Center for Quantitative Economics

A procedure for upgrading linear-convex combination forecasts with an application to volatility prediction

Verena Monschang[†] und Bernd Wilfling[†]

97/2022

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

wissen.leben
WWU Münster

A procedure for upgrading linear-convex combination forecasts with an application to volatility prediction

Verena Monschang[a], Bernd Wilfling[a,*]

[a]Department of Economics (CQE) Westfälische Wilhelms-Universität Münster, Germany

March 21, 2022

Abstract

We investigate mean-squared-forecast-error (MSE) accuracy improvements for linear-convex combination forecasts, whose components are pretreated by a procedure called 'Vector Autoregressive Forecast Error Modeling' (VAFEM). Assuming that the forecast-error series of the individual forecasts are governed by a stable VAR process under classic conditions, we obtain the following results: (i) VAFEM treatment bias-corrects all individual and linear-convex combination forecasts. (ii) Any VAFEM-treated combination has smaller theoretical MSE than its untreated analogue, if the VAR parameters are known. (iii) In empirical applications, VAFEM gains depend on (1) in-sample sizes, (2) out-of-sample forecast horizons, (3) the biasedness of the untreated forecast combination. We demonstrate the VAFEM capacity for realized-volatility forecasting, using S&P 500 data.

Keywords: Combination forecasts, mean-squared-error loss, VAR forecast-error modeling, multivariate least squares estimation.

JEL classification: C10, C32, C51, C53.

*Correspondance to: Am Stadtgraben 9, 48143 Münster, Germany.

E-mail addresses: verena.monschang@wiwi.uni-muenster.de (V. Monschang).
bernd.wilfling@wiwi.uni-muenster.de (B. Wilfling).

1 Introduction

Forecasters often strive to predict future values of a univariate target variable, while having access to multiple individual point forecasts. The econometric task then consists of optimally processing all pieces of information contained in the individual forecasts. A renowned methodology designed for this purpose is that of combining the individual forecasts to obtain a pooled univariate forecast (Bates and Granger, 1969) and, up to date, a large body of literature on forecast combinations has emerged. Prevalent articles include, *inter alia*, in-depth reviews (Clemen, 1989; Timmermann, 2006; Rossi, 2013), as well as studies that aim to predict (i) macroeconomic and monetary variables (Stock and Watson, 2004; Capistrán and Timmermann, 2009; Gaglianone and Lima, 2014; Zhang, 2019), (ii) financial quantities (Guidolin and Timmermann, 2009; Rapach et al., 2010; Pesaran and Pick, 2011; Taylor, 2020), and (iii) commodity prices (Nowotarski et al., 2014; Baumeister and Kilian, 2015; Garrat et al., 2019).

Recently, Weigt and Wilfling (2021) have proposed a 'post-processing' forecast-error modeling approach with the objective of first improving the individual forecasts, which might then lead to accuracy gains for the forecast combinations obtained from them. In a nutshell, the authors' approach, referred to as Vector Autoregressive Forecast Error Modeling (VAFEM), consists of the following three steps. *Step 1.* Modeling the individual forecast-error series as a vector autoregression (VAR), the parameters of which are estimated from past observations. *Step 2.* Using the estimated VAR model to obtain predictions of future individual forecast-errors. *Step 3.* Adapting the initial (original) individual forecasts to the predicted individual forecast-errors, obtained in Step 2.[1]

Having executed the three steps, the individual VAFEM forecasts may ultimately be combined. A follow-up question is then whether this VAFEM forecast combination gener-

[1]To distinguish the 'initial (original) individual forecasts' from the 'adapted individual forecasts' obtained in Step 3, we refer to the latter as the 'individual VAFEM forecasts' hereafter.

ates mean-squared-error (MSE) accuracy gains *vis-à-vis* the identical forecast combination scheme applied to the initial (original) individual forecasts. In view of our informal description of the three above steps, we would expect intuitively that any VAFEM accuracy gain can be thought of as stemming from two sources. (i) Systematic components that are inherent to the history of the idiosyncratic initial individual forecast-error series (e.g. autocorrelation). (ii) Existing correlation/covariation among the multiple initial individual forecast-error series. In principle, the magnitude of potential VAFEM accuracy gains should hinge on three aspects. (i) The selection of the 'correct' VAR model, specifying the interrelationships among the individual forecast-error series in Step 1 of the VAFEM procedure. (ii) The accurate estimation of the VAR parameters in Step 1. (iii) The quality of the individual forecast-error predictions in Step 2.

Weigt and Wilfling (2021) outline the VAFEM procedure within the (classic) covariance-stationary, stable VAR framework, in which (i) parameters can be consistently estimated via multivariate Least Squares (LS), and (ii) forecasting routines are well-established in the literature. In this paper, we adopt this very classic VAR framework, enabling us to use common VAR theory in our formal analysis.[2]

In the next sections, we first formalize the VAFEM approach within the classic VAR setting. We establish various theoretical results that elaborate the VAFEM benefits to both (i) the individual forecasts, and (ii) any linear-convex combination of them.[3] We obtain the following four key results. (i) Under LS estimation in Step 1, all individual VAFEM forecasts obtained in Step 3 are unbiased (in the sense of having unconditional mean-zero prediction errors). This implies that any (non-stochastic) linear-convex VAFEM combination forecast is also unbiased. (ii) If the VAR parameters in Step 1 are known with

[2]Weigt and Wilfling's (2021) article is empirical by nature. In their VAFEM-application to the well-known 7-country data set on output growth (Stock and Watson, 2004), the authors switch to a Bayesian estimation framework, which is better suited to handling large (high-dimensional) time-varying parameter VARs.

[3]Note that any particular individual forecast can be viewed as a special linear-convex combination of the set of all individual forecasts.

certainty, any linear-convex VAFEM forecast combination unambiguously outperforms its original (non-VAFEM) analogue in terms of lower (theoretical) MSE. This clear-cut accuracy gain may reverse if the VAR parameters need to be estimated (due to estimation errors). (iii) We establish settings—that are likely to bring about theoretical MSE reductions for any (non-stochastic) linear-convex VAFEM combination forecast—pertaining to (a) in-sample sizes, (b) out-of-sample forecast horizons, and (c) the biasedness of the corresponding initial (non-VAFEM) combination. (iv) We can broadly confirm our theoretical VAFEM results in an empirical application to out-of-sample realized-volatility forecasting, using daily S&P 500 data with three individual forecast series.

The paper is organized as follows. Section 2 sets out the VAFEM procedure, and establishes its theoretical forecasting properties. Section 3 reviews the individual volatility forecasting models used in the empirical application. Section 4 presents the out-of-sample realized-volatility forecasting analysis with S&P 500 data. Section 5 concludes.

2 Vector autoregressive forecast error modeling

The objective of this section is threefold. Section 2.1 outlines (i) our econometric framework, and (ii) the 3-step VAFEM procedure. Section 2.2 establishes the unbiasedness of VAFEM forecasts and compiles results on their theoretical MSEs. Section 2.3 elaborates conditions under which the VAFEM forecasts may outperform their non-VAFEM analogues.

2.1 Econometric setup, and the VAFEM procedure

For $s = 0, \pm 1, \pm 2, \ldots$, we consider the univariate target variable y_s and, at present date t, aim at forecasting the future value $y_{t+h}, h = 1, 2, \ldots$, based on information available at date t. We denote such a forecast by $\hat{y}_{t+h|t}$. We assume M given alternative h-step forecasts, $\hat{y}_{t+h|t,1}, \hat{y}_{t+h|t,2}, \ldots, \hat{y}_{t+h|t,M}$, that may stem from different sources. We collect these individual forecasts in the $M \times 1$ vector $\hat{\mathbf{y}}_{t+h|t} = (\hat{y}_{t+h|t,1}, ..., \hat{y}_{t+h|t,M})'$, and the associated individual forecast errors, $e_{t+h|t,i} = y_{t+h} - \hat{y}_{t+h|t,i}, i = 1, \ldots, M$, in the vector $\mathbf{e}_{t+h|t} = (e_{t+h|t,1}, \ldots, e_{t+h|t,M})'$. We summarize the information available at date t in the set $\mathcal{I}_t$. Explicitly, $\mathcal{I}_t$ consists of the entire histories of (i) the target variable, and (ii) the M individual h-step-ahead forecasts up to date t, that is $\mathcal{I}_t = \{\ldots, \hat{\mathbf{y}}_{t+h-1|t-1}, \hat{\mathbf{y}}_{t+h|t}, \ldots, y_{t-1}, y_t\}$.[4] We formalize the 3-step VAFEM procedure as follows.

Step 1. We assume that the forecast-error vector $\mathbf{e}_{t+h|t}$ is governed by the covariance-stationary, stable VAR(p) process,

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

with intercept vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)'$ and $M \times M$ parameter matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$. The innovation vector $\boldsymbol{\epsilon}_{t+h} = (\epsilon_{t+h,1}, ..., \epsilon_{t+h,M})'$ is assumed to be Gaussian white noise with non-singular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Under these classic assumptions, we estimate the VAR parameters in Eq. (1) consistently by multivariate LS (Lütkepohl, 2005, pp. 69-72).[5]

Step 2. For the prediction of the future individual forecast errors, we consider the conditional expectation vector $\mathbf{e}^{\mathcal{I}_t}_{t+h|t} \equiv \mathbb{E}[\mathbf{e}_{t+h|t}|\mathcal{I}_t]$ according to Eq. (1). Replacing the unknown VAR parameters in $\mathbf{e}^{\mathcal{I}_t}_{t+h|t}$ with their LS estimates from Step 1, we write the estimated forecast-error predictions as

$$\hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t} = \hat{\boldsymbol{\nu}} + \hat{\mathbf{A}}_1 \hat{\mathbf{e}}^{\mathcal{I}_t}_{(t-1)+h|t-1} + \ldots + \hat{\mathbf{A}}_p \hat{\mathbf{e}}^{\mathcal{I}_t}_{(t-p)+h|t-p}, \tag{2}$$

[4] Under this timeline, $\mathbf{e}_{s|s-h}$ is $\mathcal{I}_t$-measurable for $s \leq t$.

[5] The forecast-error modeling in Eq. (1) is particularly suited to situations in which the M individual forecasts are of a 'black-box' nature, in the sense that the specific forecasting processes are not fully formalized or even entirely unknown (e.g. judgmental, survey forecasts). The forecaster requires only the information set $\mathcal{I}_t$, from which to obtain the forecast-error sample $\{\ldots, \mathbf{e}_{t-1|t-1-h}, \mathbf{e}_{t|t-h}\}$, for estimating the VAR parameters in Eq. (1) via multivariate LS.

where $\hat{\mathbf{e}}^{\mathcal{I}_t}_{s+h|s} = \mathbf{e}_{s+h|s}$ for $s + h \leq t$. To obtain $\hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t}$ for $h > 1$, we start with the prediction of $\hat{\mathbf{e}}^{\mathcal{I}_t}_{t+1|(t+1)-h}$, and then recursively apply Eq. (2).

Step 3. We adapt the initial individual forecasts in $\hat{\mathbf{y}}_{t+h|t}$ by adding to them the predicted VAFEM errors from Step 2. Formally, we obtain our individual VAFEM forecasts as

$$\tilde{\hat{\mathbf{y}}}_{t+h|t} \equiv \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t}. \tag{3}$$

Having executed these three steps, we may principally combine the M individual VAFEM forecasts contained in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$, in the hope of increasing the forecast accuracy (in terms of lower MSEs), using any of the combination schemes suggested in the literature. In this paper, we restrict our attention to the class of time-invariant (non-stochastic) linear-convex combinations. Formally, for $h = 1, 2, \ldots$, we consider the VAFEM and non-VAFEM forecast combinations

$$\tilde{\hat{y}}^{\text{comb}}_{t+h|t} = \mathbf{w}'\tilde{\hat{\mathbf{y}}}_{t+h|t} \qquad \text{and} \qquad \hat{y}^{\text{comb}}_{t+h|t} = \mathbf{w}'\hat{\mathbf{y}}_{t+h|t}, \tag{4}$$

with weighting vector $\mathbf{w} = (w_1, \ldots, w_M)'$, satisfying $w_i \geq 0$ for $i = 1, \ldots, M$, and $\sum_{i=1}^{M} w_i = 1$. Important weights are $(1/M, \ldots, 1/M)'$—the arithmetic-mean combination—and the M Euclidian standard basis vectors $(1, 0, \ldots, 0)', \ldots, (0, \ldots, 0, 1)'$. Specifically, the ith basis vector reduces the VAFEM and non-VAFEM combination forecasts in Eq. (4) to the ith individual VAFEM and (initial) non-VAFEM forecasts $\tilde{\hat{y}}_{t+h|t,i}$ and $\hat{y}_{t+h|t,i}$, respectively.

Our empirical application in Section 4 makes extensive use of the M standard basis vectors, for conducting MSE comparisons between the individual VAFEM and the initial non-VAFEM forecasts. Besides our special cases, the popularity of general linear-convex weighting schemes stems from the fact that the associated combination forecasts (i) do not leave the co-domain of the individual forecasts, and (ii) are unbiased if the individual forecasts are unbiased (Timmermann, 2006).[6]

2.2 Unbiasedness and MSEs

2.2.1 Unbiasedness of VAFEM forecasts

We now show that the 3-step VAFEM procedure inherently generates unbiased individual forecasts, thus constituting a convenient tool for simultaneoulsy bias-correcting all initial individual forecasts (plus any linear-convex combination of them).

Proposition 1. *In the setting of Section 2.1, (i) all M individual VAFEM forecasts in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ from Eq. (3), and (ii) any (non-stochastic) linear-convex VAFEM combination forecast $\tilde{\hat{y}}^{\text{comb}}_{t+h|t}$ from Eq. (4) are unbiased.*

To obtain the result, we define the $M \times 1$ vector $\mathbf{y}_{t+h} \equiv (y_{t+h}, y_{t+h}, \ldots, y_{t+h})'$ and write the individual VAFEM prediction errors in vector form as

$$\begin{aligned} \mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t} &= \mathbf{y}_{t+h} - (\hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t}) \\ &= \mathbf{e}_{t+h|t} - \hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t}. \end{aligned} \tag{5}$$

Taking the expectations on both sides of Eq. (5), it follows that the individual VAFEM forecasts in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ are unbiased, if and only if $\mathbb{E}\left[\mathbf{e}_{t+h|t} - \hat{\mathbf{e}}^{\mathcal{I}_t}_{t+h|t}\right] = \mathbf{0}$, i.e. if and only if the predicted errors from the estimated VAR in Eq. (1) are unbiased. This latter unbiasedness follows directly from Dufour (1985), due to our use of the LS estimators and the VAFEM assumptions in Section 2.1. The unbiasedness of all individual VAFEM forecast in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$, in turn, implies the unbiasedness of any linear-convex VAFEM combination forecast $\tilde{\hat{y}}^{\text{comb}}_{t+h|t}$.

[6] The MSE-results proved below remain valid for non-stochastic, but time-varying linear-convex combination schemes $\mathbf{w}_t = (w_{t1}, \ldots, w_{tM})'$. However, the 'non-stochastic' property of $\mathbf{w}_t$ is essential in the proofs. It precludes the general validity of our theoretical MSE results for some popular combination forecasts, like the median and the trimmed mean. For the latter combinations, the linear-convex elements in $\mathbf{w}_t$ become stochastic, due to the ordering of the individual forecasts in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ and $\hat{\mathbf{y}}_{t+h|t}$, and thus are correlated with $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ and $\hat{\mathbf{y}}_{t+h|t}$.

2.2.2 MSE formulae for VAFEM forecasts

In this section, we provide theoretical MSE formulae for (i) the linear-convex VAFEM combination forecast $\tilde{\hat{y}}^{\text{comb}}_{t+h|t}$ from Eq. (4), and (ii) the individual VAFEM forecasts in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ from Eq. (3). We denote the (theoretical) univariate MSE-operator by MSE($\cdot$), and define the matrix $\text{MSEM}\left(\tilde{\hat{\mathbf{y}}}_{t+h|t}\right) \equiv \mathbb{E}\left[\left(\mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t}\right)\left(\mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t}\right)'\right]$. For the MSE of the linear-convex VAFEM combination forecast, we obtain

$$\begin{aligned}
\text{MSE}(\tilde{\hat{y}}^{\text{comb}}_{t+h|t}) &= \mathbb{E}\left[\left(y_{t+h} - \tilde{\hat{y}}^{\text{comb}}_{t+h|t}\right)^2\right] = \mathbb{E}\left[\left(y_{t+h} - \mathbf{w}'\tilde{\hat{\mathbf{y}}}_{t+h|t}\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{w}'\left\{\mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t}\right\}\right)^2\right] \\
&= \mathbf{w}'\mathbb{E}\left[\left(\mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t}\right)\left(\mathbf{y}_{t+h} - \tilde{\hat{\mathbf{y}}}_{t+h|t}\right)'\right]\mathbf{w} \\
&= \mathbf{w}'\text{MSEM}\left(\tilde{\hat{\mathbf{y}}}_{t+h|t}\right)\mathbf{w}, \qquad (6)
\end{aligned}$$

where the M elements on the main diagonal of the MSE-matrix $\text{MSEM}\left(\tilde{\hat{\mathbf{y}}}_{t+h|t}\right)$ are given by $\text{MSE}(\tilde{\hat{y}}_{t+h|t,i})$ for $i = 1, \ldots, M$ (the MSEs of the individual VAFEM forecasts).

The following proposition states a large-sample approximation of the MSEM-matrix in Eq. (6). Proofs can be found in Lütkepohl (2005, pp. 94-98), and for VAR specifications without intercept term in Reinsel (1997, pp. 155-157).

Proposition 2 (Reinsel, 1997; Lütkepohl 2005)**.** *Given a forecast-error sample of size T plus a presample of p observations, let $\mathbf{D} \equiv (\mathbf{D}_{t-T,h}, \ldots, \mathbf{D}_{t-1,h})$ with $\mathbf{D}_{s,h} \equiv \left(1, \mathbf{e}'_{s|s-h}, \ldots, \mathbf{e}'_{s-p+1|s-p+1-h}\right)'$, and consider the probability limit $\mathbf{\Gamma} \equiv \text{plim } \mathbf{DD}'/T$. Denoting the $M \times M$*

identity matrix by $\mathbf{I}_M$*, we define*

$$\boldsymbol{\mathcal{A}} \equiv \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\nu} & \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_M & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_M & \mathbf{0} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\lambda}_i \equiv \mathbf{J}\boldsymbol{\mathcal{A}}^i\mathbf{J}',$$

where $\mathbf{J} \equiv \begin{pmatrix} \mathbf{0} & \mathbf{I}_M & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}$ *is an* $M \times (Mp+1)$ *matrix. Then, under the assumptions from Section 2.1, a large-sample approximation of the MSEM-matrix in Eq. (6) is given by*

$$\text{MSEM}^{\text{approx}}\left(\tilde{\hat{\mathbf{y}}}_{t+h|t}\right) = \sum_{i=0}^{h-1} \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_i' + \frac{1}{T}\boldsymbol{\Omega}(h), \tag{7}$$

where $\boldsymbol{\Omega}(h) \equiv \sum_{i=0}^{h-1}\sum_{j=0}^{h-1} \text{trace}\left[(\boldsymbol{\mathcal{A}}')^{h-1-i}\boldsymbol{\Gamma}^{-1}\boldsymbol{\mathcal{A}}^{h-1-j}\boldsymbol{\Gamma}\right] \cdot \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_j'$*, and* $\text{trace}(\cdot)$ *denotes the trace operator.*

Remark 1. *The right-hand side of Eq. (7) consists of two major summands. It is well-known that the first major summand (*$\sum_{i=0}^{h-1} \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_i'$*) coincides with the exact MSE-matrix, if the true VAR parameters in Eq. (1) are fully known. That is, under known VAR parameters, we have*

$$\text{MSEM}\left(\tilde{\hat{\mathbf{y}}}_{t+h|t}\right) = \sum_{i=0}^{h-1} \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_i'. \tag{8}$$

The second major summand in Eq. (7), $(1/T)\boldsymbol{\Omega}(h)$*, comes into play, when the true VAR parameters are unknown and need to be estimated.*

2.3 MSE properties of VAFEM forecasts

We are now able to compare the forecasting accuracy of any linear-convex VAFEM combination forecast ($\tilde{\hat{y}}^{\text{comb}}_{t+h|t}$) with that of its (initial) non-VAFEM analogue ($\hat{y}^{\text{comb}}_{t+h|t}$). Invoking

Eqs. (6) and (8), we obtain the following result.

Proposition 3. *Under the VAFEM setup from Sections 2.1 and 2.2, assume that the true VAR parameters are known. For every (non-stochastic) linear-convex combination vector* $\mathbf{w} = (w_1, \ldots, w_m)'$*, we then have*

$$\mathrm{MSE}(\tilde{\hat{y}}^{\mathrm{comb}}_{t+h|t}) = \mathbf{w}' \left(\sum_{i=0}^{h-1} \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_i' \right) \mathbf{w} \leq \mathrm{MSE}(\hat{y}^{\mathrm{comb}}_{t+h|t}). \tag{9}$$

We prove Proposition 3 in the Appendix. The proof provides the following two clear-cut results on the accuracy gain of the VAFEM *vis-à-vis* the (initial) non-VAFEM combination forecast.

Corollary 1. *We define the accuracy gain of the linear-convex VAFEM combination forecast vis-à-vis its (initial) non-VAFEM counterpart as* $\mathrm{MSE}(\hat{y}^{\mathrm{comb}}_{t+h|t}) - \mathrm{MSE}(\tilde{\hat{y}}^{\mathrm{comb}}_{t+h|t})$*. Then, the proof of Proposition 3 establishes the following (qualitative) results:*

(i) The accuracy gain is larger, if the linear-convex (initial) non-VAFEM combination forecast $\hat{y}^{\mathrm{comb}}_{t+h|t}$ *is biased.*[7]

(ii) The accuracy gain typically decreases when the forecast horizon h *increases. In particular, we obtain* $\mathrm{MSE}(\tilde{\hat{y}}^{\mathrm{comb}}_{t+h|t}) = \mathrm{MSE}(\hat{y}^{\mathrm{comb}}_{t+h|t})$ *in Eq. (9) for* $h \to \infty$*, if all* M *initial individual forecasts in* $\hat{\mathbf{y}}_{t+h|t}$ *are unbiased.*[8]

It remains to consider the most realistic scenario, in which the true VAR parameters in Eq. (1) are unknown. Eqs. (6) and (7) provide the following MSE approximation of the linear-convex VAFEM combination forecast:

$$\mathrm{MSE}(\tilde{\hat{y}}^{\mathrm{comb}}_{t+h|t}) \approx \mathbf{w}' \left(\sum_{i=0}^{h-1} \boldsymbol{\lambda}_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\lambda}_i' \right) \mathbf{w} + \frac{1}{T} \mathbf{w}' \boldsymbol{\Omega}(h) \mathbf{w}. \tag{10}$$

[7] See Eqs. (A.4), (A.5) in the Appendix. The biasedness of $\hat{y}^{\mathrm{comb}}_{t+h|t}$ requires at least one of the individual forecasts in $\hat{\mathbf{y}}_{t+h|t}$ to be biased.

[8] See Eqs. (A.4)–(A.7) in the Appendix.

The summand $(1/T)\mathbf{w}'\mathbf{\Omega}(h)\mathbf{w}$ on the right-hand side of Eq. (10) is non-negative, since the $\mathbf{\Omega}(h)$ is positive-semidefinite. Thus, in the case of unknown VAR parameters, estimation errors might substantially increase the MSE of the VAFEM combination forecast according to the summand $(1/T)\mathbf{w}'\mathbf{\Omega}(h)\mathbf{w}$ in Eq. (10). Potentially (and not surprisingly), this can entail an underperformance of the VAFEM combination forecast.

Remark 2. *Our previous analysis indicates three circumstances in which the VAFEM procedure has the highest potential for deconvolving accuracy-improving effects on linear-convex combination forecasts:*

(i) When the linear-convex non-VAFEM combination forecast $\hat{y}^{\text{comb}}_{t|t+h}$ *is biased [Corollary 1].*

(ii) When the sample size T *is large [Eqs. (7), (10)].*

(iii) When the forecast horizon h *is small [Corollary 1].*

We end this section by recalling that, upon equating the weighting vector $\mathbf{w}$ in Eq. (4) with the Euclidian standard basis vectors, each of the M (i) individual VAFEM, and (ii) (initial) non-VAFEM forecasts in $\tilde{\hat{\mathbf{y}}}_{t+h|t}$ and $\hat{\mathbf{y}}_{t+h|t}$ represent special cases of the general linear-convex combinations $\tilde{\hat{y}}^{\text{comb}}_{t+h|t}$ and $\hat{y}^{\text{comb}}_{t+h|t}$. Thus, all our previous MSE results for the general VAFEM and non-VAFEM combination forecasts hold particularly for the individual VAFEM and non-VAFEM forecasts.

3 Individual forecast models

In Section 4, we will apply the VAFEM procedure to realized-volatility forecasting, using S&P 500 data. We base our out-of-sample analysis on a maximum number of $M = 3$ (initial) individual forecasts.[9] Two forecast series stem from models (ARFIMA, HAR)

[9]Notwithstanding our remarks in Footnote 5, we conduct our empirical analysis with three fully-specified (individual) forecasting processes, for each of which large sets of (daily) forecasts over multiple horizons are

that are frequently applied in realized-volatility prediction and have experienced a plethora of formal modifications and extensions (*inter alia*, Bollerslev et al., 2016; Audrino et al., 2019; Izzeldin et al., 2019). The other forecast series (GARCH), which we primarily use for illustrative purposes, has historical and finance-based backgrounds (e.g. Koopman et al., 2005; Taylor, 2005, Sec. 9). Since our focus is on analyzing VAFEM accuracy effects on individual and combination forecasts, and not on identifying the 'best' out-of-sample prediction model for our data set, we apply standard variants of the (initial) individual forecast specifications.

To briefly review the models, we let our target variable y_t represent the realized volatility of a financial-return variable x_t, and define realized volatility as the square-root of the sum of n (equidistantly observed) squared intraday returns $x_{t:i}$ $(i = 1, \dots, n)$:

$$y_t = \sqrt{\sum_{i=1}^{n} x_{t:i}^2}. \tag{11}$$

Our first (initial) individual forecast ($\hat{y}_{t+h|t,1}$) originates from the standard GARCH$(1,1)$ model (Bollerslev, 1986), which specifies the financial return x_t as

$$x_t = \sigma_t \cdot u_t, \tag{12}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \cdot x_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2, \tag{13}$$

with $u_t \overset{\text{i.i.d.}}{\sim} \mathbb{N}(0,1)$, and parameters $\alpha_0 > 0, \alpha_1, \beta_1 \geq 0$. Assuming $\alpha_1 + \beta_1 < 1$ (second-moment stationarity), we use the square-root of the h-step-ahead conditional-variance forecasts $\hat{\sigma}_{t+h|t}^2$ from the estimated GARCH$(1,1)$ specification as our first individual realized-volatility forecasts. For clarity of exposition, we write $\hat{y}_{t+h|t,\text{GARCH}} \equiv \hat{y}_{t+h|t,1} = \hat{\sigma}_{t+h|t}$.

In contrast to the GARCH Eqs. (12), (13), the other two forecast models directly specify

available. This enables us to verify our theoretical VAFEM results from Section 2, in terms of in-sample sizes and out-of-sample forecast horizons.

the dynamics of the realized volatility process. The Heterogeneous-Auto-Regressive (HAR) model (Corsi, 2009) specifies realized volatility as

$$y_t = \alpha_0 + \alpha_1 \cdot y_{t-1} + \alpha_2 \cdot y^{\mathrm{w}}_{t-1} + \alpha_3 \cdot y^{\mathrm{m}}_{t-1} + \sigma \cdot u_t, \tag{14}$$

where

$$y^{\mathrm{w}}_{t-1} \equiv \frac{1}{5}\sum_{i=1}^{5} y_{t-i}, \qquad y^{\mathrm{m}}_{t-1} \equiv \frac{1}{22}\sum_{i=1}^{22} y_{t-i},$$

$u_t \overset{\text{i.i.d.}}{\sim} \mathbb{N}(0,1)$, and with parameters $\alpha_0, \ldots, \alpha_3$, and $\sigma > 0$. The Auto-Regressive Fractionally Integrated Moving Average ARFIMA$(1,d,1)$ model (Granger and Joyeux, 1980; Hosking, 1981) specifies realized volatility as

$$(1-\phi\mathbb{L})(1-\mathbb{L})^d\,[y_t - \mathbb{E}(y_t)] = (1+\theta\mathbb{L}) \cdot \sigma \cdot u_t, \tag{15}$$

with lag operator $\mathbb{L}$, $u_t \overset{\text{i.i.d.}}{\sim} \mathbb{N}(0,1)$, and parameters d, ϕ, θ, σ restricted by $0 < d < 0.5$; $|\phi|, |\theta| < 1$ (stationarity and invertibility conditions), and $\sigma > 0$. Taking conditional expectations in Eqs. (14) and (15), optimal h-step-ahead forecasts $\hat{y}_{t+h|t,\mathrm{HAR}} \equiv \hat{y}_{t+h|t,2}$ and $\hat{y}_{t+h|t,\mathrm{ARFIMA}} \equiv \hat{y}_{t+h|t,3}$ subject to the MSE loss function are readily established (Corsi, 2009; Doornik and Ooms, 2004).

4 Realized-volatility forecasting with S&P 500 data

In this section, we investigate out-of-sample forecasting gains/losses from the VAFEM procedure, using S&P 500 data. We apply two types of linear-convex weighting schemes. (i) The Euclidian standard basis vectors (for comparing the individual forecasts before-and-after VAFEM treatment), and (ii) the $(1/M)$-equal-weight vector, representing the mean combination forecast.

Figure 1 about here

Table 1 about here

4.1 Data set and out-of-sample timeline

Our data set includes daily observations of the S&P 500 index between 3 January 2000 and 31 January 2018 (4539 obs.). The (log) returns were computed from closing prices, and provided by the *Realized Library of the Oxford Man Institute* (RLOMI). The *realized-volatility* data, also obtained from RLOMI, were computed from 5-minute intraday returns. Daily trading hours at the New York Stock Exchange (NYSE) were from 9:30 to 16:00 (15:30 to 22:00 CET), covering 78 5-minute intraday intervals, for computing daily realized volatility, according to Eq. (11). We implemented the entire VAFEM procedure with the EViews 12 software. Figure 1 plots the respective time-series over the entire time span. Table 1 provides some summary statistics, reflecting (i) leptokurtosis, (ii) highly significant (first-order) autocorrelation, and (iii) non-normality for the target variable (realized volatility), and the daily S&P returns.

We execute our out-of-sample analysis with alternative sets of individual forecasts (*forecast sets*). We denote these by (i) $\{\text{GARCH}\}, \{\text{HAR}\}, \{\text{ARFIMA}\}$ for $M = 1$, (ii) $\{\text{GARCH}, \text{HAR}\}, \{\text{GARCH}, \text{ARFIMA}\}, \{\text{HAR}, \text{ARFIMA}\}$ for $M = 2$, and $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$ for $M = 3$. To compare the (initial) non-VAFEM (individual and combination) forecasts with their VAFEM analogues, we implement the following timeline (for each forecast set).

(a) *Out-of-sample initial individual forecasts* ($\hat{\mathbf{y}}_{t+h|t}$) *and forecast errors* ($\mathbf{e}_{t+h|t}$)*:*

We initialize a rolling window of the fixed observation length $N = 500$ (trading days) for estimating the parameters of the individual forecast models (GARCH, HAR, ARFIMA). The initializing window starts on Obs. #23 (03/FEB/2000) and ends on Obs. #522 (07/FEB/2002).[10] We obtain the first initial individual forecasts at forecast horizon h ($h = 1, 2, 3, 5, 10$ trading days) for Obs. #$(522 + h)$ (22/FEB/2002 for $h = 10$). Rolling the estimation window 1-day-ahead and proceeding in the same fashion, we obtain $4018 - h$ individual forecasts, and associated forecast-errors. For $h = 10$, this amounts to 4008 forecasts and errors, respectively, over the period 22/FEB/2002 – 31/JAN/2018.

(b) *Out-of-sample VAFEM forecasts ($\tilde{\tilde{\mathbf{y}}}_{t+h|t}$) and errors*

We initialize a second rolling window of the fixed length $T_p \equiv T + p$, containing the first T_p errors of the M initial individual forecasts ($\mathbf{e}_{t+h|t}$) obtained in (a). With these, we estimate the VAR(p) specification (1) via multivariate LS (for $p = 1, 2, 3, 4$). Using these estimates, we obtain the first M h-period-ahead VAFEM forecasts ($\tilde{\tilde{\mathbf{y}}}_{t+h|t}$) according to Eq. (3), and compute the associated first M VAFEM forecast errors ($\mathbf{y}_{t+h} - \tilde{\tilde{\mathbf{y}}}_{t+h|t}$). We roll the estimation window, day-by-day, and for each window position, reestimate the VAR(p) specification (1) to obtain the h-period-ahead VAFEM forecasts, along with the associated VAFEM forecast errors. For instance, for $T = 50$ ($T = 750$), $h = 10, p = 4$, the initializing VAR(4) estimation window covers the period 22/FEB/2002 – 09/MAY/2002 (28/FEB/2005). For these two settings of (T, h, p), the rolling procedure generates 3945 and 3245 (M-tuple) VAFEM forecasts, along with the associated VAFEM forecast errors (until the sample ending date 31/JAN/2018).

(c) *Out-of-sample MSEs and MDM p-values*

We consider (i) the M-tuple VAFEM forecast errors obtained in (b), and (ii) the contemporaneous individual non-VAFEM forecast errors ($\mathbf{e}_{t+h|t}$) obtained in (a). For

[10] We need the first 22 observations (03/JAN/2000 – 02/FEB/2000) to initialize the HAR model; see Eq. (14).

all settings of (T, h, p), we conduct the evaluation over the same period covering the last 3245 observations (14/MAR/2005 – 31/JAN/2018). Over this evaluation period, we initialize a third rolling window of the fixed length $E = 500$, containing the first E VAFEM forecast-error M-tuples and their corresponding individual non-VAFEM analogues. For both of these, we compute the out-of-sample (sampling) MSEs over the E observations in the window. We apply the modified Diebold-Mariano test (MDM; Harvey et al., 1997) for equal predictive ability to the corresponding VAFEM and non-VAFEM forecast combinations, and compute the MDM p-values on the basis of the window observations.[11] We roll the window day-by-day, each time proceeding in the same fashion. This leads to 2746 out-of-sample MSEs per forecast-errors series, and the same number of MDM p-values for the chosen pairs of VAFEM and non-VAFEM forecast-error series over the S&P-500 sampling period.

Figure 2 about here

4.2 VAFEM forecasting results

4.2.1 VAFEM bias-correction

We start the out-of-sample analysis by illustrating the VAFEM forecast-bias correction (Proposition 1). Figure 2 illustrates this finite-sample property, using the 3-element forecast set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$, with $T = 50, p = 1, h = 1$. The left panels display the forecast errors of the non-VAFEM individual forecasts, the right panels the respective VAFEM analogues. In contrast to non-VAFEM HAR and ARFIMA (left panels), the non-VAFEM GARCH forecasts exhibit pronounced downward bias. In line with Proposition 1, the VAFEM treatment removes any bias from the individual forecasts (right panels).

Figure 3 about here

[11] We implemented the left-tailed version of the MDM test.

4.2.2 Forecast set $\{\textbf{GARCH}, \textbf{HAR}, \textbf{ARFIMA}\}$

Next, we analyze the (rolling) out-of-sample MSEs, and the MDM p-values (Section 4.1, Step (c)). Figure 3 shows the mean- and forecast-specific VAFEM MSEs relative to their non-VAFEM MSE-analogues (MSE-ratios 'VAFEM-MEAN-MSE/non-VAFEM-MEAN-MSE', and so forth), obtained for the 3-element forecast set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$, using the common parameters $p = 1, h = 1$, and the two sample sizes $T = 50$ (left panels), $T = 750$ (right panels).[12] In each panel, the MSE-ratios (blue lines) are assigned to the left axis. MSE-ratios falling below (exceeding) 1 indicate VAFEM accuracy gains (losses). The p-values of the MDM test for differences in VAFEM and non-VAFEM predicitve ability are assigned to the right axes (red lines). The grey-shaded areas contain those p-values that fall below the 5% level (significant accuracy gains via the VAFEM treatment).

Panel-rows 1 and 2 in Figure 3 indicate exceptional VAFEM accuracy gains for (i) the MEAN combination, and (ii) the GARCH forecasts. For $T = 50$ (left panels), both ratio-series exhibit MSE-reductions of 90% (and higher) over (quasi) the entire out-sample period. Even larger MSE-reductions emerge for $T = 750$ (right panels). All VAFEM accuracy gains are highly significant, with MDM p-values virtually equal to zero within the grey-shaded areas. *Prima facie*, the accuracy gains/losses for the HAR and ARFIMA forecasts appear less distinctive (Panel-rows 3, 4). For both forecasts, the sample-size increase from $T = 50$ to $T = 750$ is associated (on trend) with lower MSE-ratios. For $T = 750$, the ARFIMA forecasts exhibit VAFEM accuracy gains for 75.97% of the computed MSEs, where, for 45.12% of the MSEs, the accuracy gains are significant (MDM p-values within the grey-shaded area). By contrast, the HAR forecasts incur accuracy losses for 88.97% of the MSEs, and exhibit very few significant VAFEM accuracy gains for $T = 750$.

Tables 2 – 5 about here

[12]For convenience, we refer to the 'in-sample size' T as the 'sample size' throughout Section 4.

Tables 2 – 5 condense the MSE-results from Figure 3, and report summary statistics for a broad range of parameter settings, based on the set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$. Each table displays the results for the sample sizes $T \in \{50, 100, 250, 500, 750\}$, VAR lag-lengths $p = 1, 2, 3, 4$, and forecast horizons $h \in \{1, 2, 3, 5, 10\}$.[13] For each parameter setting, we consider the following three summary statistics, observed during the out-of-sample period:

(i) The percentage of (rolling) MSE ratios (VAFEM relative to non-VAFEM) falling below 1 (termed 'MSE ratios < 1 (in %)').

(ii) The percentage of (rolling) VAFEM predictive-ability improvements, significant at least at the 5% level (termed 'MDM p-values ≤ 0.05 (in %)').

(iii) The (single) MSE ratio (VAFEM/non-VAFEM) computed for the full evaluation period (14/MAR/2005 – 31/JAN/2018; 3245 observations) from Step (c) of the out-of-sample timeline in Section 4.1 (termed 'MSE ratio (full period)'). This MSE ratio serves as an aggregated measure of VAFEM accuracy gains/losses over the out-of-sample period. We also conducted the modified Diebold-Mariano tests for the full evaluation period. In Tables 2 – 5, we indicate significantly different MSEs (VAFEM MSE $<$ non-VAFEM MSE) at the 10, 5, and 1% levels by *,**,***, respectively.

We start our detailed analysis with the out-of-sample VAFEM results for the mean combination (Table 2) and the individual GARCH forecasts (Table 3), which appear similar. In both tables, we find highly significant accuracy gains with out-of-sample percentages, in terms of 'MSE-ratios ≤ 1' and 'MDM p-values $\leq 0.05'$, equal to 100% across the vast majority of the tabulated (T, p, h)-constellations. More concretely, we restrict our analysis to the VAR lag length $p = 1$, and ignore the two settings $(T = 50, h = 5)$, $(T = 50, h = 10)$ in Tables 2 and 3. For the remaining 23 $(T, p = 1, h)$ settings in Table 2, we find full-period

[13] In the main text, we only report results for the VAR lag-length $p = 1$. Results for $p = 2, 3, 4$ are provided in the supplementary file 'Tables_2-5'.

MSE ratios ranging between 0.05 and 0.28 (implying VAFEM-induced MSE reductions between 72% and 95%) for the mean combination. Analogously, for the GARCH forecasts in Table 3, the 23 full-period MSE ratios range between 0.04 and 0.20, indicating VAFEM-induced MSE reductions between 80% and 96%. We ascribe these comprehensive VAFEM accuracy gains to the severe biasedness of the initial (non-VAFEM) GARCH forecast series (Remark 2(i), Section 2.3).[14]

Next, we discuss VAFEM results for the individual ARFIMA forecasts. Figure 3 displays the (rolling) ARFIMA MSE ratios for the special cases ($p = 1, h = 1$) and $T = 50, 750$ (left and right panels in Row 4). Mere visual inspection suggests that, *ceteris paribus*, the sample-size increase entails clear-cut VAFEM accuracy gains (in the form of reduced ARFIMA MSE-ratios). Table 5 corroborates this impression. For brevity, we only analyze the case $p = 1$.[15]

(i) Let us consider the forecast horizon h as fixed. Then, the stepwise sample-size increases ($T = 50, T = 100, \ldots$) are accompanied by reductions in the full-period ARFIMA MSE-ratios. For $h = 1$, the sequence of decreasing (full-period) MSE ratios is 1.02, 0.83, 0.77, 0.76, 0.75 (implying a 25% VAFEM MSE-reduction for $T = 750$).[16] This tendency of observing increased VAFEM accuracy gains (decreasing full-period MSE-ratios) with increasing sample sizes, *ceteris paribus*, is strictly consistent with our theoretical results in Section 2.3; see Remark 2(ii).

(ii) Let us consider the sample size T as fixed. Then, the stepwise increases in the forecast horizons ($h = 1, h = 2, \ldots$) entail increasing full-period ARFIMA MSE-ratios; e.g. 0.75, 0.87, 1.08, 1.18, 1.37 for $T = 750$. This pattern of observing higher VAFEM accuracy gains for nearby forecast horizons coincides with our theoretical result in Remark

[14]The biasedness of the non-VAFEM GARCH forecasts is shown in Figure 2 (upper left panel). This GARCH-biasedness induces biased non-VAFEM mean-combination forecasts (not shown here).

[15]The same qualitative features also hold for $p = 2, 3, 4$; see Table 5 in the supplementary file.

[16]We note that the other two summary statistics in Table 5 typically increase with increasing sample size.

2(iii), Section 2.3.

Finally, we present the VAFEM results for the HAR forecasts. Figure 3 (panels in Row 3) gives a first impression. For the two parameter settings, $p = 1, h = 1, T \in \{50, 750\}$, the (rolling) HAR MSE-ratios exceed 1 most of the time. Table 4 substantiates this indication of VAFEM accuracy losses. Irrespective of the parameter setting, all full-period HAR MSE-ratios exceed 1. We emphasize, however, that the VAFEM results for the HAR forecasts are still widely consistent with our theoretical findings presented in Remark 2 (Section 2.3).[17] We give two complementary explanations of the VAFEM accuracy losses for the HAR forecasts. (1) The initial non-VAFEM HAR forecasts are unbiased (Figure 2, Row 2, left panel). Thus, a promotive precondition for realizing VAFEM accuracy gains according to Remark 2(i) (Section 2.3) does not apply. (2) In view of VAR Eq. (1), we would expect to find (substantial) VAFEM accuracy gains, if the (initial) non-VAFEM HAR forecast-errors (i) featured autocorrelation, and/or (ii) interrelated significantly with the errors of the remaining non-VAFEM forecasts from the set.

For our S&P 500 data, the initial non-VAFEM HAR forecast-errors (i) do not exhibit autocorrelation (see Section 4.2.3). (ii) Also, the non-VAFEM HAR forecast-errors do not effectively interrelate with the errors of the GARCH and ARFIMA forecasts in the joint VAR modeling. Overall, the non-VAFEM HAR forecasts already feature sufficiently good out-of-sample properties, leaving virtually no room for further VAFEM accuracy improvement (under the forecast set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$).

Figure 4 about here

[17]For fixed forecast horizon h (sample size T), the HAR forecasts (largely) show decreasing (increasing) full-period MSE-ratios under increases in T (h).

4.2.3 Smaller sets

We briefly outline results for some selected 2- and 1-element forecast sets.[18] Figure 4 displays (rolling) MSE-ratios and MDM p-values for the sets $\{\text{HAR}, \text{ARFIMA}\}$ [Panels (a)–(c)], and $\{\text{GARCH}, \text{ARFIMA}\}$ [Panels (d)–(f)], for $T = 750, p = 1, h = 1$. Under $\{\text{HAR}, \text{ARFIMA}\}$, we observe (i) balanced VAFEM accuracy gains/losses for the combination mean [Panel (a)], (ii) persistent losses for the HAR forecasts [Panel (b)], and (iii) substantial VAFEM accuracy gains for the ARFIMA forecasts [Panel (c); full-period MSE ratio: 0.76, MDM p-value: 0.07].

Under the set $\{\text{GARCH}, \text{ARFIMA}\}$, we obtain clear-cut VAFEM accuracy gains for (i) both individual forecasts [GARCH, Panel (e); ARFIMA, Panel (f)], plus (ii) their mean combination [Panel (d)]. While the accuracy gains in Panels (d) and (e) are unambiguous, the gains for the ARFIMA forecasts in Panel (f) are less extreme, but nonetheless significant (full-period MSE ratio: 0.82, MDM p-value: 0.08).

Figure 5 about here

Figure 5 contains the (rolling) MSE-ratios under the single-element ($M = 1$) forecast sets $\{\text{HAR}\}$ [Panels (a), (b)] and $\{\text{ARFIMA}\}$ [Panels (c), (d)], using $p = 1, h = 1$, and the two sample sizes $T = 50, 750$. For $M = 1$, the forecast-error VAR in Eq. (1) reduces to a univariate autoregression of order p. The VAFEM procedure then confines itself to adjusting the forecasts in Eq. (3) for autocorrelation in past forecast errors. As shown in Panels (c) and (d), the ARFIMA forecasts realize VAFEM accuracy gains through this autocorrelation adjustment (full-period MSE ratio for $T = 750$ [$T = 50$]: 0.83 [0.89]; MDM p-values: 0.09 [0.22]). By contrast, the VAFEM-HAR forecasts in Panels (a), (b) do not exhibit accuracy improvements. Obviously, the initial non-VAFEM HAR model is already capable of generating (first-order) autocorrelation-free forecast errors.

[18] We only show graphical results. Full graph and table disclosure for all forecast sets, analogous to Figures 3–5 and Tables 2–5, is available upon request.

5 Concluding remarks

In this paper, we formalize a 'post-processing' 3-step procedure (called VAFEM) for improving the MSE-accuracy of linear-convex forecast combinations in a classic VAR framework, in which the multivariate LS estimators are consistent. We (i) prove probabilistic characteristics of the procedure (forecast bias-correction, MSE-reductions under known VAR parameters), and (ii) establish settings under which the VAFEM approach has the highest potential to generate forecasting improvements in real-world applications. Accuracy gains are most likely to occur, if (i) some of the original individual forecasts are biased, (ii) a sufficiently large number of historical forecast errors are available (for accurately estimating the VAR parameters), and (iii) the forecast horizons are small. Our theoretical results are to a large extent confirmed empirically in an out-of-sample volatility-forecasting analysis, using daily S&P 500 data over an 18-year time span.

At various points of our analysis, we emphasize that 'large VAR-parameter estimation errors' (in Step 1 of the VAFEM procedure) might ultimately bring about accuracy-losses. A closely related practical issue concerns the adequate in-sample VAR model selection. Here, two crucial aspects include (i) the VAR lag-length selection (p), and (ii) the 'correct' VAR specification in Eq. (1). Pertaining to (i), there are several manifest techniques to experiment with, like the Akaike/Bayesian information criteria, and the discounted MSFE(δ) criterion (Stock and Watson, 2004, Eq. (4)). With respect to (ii), Weigt and Wilfling (2021) successfully apply a heteroskedastic VAR with time-varying parameters, using a Bayesian estimation technique. Additionally, it remains to investigate the accuracy-effects of the VAFEM methodology on the numerous other (non-linear-convex) combination forecasts suggested in the literature. We leave all these topics for future research.

References

Audrino, F., Huang, C. & Okhrin, O. (2019). Flexible HAR model for realized volatility. Studies in Nonlinear Dynamics & Econometrics, 23(3), 1-22.

Bates, J.M. & Granger, C. (1969). The combination of forecasts. Journal of the Operational Research Society, 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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-328.

Bollerslev, T., Patton, A.J. & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for volatility forecasting. Journal of Econometrics, 192, 1-18.

Capistrán, C. & Timmermann, A. (2009). Forecast combination with entry and exit of experts. Journal of Business & Economic Statistics, 27, 428-440.

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5, 559-583.

Corsi, F. (2009). A simple long-memory model of realized volatility. Journal of Financial Econometrics, 7, 174-196.

Doornik, J.A. & Ooms, M. (2004). Inference and forecasting for ARFIMA models with an application to US and UK inflation. Studies in Nonlinear Dynamics & Econometrics, 8(2), Article 14, 1-23.

Dufour, J. (1985). Unbiasedness of predictions from estimated vector autoregressions. Econometric Theory, 1, 387-402.

Gaglianone, W.P. & Lima, L.R. (2014). Constructing optimal density forecasts from point forecast combinations. Journal of Applied Econometrics, 29, 736-757.

Garrat, A., Vahey, S.P. & Zhang, Y. (2019). Real-time forecast combinations for the oil price. Journal of Applied Econometrics, 34, 456-462.

Guidolin, M. & Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. Journal of Econometrics, 150, 297-311.

Granger, C.W. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis, 1, 15-39.

Harvey, D., Leybourne, S. & Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13, 281-291.

Hosking, J.R.M. (1981). Fractional differencing. Biometrica, 68, 165-176.

Izzeldin, M., Hassan, M.K., Pappas, V. & Tsionas, M. (2019). Forecasting realised volatility using ARFIMA and HAR models. Quantitative Finance, 19, 1627-1638.

Koopman, S.J., Jungbacker, B. & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance, 12, 445-475.

Lütkepohl, H. (2005). New introduction to multiple time series analysis. Berlin: Springer.

Nowotarski, J., Raviv, E., Trueck, S. & Weron, R. (2014). An empirical comparison of alternative schemes for combining electricity spot price forecasts. Energy Economics, 46, 395-412.

Pesaran, M.H. & Pick, A. (2011). Forecast combination across estimation windows. Journal of Business & Economic Statistics, 29, 307-318.

Rapach, D.E., Strauss, J.J. & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies, 23, 821-862.

Reinsel, G.C. (1997). Elements of multivariate time series analysis. Second Edition. New York: Springer.

Rossi, B. (2013). Advances in forecasting under instability. In: G. Elliott & A., Timmermann (Eds.) Handbook of economic forecasting (Vol. 2B). Amsterdam: Elsevier, pp. 1203-1324.

Stock, J.H. & Watson, M.W. (2004). Combination forecasts of output growth in a seven-country data set. Journal of Forecasting, 23, 405-430.

Taylor, S.J. (2005). Asset price dynamics, volatility, and prediction. Princeton: Princeton University Press.

Taylor, J.W. (2020). Forecast combinations for value at risk and expected shortfall. International Journal of Forecasting, 36, 428-441.

Timmermann, A. (2006). Forecast combinations. In: G. Elliott et al. (Eds.) Handbook of economic forecasting (Vol. 1). Amsterdam: Elsevier, pp. 135-196.

Weigt, T. & Wilfling, B. (2021). An approach to increasing forecast-combination accuracy through VAR error modeling. Journal of Forecasting, 40, 686-699.

Zhang, B. (2019). Real-time inflation forecast combination for time-varying coefficient models. Journal of Forecasting, 38, 175-191.

SUPPLEMENTARY MATERIAL

Pdf-file 'Tables_2-5': The file continues the Tables 2–5 from the main text, reporting results for the VAR lag-lengths $p = 2, 3, 4$.

Appendix: A. Proof(s) and remark(s)

Proof of Proposition 3. We first express the VAR(p) model from Eq. (1) in the VAR(1) (companion) form. Defining the $Mp \times 1$ column vectors

$$\begin{aligned}
\mathbf{E}_{t,h} &\equiv \left(\mathbf{e}'_{t+h|t}, \mathbf{e}'_{(t-1)+h|t-1}, \ldots, \mathbf{e}'_{(t-p+1)+h|t-p+1}\right)', \\
\boldsymbol{\xi}_{t+h} &\equiv \left(\boldsymbol{\epsilon}'_{t+h}, \mathbf{0}', \ldots, \mathbf{0}'\right)', \\
\mathbf{N} &\equiv \left(\boldsymbol{\nu}', \mathbf{0}', \ldots, \mathbf{0}'\right)',
\end{aligned}$$

and the $Mp \times Mp$ matrix

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_M & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_M & \mathbf{0} \end{pmatrix},$$

we obtain the companion form as

$$\mathbf{E}_{t,h} = \mathbf{N} + \mathbf{A}\mathbf{E}_{t-1,h} + \boldsymbol{\xi}_{t+h}. \tag{A.1}$$

Given our stability assumption in Section 2.1, the VAR process in Eq. (A.1) has the moving average representation

$$\mathbf{E}_{t,h} = \mathbb{E}(\mathbf{E}_{t,h}) + \sum_{i=0}^{\infty} \mathbf{A}^i \boldsymbol{\xi}_{t+h-i}. \tag{A.2}$$

Next, we define the $M \times Mp$ matrix $\mathbf{J}_1 \equiv \begin{pmatrix} \mathbf{I}_M & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}$, and the $M \times M$ matrix $\boldsymbol{\lambda}_i \equiv \mathbf{J}_1 \mathbf{A}^i \mathbf{J}_1'$. It is straightforward to verify that the matrix $\boldsymbol{\lambda}_i$ thus defined is identical to the matrix $\boldsymbol{\lambda}_i$ from Proposition 2, so that $\boldsymbol{\lambda}_i$ has the two representations $\boldsymbol{\lambda}_i = \mathbf{J}_1 \mathbf{A}^i \mathbf{J}_1' = \mathbf{J}\boldsymbol{\mathcal{A}}^i\mathbf{J}'$ (with $\mathbf{J}$ and $\boldsymbol{\mathcal{A}}$ from Proposition 2). We now have

$$\begin{aligned} \mathbf{e}_{t+h|t} &= \mathbf{J}_1 \mathbf{E}_{t,h} \\ &= \mathbf{J}_1 \mathbb{E}(\mathbf{E}_{t,h}) + \sum_{i=0}^{\infty} \mathbf{J}_1 \mathbf{A}^i \boldsymbol{\xi}_{t+h-i} \\ &= \mathbb{E}(\mathbf{e}_{t+h|t}) + \sum_{i=0}^{\infty} \boldsymbol{\lambda}_i \boldsymbol{\epsilon}_{t+h-i}. \end{aligned} \tag{A.3}$$

We note that the matrices $\mathbb{E}(\mathbf{e}_{t+h|t})\mathbb{E}(\mathbf{e}_{t+h|t})'$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ from Eq. (1) are positive semidefinite. Then, using the $M \times 1$ vector $\mathbf{y}_{t+h} = (y_{t+h}, \dots, y_{t+h})'$ and the representation in

Eq. (A.3), we obtain

$$
\begin{aligned}
\text{MSE}(\hat{y}^{\text{comb}}_{t+h|t}) &= \mathbb{E}\left[\left(y_{t+h} - \mathbf{w}'\hat{\mathbf{y}}_{t+h|t}\right)^2\right] \\
&= \mathbf{w}'\,\mathbb{E}\left[\left(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}\right)\left(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}\right)'\right]\mathbf{w} \\
&= \mathbf{w}'\,\mathbb{E}\left[\mathbf{e}_{t+h|t}\mathbf{e}'_{t+h|t}\right]\mathbf{w} \\
&= \mathbf{w}'\,\mathbb{E}\left[\left(\mathbb{E}(\mathbf{e}_{t+h|t}) + \sum_{i=0}^{\infty}\boldsymbol{\lambda}_i\boldsymbol{\epsilon}_{t+h-i}\right)\left(\mathbb{E}(\mathbf{e}_{t+h|t}) + \sum_{i=0}^{\infty}\boldsymbol{\lambda}_i\boldsymbol{\epsilon}_{t+h-i}\right)'\right]\mathbf{w} \\
&= \mathbf{w}'\left\{\mathbb{E}(\mathbf{e}_{t+h|t})\mathbb{E}(\mathbf{e}_{t+h|t})' + \mathbb{E}\left[\left(\sum_{i=0}^{\infty}\boldsymbol{\lambda}_i\boldsymbol{\epsilon}_{t+h-i}\right)\left(\sum_{i=0}^{\infty}\boldsymbol{\lambda}_i\boldsymbol{\epsilon}_{t+h-i}\right)'\right]\right\}\mathbf{w} \\
&= \mathbf{w}'\,\mathbb{E}(\mathbf{e}_{t+h|t})\mathbb{E}(\mathbf{e}_{t+h|t})'\,\mathbf{w} + \mathbf{w}'\left(\sum_{i=0}^{\infty}\boldsymbol{\lambda}_i\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\boldsymbol{\lambda}'_i\right)\mathbf{w} \qquad \text{(A.4)} \\
&\geq \sum_{i=0}^{\infty}\mathbf{w}'\boldsymbol{\lambda}_i\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\left(\mathbf{w}'\boldsymbol{\lambda}_i\right)' \qquad \text{(A.5)} \\
&= \sum_{i=0}^{h-1}\mathbf{w}'\boldsymbol{\lambda}_i\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\left(\mathbf{w}'\boldsymbol{\lambda}_i\right)' + \sum_{i=h}^{\infty}\mathbf{w}'\boldsymbol{\lambda}_i\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\left(\mathbf{w}'\boldsymbol{\lambda}_i\right)' \qquad \text{(A.6)} \\
&\geq \mathbf{w}'\left(\sum_{i=0}^{h-1}\boldsymbol{\lambda}_i\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\boldsymbol{\lambda}'_i\right)\mathbf{w}. \qquad \text{(A.7)}
\end{aligned}
$$

In view of Eq. (6) and Eq. (8) in Remark 1, this completes the proof. □

Remark 3. *We note the following two issues.*

(i) If all M (initial) individual forecast models are unbiased, we have $\mathbb{E}(\mathbf{e}_{t+h|t}) = \mathbf{0}$ in Eq. (A.4), and obtain equality in (A.5).

(ii) When the forecast horizon h becomes infinitely large ($h \to \infty$), there is no need to split up the sum in (A.6), and we obtain equality in (A.7).

Figures and Tables

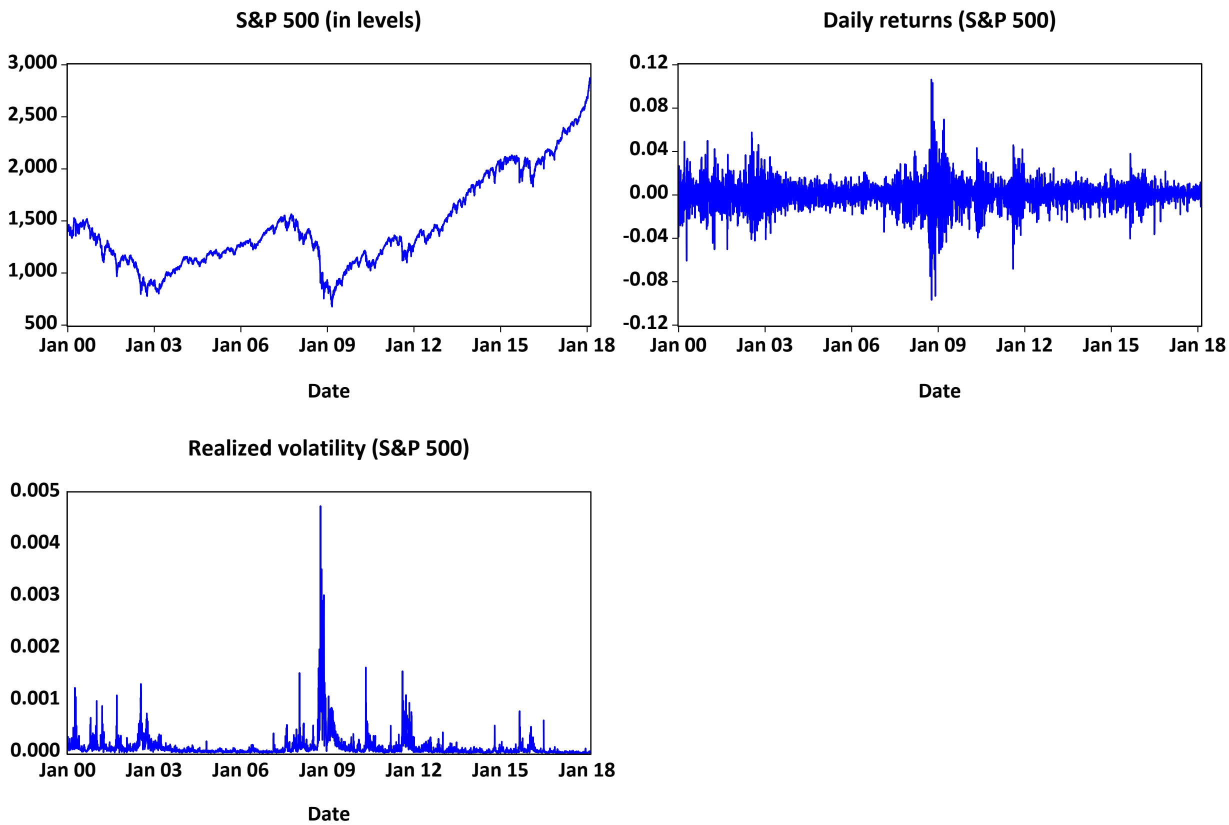

Figure 1: S&P 500 (in levels), daily returns, and realized volatility

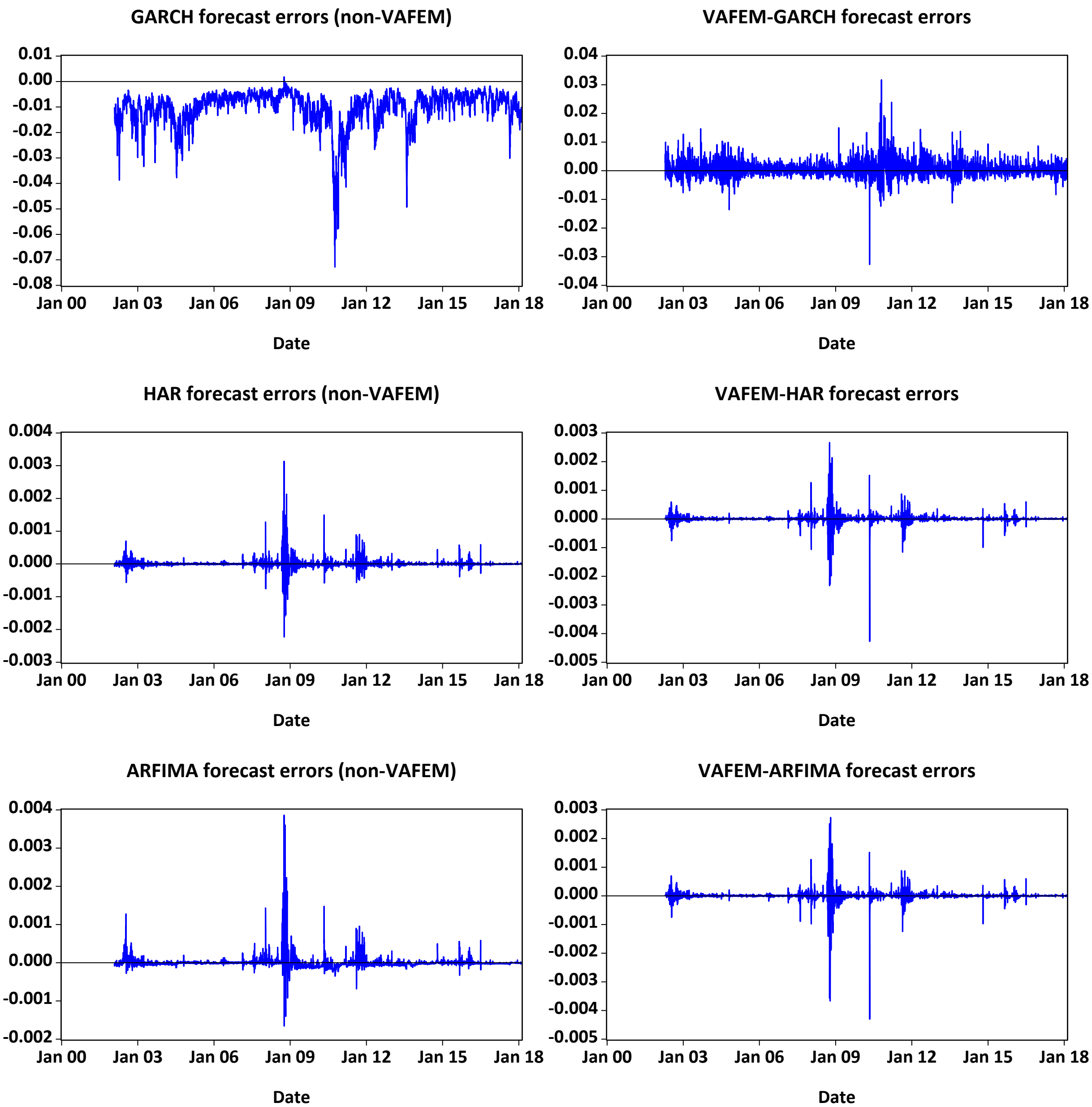

Figure 2: VAFEM forecast-bias correction for the forecast set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$ and $(T = 50, p = 1, h = 1)$

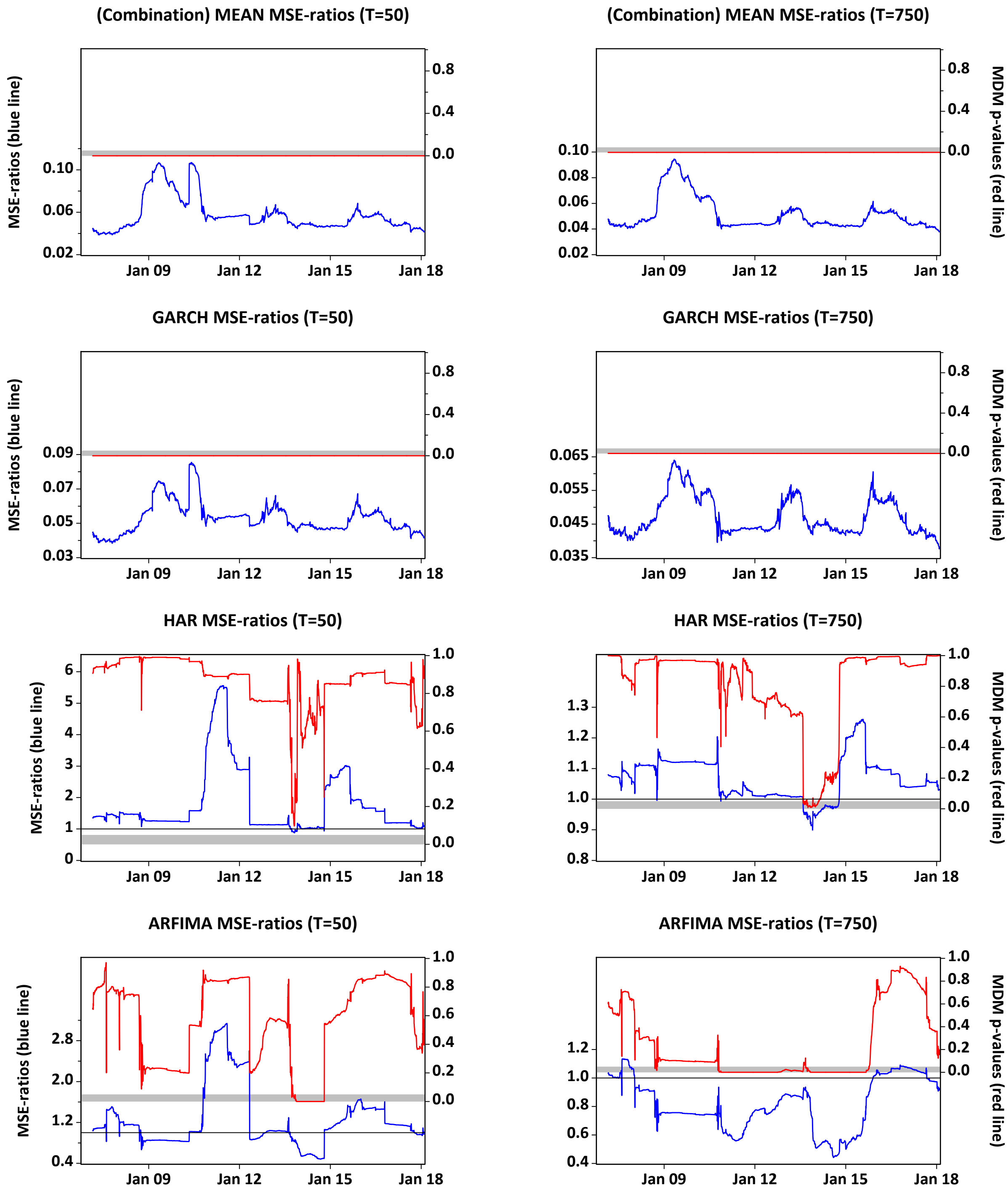

Figure 3: Rolling MSE-ratios (VAFEM/non-VAFEM; blue lines; left axes) and modified Diebold-Mariano (MDM) p-values (red lines; right axes) for the forecast set $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$, $T \in \{50, 750\}, p = 1, h = 1$

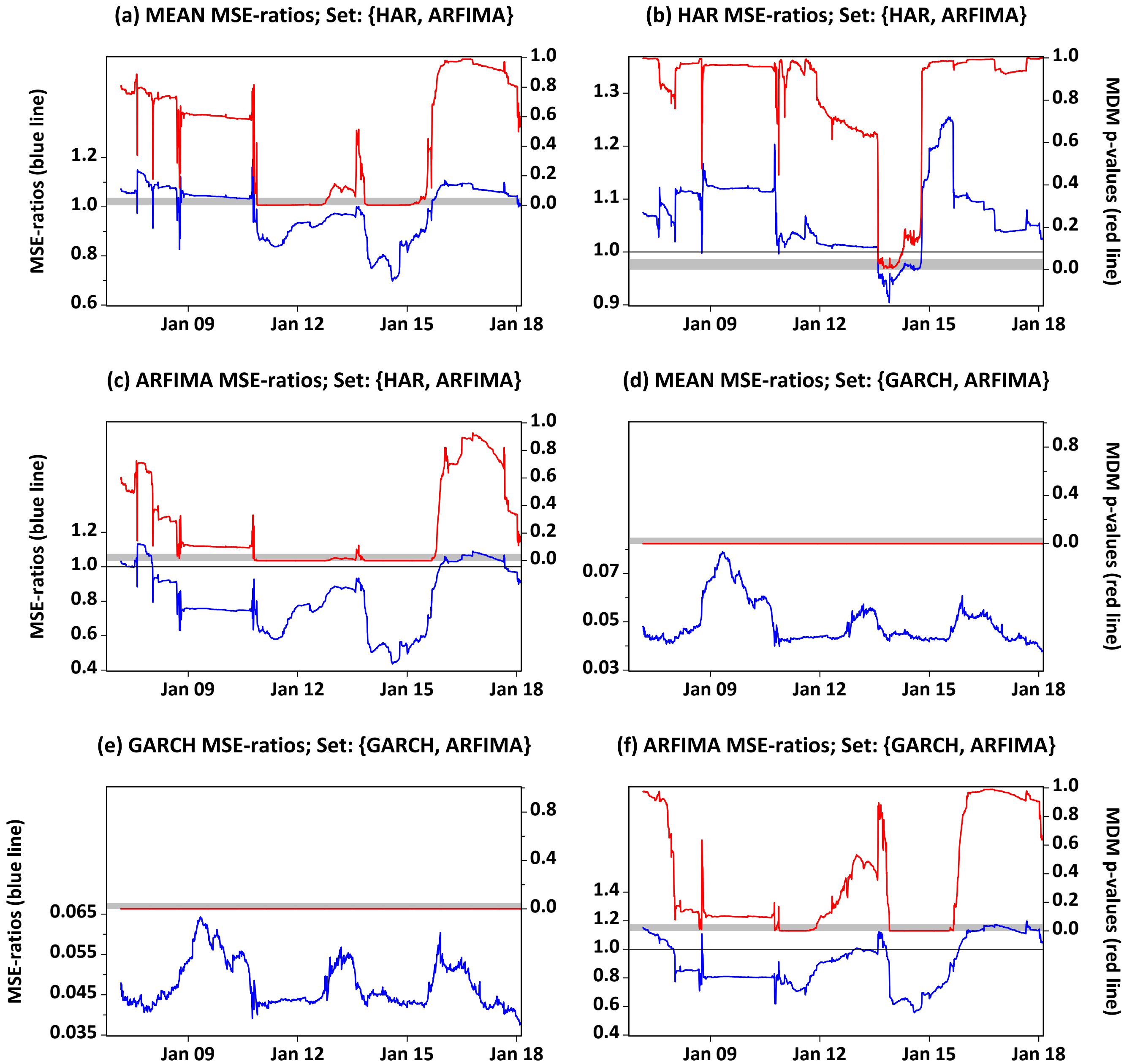

Figure 4: Rolling MSE-ratios (VAFEM/non-VAFEM) and MDM p-values for the forecast sets $\{\text{HAR}, \text{ARFIMA}\}$ [Panels (a)–(c)], $\{\text{GARCH}, \text{ARFIMA}\}$ [Panels (d)–(f)], for $T = 750, p = 1, h = 1$

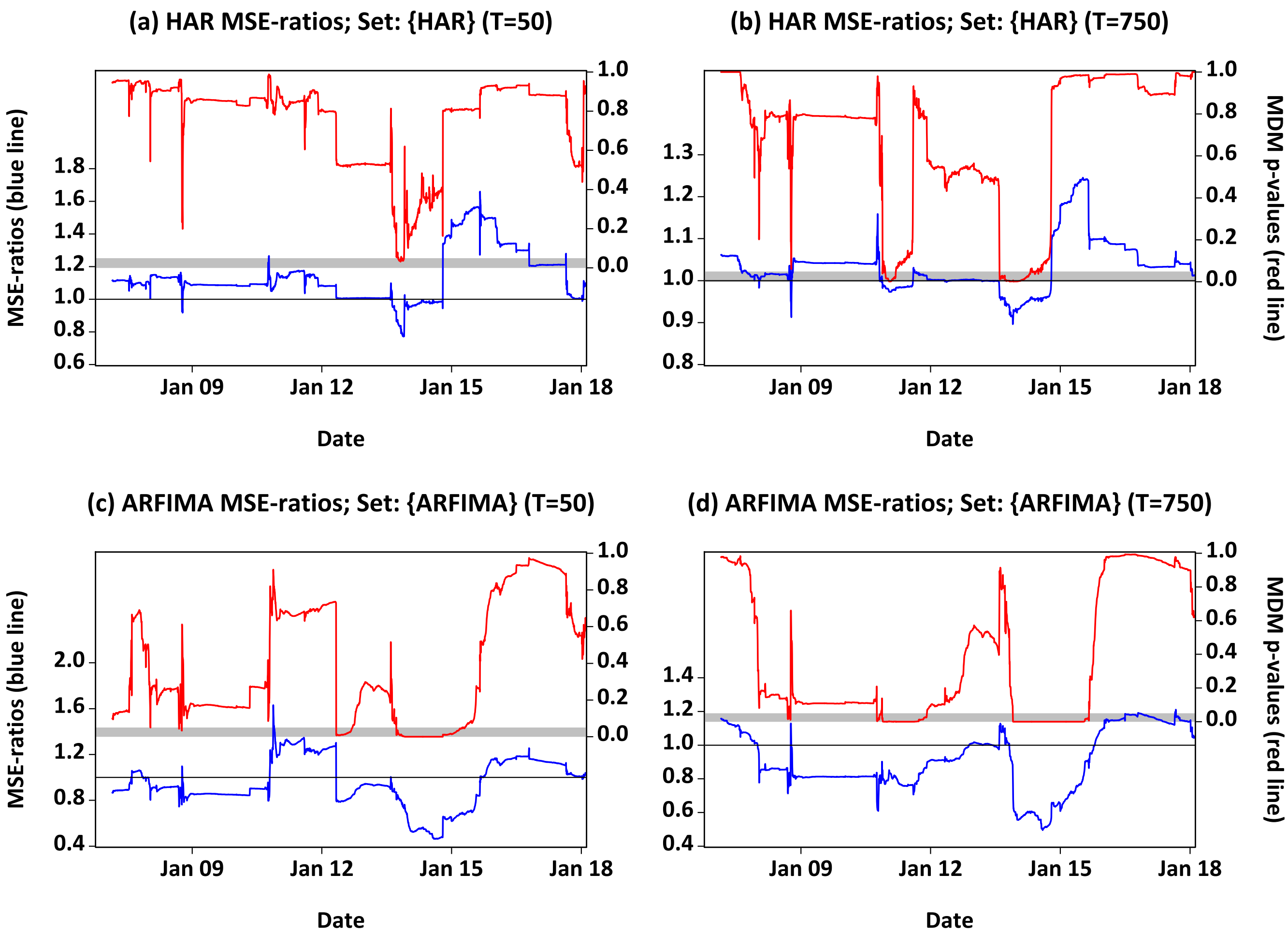

Figure 5: Rolling MSE-ratios (VAFEM/non-VAFEM) and MDM p-values for the forecast sets {HAR} [Panels (a), (b)], {ARFIMA} [Panels (c), (d)], for $p = 1, h = 1$, and $T = 50$ (left panels), $T = 750$ (right panels)

Table 1: Summary statistics (daily observations, 1 January 2000 until 31 January 2018)

	Realized volatility (S&P 500)	Daily returns (S&P 500)
Number of obs.	4539	4538
Minimum	1.22×10^{-6}	-0.0969
Maximum	0.0047	0.1064
Mean	0.0001	0.0001
Skewness	8.1957	-0.1960
Kurtosis	103.0757	11.3045
ACF(1)	0.7670	-0.0708
$Q(1)$ (p-value)	0.0000	0.0000
JB (p-value)	0.0000	0.0000

Note: ACF(1) denotes the value of the (sample) autocorrelation function at lag 1. $Q(1)$ is the Ljung-Box test for autocorrelation at lag 1. JB denotes the Jarque-Bera normality test.

Table 2: Out-of-sample forecast evaluation — Combination: MEAN (VAFEM versus non-VAFEM), Forecast set: $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 1$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.06***	0.05***	0.05***	0.05***	0.05***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.12***	0.08***	0.07***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	96.14	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.20***	0.10***	0.09***	0.09***	0.09***
	$h = 5$	MSE ratios < 1 (in %)	81.79	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	81.79	100.00	100.00	100.00	100.00
		MSE ratio (full period)	1.65	0.14***	0.11***	0.10***	0.10***
	$h = 10$	MSE ratios < 1 (in %)	56.23	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	41.84	81.79	100.00	100.00	100.00
		MSE ratio (full period)	99.24	0.28***	0.14***	0.13***	0.13***

Note: 'MSE ratios < 1 (in %)' represents the percentage of MSE ratios (VAFEM relative to non-VAFEM) less than 1 (indicating lower VAFEM than non-VAFEM MSEs), observed during the out-sample period. 'MDM p-values ≤ 0.05 (in %)' is the percentage of VAFEM-MSE improvements, significant at least at the 5% level, according to the left-tailed version of the modified Diebold-Mariano test for equal predictive ability, observed during the out-of-sample period. 'MSE ratio (full period)' represents the MSE ratio (VAFEM/non-VAFEM), computed for the full evaluation period from Step (c) of the timeline in Section 4.1 (3245 observations between 14/MAR/2005 and 31/JAN/2018). *, **, *** denote significantly different MSEs according to the (left-tailed) modified Diebold-Mariano test at 10, 5, and 1% levels, respectively.

Table continued (with results for the VAR lag-lengths $p = 2, 3, 4$) in the supplementary file 'Tables_2-5'.

Table 3: Out-of-sample forecast evaluation — Individual forecast: GARCH (VAFEM versus non-VAFEM), Forecast set: {GARCH, HAR, ARFIMA}

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 1$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.05***	0.05***	0.05***	0.05***	0.04***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.11***	0.08***	0.07***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	96.61	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.17***	0.10***	0.09***	0.09***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	81.79	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	81.79	100.00	100.00	100.00	100.00
		MSE ratio (full period)	1.24	0.13***	0.11***	0.10***	0.10***
	$h = 10$	MSE ratios < 1 (in %)	81.72	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	47.27	100.00	100.00	100.00	100.00
		MSE ratio (full period)	67.46	0.20***	0.14***	0.13***	0.13***

Notes: Analogous to the notes for Table 2.

Table continued (with results for the VAR lag-lengths $p = 2, 3, 4$) in the supplementary file 'Tables_2-5'.

Table 4: Out-of-sample forecast evaluation — Individual forecast: HAR (VAFEM versus non-VAFEM), Forecast set: $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$

Lag length (p) /	Horizon (h)		Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 1$	$h = 1$	MSE ratios < 1 (in %)	2.00	7.36	7.28	5.79	11.03
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.04	4.33
		MSE ratio (full period)	1.52	1.21	1.12	1.10	1.10
	$h = 2$	MSE ratios < 1 (in %)	0.95	10.89	7.25	4.95	7.36
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.22
		MSE ratio (full period)	2.41	1.40	1.16	1.12	1.10
	$h = 3$	MSE ratios < 1 (in %)	0.22	8.38	11.54	12.31	14.13
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.33	7.32
		MSE ratio (full period)	4.15	1.75	1.30	1.22	1.20
	$h = 5$	MSE ratios < 1 (in %)	0.58	10.20	3.35	0.11	8.99
		MDM p-values ≤ 0.05 (in %)	0.00	0.07	0.00	0.00	8.05
		MSE ratio (full period)	32.88	2.34	1.31	1.23	1.22
	$h = 10$	MSE ratios < 1 (in %)	3.64	10.16	7.32	1.17	9.29
		MDM p-values ≤ 0.05 (in %)	0.00	2.48	0.25	0.00	4.66
		MSE ratio (full period)	> 100	20.52	1.77	1.56	1.55

Notes: Analogous to the notes for Table 2.

Table continued (with results for the VAR lag-lengths $p = 2, 3, 4$) in the supplementary file 'Tables_2-5'.

Table 5: Out-of-sample forecast evaluation — Individual forecast: ARFIMA (VAFEM versus non-VAFEM), Forecast set: $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$

Lag length (p) /			Sample size				
	Horizon (h)		$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 1$	$h = 1$	MSE ratios < 1 (in %)	32.81	58.81	73.78	73.89	75.97
		MDM p-values ≤ 0.05 (in %)	9.58	16.39	36.71	38.09	45.12
		MSE ratio (full period)	1.02	0.83	0.77*	0.76*	0.75*
	$h = 2$	MSE ratios < 1 (in %)	11.54	39.55	56.52	69.56	72.54
		MDM p-values ≤ 0.05 (in %)	10.45	14.06	26.44	30.15	36.89
		MSE ratio (full period)	1.87	1.05	0.90	0.88	0.87
	$h = 3$	MSE ratios < 1 (in %)	10.82	14.97	32.01	34.01	53.68
		MDM p-values ≤ 0.05 (in %)	10.82	13.84	19.63	27.82	28.22
		MSE ratio (full period)	3.14	1.46	1.15	1.10	1.08
	$h = 5$	MSE ratios < 1 (in %)	10.85	14.79	31.79	35.51	53.13
		MDM p-values ≤ 0.05 (in %)	10.82	12.20	16.02	29.06	35.32
		MSE ratio (full period)	26.02	2.08	1.27	1.19	1.18
	$h = 10$	MSE ratios < 1 (in %)	9.10	12.31	24.40	34.63	50.76
		MDM p-values ≤ 0.05 (in %)	0.04	11.00	17.92	22.29	32.67
		MSE ratio (full period)	> 100	18.54	1.71	1.40	1.37

Notes: Analogous to the notes for Table 2.

Table continued (with results for the VAR lag-lengths $p = 2, 3, 4$) in the supplementary file 'Tables_2-5'.

Supplementary Material

A procedure for upgrading linear-convex combination forecasts with an application to volatility prediction

- This supplement continues the Tables 2–5 from the main text, reporting results for the VAR lag-lengths $p = 2, 3, 4$.

Table 2: Continued. Out-of-sample forecast evaluation — Combination: MEAN (VAFEM versus non-VAFEM), Forecast set: $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$

Lag length (p) /	Horizon (h)		Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 2$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.06***	0.05***	0.05***	0.05***	0.05***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.12***	0.09***	0.07***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	96.21	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	88.82	81.79	100.00	100.00	100.00
		MSE ratio (full period)	0.35***	0.60	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	70.03	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	70.03	81.79	100.00	100.00	100.00
		MSE ratio (full period)	12.04	2.12	0.11***	0.10***	0.10***
	$h = 10$	MSE ratios < 1 (in %)	51.64	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	43.59	60.23	100.00	100.00	100.00
		MSE ratio (full period)	> 100	> 100	0.17***	0.12***	0.12***
$p = 3$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.07***	0.06***	0.05***	0.05***	0.05***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.15***	0.11***	0.08***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	96.14	95.92	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	77.79	92.13	100.00	100.00	100.00
		MSE ratio (full period)	0.44***	0.27***	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	70.03	81.76	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	61.58	81.76	100.00	100.00	100.00
		MSE ratio (full period)	34.89	26.41	0.11***	0.10***	0.09***
	$h = 10$	MSE ratios < 1 (in %)	34.05	60.09	81.79	81.79	81.79
		MDM p-values ≤ 0.05 (in %)	15.80	38.86	81.79	81.79	81.79
		MSE ratio (full period)	> 100	> 100	16.65	1.66	1.48

Continued on next page.

Table 2: Continued.

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 4$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.09***	0.06***	0.05***	0.05***	0.05***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.20***	0.11***	0.08***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	91.88	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	67.33	93.59	100.00	100.00	100.00
		MSE ratio (full period)	0.55**	0.22***	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	63.04	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	51.78	74.22	100.00	100.00	100.00
		MSE ratio (full period)	40.02	33.60	0.11***	0.09***	0.09***
	$h = 10$	MSE ratios < 1 (in %)	15.29	57.06	81.76	81.79	81.79
		MDM p-values ≤ 0.05 (in %)	8.99	43.26	78.33	81.79	81.79
		MSE ratio (full period)	> 100	> 100	> 100	3.75	2.98

Notes: Analogous to the notes for Table 2 in the main text.

Table 3: Continued. Out-of-sample forecast evaluation — Individual forecast: GARCH (VAFEM versus non-VAFEM), Forecast set: {GARCH, HAR, ARFIMA}

Lag length (p) /	Horizon (h)		Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 2$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.06***	0.05***	0.05***	0.04***	0.04***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.11***	0.09***	0.07***	0.07***	0.07***
	$h = 3$	MSE ratios < 1 (in %)	96.69	92.24	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	95.19	81.79	100.00	100.00	100.00
		MSE ratio (full period)	0.26***	0.47*	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	70.03	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	70.03	81.79	100.00	100.00	100.00
		MSE ratio (full period)	7.32	1.05	0.10***	0.10***	0.09***
	$h = 10$	MSE ratios < 1 (in %)	69.85	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	43.77	81.79	100.00	100.00	100.00
		MSE ratio (full period)	> 100	> 100	0.16***	0.12***	0.11***
$p = 3$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.07***	0.06***	0.05***	0.05***	0.04***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.13***	0.10***	0.07***	0.07***	0.06***
	$h = 3$	MSE ratios < 1 (in %)	96.50	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	94.94	93.41	100.00	100.00	100.00
		MSE ratio (full period)	0.32***	0.20***	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	70.03	81.76	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	70.03	81.76	100.00	100.00	100.00
		MSE ratio (full period)	24.29	22.37	0.10***	0.09***	0.09***
	$h = 10$	MSE ratios < 1 (in %)	34.16	60.12	81.79	81.79	81.79
		MDM p-values ≤ 0.05 (in %)	16.02	60.09	81.79	81.79	81.79
		MSE ratio (full period)	> 100	> 100	6.52	0.85	0.82

Continued on next page.

Table 3: Continued.

Lag length (p) /	Horizon (h)		Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 4$	$h = 1$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.08***	0.06***	0.05***	0.04***	0.04***
	$h = 2$	MSE ratios < 1 (in %)	100.00	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	100.00	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.16***	0.10***	0.07***	0.07***	0.06***
	$h = 3$	MSE ratios < 1 (in %)	93.48	100.00	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	81.46	100.00	100.00	100.00	100.00
		MSE ratio (full period)	0.44***	0.16***	0.09***	0.08***	0.08***
	$h = 5$	MSE ratios < 1 (in %)	69.99	81.79	100.00	100.00	100.00
		MDM p-values ≤ 0.05 (in %)	57.65	81.76	100.00	100.00	100.00
		MSE ratio (full period)	60.01	35.30	0.10***	0.09***	0.09***
	$h = 10$	MSE ratios < 1 (in %)	34.27	60.12	81.79	81.79	81.79
		MDM p-values ≤ 0.05 (in %)	15.00	52.44	78.37	81.79	81.79
		MSE ratio (full period)	> 100	> 100	33.10	1.10	0.98

Notes: Analogous to the notes for Table 2 in the main text.

Table 4: Continued. Out-of-sample forecast evaluation — Individual forecast: HAR (VAFEM versus non-VAFEM), Forecast set: $\{\text{GARCH}, \text{HAR}, \text{ARFIMA}\}$

Lag length (p) /	Horizon (h)		Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 2$	$h = 1$	MSE ratios < 1 (in %)	1.02	0.55	1.68	10.16	18.75
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.62
		MSE ratio (full period)	1.90	1.49	1.23	1.16	1.15
	$h = 2$	MSE ratios < 1 (in %)	0.07	0.15	1.71	5.39	12.35
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	2.33
		MSE ratio (full period)	3.53	1.81	1.21	1.16	1.14
	$h = 3$	MSE ratios < 1 (in %)	0.51	0.11	1.82	0.29	10.05
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.66
		MSE ratio (full period)	12.08	8.05	1.39	1.29	1.26
	$h = 5$	MSE ratios < 1 (in %)	0.00	0.66	0.47	0.58	2.88
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.18
		MSE ratio (full period)	> 100	> 100	1.68	1.49	1.47
	$h = 10$	MSE ratios < 1 (in %)	0.04	5.68	2.08	0.15	5.75
		MDM p-values ≤ 0.05 (in %)	0.00	0.66	0.15	0.00	1.86
		MSE ratio (full period)	> 100	> 100	3.03	1.82	1.79
$p = 3$	$h = 1$	MSE ratios < 1 (in %)	0.00	0.04	0.51	4.01	12.75
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.04
		MSE ratio (full period)	2.79	2.24	1.57	1.35	1.33
	$h = 2$	MSE ratios < 1 (in %)	0.00	0.07	0.15	0.51	6.99
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.00
		MSE ratio (full period)	6.01	2.92	1.44	1.26	1.23
	$h = 3$	MSE ratios < 1 (in %)	0.00	0.11	0.22	0.04	9.25
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	8.05
		MSE ratio (full period)	25.03	7.26	1.46	1.34	1.31
	$h = 5$	MSE ratios < 1 (in %)	0.00	0.00	0.47	4.12	9.03
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	8.05
		MSE ratio (full period)	> 100	> 100	1.95	1.43	1.42
	$h = 10$	MSE ratios < 1 (in %)	0.00	2.77	0.80	0.51	7.98
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	2.04
		MSE ratio (full period)	> 100	> 100	> 100	> 100	> 100

Continued on next page.

Table 4: Continued.

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 4$	$h = 1$	MSE ratios < 1 (in %)	0.00	0.00	0.07	2.62	8.16
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.00
		MSE ratio (full period)	4.57	3.38	2.18	1.61	1.57
	$h = 2$	MSE ratios < 1 (in %)	0.00	0.07	0.07	0.15	3.57
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	0.00
		MSE ratio (full period)	11.49	4.21	1.68	1.37	1.33
	$h = 3$	MSE ratios < 1 (in %)	0.00	0.11	0.04	0.00	8.78
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	3.97
		MSE ratio (full period)	33.22	8.23	1.46	1.37	1.33
	$h = 5$	MSE ratios < 1 (in %)	0.00	0.18	0.18	5.68	8.96
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.22	8.05
		MSE ratio (full period)	> 100	> 100	2.68	1.67	1.63
	$h = 10$	MSE ratios < 1 (in %)	0.00	2.40	1.78	0.36	6.77
		MDM p-values ≤ 0.05 (in %)	0.00	0.00	0.00	0.00	2.04
		MSE ratio (full period)	> 100	> 100	> 100	> 100	> 100

Notes: Analogous to the notes for Table 2 in the main text.

Table 5: Continued. Out-of-sample forecast evaluation — Individual forecast: ARFIMA (VAFEM versus non-VAFEM), Forecast set: {GARCH, HAR, ARFIMA}

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 2$	$h = 1$	MSE ratios < 1 (in %)	24.58	44.65	68.54	77.20	77.90
		MDM p-values ≤ 0.05 (in %)	8.78	9.21	32.92	36.45	37.11
		MSE ratio (full period)	1.14	0.90	0.82	0.79*	0.79*
	$h = 2$	MSE ratios < 1 (in %)	10.74	14.71	35.94	58.38	68.50
		MDM p-values ≤ 0.05 (in %)	9.03	11.00	19.88	27.09	33.76
		MSE ratio (full period)	2.93	1.55	1.01	0.98	0.96
	$h = 3$	MSE ratios < 1 (in %)	10.56	14.53	26.47	33.87	35.32
		MDM p-values ≤ 0.05 (in %)	10.56	13.55	16.68	21.70	27.31
		MSE ratio (full period)	15.55	10.30	1.30	1.21	1.19
	$h = 5$	MSE ratios < 1 (in %)	10.78	14.09	20.58	25.27	29.72
		MDM p-values ≤ 0.05 (in %)	9.47	11.51	13.69	15.77	26.37
		MSE ratio (full period)	> 100	25.83	1.66	1.50	1.46
	$h = 10$	MSE ratios < 1 (in %)	1.86	11.80	20.76	21.27	29.39
		MDM p-values ≤ 0.05 (in %)	0.00	11.00	12.09	9.29	17.15
		MSE ratio (full period)	> 100	> 100	7.85	3.17	3.04
$p = 3$	$h = 1$	MSE ratios < 1 (in %)	9.58	17.33	44.72	73.63	75.64
		MDM p-values ≤ 0.05 (in %)	8.16	8.81	28.59	34.12	36.64
		MSE ratio (full period)	1.62	1.33	1.04	0.96	0.95
	$h = 2$	MSE ratios < 1 (in %)	9.76	14.57	28.70	31.35	49.38
		MDM p-values ≤ 0.05 (in %)	8.19	10.63	18.79	25.31	28.59
		MSE ratio (full period)	5.38	2.83	1.39	1.13	1.09
	$h = 3$	MSE ratios < 1 (in %)	10.56	12.64	25.86	31.83	45.34
		MDM p-values ≤ 0.05 (in %)	9.69	10.71	14.06	27.28	28.04
		MSE ratio (full period)	24.93	9.25	1.31	1.20	1.17
	$h = 5$	MSE ratios < 1 (in %)	4.04	13.73	21.23	28.73	36.31
		MDM p-values ≤ 0.05 (in %)	0.00	10.89	15.84	23.38	22.76
		MSE ratio (full period)	> 100	6.63	2.38	1.71	1.67
	$h = 10$	MSE ratios < 1 (in %)	3.13	8.12	18.43	19.08	27.53
		MDM p-values ≤ 0.05 (in %)	0.00	7.25	12.31	13.51	17.70
		MSE ratio (full period)	> 100	> 100	> 100	> 100	> 100

Continued on next page.

Table 5: Continued.

Lag length (p) / Horizon (h)			Sample size				
			$T = 50$	$T = 100$	$T = 250$	$T = 500$	$T = 750$
$p = 4$	$h = 1$	MSE ratios < 1 (in %)	8.16	14.35	39.44	51.24	53.57
		MDM p-values ≤ 0.05 (in %)	7.47	8.27	25.09	33.07	35.07
		MSE ratio (full period)	2.55	1.97	1.42	1.21	1.19
	$h = 2$	MSE ratios < 1 (in %)	8.19	8.96	27.31	30.63	50.55
		MDM p-values ≤ 0.05 (in %)	6.88	8.16	16.64	24.62	28.30
		MSE ratio (full period)	10.41	4.49	1.92	1.26	1.21
	$h = 3$	MSE ratios < 1 (in %)	10.52	8.99	21.70	29.17	32.74
		MDM p-values ≤ 0.05 (in %)	8.99	7.32	13.11	25.46	27.86
		MSE ratio (full period)	34.44	11.43	1.50	1.19	1.14
	$h = 5$	MSE ratios < 1 (in %)	0.00	11.25	18.65	22.14	26.66
		MDM p-values ≤ 0.05 (in %)	0.00	8.92	12.09	21.19	21.49
		MSE ratio (full period)	> 100	> 100	3.15	2.03	1.96
	$h = 10$	MSE ratios < 1 (in %)	0.00	7.68	14.71	24.65	26.29
		MDM p-values ≤ 0.05 (in %)	0.00	7.43	12.45	13.22	17.59
		MSE ratio (full period)	> 100	> 100	> 100	7.94	3.64

Notes: Analogous to the notes for Table 2 in the main text.