Center for Quantitative Economics

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

Till Weigt[†] und Bernd Wilfling[†]

68/2018

[†] Department of Economics, University of Münster, Germany

wissen.leben
WWU Münster

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

Till Weigt [a], Bernd Wilfling [a*]

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

(Date of this version: February 14, 2018)

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 of 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 ranging between 5.1% up to 18%) for the simple average and three time-varying-parameter combination forecasts.

Keywords: Forecast combinations, large time-varying parameter VARs, Bayesian VAR estimation, state-space model, forgetting factors, dynamic model averaging.

JEL classification: C53, C32, C11

*Corresponding author. Tel.: +49 251 83 25040; fax: +49 251 83 25042.
E-mail addresses: till.weigt@wiwi.uni-muenster.de (T. Weigt), bernd.wilfling@wiwi.uni-muenster.de (B. Wilfling).

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 Diebold and Pauly, 1987; Stock and Watson, 2004; Aiolfi and Timmermann, 2006; Timmermann 2006; Pesaran and Pick, 2011; Baumeister and Kilian, 2015; 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-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 idea behind our VAR forecast-error-modeling 3-step procedure. At this stage, we consider

VAR modeling on the basis of the classical (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 from Koop and Korobilis (2013), designed to handle large time-varying parameter VARs.[1] At this point, we extend the classical 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 and Zellner 1992, p. 699; Timmermann 2006, pp. 181-182; Rossi, 2013, p. 1213).

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 17.9% for the adapted mean combination over the original mean combination forecast (in terms of a 17.9% mean-squared-forecast-error reduction). However, since the improvement stems crucially from an *ex-post* perspective on the data, this sizable gain does not reflect realistic potential that is generally available to the forecaster. We rather interpret these $\approx 18\%$ as an upper bound 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

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

the VAR specification in our 3-step procedure, we still obtain robust and substantial accuracy gains. For example, under a rigidly standardized VAR specification (with fixed lag-length) we still find a notable accuracy gain of 5.1% for the mean combination, and even considerably larger gains ranging between 7.7 and 17.6% 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 from 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 $t = 0, \pm 1, \pm 2, \ldots$ 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 $\widehat{y}_{t+h|t}$. We assume given M alternative forecast models and, associated with each model, the corresponding individual forecasts $\widehat{y}_{t+h|t,1}, \ldots, \widehat{y}_{t+h|t,M}$, which we collect in the $M \times 1$ vector $\widehat{\mathbf{y}}_{t+h|t} = (\widehat{y}_{t+h|t,1}, \ldots, \widehat{y}_{t+h|t,M})'$. The information set at date t, $\mathcal{F}_t$, consists of (i) these M forecasts, (ii) the entire history of these M forecasts, and (iii) the entire history of our time-series variable, i.e. $\mathcal{F}_t = \{\ldots, \widehat{\mathbf{y}}_{t-1+h|t-1}, \widehat{\mathbf{y}}_{t+h|t}, \ldots, y_{t-1}, y_t\}$. We collect the forecast errors $e_{t+h|t,i} = y_{t+h} - \widehat{y}_{t+h|t,i}$ for $i = 1, \ldots, M$, in the $M \times 1$ vector $\mathbf{e}_{t+h|t} = (e_{t+h|t,1}, \ldots, e_{t+h|t,M})'$.

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

$$\mathbf{e}_{t+h|t} = \boldsymbol{\nu} + \mathbf{A}_1 \mathbf{e}_{t+h-1|t-1} + \ldots + \mathbf{A}_p \mathbf{e}_{t+h-p|t-p} + \boldsymbol{\epsilon}_{t+h}, \tag{1}$$

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

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

for $i = 1, \ldots M$.[2] From Eq. (1), it follows directly that

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

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 base these forecast-error predictions, which we collect in the $M \times 1$ vector $\widehat{\mathbf{e}}_{t+h|t}$, on the optimal weights $\nu_i^*, \mathbf{A}_{1,i}^*, \ldots, \mathbf{A}_{p,i}^*$ (and, accordingly, on their estimates) from the M minimization problems in Eq. (2).

In the final Step 3, we adapt the inital M individual forecasts in $\widehat{\mathbf{y}}_{t+h|t}$ by adding to them the predicted VAFEM errors $\widehat{\mathbf{e}}_{t+h|t}$ from Step 2, i.e. we compute the adapted M individual forecasts as $\widetilde{\widehat{\mathbf{y}}}_{t+h|t} = \widehat{\mathbf{y}}_{t+h|t} + \widehat{\mathbf{e}}_{t+h|t}$. According to Eq. (3), the MSEs of our adapted M individual forecasts in $\widetilde{\widehat{\mathbf{y}}}_{t+h|t}$ should not, on average, exceed the MSEs of their M initial counterparts in $\widehat{\mathbf{y}}_{t+h|t}$. We note that the above-stated classical 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.

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

$$\mathbf{Z}_{t+h} \equiv \begin{pmatrix} \mathbf{z}_{1t+h}' & 0 & \ldots & 0 \\ 0 & \mathbf{z}_{2t+h}' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{z}_{Mt+h}' \end{pmatrix}$$

with entries $\mathbf{z}_{it+h} \equiv (1, \mathbf{e}_{t+h-1|t-1}', \ldots, \mathbf{e}_{t+h-p|t-p}')'$ for $i = 1, \ldots, M$, and (ii) the $M(1+$

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

$p \cdot M) \times 1$ vector $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M)'$, where $\boldsymbol{\beta}_i \equiv (\nu_i, \mathbf{A}_{1,i}, \ldots, \mathbf{A}_{p,i})$ for $i = 1, \ldots, M$. With this notation, we rewrite Eq. (1) in Step 1 of our VAFEM procedure as

$$\mathbf{e}_{t+h|t} = \mathbf{Z}_{t+h}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{t+h}, \tag{4}$$

where we estimate $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$ with techniques described in Section 3.

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

$$\mathbb{E}\left\{\mathbf{e}_{t+1|t}|\mathcal{F}_t\right\} = \mathbf{Z}_{t+1}\boldsymbol{\beta}. \tag{5}$$

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

$$\mathbf{e}_{t+h|t} = \mathbf{Z}_{t+1}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{t+h}. \tag{6}$$

Using the estimates in $\widehat{\boldsymbol{\beta}}$ mentioned in Step 1, we predict the VAFEM forecast errors by

$$\widehat{\mathbf{e}}_{t+h|t} = \mathbf{Z}_{t+1}\widehat{\boldsymbol{\beta}}. \tag{7}$$

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

$$\widetilde{\widehat{\mathbf{y}}}_{t+h|t} = \widehat{\mathbf{y}}_{t+h|t} + \widehat{\mathbf{e}}_{t+h|t} = \widehat{\mathbf{y}}_{t+h|t} + \mathbf{Z}_{t+1}\widehat{\boldsymbol{\beta}}. \tag{8}$$

3 Data and large VARs

3.1 G7 data set, individual forecasts, combinations

Our goal 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 ith individual forecast, $\widehat{y}_{t+h|t,i}$, is based on a single country-specific predictor variable, say x_t, and is obtained using h-step-ahead projections for the quarterly horizons $h = 2, 4, 8$. Formally, the country's ith forecast is made by using the h-step-ahead regression model

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

where u_{t+h} represents the error term and $\beta_1(\mathrm{L}), \beta_2(\mathrm{L})$ 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 $1/M \sum_{i=1}^{M} \widehat{y}_{t+h|t,i}$ 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 $\mathrm{tvp}(0.1), \mathrm{tvp}(0.2)$ and $\mathrm{tvp}(0.4)$.

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 (Canova and Ciccarelli, 2009; Banbura et al., 2010; Koop, 2013; Koop and 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 et al. (2010).

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

$$\mathbf{e}_{t+h|t} = \mathbf{Z}_{t+h}\boldsymbol{\beta}_{t+h} + \boldsymbol{\epsilon}_{t+h}, \tag{10}$$

$$\boldsymbol{\beta}_{t+h+1} = \boldsymbol{\beta}_{t+h} + \mathbf{u}_{t+h+1}, \tag{11}$$

where (i) $\boldsymbol{\epsilon}_{t+h}$ and $\mathbf{u}_{t+h}$ are i.i.d. $N(\mathbf{0}, \boldsymbol{\Sigma}_{t+h})$ and $N(\mathbf{0}, \mathbf{Q}_{t+h})$, respectively, and (ii) $\boldsymbol{\epsilon}_t$ 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 Eq. (11), and $\boldsymbol{\Sigma}_{t+h} \equiv \boldsymbol{\Sigma}$ in observation Eq. (10) renders the state-space representation equivalent to VAFEM Eq. (4), in which the parameter vector $\boldsymbol{\beta}$ 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 Eq. (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 Eqs. (10) and (11). The prior on the parameter vector $\boldsymbol{\beta}_{t+h}$ is a Minnesota prior with a large variance for the intercepts in $\boldsymbol{\nu}$, and only one fixed scalar $\gamma \in [0, 1]$ representing the precision of the coefficients in $\boldsymbol{\beta}_{t+h}$. The Minnesota prior is defined as γ/r^2, where r is the lag of the parameter. Given the information set $\mathbf{e}^{t+h-1|t-1} \equiv \{\mathbf{e}'_{t+h-1|t-1}, \ldots, \mathbf{e}'_{t+h-p|t-p}\}$, the update and prediction steps in Kalman filtering are based on $\boldsymbol{\beta}_{t+h-1}|\mathbf{e}^{t+h-1|t-1} \sim N(\boldsymbol{\beta}_{t+h-1|t+h-1}, \mathbf{V}_{t+h-1|t+h-1})$ and $\boldsymbol{\beta}_{t+h}|\mathbf{e}^{t+h-1|t-1} \sim N(\boldsymbol{\beta}_{t+h|t+h-1}, \mathbf{V}_{t+h|t+h-1})$, respectively. The estimated covariance matrix in Eq. (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. $\widehat{\mathbf{\Sigma}}_{t+h} = \kappa\widehat{\mathbf{\Sigma}}_{t+h-1} + (1-\kappa)\widehat{\boldsymbol{\epsilon}}_{t+h}\widehat{\boldsymbol{\epsilon}}'_{t+h}$. 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 $\boldsymbol{\beta}_{t+h}$ 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 $\mathbf{\Sigma}_{t+h}$ and $\mathbf{Q}_{t+h}$, rather than using the simplifying covariance structures described above. Owing to the normal distributions specified in Eqs. (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 Eq. (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). Throughout the expanding window, we keep the VAR lag-length p constant [VAFEM(p)]. Our adapted individual VAFEM forecasts, plus the VAFEM combination forecasts obtained from them, are then available from 1981:I+h onwards.

Table 1 about here

For our (recursively expanding) Bayesian in-sample VAR estimation, we adopt the grid values from Koop and Korobilis (2013): $\alpha = 0.99, \gamma \in \{0.0001, 0.001, 0.005, 0.01, 0.05\}, \kappa \in \{0.94, 0.96, 0.98\}$, $\lambda \in \{0.97, 0.98, 0.99, 1\}$.[3] 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 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 remain in our VAFEM analysis after executing this model-reducing mechanism.

Table 2 about here

Table 3 about here

Table 4 about here

[3] The 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).

Table 5 about here

Table 6 about here

Table 7 about here

4.2 Out-of-sample evaluation of VAFEM mean combinations

Tables 2–7 report the country-specific out-of-sample forecasting results grouped by the two target variables (real GDP, Industrial Production) and the three forecast horizons $h = 2, 4, 8$ quarters. The six tables present results for 39 cases: seven countries, two target variables, three forecast horizons, 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). 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 $\widetilde{\widetilde{\mathbf{y}}}_{t+h|t}$ according to Eq. (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 39 cases. We estimated two in-sample VAR variants: (i) a homoscedastic VAR with constant parameter vector $\boldsymbol{\beta}$, and (ii) a heteroscedastic VAR with time-varying parameter vector $\boldsymbol{\beta}_t$. We label the MSFE of a VAFEM mean combination forecast with *'s, 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% (***).[4]

[4]It might seem appealing to report statistical significance for the improvements of the VAFEM mean combinations on the non-VAFEM mean combination forecasts. In principle, two types of statistical tests could be applied to each of the 39 cases separately. (i) We could define the non-VAFEM mean from Block (2) as the benchmark forecast against which we compare all VAFEM mean combination forecasts via the methods proposed in White (2000), and with Hansen's (2005) test for superior

Throughout the 39 cases analyzed in Block (3) of Tables 2–7, we find that under the constant-parameter in-sample VAR specification (line '$\boldsymbol{\beta}$ (homoscedastic VAR)') the VAFEM mean combination outperforms its non-VAFEM counterpart from Block (2) in 16 out of 39 cases (41,0%). Among these 16 improvements, we observe 3 '*', 1 '**', and 12 '***' MSFE reductions. *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 (line '$\boldsymbol{\beta}_t$ (heteroscedastic VAR)'). Here, the VAFEM mean combinations outperform their non-VAFEM counterparts in 21 out of 39 cases (53.9 %), with 5 '*', 2 '**', and 14 '***' MSFE reductions. Among this total of 21 MSFE reductions, we have 11 out of 26 (42.3%) accuracy gains for the short- and medium-term forecast horizons $h = 2, 4$, while for the longer-term horizon $h = 8$, we observe substantial VAFEM gains in 10 (2 '*', 2 '**', 6 '***') out of 13 cases (76.9 %).

Table 8 about here

Table 8 provides an overview of VAFEM accuracy gains by means of an overall performance ranking involving all combination forecasts from Tables 2-7. The upper part of Table 8 ranks the alternative (non-VAFEM and VAFEM) combination forecasts according to their overall average losses aggregated over the 39 cases.[5] The reference combination forecast is the non-VAFEM mean (Rank 7) with average loss 0.609. The VAFEM mean combination using the rigid $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}$ (constant-parameter) in-sample VAR specification performs worse than the non-VAFEM mean, in terms of a 3.9% higher average loss (Rank 8). By contrast, the VAFEM mean combination induced by the $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}_t$ (time-varying parameter) in-sample VAR specification clearly outperforms the non-VAFEM mean in terms of a sizable 5.1% accuracy

predictive ability. (ii) We could compare all (non-VAFEM and VAFEM) mean combination forecasts without focusing on a benchmark, using the model confidence set from Hansen et al. (2011). However, in the subsequent Tables 8 and 10, we condense the information from Tables 2-7 by ranking the alternative non-VAFEM and VAFEM combination forecasts according to their average losses (aggregated over the 39 cases). No statistical tests for comparing these average losses are currently available in the literature, leaving this an important issue for future research.

[5]The concept of this ranking is from Stock and Watson (2004). The authors suggest computing the aggregated losses over all 39 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).

gain (Rank 5).

Table 8 addresses additional VAFEM mean combinations exhibiting a large accuracy improvement on the non-VAFEM mean. These VAFEM mean combinations refer to Block (4) of Tables 2-7, where we report out-of-sample MSFEs of VAFEM means obtained from (CT, p)-flexibilized in-sample VAR specifications, each estimated as constant-parameter ('$\boldsymbol{\beta}$ (homoscedastic VAR)') and time-varying parameter ('$\boldsymbol{\beta}_t$ (heteroscedastic VAR)') variants. More precisely, for all 39 cases and both VAR-parameter variants, we disclose the MSFEs of the best VAFEM mean combinations, obtained from *ex-post* searching for that tuple $(\text{CT}, p) \in \{0.85, 0.90, 0.95, 0.99\} \times \{1, 2, 3, 4\}$ producing the (case-specific) minimal out-of-sample MSFE. (The optimizing tuple is displayed in the lines 'Specification (CT, p)' in Block (4).) In Table 8, we denote these two in-sample VAR specifications as (CT, p)-$\boldsymbol{\beta}$ (constant-parameter VAR) and (CT, p)-$\boldsymbol{\beta}_t$ (time-varying parameter VAR). The table shows that the (CT, p)-$\boldsymbol{\beta}$ VAFEM mean combination yields a 7.6% accuracy gain (Rank 4), compared to the non-VAFEM mean. The (CT, p)-$\boldsymbol{\beta}_t$ VAFEM mean combination even outperforms the non-VAFEM mean combination forecast in terms of a 13.6% loss reduction (Rank 2). Additionally, we consider the VAFEM mean combination forecasts under those scenarios, in which—for each of the 39 cases—we *ex post* select the better of the two in-sample VAR specifications (CT, p)-$\boldsymbol{\beta}$ and (CT, p)-$\boldsymbol{\beta}_t$. Here, the VAFEM mean combinations outperform the non-VAFEM means by nearly 18% (Rank 1).

We note that the latter three VAFEM specifications (Ranks 1, 2, 4) 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 accuracy gain of approximately 18 % as a practical upper bound for the accuracy 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 the VAFEM framework. Overall, it is worth highlighting that even the simple, rigidly standardized in-sample VAR specification $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}_t$ leads to a 5.1% accuracy gain for the mean combination forecast (Rank 5).

Besides analyzing mean combination forecasts, Table 8 also reports on the three original (non-VAFEM) time-varying-parameter combination forecasts tvp(0.1), tvp(0.2) and tvp(0.4), as in Stock and Watson (2004, Table VIII). In their analysis, the non-VAFEM tvp(0.1)- and tvp(0.2)-combination forecasts (Ranks 3, 6) both outperform the non-VAFEM mean (Rank 7). We note that several VAFEM mean combinations now outperform the non-VAFEM tvp($\cdot$)-combination forecasts. In particular, the rigid ($\mathrm{CT} = 0.85, p = 4$)-$\boldsymbol{\beta}_t$ VAFEM mean (Rank 5) outperforms the non-VAFEM tvp(0.2) combination, while the (CT, p)-$\boldsymbol{\beta}_t$ VAFEM mean combination on Rank 2 beats all 3 non-VAFEM tvp($\cdot$) combination forecasts (as does the idealized VAFEM mean combination forecast on Rank 1).

Table 9 about here

Finally, Table 8 provides a ranking of several out-of-sample forecasts stemming directly from VARs (of different dimensions), which include alternative macroeconomic and financial time series from the G7 data set.[6] We estimated (in-sample) constant- and time-varying-parameter VARs, each with fixed lag-length $p = 4$, containing alternative sets of variables. In the lower block of Table 8, the ending '_small' indicates the smallest set of variables used in the VAR, while VARs with the endings '_medium' and '_large' contain additional time series. The ending '_DMA' represents VARs resulting from DMA among the corresponding small, medium and large VARs. Table 9 gives an overview of the economic and financial variables (and their transformations involved) that we included in the VAR specifications. Obviously, these direct out-of-sample VAR forecasts, obtained from 8 alternative specifications, all vastly underperform any of the 9 non-VAFEM and VAFEM combination forecasts from the upper block of Table 8.

4.3 VAR selection issues and tvp($\cdot$)-combination forecasts

As mentioned above, the in-sample VAR specifications associated with the VAFEM mean combination forecasts for the Ranks 1, 2, and 4 in Table 8 can be criticized on the grounds of their *ex-post* nature, since the practitioner—being equipped solely with information as of date t—does not know *ex ante* which explicit in-sample VAR

[6]This direct VAR forecasting methodology is adopted from Koop and Korobilis (2013).

specification will produce the lowest MSFEs for future observations at $t+1, t+2, \ldots$. Viewed from this angle, we interpret the out-of-sample forecasting results for these in-sample VAR specifications as a guideline to the maximal accuracy gains obtainable from VAFEM mean combination forecasting (with the G7 data set), provided that the practitioner had accurate knowledge of the data-generating process. The question of whether such potential accuracy gains are generally exploitable via our VAFEM approach to real-world data sets 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 impose 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, where model averaging is based on the predictive likelihood of a set of variables included in a baseline VAR of minimal dimension. The baseline set of variables, which are known from past research to have a macroeconomic predictive content, are also included in VARs of higher dimensions. The forecasts of the baseline-set variables obtained from the VAR of minimal dimension are then averaged with their own forecasts obtained from VARs of higher dimensions. The averaging weights are built upon the predictive likelihoods from the VARs of different dimensions, but applied only to the baseline-set of variables. Proceeding in this fashion, the additional variables from the higher-dimensional VARs implicitly improve the forecasts of the baseline-set variables. Two aspects are worth mentioning at this point. (i) Determining the number of higher-dimensional VARs used in the averaging procedure is left to the econometrician (and is thus somewhat arbitrary). (ii) The procedure rests on an ordering of the variables according to their predictive ability.

Generally, our VAFEM framework does not offer any clear-cut set of undisputed baseline forecast models, whose forecast-error series could take on the role of the baseline-set variables. In particular, the heuristic nature of our CT-forecast-model selection mechanism allows for the (realistic) scenario that a specific individual forecast-error series, which has predictive content, is excluded from the in-sample VAR specification under a given (comparably low) CT value. Apart from that, a reasonable ordering of the variables is also difficult to justify, so that DMA over different VAR dimensions, as performed by Koop and Korobilis (2013), 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]

Table 10 about here

We present a final empirical result, demonstrating that our VAFEM approach has the potential to produce large accuracy-gains, even without 'optimizing' the in-sample VAR specification. Table 10 displays the average MSFEs of several non-VAFEM and VAFEM time-varying parameter forecast combinations (relative to autoregression) aggregated over all 39 cases of the G7 data set. Block (1) reports the results for the non-VAFEM tvp(0.1)-, tvp(0.2)- and tvp(0.4)-combination forecasts (Stock and Watson, 2004, Table XI). Blocks (2) and (3) display the average (aggregated) MSFEs of the corresponding VAFEM tvp($\cdot$)-forecast combinations under the rigid in-sample VAR specifications $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}$ and $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}_t$. Based on these simple VAR specifications—which are not subject to hindsight criticism—our VAFEM approach provides far larger aggregated accuracy gains for tvp($\cdot$)-combination forecasts than for mean combination forecasts (see Table 8). In particular, the constant-parameter $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}$ specification produces aggregated accuracy gains

[7]Recently, Onorante and Raftery (2016) proposed a heuristic approach to (practically) handling a large number of model combinations in DMA. Their main idea is to restrict attention to an adequately defined subset of models and to dynamically optimize the choice of models at each point in time. Checking this procedure for validity, within our VAFEM framework, should be tackled in future research.

between 12.1% [tvp(0.1)] and 17.6 % [tvp(0.4)]. Interestingly, the accuracy gains under the time-varying-parameter $(\mathrm{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}_t$ specification are slightly lower, but—ranging between 7.7% [tvp(0.1)] and 16.7 % [tvp(0.4)]—are still substantial.

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 be 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 into 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 5.1 to 17.9%, depending on the selected in-sample VAR specification. For the time-varying parameter combinations, we find substantial MSFE reductions, ranging between 12.1 and 17.6% under a simply-structured (non-optimized) in-sample VAR specification.

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 8 and 10 (see Footnote 3). 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 conceptually appealing extension could consist of investigating alternative methods for selecting 'optimal' in-sample VAR specifications. In Footnote 6, we mention the heuristic approach of Onorante and Raftery (2016) for dealing with a large number of model combinations in DMA. Their approach merits careful analysis within a simulation framework.

References

Aiolfi, M., Timmermann, A., 2006. Persistance 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 Applied Econometrics 25, 71-92.

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.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Operational Research Quarterly 20, 451-468.

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.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365-380.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453-497.

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.

LeSage, J.P., Magura, M., 1992. A mixture-model approach to combining forecasts. Journal of Business and Economic Statistics 3, 445-452.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer-Verlag Berlin Heidelberg, Germany.

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., Ettler, 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, Vol. 2B, pp. 1203-1324. Elsevier, Amsterdam.

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.F. (Eds.), The Oxford Handbook of Economic Forecasting, pp. 35-60. Oxford University Press, New York, NY.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 1, pp. 135-196. Elsevier, Amsterdam.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097-1126.

Tables and Figures

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 three 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.

Table 2: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of two-quarter growth of real GDP ($h = 2$)

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 forecast (Stock & Watson, 2004)							
AR RMSFE	0.016	–	0.013	0.011	0.013	0.010	0.011
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	0.90	–	0.92	0.99	0.99	0.95	0.95
tvp(0.1)	0.79	–	0.86	0.76	0.80	0.99	0.96
tvp(0.2)	0.78	–	0.86	0.70	0.81	1.04	0.99
tvp(0.4)	0.76	–	0.87	0.70	0.83	1.05	1.04
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.76***	–	0.95	0.75***	0.96*	1.07	1.05
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.75***	–	1.02	0.76***	1.01	0.93*	0.99
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.90, 1)$	–	$(0.99, 1)$	$(0.90, 1)$	$(0.99, 3)$	$(0.99, 4)$	$(0.99, 3)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.73***	–	0.89*	0.67***	0.87***	0.95	0.98
Specification (CT, p)	$(0.95, 2)$	–	$(0.85, 4)$	$(0.99, 1)$	$(0.95, 3)$	$(0.85, 2)$	$(0.95, 3)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.73***	–	1.02	0.68***	0.87***	0.86**	0.91*

Notes:

(i) In Block (1), AR RMSFE denotes the root-mean-squared forecast error of the benchmark autoregressive model (in decimal values of the h-period growth, i.e. not an annual rate) 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' and 'tvp($\cdot$)' denote the simple average and the time-varying-parameter combination forecasts of the original M individual forecasts. All these MSFEs were compiled from Stock and Watson (2004, Table II).

(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 tuple $(\mathrm{CT}, p) \in \{0.85, 0.90, 0.95, 0.99\} \times \{1, 2, 3, 4\}$ in the constant- and time-varying-parameter in-sample VAR specifications, which produces minimal MSFEs.

(iv) *, **, and *** 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.

Table 3: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of four-quarter growth of real GDP ($\boldsymbol{h} = 4$)

	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 forecast (Stock & Watson, 2004)							
AR RMSE	0.025	–	0.018	0.019	0.023	0.018	0.016
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	0.96	–	1.05	1.06	0.98	0.94	0.89
tvp(0.1)	0.85	–	0.91	0.55	0.63	1.11	0.98
tvp(0.2)	0.97	–	0.98	0.49	0.63	1.27	1.15
tvp(0.4)	1.07	–	1.11	0.52	0.65	1.33	1.41
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.90***	–	1.19	0.52***	0.80***	1.17	1.03
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.73***	–	1.07	0.51***	0.83***	1.04	1.03
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.90, 1)$	–	$(0.99, 1)$	$(0.90, 1)$	$(0.85, 3)$	$(0.90, 2)$	$(0.90, 3)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.83***	–	1.03*	0.46***	0.79***	1.08	0.99
Specification (CT, p)	$(0.95, 3)$	–	$(0.85, 4)$	$(0.95, 3)$	$(0.85, 3)$	$(0.85, 4)$	$(0.90, 3)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.72***	–	1.07	0.40***	0.82***	1.04	0.90

Notes: Analogous to the notes to Table 2.

Table 4: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of eight-quarter growth of real GDP ($h = 8$)

	Canada	France	Germany	Italy	Japan	UK	USA
Forecast Period	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV
(1) Benchmark forecast (Stock & Watson, 2004)							
AR RMSE	0.046	–	0.030	0.038	0.046	0.034	0.025
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	1.00	–	1.05	0.96	0.99	1.06	0.98
tvp(0.1)	0.87	–	1.01	0.32	0.65	1.53	1.15
tvp(0.2)	1.08	–	1.24	0.37	0.69	1.94	1.41
tvp(0.4)	1.22	–	1.46	0.45	0.69	2.31	1.72
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	1.12	–	1.13	0.30***	0.79***	1.66	1.55
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.71***	–	0.90***	0.32***	0.82***	1.38	0.92**
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.85, 1)$	–	$(0.99, 1)$	$(0.85, 1)$	$(0.99, 1)$	$(0.85, 4)$	$(0.85, 1)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.94**	–	0.44***	0.30***	0.74***	1.66	1.12
Specification (CT, p)	$(0.95, 4)$	–	$(0.90, 4)$	$(0.90, 4)$	$(0.85, 4)$	$(0.85, 4)$	$(0.90, 2)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.65***	–	0.87***	0.23***	0.82***	1.38	0.90**

Notes: Analogous to the notes to Table 2.

Table 5: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of two-quarter growth of 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 forecast (Stock & Watson, 2004)							
AR RMSE	0.031	0.018	0.026	0.028	0.026	0.018	0.019
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	0.93	1.08	0.90	1.00	1.02	0.98	0.89
tvp(0.1)	0.92	0.97	0.88	0.93	0.92	0.97	0.89
tvp(0.2)	0.92	0.93	0.86	0.88	0.88	0.97	0.90
tvp(0.4)	0.94	0.96	0.84	0.87	0.90	0.98	0.91
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.90^{*}	1.18	0.98	0.95^{**}	0.88^{***}	1.16	0.99
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.91^{*}	1.24	1.03	0.96^{*}	0.89^{***}	1.20	0.93
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.85, 4)$	$(0.99, 1)$	$(0.99, 4)$	$(0.99, 2)$	$(0.99, 1)$	$(0.95, 3)$	$(0.99, 4)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	0.90^{*}	0.98^{**}	0.91	0.81^{***}	0.72^{***}	1.01	0.96
Specification (CT, p)	$(0.99, 1)$	$(0.85, 1)$	$(0.90, 1)$	$(0.95, 1)$	$(0.99, 1)$	$(0.95, 2)$	$(0.85, 4)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.85^{**}	1.09	1.00	0.85^{***}	0.87^{***}	0.95^{*}	0.93

Notes: Analogous to the notes to Table 2.

Table 6: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of four-quarter growth of IP ($h = 4$)

Forecast Period	Canada 82:I – 98:IV	France 82:I – 98:IV	Germany 82:I – 98:IV	Italy 82:I – 98:IV	Japan 82:I – 98:IV	UK 82:I – 98:IV	USA 82:I – 98:IV
(1) Benchmark forecast (Stock & Watson, 2004)							
AR RMSE	0.047	0.031	0.037	0.041	0.052	0.026	0.029
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	0.96	1.16	0.98	1.03	1.02	0.95	0.86
tvp(0.1)	0.95	0.92	0.93	0.85	0.83	0.95	0.88
tvp(0.2)	1.02	0.89	0.90	0.81	0.80	0.98	0.92
tvp(0.4)	1.17	0.95	0.91	0.86	0.85	1.04	1.02
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	1.03	1.28	1.10	0.77^{***}	0.83^{***}	1.06	1.17
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	1.05	1.40	1.06	0.82^{***}	0.87^{***}	1.08	1.41
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.85, 4)$	$(0.99, 4)$	$(0.99, 1)$	$(0.95, 1)$	$(0.99, 1)$	$(0.85, 4)$	$(0.95, 2)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	1.03	1.01^{***}	0.97^{*}	0.74^{***}	0.78^{***}	1.06	0.96
Specification (CT, p)	$(0.85, 4)$	$(0.95, 4)$	$(0.90, 4)$	$(0.90, 4)$	$(0.85, 2)$	$(0.95, 1)$	$(0.95, 4)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	1.05	1.17	0.94^{*}	0.77^{***}	0.85^{***}	0.90^{**}	0.93

Notes: Analogous to the notes to Table 2.

Table 7: MSFEs of VAFEM mean combination forecasts (relative to autoregression): out-of-sample forecasts of eight-quarter growth of IP ($h = 8$)

	Canada	France	Germany	Italy	Japan	UK	USA
Forecast Period	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I – 97:IV	83:I - 97:IV
(1) Benchmark forecast (Stock & Watson, 2004)							
AR RMSE	0.070	0.050	0.054	0.059	0.111	0.041	0.042
(2) Non-VAFEM combination forecasts (Stock & Watson, 2004, balanced panel subset)							
mean	1.00	1.04	0.99	0.98	0.99	1.03	0.89
tvp(0.1)	1.00	0.82	0.90	0.68	0.59	1.13	0.93
tvp(0.2)	1.16	0.95	0.91	0.79	0.56	1.37	0.97
tvp(0.4)	1.39	1.10	0.96	1.00	0.62	1.85	1.04
(3) VAFEM mean combination forecasts: CT $= 0.85, p = 4$							
$\boldsymbol{\beta}$ (homoscedastic VAR)	1.32	1.00*	1.07	0.87***	0.45***	1.89	1.18
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.99*	0.98**	0.97*	0.87***	0.45***	1.41	1.02
(4) VAFEM mean combination forecasts: best (*ex post*) in-sample VAR specification							
Specification (CT, p)	$(0.90, 2)$	$(0.85, 2)$	$(0.95, 4)$	$(0.99, 4)$	$(0.85, 4)$	$(0.85, 1)$	$(0.95, 1)$
$\boldsymbol{\beta}$ (homoscedastic VAR)	1.25	0.99*	1.06	0.76***	0.45***	1.67	0.33***
Specification (CT, p)	$(0.90, 4)$	$(0.95, 3)$	$(0.99, 1)$	$(0.95, 2)$	$(0.85, 4)$	$(0.99, 1)$	$(0.90, 2)$
$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	0.86***	0.81***	0.85***	0.59***	0.45***	0.96**	0.93

Notes: Analogous to the notes to Table 2.

Table 8: Combination forecasts ranked by average losses: both output measures (real GDP, IP), all horizons ($h = 2, 4, 8$)

Rank	VAFEM specification (of mean combination forecast) / Forecast combination	VAFEM / Non-VAFEM	Average loss	Deviation from loss of non-VAFEM mean combination (in %)
1	Best of (CT, p)-$\boldsymbol{\beta}_t$ and (CT, p)-$\boldsymbol{\beta}$	VAFEM	0.500	−17.9
2	(CT, p)-$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	VAFEM	0.526	−13.6
3	tvp(0.1)	Non-VAFEM	0.558	−8.4
4	(CT, p)-$\boldsymbol{\beta}$ (homoscedastic VAR)	VAFEM	0.563	−7.6
5	(CT $= 0.85, p = 4$)-$\boldsymbol{\beta}_t$ (heteroscedastic VAR)	VAFEM	0.578	−5.1
6	tvp(0.2)	Non-VAFEM	0.603	−1.0
7	mean	Non-VAFEM	0.609	
8	(CT $= 0.85, p = 4$)-$\boldsymbol{\beta}$ (homoscedastic VAR)	VAFEM	0.633	3.9
9	tvp(0.4)	Non-VAFEM	0.665	9.2
Direct VAR forecasting:	VAR specification			
10	VAR($\boldsymbol{\beta}_t, p = 4$)_DMA		0.794	30.4
11	VAR($\boldsymbol{\beta}_t, p = 4$)_small		0.803	31.9
12	VAR($\boldsymbol{\beta}, p = 4$)_small		0.807	32.5
13	VAR($\boldsymbol{\beta}, p = 4$)_medium		0.838	37.6
14	VAR($\boldsymbol{\beta}_t, p = 4$)_medium		0.886	45.5
15	VAR($\boldsymbol{\beta}, p = 4$)_DMA		0.924	51.7
16	VAR($\boldsymbol{\beta}, p = 4$)_large		0.989	62.4
17	VAR($\boldsymbol{\beta}_t, p = 4$)_large		1.926	216.3

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 39 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 2–7. (CT $= 0.85, p = 4$)-$\boldsymbol{\beta}_t$ and (CT $= 0.85, p = 4$)-$\boldsymbol{\beta}$ refer to the in-sample VAR specifications from Block (3) of Tables 2–7, (CT, p)-$\boldsymbol{\beta}_t$ and (CT, p)-$\boldsymbol{\beta}$ refer to those from Block (4) of Tables 2–7. 'mean' and 'tvp(·)' denote the non-VAFEM combination forecasts from Block (2) of Tables 2–7.

(iii) 'Best of (CT, p)-$\boldsymbol{\beta}_t$ and (CT, p)-$\boldsymbol{\beta}$' represents the VAFEM mean combination forecasts with the (*ex post*) smaller MSFE, when specifying the in-sample VAR either as (CT, p)-$\boldsymbol{\beta}_t$ or (CT, p)-$\boldsymbol{\beta}$.

(iv) In the Block 'Direct VAR forecasting', VAR($\boldsymbol{\beta}, p = 4$) and VAR($\boldsymbol{\beta}_t, p = 4$) denote VAR specifications with constant and time-varying parameters, respectively, each with fixed VAR lag-length $p = 4$. The ending '_small' indicates the smallest set of time series used in the VAR, as shown in Table 9. VARs with the ending '_medium' contain some additional variables, while VARs with the ending '_large' include the largest set of variables. Country-specific details on the time series used are given in Table 9. VARs with the ending '_DMA' are the result of Dynamic Model Averaging among the corresponding small, medium, and large VARS.

Table 9: Time series used in small, medium and large VARs

Series	Country						
	Canada	France	Germany	Italy	Japan	UK	USA
rgdp_5	s, m, l		s, m, l	s, m, l	s, m, l	s, m, l	s, m, l
cpi_6	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l
ip_5	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l	s, m, l
rovnght_2		s, m, l					s, m, l
rtbill_2	s, m, l						
rbndm_2				s, m, l			
rbndl_2		s, m, l				s, m, l	
comod_5	m, l	m, l	m, l	m, l	m, l	m, l	m, l
stockp_5	m, l	m, l	m, l	m, l	m, l	m, l	m, l
mon1_6	m, l						l
mon2_6		m, l					m, l
mon3_6		m, l					l
oil_5	l	l	l	l	l	l	l
unemp_1		l					
unemp_2	l			l	l	l	l
gold_5	l	l	l	l	l	l	l
capu_1	l						
emp_5	l		l		l	l	l
pgdp_ 5	l		l	l	l	l	l
ppi_5	l				l	l	l
earn_5	l	l			l		l

Note: The identifiers of the time series (string in front of '_') were adopted from Stock and Watson (2004, Table Ia). The digit behind '_' indicates the transformation of the time series used. '1' means no transformation (series in levels). '2' represents the first difference, '5' the series in logs, '6' the first-difference of the logarithm. 's', 'm', and 'l' indicate that the series is used in small, medium and large VAR specifications.

Table 10: Average (equally weighted) MSFEs of tvp(·)-combination forecasts relative to autoregression over all 39 cases (forecast period: 90:III – 99:IV)

tvp(·)-combination	Average MSFE	Deviation from MSFE of non-VAFEM tvp(·)-combination (in %)
(1) Non-VAFEM tvp(·)-combination forecasts, balanced panel (Stock & Watson, 2004)		
tvp(0.1)	0.91	
tvp(0.2)	0.99	
tvp(0.4)	1.08	
(2) VAFEM tvp(·)-combination forecasts: In-sample VAR specification: $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}$ (homoscedastic VAR)		
tvp(0.1)	0.80	−12.1
tvp(0.2)	0.83	−16.2
tvp(0.4)	0.89	−17.6
(3) VAFEM tvp(·)-combination forecasts: In-sample VAR specification: $(\text{CT} = 0.85, p = 4)$-$\boldsymbol{\beta}_t$ (heteroscedastic VAR)		
tvp(0.1)	0.84	−7.7
tvp(0.2)	0.86	−13.1
tvp(0.4)	0.90	−16.7

Notes:

(i) In Block (1), the MSFEs of the non-VAFEM time-varying-parameter combination forecasts of the original individual forecasts were compiled from Stock and Watson (2004, Table XI).

(ii) In Blocks (2) and (3), the in-sample VAR specifications are denoted as in Tables 2–8.

(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 $\phi \in \{0.1, 0.2, 0.4\}$.