time-varying parameter combinations, we find substantial accuracy gains for both combination schemes. On the basis of the data, we argue that our 3-step procedure has the potential to reduce the (aggregated) MSFE of the mean combination forecast by up to 18%, depending on the selected in-sample VAR specification. For the time-varying parameter combinations, we find substantial MSFE reductions, ranging between 5.7 and 18% under simply-structured (non-optimized) in-sample VAR specifications.

Our analysis leaves a number of issues to be tackled in future research. A first important point concerns the lack of statistical tests for comparing the aggregated losses among alternative VAFEM and non-VAFEM combination forecasts from Tables 6 and 7 (see Footnote 6). These aggregated results appear particularly useful, since they may quantify overall VAFEM accuracy gains, collected over a broad range of characteristics (*inter alia*, over different forecast horizons). A second extension could consist of systematically investigating potential VAFEM accuracy gains in alternative data settings and for forecast combinations other than the mean and tvp(\cdot).

ACKNOWLEDGEMENTS

We are grateful to Siem Jan Koopman and an anonymous reviewer for their constructive and extensive comments, which greatly improved the paper. The usual disclaimer applies. Open access funding enabled and organized by Projekt DEAL.

ORCID

Till Weigt  <https://orcid.org/0000-0002-7819-6328>

Bernd Wilfling  <https://orcid.org/0000-0002-5332-6878>

REFERENCES

- Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135, 31–53.
- Banbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Business & Economic Statistics*, 25, 71–92.
- Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20, 451–468.
- Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, 33, 338–351.

- Canova, F., & Ciccarelli, M. (2009). Estimating multicountry VAR models. *International Economic Review*, 50, 929–959.
- De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146, 318–328.
- Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6, 21–40.
- George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142, 553–580.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281–291.
- Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178, 294–309.
- Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177–203.
- Koop, G., & Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267–358.
- Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177, 185–198.
- Koop, G., & Korobilis, D. (2016). Model uncertainty in Panel Vector Autoregressive models. *European Economic Review*, 81, 115–131.
- Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28, 204–230.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Heidelberg, Germany: Springer-Verlag Berlin.
- LeSage, J. P., & Magura, M. (1992). A mixture-model approach to combining forecasts. *Journal of Business and Economic Statistics*, 3, 445–452.
- Onorante, L., & Raftery, A. (2016). Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review*, 81, 2–14.
- Palm, F. C., & Zellner, A. (1992). To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting*, 11, 687–701.
- Pesaran, M. H., & Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29, 307–318.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72, 821–852.
- Raftery, A. E., Kárný, M., & Ettlér, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1), 52–66.
- Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G., & Timmermann, A. (Eds.), *Handbook of Economic Forecasting* (pp. 1203–1324), Vol. 2B. Amsterdam: Elsevier.
- Sessions, D. N., & Chatterjee, S. (1989). The combining of forecasts using recursive techniques with non-stationary weights. *Journal of Forecasting*, 8, 239–251.
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405–430.
- Stock, J. H., & Watson, M. W. (2011). Dynamic factor models. In Clements, M. P., & Hendry, D. (Eds.), *The Oxford Handbook of Economic Forecasting* (pp. 35–60). New York: Oxford University Press.
- Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C. W. J., & Timmermann, A. (Eds.), *Handbook of Economic Forecasting* (pp. 135–196), Vol. 1. Amsterdam: Elsevier.

AUTHOR BIOGRAPHIES

Till Weigt studied at Darmstadt University of Technology and graduated as Mathematician in 2012. Currently, he is research associate and PhD candidate at the Chair of Empirical Economics at the University of Münster (WWU). His research focuses on Econometric Time Series Analysis, Probabilistic Forecasting, and Bayesian Filtering.

Bernd Wilfling is Full Professor for Empirical Economics at the University of Münster (WWU). His research interests include Time Series Analysis, Financial Econometrics, Volatility Modeling and Forecasting, and International Financial Markets.

How to cite this article: Weigt T, Wilfling B. An approach to increasing forecast-combination accuracy through VAR error modeling. *Journal of Forecasting*. 2021;40:686–699. <https://doi.org/10.1002/for.2733>