2-2021

Depth-Weighted Forecast Combination: Application to COVID-19 Cases

Yoonseok Lee
Syracuse University, ylee41@maxwell.syr.edu

Donggyu Sul
UT at Dallas, d.sul@utdallas.edu

CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Depth-Weighted Forecast Combination: Application to COVID-19 Cases

Yoonseok Lee and Donggyu Sul

Paper No. 238
February 2021

ISSN: 1525-3066

Maxwell Syracuse University | Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/ email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH – Spring 2021

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Peter Wilcoxen, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Graham Ambrose, PAIA
Mariah Brennan, SOC. SCI.
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, ECON
Stephanie Coffey, PAIA
William Clay Fannin, PAIA
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Jeehee Han, PAIA
Mary Helander, SOC. SCI.
Amra Kandic, SOC
Sujung Lee, SOC
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Nicholas Oesterling, PAIA
Claire Pendergrast, SOC
Lauryn Quick, PAIA
Krushna Ranaware, SOC
Sarah Reilly, SOC
Christopher Rick, PAIA
Spencer Shanholtz, PAIA
Sarah Souders, PAIA
Joaquin Urrego, ECON
Yao Wang, ECON
Yi Yang, ECON
Xiaoyan Zhang, Human Dev.
Bo Zheng, PAIA
Dongmei Zuo, SOC. SCI.

STAFF

Ute Brady, Postdoctoral Scholar
Willy Chen, Research Associate
Katrina Fiacchi, Administrative Specialist
Michelle Kincaid, Senior Associate, Maxwell X Lab
Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Samantha Trajkovski, Postdoctoral Scholar
Laura Walsh, Administrative Assistant

Abstract

We develop a novel forecast combination based on the order statistics of individual predictability when many forecasts are available. To this end, we define the notion of forecast depth, which measures the size of forecast errors during the training period and provides a ranking among different forecast models. The forecast combination is in the form of a depth-weighted trimmed mean, where the group of models with the worst forecasting performance during the training period is dropped. We derive the limiting distribution of the depth-weighted forecast combination, based on which we can readily construct forecast confidence intervals. Using this novel forecast combination, we forecast the national level of new COVID-19 cases in the U.S. and compare it with other approaches including the ensemble forecast from the Centers for Disease Control and Prevention. We find that the depth-weighted forecast combination yields more accurate predictions compared with other forecast combinations.

JEL No.: C32, C53

Keywords: Forecast Combination, Forecast depth, Depth-weighted trimmed mean, COVID-19

Authors: Yoonseok Lee, Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, ylee41@maxwell.syr.edu; Donggyu Sul, Department of Economics, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, d.sul@utdallas.edu

1 Introduction

Since the seminal work by Bates and Granger (1969), forecast combinations have been successfully used in many empirical studies when multiple forecasts of the same variable are available. The forecast combination becomes one of the core methods of forecasting practice, as the cost of collecting real-time forecasts has been dropped significantly. It is also known that combined forecasts often produce better forecasts than the ex ante best single forecasting model. See Clemen (1989), Stock and Watson (2001), and Timmermann (2006), for instance, for survey of this literature.

When we know the individual forecasting models and their information sets (i.e., the predictors), we can readily find the optimal weights by minimizing the forecast mean squared error loss. We can also apply ensemble methods in machine learning such as bagging and boosting to combine forecasts. Such approaches, however, require a long training period but a small number of forecasting models, so that the weights can be estimated. In practice, it is often the case that the individual forecasting models are not fully known and only the individual forecast reports are available. Even when the individual forecasting models and the information sets are known, pooling of information sets is either impossible or prohibitively costly particularly in real-time practice (e.g., Diebold and Pauly, 1990). In such cases, it is common to use the equally weighted average of forecasts or the weighted average based on the inverse squared forecast error values (e.g., Stock and Watson, 2001). The equal weight approach is the most popular one because it is simple but outperforms the estimated optimal weights, which is often called the "forecast combination puzzle".

In this paper, we propose a forecast combination based on the order statistics of individual predictability when many forecasts are available. We assume data rich environment of the forecasts but we do not require the knowledge of each forecasting model nor its information set. The weights can be obtained using the cross-section information and hence we do not need a long training period to estimate the weights.

More precisely, we modify the notion of data depth (e.g., Zuo and Serfling, 2000; Lee and Sul, 2019) in the context of forecast combination and develop the *forecast depth*, which measures the nearness of each vector of forecasts toward the vector of observed values over the training period. The forecast depth provides a ranking among the forecasting models. The weights for forecast combination are proportional to the forecast depth after trimming out the forecasts from the group of the worst performers. In this sense, this novel weighting scheme shares the idea of the rank-based approach (e.g., Aiolfi and Timmermann, 2006) and the idea of trimming (e.g., Granger and Jeon, 2004) in forecast combination. The depth-weighted forecast combination is in the form of the L-statistic, and thus it is more robust toward very bad forecasts than the equally-weighted combination. One limitation of the depth-weighted forecast combination is that it ignores the correlation structure among the forecast errors, though it is quite a common

practice in the literature.

The main contribution of this paper can be summarized in three folds. First, we develop the forecast depth, based on which we can readily rank the forecasting performance over multiple periods and construct a robust forecast combination in the form of the depth-weighted trimmed mean. This approach only requires the forecast values and does not need to know each forecasting model. The number of forecasts can be large and the training period can be very short. Second, we derive the limiting distribution of the depth-weighted forecast combination with trimming, which can be used to construct a confidence interval of the prediction without relying on subsampling. Both the weight and the trimming threshold depend on the level of each forecast depth, and hence they are treated random in deriving the limiting distribution. Since the proposed forecast combination can cover popular ones as its special cases, such as the equal-weight forecast combination, the trimmed forecast combination, and the median forecast, this result also provides limiting distributions of those forecast combinations as well. Third, we apply the depth-weighted forecast combination to predict the national level of new COVID-19 cases in the United States. Using the past forecasts, we find that the new forecast combination yields around 80% of the forecast mean squared error than the ensemble forecast reported by the Centers for Disease Control and Prevention. As of January 14, 2021, we also report the forecasts for the next four weeks, which shows a more volatile trend than the ensemble forecast and predicts up to 10% higher new cases in the next four weeks.

It is worthy noting that the depth-weighted forecast combination uses cross-sectional distribution information in prediction, which can be time-varying. In the recent works, Joon Y. Park has developed novel approaches to estimate distributional dynamics and unknown trends in time series distribution. For example, see Chang, Kim, and Park (2016), Hu, Park, and Qian (2017), and Chang, Kaufmann, Kim, Miller, Park, and Park (2020). Once the unknown stochastic trend in the distribution of many forecasts is estimated, this information can be used to model the dynamics of forecast depth and reinforce forecast combination especially in long horizon forecasting. We do not consider this method here, but it will be a very promising and interesting topic for future research.

The rest of the paper is organized as follows. Section 2 defines the forecast depth and develops the depth-weighted trimmed forecast combination. Section 3 examines the performance of the proposed forecast combination with the new COVID-19 cases in the U.S. and reports the forecasts for the next four weeks. Section 4 summarizes the asymptotic properties of the depth-weighted forecast combination. Section 5 concludes with some remarks.

2 Forecast Depth and Depth-Weighted Forecast Combination

As in Bates and Granger (1969), we suppose there are n forecasts $y_{i,t+1}$ of y^0_{t+1} for $i = 1, \ldots, n$. Such multiple forecasts can come from different forecasting models or from different agents. We consider forecast combination in the form of

$$\widehat{y}_{t+1} = \sum_{i=1}^{n} \pi_{i,t} y_{i,t+1} \tag{1}$$

for some potentially time-varying weights $\pi_{i,t}$. In particular, we consider the weight based on the data depth (e.g., Zuo and Serfling, 2000; Lee and Sul, 2019) of forecast error vector at given t.

More precisely, we let $Y_{i,t} = (y_{i,t-k+1}, \ldots, y_{i,t-1}, y_{i,t})'$ be the $k \times 1$ vector of forecasts in the training period, which is the most recent k observations, and $Y^0_t = (y^0_{t-k+1}, \ldots, y^0_{t-1}, y^0_t)'$ be the $k \times 1$ vector of the observed values during this period. We denote the forecast error vector of i as

$$e_{i,t} = Y_{i,t} - Y^0_t.$$

We let m be a $k \times 1$ vector with $\|m\| = 1$. Given m, we define the normalized forecast error distance between $Y_{i,t}$ and Y_t (or the forecast outlyingness) as

$$\mathcal{O}_{i,t} = \frac{|m' e_{i,t}|}{s_t} = \frac{\left|m'(Y_{i,t} - Y^0_t)\right|}{s_t},$$

where $s_t \in (0, \infty)$ is some dispersion measure of $m' e_{i,t}$ that is affine invariant and measurable to the information set at t, $\mathcal{I}_t = \sigma(\cup_{i=1}^n \{y_{i,r}, y^0_r\}_{r \le t})$.[1] For instance, under the assumption that $\{e_{i,t}\}$ is a random sample from a common distribution across i, we can consider the square root of the forecast mean squared error (FMSE),

$$s_t = (\mathbb{E}[(m' e_{i,t})^2 | \mathcal{I}_t])^{1/2}, \tag{2}$$

or the forecast median absolute deviation (FMAD),[2]

$$s_t = \inf \left\{ v : \mathbb{P}\left(\left| m' e_{i,t} \right| \le v | \mathcal{I}_t \right) \ge 1/2 \right\}. \tag{3}$$

[1] When $m' e_{i,t} = s_t = 0$, we define $\mathcal{O}_{i,t} = 0$.

[2] When the distribution of $\{e_{i,t}\}$ is heterogeneous across i, we can even consider $\mathcal{O}_{i,t} = |m_i' e_{i,t}| / s_t$, where $m_i = (m_{i1}, \ldots, m_{ik})'$ be a $k \times 1$ vector with $\|m_i\| = 1$ for all i. In such cases, s_t can be alternatively defined as $(\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathbb{E}\left[(m_i' e_{i,t})^2\right])^{1/2}$ and $\lim_{n\to\infty} \mathrm{med}_{1 \le i \le n} \inf \{v_i : \mathbb{P}(|m_i' e_{i,t}| \le v_i) \ge 1/2\}$, respectively.

The *forecast depth* is defined as[3]

$$\mathcal{D}_{i,t} = \frac{1}{1+\mathcal{O}_{i,t}}.$$

By construction, the forecast depth $\mathcal{D}_{i,t}$ takes values between zero and one; it is one when $e_{i,t} = 0$. It also satisfies the typical properties of the data depth (e.g., Zuo and Serfling, 2000). In particular, for a given m, $\mathcal{D}_{i,t}$ does not change from any location shift or rescaling of $e_{i,t}$ (*Affine Invariance*); $\mathcal{D}_{i,t}$ reaches the maximal value 1 if the model i makes perfect prediction (*Maximality at Center*); $\mathcal{D}_{i,t}$ decreases monotonically as it moves away from the maximal depth location, the deepest point (*Monotonicity Relative to the Deepest Point*); and $\mathcal{D}_{i,t}$ reaches to the minimal value 0 as the forecast error diverges (*Vanishing at Infinity*). The monotonicity yields a well-defined quantile function of $m'e_{i,t}$ since it excludes any quantile-crossing problem, which is to be the key to construct a depth-based trimming in the weight $\pi_{i,t}$. The last property is important for the forecast robustness against very under-performed forecasters or outliers.

The forecast depth $\mathcal{D}_{i,t}$ is in a similar form as the "projection depth".[4] However, unlike the projection depth, it considers the distance from each forecast toward the observed value (i.e., the forecast error) instead of the distance toward a central location parameter of the distribution of $Y_{i,t}$, such as the mean or the median. More importantly, we preset the normalization vector m in our case that can be potentially heterogeneous, whereas the typical projection depth needs to search for the m vector so that the outlyingness is maximized. The latter feature is unique in the forecasting problem, which is valid because one often has ordering of the importance among the forecast errors during the training period $e_{i,t} = (y_{i,t-k+1} - y^0_{t-k+1}, \ldots, y_{i,t-1} - y^0_{t-1}, y_{i,t} - y^0_t)'$.

Several examples of the normalization vector m can be considered. If we consider the forecast performance in the most recent observations are more important than the distant ones, we can let the jth element of $m = (m_1, \ldots, m_k)'$ as $m_j = K(j/k)/\sum_{\ell=1}^{k} K(\ell/k)$ for $j = 1, \ldots, k$, where $K(\cdot)$ is some non-increasing one-side kernel function. Examples include the discount factor approach by Bates and Granger (1969) such that $K(j/k) = \gamma^{k-j}$ for some $\gamma > 1$, and the Box–Cox transform weights by Diebold and Pauly (1987). If we treat all the forecast errors during the training period equally important, then $m_j = 1/k$. Note that such choices of m do not require

[3] It should be emphasized that the original notion of depth is mainly motivated to define a robust central location of multi-dimensional variables, which is not straightforward to define by simply extending the standard notion of median for univariate variables. In our case, on the other hand, the central location is already given as the vector of observed values Y^0_t, and hence the forecast depth provides a normalized distance from a vector of forecasts $Y_{i,t}$ toward the vector of observed values Y^0_t.

[4] The projection depth is defined as follow. For an i.i.d. sample $\{Y_i\}$, the projection depth of Y_i is given as $\mathcal{D}_i = 1/[1+\mathcal{O}_i]$, where the outlyingness is defined as

$$\mathcal{O}_i = \sup_{m \in \mathbb{R}^k : \|m\|=1} |m'Y_i - \mu(m'Y_i)|/\sigma(m'Y_i)$$

for some location and dispersion parameters $\mu(m'Y_i)$ and $\sigma(m'Y_i)$ of the distribution of $m'Y_i$.

a balanced longitudinal data structure in $Y_{i,t}$, that is all forecasts are not necessarily available over the given training period. Therefore, we can even define heterogeneous normalization vector $m_i = (m_{i1}, \ldots, m_{ik_i})'$ by letting $m_{ij} = K(j/k_i)/\sum_{\ell=1}^{k} K(\ell/k_i)$ for $j = 1, \ldots, k_i$ with $k_i \leq k$ for each i.

When a balanced data of $Y_{i,t}$ is available and if we consider that individual forecast $y_{i,t+1}$ follows a homogeneous autoregressive structure, then m_j can be obtained from the normalized fitted $AR(k)$ coefficients in $e_{i,t+1} = \sum_{j=1}^{k} m_j e_{i,t+1-j} + \varepsilon_{i,t+1}$. We can also let m_j be the normalized inverse of cross-sectional FMSE, $(1/n)\sum_{i=1}^{n}(y_{i,t-j+1} - y^0_{t-j+1})^2$, at each time $\{t-k+1, \ldots, t-1, t\}$ during the k training period, which is of a similar normalization scheme as the standard Mahalanobis distance.

The forecast depth provides a ranking of predictability among the forecasting models or agents, since the better performing models have higher levels of forecast depth. We can also use the forecast depth as a tool to detect very under-performed forecasting models. We hence propose to define the weight for the forecast combination in (1) that is proportional to the individual forecast depth. More precisely, we set some trimming parameter $\tau \in (0, 1)$ and let

$$\pi_{i,t} = \frac{1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t})}{\sum_{j=1}^{n} 1\{\widehat{\mathcal{D}}_{j,t} \geq \tau\} W(\widehat{\mathcal{D}}_{j,t})}, \tag{4}$$

where $1\{\cdot\}$ is the binary indicator, $W(\cdot)$ is some scalar weight function, and the forecast depth estimator $\widehat{\mathcal{D}}_{i,t}$ is defined as

$$\widehat{\mathcal{D}}_{i,t} = \frac{1}{1 + \widehat{\mathcal{O}}_{i,t}} \quad \text{with} \quad \widehat{\mathcal{O}}_{i,t} = \frac{\left|m_i'(Y_{i,t} - Y_t^0)\right|}{\widehat{s}_t}$$

for some consistent estimator $\widehat{s}_t$. For each of the aforementioned examples, we can use

$$\widehat{s}_t = \left(\frac{1}{n}\sum_{i=1}^{n} m'(Y_{i,t} - Y_t^0)(Y_{i,t} - Y_t^0)'m\right)^{1/2}$$

and

$$\widehat{s}_t = \mathrm{med}_{1 \leq i \leq n} \left|m'(Y_{i,t} - Y_t^0)\right|,$$

respectively. Since we define the training period over a rolling window of the most recent k periods, the weight in (4) is naturally time-varying.

From (1) and (4), the proposed combined forecast is then defined as a form of the trimmed

depth-weighted mean given as

$$\widehat{y}_{t+1} = \frac{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t}) y_{i,t+1}}{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t})}, \tag{5}$$

where $\{y_{i,t+1}\}_{i=1}^{n}$ are the individual forecasts of y_{t+1} at the current time t.[5] This combined forecast assigns the weight on $y_{i,t+1}$ based on its forecast depth during the training period: if the ith model's forecast error $m'e_{i,t}$ is near zero and hence the forecast depth estimator $\widehat{\mathcal{D}}_{i,t}$ is near maximum, then its forecast $y_{i,t+1}$ gets a high weight; if its forecast error is too large, on the other hand, it is trimmed and gets a zero weight. In this sense, unlike the forecast combination based on the equal weights or forecast mean squared errors, the depth-weighted trimmed forecast combination $\widehat{y}_{t+1}$ in (5) is robust toward very under-performed (or outlying) forecasts over the training period. Note that the trimming scheme is random as it depends on $\widehat{\mathcal{D}}_{i,t}$, so $\widehat{y}_{t+1}$ is a randomly trimmed forecast combination.

Based on the choice of τ and $W(\cdot)$, $\widehat{y}_{t+1}$ in (5) can cover popular forecast combinations. For instance, when $W(\cdot) = 1$, $\widehat{y}_{t+1}$ is the equally weighed combined forecast with trimming, which converges to the trimmed mean of $y_{i,t+1}$ if $\mathbb{E}[y_{i,t+1}]$ exists; when $\tau = 0$ in addition, it is simply the equally weighed combined forecast. When $\tau = \max_{1 \leq i \leq n} \widehat{\mathcal{D}}_{i,t}$, $\widehat{y}_{t+1}$ is the same as $y_{i,t+1}$ whose forecast depth is the maximal. For the case with $m_j = 1/k$, this maximal depth forecast corresponds to the forecast of agent i whose forecasts has been the most accurate (i.e., ex ante the best single forecast). When $y_{i,t+1}$ has a density function that is elliptically symmetric about its mode, this maximal depth forecast becomes the median combination forecast.

3 Forecasting New COVID-19 Cases

We apply the depth-weighted forecast combination (5) to predict weekly COVID-19 cases in the United States. The data is collected from the Centers for Disease Control and Prevention (CDC) COVID Data Tracker (https://covid.cdc.gov/covid-data-tracker/#forecasting_weekly-deaths) as of January 11, 2021, which is update on January 14, 2021. The dataset includes weekly forecast history from 34 individual modeling groups.[6] It also includes the ensemble forecast that is reported in the weekly forecast digest by the CDC, where the confidence interval at each prediction point is calculated as the average of the corresponding quantiles across all

[5] For $h \geq 1$, the h-step ahead combined forecast can be similarly defined as

$$\widehat{y}_{t+h} = \frac{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t}) y_{i,t+h}}{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t})}.$$

[6] The list of the agents and details of the forecasting models can be found at https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html.

individual model forecasts (Busetti 2017).[7]

First, we compare different forecast combination approaches, including the ensemble forecast reported by the CDC (Ensemble), equally-weighted average (Equal), inverse FMSE based average (iMSE), and the forecast-depth based average (FD) developed in this paper. For the forecast-depth based combination, we consider 4 different cases by combining the following choices. For a given training period size k, FD_A is the forecast-depth using the normalization vector $m_j = 1/k$ and FD_B is one using (normalized) $m_j = 1 - ((j-1)/k)$. For each forecast depth, the type of $\widehat{s}_t$ is specified in the parentheses: "(mse)" uses the square root of the sample FMSE, $\widehat{s}_t = (n^{-1}\sum_{i=1}^{n}(m'e_{i,t})^2)^{1/2}$, and "(mad)" uses the sample FMAD, $\widehat{s}_t = \text{med}_{1\le i\le n}|m'e_{i,t}|$. Note that we do not have a balanced longitudinal data structure in this exercise since each modeling group has a different length of forecast history. Instead of artificially forming a balanced longitudinal data, we define the k training period as the most recent k forecasts from the target forecasting week. For a given target forecasting week, any modeling group i is dropped in the average if it does not have at least k most recent forecasts to form the training period. Modeling groups often report several forecasts for the same week; in such cases, we use the very last forecast (the most recent one) for that week.

Tables 1 to 4 report the FMSE comparisons among different forecast combinations for h-week ahead forecasting for $h = 1, \ldots, 4$, respectively. Values in the tables are the ratio of the FMSE of each combined forecast to that of the ensemble forecast by the CDC, that are averaged over the most recent 15 weeks of forecasting performance, from the week ending on 10/3/2020 to the week ending on 1/9/2021. So, values less than 1 implies that the FMSE is smaller than that of the ensemble forecast; smaller values implies better performance. At the current time T, the FMSE is calculated as $(1/15)\sum_{t=T-15}^{T-15}\left(\widehat{y}^c_{t+1} - y^0_{t+1}\right)^2$ for each combined forecast $\widehat{y}^c_{t+1}$. In each table, k is the training period size, "trim" is the trimming proportion from the entire cross-section, and $\overline{n}$ is the average sample size having k training periods. The forecast-depth based combination shows good performance with about 80% of the FMSE of the ensemble forecast. It also outperforms other methods in general. For some cases, the inverse FMSE based average outperforms the forecast-depth based method, but its performance is not stable. We can conclude that the forecast-depth based combination yields quite stable and outperforming prediction in general.

Second, based on the available forecasts up to the week ending 1/9/2021, we report predictions for the next 4 weeks (ending on 1/16/2021, 1/23/2021, 1/30/2021, 2/6/2021) as of January 14, 2021. For each h-week ahead prediction ($h = 1, \ldots, 4$), we consider 30 different cases of $k = 1, \ldots, 5$ and $trim = 0.0, 0.1, \ldots, 0.5$ for each of the four forecast combinations, FD_A(mse), FD_A(mad), FD_B(mse), and FD_B(mad). Figure 1 reports the predictions of the average among these depth-weighted forecast combinations in black square line (avgFD) and the

[7] For further details about the ensemble forecast, see medRxiv 2020.08.19.20177493; doi: https://doi.org/10.1101/2020.08.19.20177493.

Table 1: 1-step ahead FMSE ratio to the Ensemble forecast

k	trim	$\overline{n}$	equal	iMSE	FD$_A$(mse)	FD$_A$(mad)	FD$_B$(mse)	FD$_B$(mad)
1	0.0	24.9	0.8772	0.8438	0.8463	0.8425	0.8463	0.8425
	0.1				0.8291	0.8265	0.8291	0.8265
	0.2				0.8163	0.8131	0.8163	0.8131
	0.3				0.8061	0.8026	0.8061	0.8026
	0.4				0.8193	0.8167	0.8193	0.8167
	0.5				0.8323	0.8313	0.8323	0.8313
2	0.0	24.2	0.8835	0.8287	0.8500	0.8476	0.8493	0.8458
	0.1				0.8029	0.8022	0.8026	0.8012
	0.2				0.8127	0.8115	0.8335	0.8309
	0.3				0.7816	0.7818	0.8195	0.8179
	0.4				0.8396	0.8395	0.8505	0.8483
	0.5				0.8432	0.8443	0.8308	0.8296
3	0.0	23.6	0.8867	0.7640	0.8465	0.8433	0.8506	0.8472
	0.1				0.7951	0.7934	0.8023	0.8005
	0.2				0.7951	0.7941	0.8033	0.8022
	0.3				0.7658	0.7654	0.7848	0.7846
	0.4				0.7813	0.7809	0.8451	0.8438
	0.5				0.8268	0.8259	0.8496	0.8479
4	0.0	22.9	0.8811	0.7745	0.8356	0.8320	0.8412	0.8383
	0.1				0.7805	0.7786	0.7902	0.7887
	0.2				0.7400	0.7391	0.7877	0.7873
	0.3				0.7708	0.7696	0.7967	0.7962
	0.4				0.8160	0.8144	0.8421	0.8410
	0.5				0.8349	0.8333	0.8275	0.8266
5	0.0	22.1	0.8842	0.7853	0.8372	0.8329	0.8400	0.8364
	0.1				0.7814	0.7789	0.7843	0.7824
	0.2				0.7628	0.7611	0.7814	0.7806
	0.3				0.7383	0.7368	0.7804	0.7795
	0.4				0.8170	0.8146	0.8455	0.8440
	0.5				0.8187	0.8167	0.8219	0.8204
	avg	23.6	0.8826	0.7993	0.8074	0.8057	0.8203	0.8185

Note: Values in the table are the ratio of the forecast mean square error of each combined forecast to that of the ensemble forecast, that are averaged over the most recent 15 weeks of forecasting, 10/3/2020 - 1/9/2021.

Table 2: 2-step ahead FMSE ratio to the Ensemble forecast

k	trim	$\overline{n}$	equal	iMSE	FD$_A$(mse)	FD$_A$(mad)	FD$_B$(mse)	FD$_B$(mad)
1	0.0	24.7	0.8799	1.0812	0.8455	0.8439	0.8455	0.8439
	0.1				0.8170	0.8155	0.8170	0.8155
	0.2				0.7928	0.7925	0.7928	0.7925
	0.3				0.7941	0.7939	0.7941	0.7939
	0.4				0.8057	0.8066	0.8057	0.8066
	0.5				0.8108	0.8120	0.8108	0.8120
2	0.0	23.9	0.8825	0.9728	0.8544	0.8541	0.8491	0.8482
	0.1				0.8261	0.8262	0.8243	0.8235
	0.2				0.7939	0.7950	0.7897	0.7902
	0.3				0.7696	0.7716	0.7700	0.7716
	0.4				0.8127	0.8140	0.8236	0.8245
	0.5				0.8223	0.8227	0.8017	0.8028
3	0.0	23.3	0.8769	0.7556	0.8361	0.8319	0.8397	0.8366
	0.1				0.7932	0.7912	0.8078	0.8052
	0.2				0.7594	0.7581	0.7716	0.7703
	0.3				0.7543	0.7526	0.7659	0.7644
	0.4				0.7577	0.7556	0.7831	0.7812
	0.5				0.7668	0.7641	0.7974	0.7944
4	0.0	22.7	0.8804	0.7675	0.8382	0.8337	0.8376	0.8334
	0.1				0.8002	0.7976	0.8055	0.8024
	0.2				0.7595	0.7581	0.7657	0.7641
	0.3				0.7493	0.7479	0.7673	0.7651
	0.4				0.7367	0.7354	0.7795	0.7764
	0.5				0.7757	0.7737	0.7818	0.7789
5	0.0	21.9	0.8834	0.7669	0.8422	0.8363	0.8435	0.8390
	0.1				0.8039	0.7998	0.8074	0.8048
	0.2				0.7587	0.7557	0.7724	0.7707
	0.3				0.7780	0.7754	0.7677	0.7660
	0.4				0.7882	0.7845	0.7717	0.7695
	0.5				0.7670	0.7635	0.7743	0.7720
	avg	23.3	0.8806	0.8688	0.7937	0.7921	0.7988	0.7973

Note: Values in the table are the ratio of the forecast mean square error of each combined forecast to that of the ensemble forecast, that are averaged over the most recent 15 weeks of forecasting, 10/3/2020 - 1/9/2021.

Table 3: 3-step ahead FMSE ratio to the Ensemble forecast

k	trim	$\overline{n}$	equal	iMSE	FD$_A$(mse)	FD$_A$(mad)	FD$_B$(mse)	FD$_B$(mad)
1	0.0	24.3	0.8789	1.3785	0.8613	0.8638	0.8613	0.8638
	0.1				0.8390	0.8423	0.8390	0.8423
	0.2				0.8115	0.8170	0.8115	0.8170
	0.3				0.8047	0.8111	0.8047	0.8111
	0.4				0.8047	0.8120	0.8047	0.8120
	0.5				0.7712	0.7809	0.7712	0.7809
2	0.0	23.6	0.8663	1.1753	0.8530	0.8596	0.8498	0.8533
	0.1				0.8215	0.8299	0.8224	0.8270
	0.2				0.8035	0.8138	0.8049	0.8111
	0.3				0.8046	0.8155	0.8157	0.8217
	0.4				0.8234	0.8345	0.8076	0.8136
	0.5				0.8617	0.8718	0.8146	0.8215
3	0.0	23.1	0.8691	0.7889	0.8431	0.8409	0.8476	0.8487
	0.1				0.8061	0.8057	0.8090	0.8117
	0.2				0.8004	0.7999	0.8102	0.8128
	0.3				0.7843	0.7843	0.8065	0.8085
	0.4				0.7894	0.7893	0.7943	0.7973
	0.5				0.7758	0.7766	0.7843	0.7890
4	0.0	22.5	0.8727	0.8234	0.8584	0.8558	0.8549	0.8543
	0.1				0.8365	0.8359	0.8212	0.8222
	0.2				0.8238	0.8245	0.8189	0.8203
	0.3				0.8273	0.8285	0.8121	0.8139
	0.4				0.8301	0.8312	0.8165	0.8179
	0.5				0.8337	0.8351	0.8423	0.8424
5	0.0	21.5	0.8798	0.8203	0.8669	0.8642	0.8680	0.8684
	0.1				0.8475	0.8469	0.8410	0.8436
	0.2				0.8285	0.8298	0.8367	0.8401
	0.3				0.8477	0.8488	0.8362	0.8399
	0.4				0.8255	0.8273	0.8218	0.8268
	0.5				0.9286	0.9294	0.8914	0.8937
	avg	23.0	0.8733	0.9973	0.8271	0.8302	0.8240	0.8276

Note: Values in the table are the ratio of the forecast mean square error of each combined forecast to that of the ensemble forecast, that are averaged over the most recent 15 weeks of forecasting, 10/3/2020 - 1/9/2021.

Table 4: 4-step ahead FMSE ratio to the Ensemble forecast

k	trim	$\overline{n}$	equal	iMSE	FD$_A$(mse)	FD$_A$(mad)	FD$_B$(mse)	FD$_B$(mad)
1	0.0	24.0	0.8566	1.2269	0.8251	0.8205	0.8251	0.8205
	0.1				0.8095	0.8068	0.8095	0.8068
	0.2				0.7891	0.7878	0.7891	0.7878
	0.3				0.7892	0.7885	0.7892	0.7885
	0.4				0.7853	0.7852	0.7853	0.7852
	0.5				0.7774	0.7778	0.7774	0.7778
2	0.0	23.3	0.8584	1.0379	0.8308	0.8274	0.8309	0.8274
	0.1				0.8099	0.8085	0.8147	0.8131
	0.2				0.8183	0.8154	0.8145	0.8120
	0.3				0.7961	0.7943	0.8262	0.8235
	0.4				0.7962	0.7950	0.7995	0.7981
	0.5				0.7872	0.7869	0.7854	0.7850
3	0.0	22.8	0.8616	0.8530	0.8474	0.8444	0.8406	0.8391
	0.1				0.8367	0.8359	0.8162	0.8170
	0.2				0.8086	0.8099	0.8066	0.8082
	0.3				0.8015	0.8038	0.8161	0.8171
	0.4				0.8181	0.8204	0.8201	0.8215
	0.5				0.8606	0.8625	0.8470	0.8488
4	0.0	22.2	0.8641	0.8139	0.8483	0.8438	0.8497	0.8478
	0.1				0.8313	0.8288	0.8389	0.8390
	0.2				0.8146	0.8139	0.8155	0.8178
	0.3				0.8156	0.8157	0.8152	0.8183
	0.4				0.8316	0.8310	0.8241	0.8274
	0.5				0.8269	0.8273	0.8778	0.8802
5	0.0	21.2	0.8815	0.8254	0.8620	0.8580	0.8666	0.8656
	0.1				0.8478	0.8460	0.8575	0.8587
	0.2				0.8382	0.8380	0.8413	0.8443
	0.3				0.8314	0.8324	0.8311	0.8354
	0.4				0.8643	0.8648	0.8222	0.8279
	0.5				0.8807	0.8813	0.8829	0.8877
	avg	22.7	0.8644	0.9514	0.8227	0.8217	0.8239	0.8242

Note: Values in the table are the ratio of the forecast mean square error of each combined forecast to that of the ensemble forecast, that are averaged over the most recent 15 weeks of forecasting, 10/3/2020 - 1/9/2021.

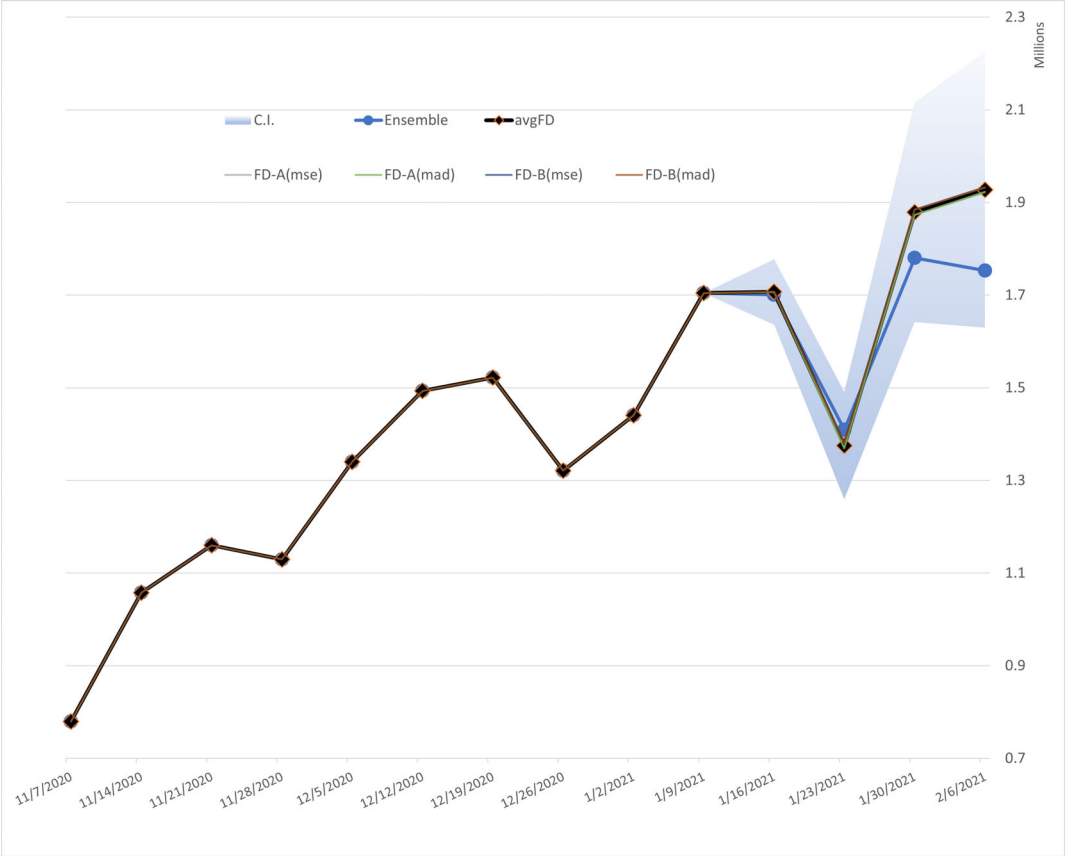

Figure 1: COVID-19 New Cases Forecast (US National)

ensemble forecast reported by the CDC in blue circle line (Ensemble). It also includes the average of each forecast combination ($\mathrm{FD}_A(\mathrm{mse})$, $\mathrm{FD}_A(\mathrm{mad})$, $\mathrm{FD}_B(\mathrm{mse})$, $\mathrm{FD}_B(\mathrm{mad})$) over 30 different choices of $(k, trim)$, but they are very close to each other and hence almost indistinguishable from their average, avgFD. The pointwise 95% confidence interval of the avgFD forecast point is obtained as the average of the confidence intervals of each depth-weighted forecast combinations using the normal approximation based on (9) in the following section. The prediction based on the depth-weighted forecast combination shows a more volatile trend than the ensemble forecast does and predicts up to 10% higher new cases in the next four weeks. But the ensemble forecast values are within the 95% confidence interval of the forecast depth combination. In particular, the forecasts and the 95% confidence intervals are given in Table 5.

Finally, we report some of the forecast depth values. The forecast depth naturally gives a ranking among the forecasting models based on their past performance. Therefore, we can use this information to tell the performance of each model. As an illustration, we report the forecast depth of $\mathrm{FD}_A(\mathrm{mse})$ for the length of training period $k = 1, \ldots, 5$ ending at 1/9/2021. Figures 2 to 6 depict the scatter plots of the forecast depth values to the forecasting error of all the forecast

Table 5: New COVID-19 cases prediction

forecasting week	avgFD	95% CI of avgFD	Ensemble
1/16/2021	1706930	[1636061, 1777798]	1700701
1/23/2021	1375329	[1259383, 1491274]	1409573
1/30/2021	1878847	[1642163, 2115530]	1780502
2/6/2021	1927617	[1629979, 2225256]	1753039

Note: Values in the table are predictions of the new COVID-19 case in the following four weeks, as of January 14, 2021.

modeling groups in the sample for each k. Interestingly, during this recent training periods, we can find that some models consistently report more accurate predictions than others.[8]

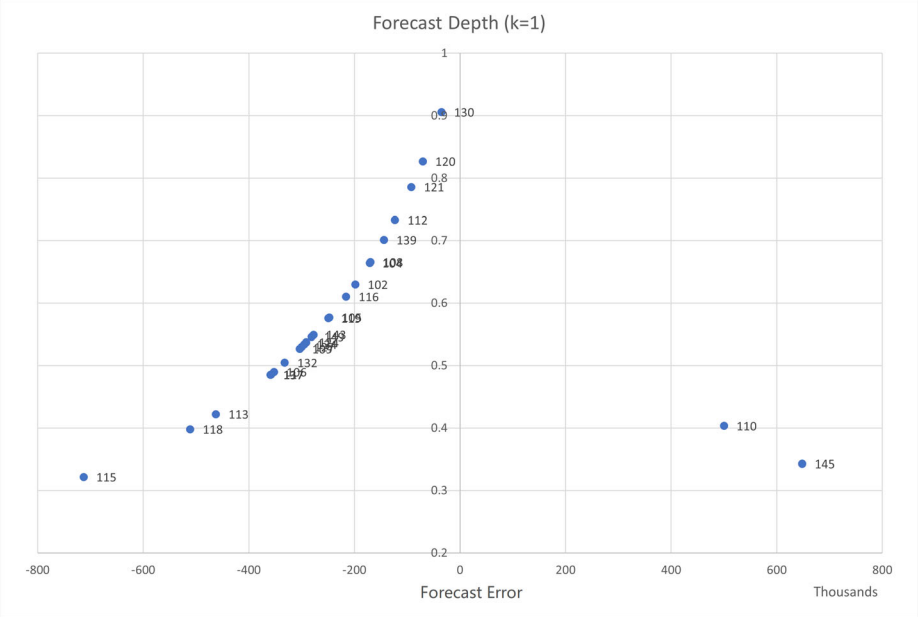

Figure 2: Forecast Depth (k=1)

4 Limiting Distribution of Combined Forecast

We now derive limiting distribution of the depth-weighted forecast combination (5), based on which we can conduct further inferences. For each t, we define a (normalized) forecasting error

[8] The lable on each dot represents the modeling group. The list of them is availabe from the authors upon request.

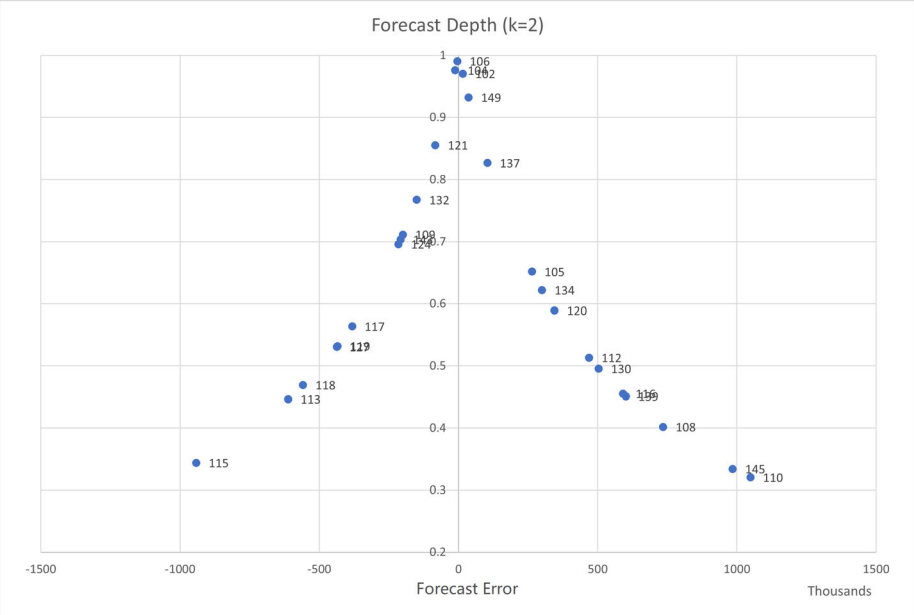

Figure 3: Forecast Depth (k=2)

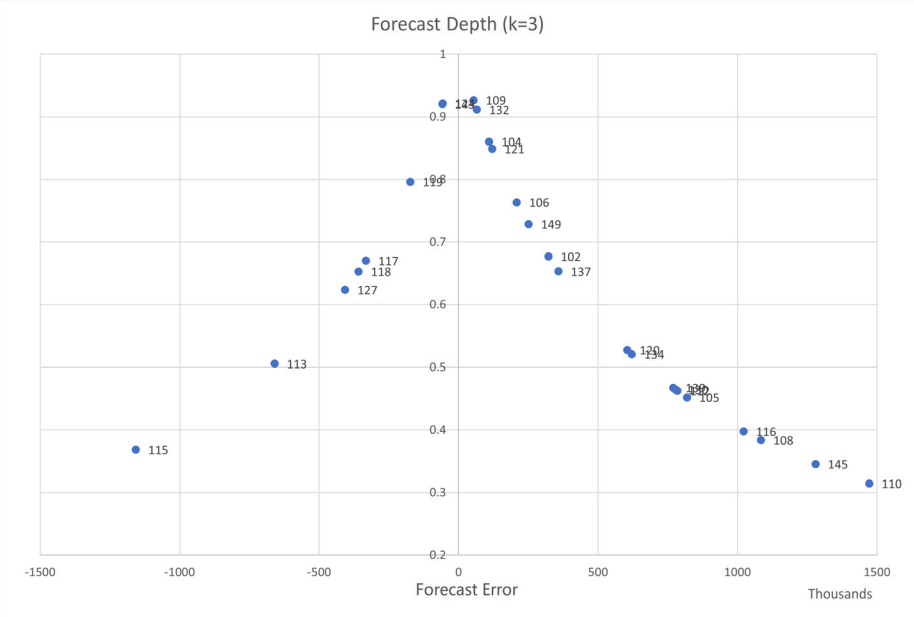

Figure 4: Forecast Depth (k=3)

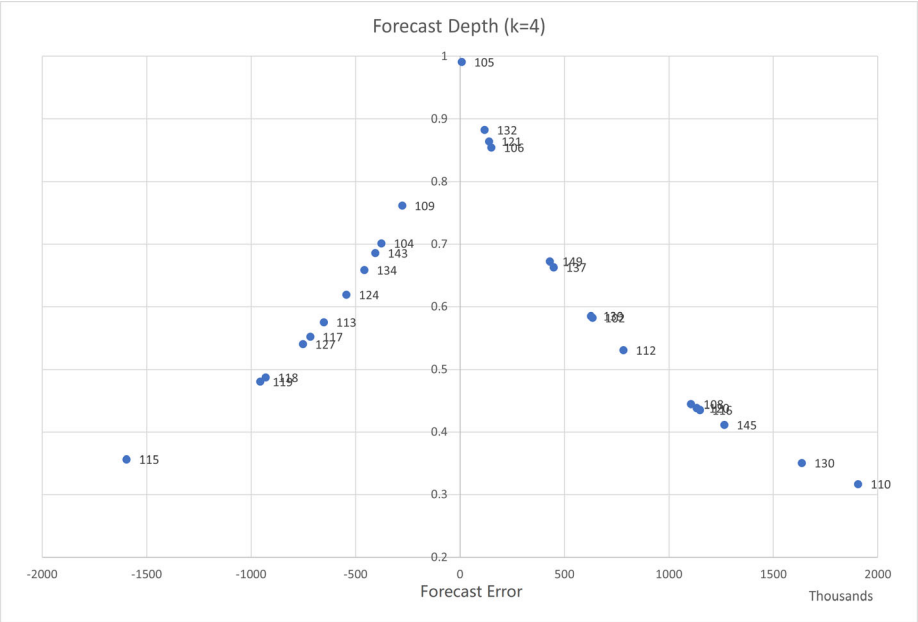

Figure 5: Forecast Depth (k=4)

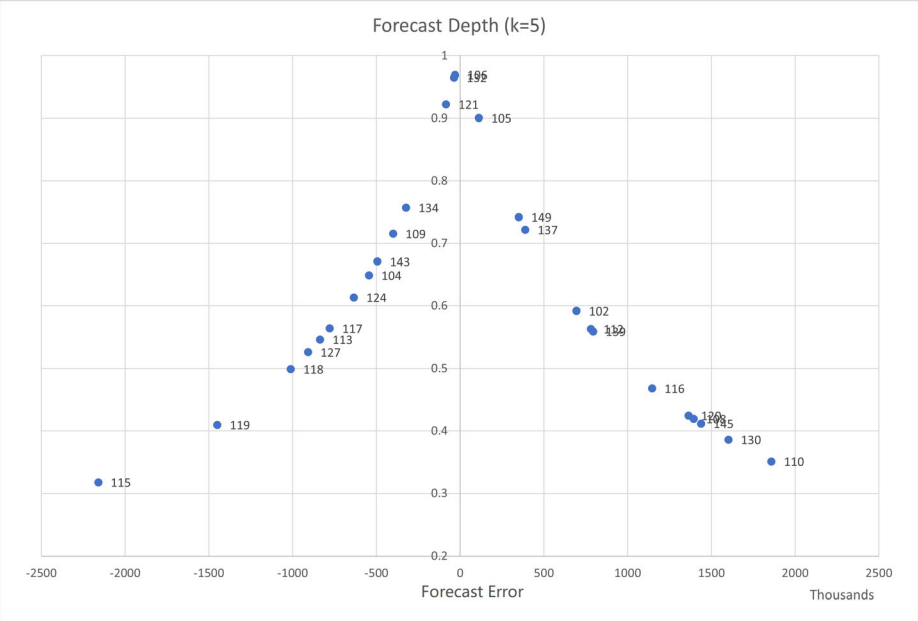

Figure 6: Forecast Depth (k=5)

vector given as

$$z_{i,t+1} = \begin{pmatrix} m'e_{i,t} \\ e_{i,t+1} \end{pmatrix} = \begin{pmatrix} m'(Y_{i,t} - Y_t^0) \\ y_{i,t+1} - y_{t+1}^0 \end{pmatrix} \in \mathbb{R}^2,$$

which is a random sample from an underlying distribution $F_t = (F_{1t}, F_{2t})'$ for all i. To simplify the notations, we drop the subscript "t" in the distribution notations in what follows. When we further suppose that $(e_{i,t}', e_{i,t+1})' = (y_{i,t-k+1}, \ldots, y_{i,t}, y_{i,t+1})' \in \mathbb{R}^{k+1}$ is stationary over t, though it is not required to derived the results below, this simplification is natural. Note that, however, imposing stationarity of the forecasting error does not exclude the potential nonstationarity of the observed series $(Y_t^{0\prime}, y_{t+1}^0)'$ and the forecast series $(Y_{i,t}', y_{i,t+1})'$. The depth estimator $\widehat{\mathcal{D}}_{i,t}$ is only based on the forecasting errors in the training set $e_{i,t}$ through the form of $m'e_{i,t} = m'(Y_{i,t} - Y_t^0)$, and hence we write $\widehat{\mathcal{D}}_{i,t} = \mathcal{D}(m'e_{i,t}, \widehat{F}_1)$ and $\mathcal{D}_{i,t} = \mathcal{D}(m'e_{i,t}, F_1)$, where $\widehat{F} = (\widehat{F}_1, \widehat{F}_2)'$ denotes the empirical distribution of $(m'e_{i,t}, e_{i,t+1})'$.[9] Similarly, we denote $\widehat{\mathcal{O}}_{i,t} = \mathcal{O}(m'e_{i,t}, \widehat{F}_1)$, $\mathcal{O}_{i,t} = \mathcal{O}(m'e_{i,t}, F_1)$, $s_t = s(F_1)$, and $\widehat{s}_t = s(\widehat{F}_1)$.

Using these notations, we can rewrite the sample forecast error of the depth-weighted trimmed combined forecast $\widehat{y}_{t+1}$ in (5) as

$$\begin{aligned} \theta(\widehat{F}) &\equiv \widehat{y}_{t+1} - y_{t+1}^0 \\ &= \frac{\sum_{i=1}^n \left(y_{i,t+1} - y_{t+1}^0\right) 1\{\mathcal{D}(m'e_{i,t}, \widehat{F}_1) \geq \tau\} W(\mathcal{D}(m'e_{i,t}, \widehat{F}_1))}{\sum_{i=1}^n 1\{\mathcal{D}(m'e_{i,t}, \widehat{F}_1) \geq \tau\} W(\mathcal{D}(m'e_{i,t}, \widehat{F}_1))} \\ &= \frac{\int u_2 1\{\mathcal{D}(u_1, \widehat{F}_1) \geq \tau\} W(\mathcal{D}(u_1, \widehat{F}_1)) d\widehat{F}\,(u)}{\int 1\{\mathcal{D}(u_1, \widehat{F}_1) \geq \tau\} W(\mathcal{D}(u_1, \widehat{F}_1)) d\widehat{F}_1\,(u_1)}, \end{aligned} \tag{6}$$

where $u = (u_1, u_2)' \in \mathbb{R}^2$. As the number of forecasts n increases, $\theta(\widehat{F})$ will converge to a depth-weighted trimmed mean forecast error given by[10]

$$\theta(F) = \frac{\int u_2 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF\,(u)}{\int 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF_1\,(u_1)} \tag{7}$$

provided $\sup_{m \in \mathbb{R}^k} |\widehat{s}_t - s_t| = o_p(1)$ and $\sup_{u \in \mathbb{R}^2} |\widehat{F}(u) - F(u)| = o_p(1)$, which holds in general from the standard results. We can rewrite the numerator of $\theta(F)$ in (7) as

$$\int \left\{ \int u_2 F_{2|1}\,(du_2) \right\} 1\left\{\mathcal{D}(u_1, F_1) \geq \tau\right\} W(\mathcal{D}(u_1, F_1)) dF_1\,(u_1)\,, \tag{8}$$

[9] By the affine invariance property of the depth, we have $\mathcal{D}(m'e_{i,t}, F_1) = \mathcal{D}(e_{i,t}, \underline{F}_1)$, where $\underline{F}_1$ is the joint distribution of $e_{i,t} \in \mathbb{R}^k$.

[10] Since $\int 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF_1\,(u_1) = \int 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF\,(u)$, $\theta(F)$ can be also written as

$$\theta(F) = \frac{\int u_2 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF\,(u)}{\int 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) dF\,(u)}.$$

where $\int u_2 dF_{2|1}(u_2) = \mathbb{E}[e_{i,t+1}|m'e_{i,t}]$. This expression implies that the mean forecast error $\theta(F)$ in (7) is a weighted average of the projection of $e_{i,t+1}$ on a linear combination of the past forecast errors $m'e_{i,t}$, where the weights are given by $1\{\mathcal{D}(\cdot, F_1) \geq \tau\}W(\mathcal{D}(\cdot, F_1))$ that is based on the linear combination of the forecast error during the training period.

To obtain the asymptotic representation of the depth-weighted forecast combination, we define the influence function of $\theta(F)$ in (7). We let δ_x be the point-mass distribution at $x \in \mathbb{R}^2$ and $F(\varepsilon, \delta_x) = (1-\varepsilon)F + \varepsilon\delta_x$ be a version of F that is contaminated by an ε amount of an arbitrary point-mass distribution at x, where $0 \leq \varepsilon \leq 1$. Then, the influence function of $\theta(F)$ is defined as

$$\phi(x; \theta(F)) = \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \{\theta(F(\varepsilon, \delta_x)) - \theta(F)\}$$

and the limiting distribution of $\theta(\widehat{F})$ can be obtained from

$$\sqrt{n}\left(\theta(\widehat{F}) - \theta(F)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(z_{i,t+1}; \theta(F)) + o_p(1)$$

for given t. To this end, we first assume the following conditions, similarly as Wu and Zuo (2009) and Lee and Sul (2019). We define

$$\begin{aligned} \widehat{\nu}(\cdot) &= \sqrt{n}(\widehat{F}_1(\cdot) - F_1(\cdot)), \\ \widehat{H}(\cdot) &= \sqrt{n}(\mathcal{D}(\cdot, \widehat{F}_1) - \mathcal{D}(\cdot, F_1)). \end{aligned}$$

We also let $L(\cdot) = -((1-\tau)/\tau)s(\cdot)$ and $U(\cdot) = ((1-\tau)/\tau)s(\cdot)$.

Assumption 1 *(i) $W(\cdot)$ is continuously differentiable with a bounded derivative $\dot{W}(\cdot)$. (ii) $\int 1\{\mathcal{D}(u_1, F_1) \geq \tau\}W(\mathcal{D}(u_1, F_1))dF_1(u_1) > 0$ and $\int |u_2| 1\{\mathcal{D}(u_1, F_1) \geq \tau\}W(\mathcal{D}(u_1, F_1))dF(u) < \infty$. (iii) $s(\widehat{F}_1), s(F_1) \in (0, \infty)$ with satisfying $\sup_{m \in \mathbb{R}^k} |s(\widehat{F}_1) - s(F_1)| = o_p(1)$. (iv) $\sup_{u \in \mathbb{R}^2} |\widehat{F}(u) - F(u)| = o_p(1)$. (v) The joint density function $f(u_1, u_2)$ of $z_{i,t+1}$ exists and satisfies $\int (u_2 - \theta(F))f(u_1, u_2)du_2 < \infty$ at $u_1 = U(F_1), L(F_1)$.*

The following theorem summarizes the asymptotic properties of $\theta(\widehat{F})$ in (6). Note that $\pi_{i,t}, \widehat{\mathcal{D}}_{i,t}, \mathcal{D}_{i,t} \in \mathcal{I}_t$ but we allow that $e_{i,t+1} \in \mathcal{I}_{t+1}$.

Theorem 1 *Suppose Assumption 1 holds. Then, for given m and t, $\theta(\widehat{F}) \to_p \theta(F)$ as $n \to \infty$. Furthermore,*

$$\sqrt{n}\left(\theta(\widehat{F}) - \theta(F)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(z_{i,t+1}; \theta(F)) + o_p(1),$$

where

$$\phi(z_{i,t+1};\theta(F)) = \frac{\phi_1(z_{i,t+1};\theta(F)) + \phi_2(z_{i,t+1};\theta(F)) + \phi_3(z_{i,t+1};\theta(F))}{\int 1(u_1,F_1)W(\mathcal{D}(u_1,F_1))dF_1(u_1)}$$

with

$$\begin{aligned}
\phi_1(z_{i,t+1};\theta(F)) &= (e_{i,t+1}-\theta(F))1(m'e_{i,t},F_1)W(\mathcal{D}(m'e_{i,t},F_1)),\\
\phi_2(z_{i,t+1};\theta(F)) &= \int (u_2-\theta(F))1(u_1,F_1)\dot{W}(\mathcal{D}(u_1,F_1))\phi(m'e_{i,t};\mathcal{D}(u_1,F_1))dF(u),\\
\phi_3(z_{i,t+1};\theta(F)) &= \frac{1-\tau}{\tau}W(\tau)\phi(m'e_{i,t};s(F_1))\int (u_2-\theta(F))f(U(F_1),u_2)du_2\\
&\quad -\frac{1-\tau}{\tau}W(\tau)\phi(m'e_{i,t};s(F_1))\int (u_2-\theta(F))f(L(F_1),u_2)du_2,
\end{aligned}$$

and

$$\begin{aligned}
\phi(\cdot;\mathcal{D}(u_1,F_1)) &= \frac{\mathcal{O}(u_1,F_1)\phi(\cdot;s(F_1))}{s(F_1)(1+\mathcal{O}(u_1,F_1))^2},\\
1(u_1,F_1) &= 1\{L(F_1)\le u_1\le U(F_1)\}.
\end{aligned}$$

$\phi(\cdot;s(F_1))$ is the influence function of $s(F_1)$. Consequently, $\sqrt{n}(\theta(\widehat{F})-\theta(F)) \to_d \mathcal{N}(0,\sigma_{t+1}^2)$ as $n\to\infty$, where $\sigma_{t+1}^2 = \mathbb{E}\left[\phi(z_{i,t+1};\theta(F))^2\right]$.

Proof The consistency follows from Lemma A.5 and Theorem 6 of Zuo (2006) since we have

$$\begin{aligned}
\sup_{u_1\in\mathbb{R}}|\mathcal{D}(u_1,\widehat{F}_1)-\mathcal{D}(u_1,F_1)| &= \sup_{m\in\mathbb{R}^k}\left|\frac{1}{1+|m'e_{i,t}|/s(\widehat{F}_1)}-\frac{1}{1+|m'e_{i,t}|/s(F_1)}\right|\\
&= \sup_{m\in\mathbb{R}^k}\frac{\left|s(\widehat{F}_1)-s(F_1)\right|}{(1+|m'e_{i,t}|/s(F_1))^2}+o_p(1)=o_p(1)
\end{aligned}$$

as we assume $\sup_{m\in\mathbb{R}^k}|s(\widehat{F}_1)-s(F_1)|=o_p(1)$. The asymptotic normality follows similarly as the proof of Theorem 4.1 in Wu and Zuo (2009), so we sketch the proof here. Since $L(\cdot)<U(\cdot)$, we write

$$\sqrt{n}\left(\theta(\widehat{F})-\theta(F)\right) = \frac{\sqrt{n}\int u_2^* 1\{\mathcal{D}(u_1,\widehat{F}_1)\ge\tau\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}(u)}{\int 1\{\mathcal{D}(u_1,\widehat{F}_1)\ge\tau\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)}$$

$$= \frac{\sqrt{n}\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)}\left\{\int u_2^* d\widehat{F}_{2|1}(u_2|u_1)\right\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)}{\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)},$$

where $u_2^* = u_2 - \theta(F)$. We decompose the numerator into

$$\begin{aligned}
N_{1n} &= \sqrt{n}\int_{L(F_1)}^{U(F_1)} \widehat{g}(u_1)W(\mathcal{D}(u_1,F_1))d\widehat{F}_1(u_1),\\
N_{2n} &= \int_{L(F_1)}^{U(F_1)} \widehat{g}(u_1)\sqrt{n}\left\{W(\mathcal{D}(u_1,\widehat{F}_1)) - W(\mathcal{D}(u_1,F_1))\right\}d\widehat{F}_1(u_1),\\
N_{2n} &= \sqrt{n}\left\{\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)} - \int_{L(F_1)}^{U(F_1)}\right\}\widehat{g}(u_1)W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1),
\end{aligned}$$

where $\widehat{g}(u_1) = \int u_2^* d\widehat{F}_{2|1}(u_2|u_1)$.

For N_{1n}, we immediately have

$$N_{1n} = \sqrt{n}\int u_2^* 1\{\mathcal{D}(u_1,F_1)\geq\tau\}W(\mathcal{D}(u_1,F_1))d\widehat{F}(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_{1i},$$

where

$$\eta_{1i} = (e_{i,t+1} - \theta(F))1\left\{L(F_1)\leq m'e_{i,t}\leq U(F_1)\right\}W(\mathcal{D}(m'e_{i,t},F_1)).$$

For N_{2n}, we note that $\sup_{u_1\in\mathbb{R}}|\widehat{g}(u_1) - g(u_1)| = o_p(1)$ with $g(u_1) = \int u_2^* dF_{2|1}(u_2|u_1)$ from the standard results of nonparametric conditional expectation estimators. We thus have

$$N_{2n} = \int_{L(F_1)}^{U(F_1)} g(u_1)\dot{W}(\widehat{\Delta}(u_1))\widehat{H}(u_1)d\widehat{F}_1(u_1) + o_p(1)$$

for some $\widehat{\Delta}(u_1)$ between $\mathcal{D}(u_1,\widehat{F}_1)$ and $\mathcal{D}(u_1,F_1)$, where $\sup_{u_1\in[L(F_1),U(F_1)]}|\widehat{\Delta}(u_1) - \mathcal{D}(u_1,F_1)| \leq \sup_{u_1\in[L(F_1),U(F_1)]}|\mathcal{D}(u_1,\widehat{F}_1) - \mathcal{D}(u_1,F_1)| \leq C\sup_{m\in\mathbb{R}^k}|s(\widehat{F}_1) - s(F_1)| = o_p(1)$ for some positive $C<\infty$. By Lemma A.3 of Wu and Zuo (2009), $\sup_{u_1\in[L(F_1),U(F_1)]}(1+|u_1|)|\widehat{H}(u_1)| = O_p(1)$ and there exists $\phi(x_1;\mathcal{D}(u_1,F_1))$ such that $\widehat{H}(u_1) = \int\phi(x_1;\mathcal{D}(u_1,F_1))d\widehat{\nu}(x_1) + o_p(1)$ uniformly over $x_1\in[L(F_1),U(F_1)]$. Similarly as the proof of Theorem 2 in Lee and Sul (2019), therefore, we can verify that

$$\begin{aligned}
N_{2n} &= \int_{L(F_1)}^{U(F_1)} g(u_1)\dot{W}(\mathcal{D}(u_1,F_1))\left(\int\phi(x_1;\mathcal{D}(u_1,F_1))d\widehat{v}(x_1)\right)dF_1(u_1) + o_p(1)\\
&= \iint u_2^* 1\left\{L(F_1)\leq u_1\leq U(F_1)\right\}\dot{W}(\mathcal{D}(u_1,F_1))\phi(x_1;\mathcal{D}(u_1,F_1))dF(u)d\widehat{v}(x_1) + o_p(1)
\end{aligned}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_{2i} + o_p(1),$$

where

$$\eta_{2i} = \int u_2^* 1\{L(F_1) \le u_1 \le U(F_1)\}\dot{W}(\mathcal{D}(u_1,F_1))\phi(m'e_{i,t};\mathcal{D}(u_1,F_1))dF(u)$$

and $\phi(\cdot;\mathcal{D}(u_1,F_1))$ is the influence function of $\mathcal{D}(\cdot,F_1)$ given by

$$\phi(x_1;\mathcal{D}(u_1,F_1)) = \frac{\mathcal{O}(u_1,F_1)\phi(x_1;s(F_1))}{s(F_1)(1+\mathcal{O}(u_1,F_1))^2}$$

with $\phi(\cdot;s(F_1))$ being the influence function of $s(F_1)$. For N_{3n}, we similarly have

$$\begin{aligned}
N_{3n} &= \sqrt{n}\int_{U(F_1)}^{U(\widehat{F}_1)} g(u_1)W(\mathcal{D}(u_1,F_1))dF_1(u_1) \\
&\quad -\sqrt{n}\int_{L(F_1)}^{L(\widehat{F}_1)} g(u_1)W(\mathcal{D}(u_1,F_1))dF_1(u_1) + o_p(1) \\
&= \sqrt{n}\iint_{U(F_1)}^{U(\widehat{F}_1)} u_2^* W(\mathcal{D}(u_1,F_1))f(u_1,u_2)du_1du_2 \\
&\quad -\sqrt{n}\iint_{L(F_1)}^{L(\widehat{F}_1)} u_2^* W(\mathcal{D}(u_1,F_1))f(u_1,u_2)du_1du_2 + o_p(1) \\
&= \int u_2^*\sqrt{n}\left\{U(\widehat{F}_1) - U(F_1)\right\}W(\mathcal{D}(U(F_1),F_1))f(U(F_1),u_2)du_2 \\
&\quad -\int u_2^*\sqrt{n}\left\{L(\widehat{F}_1) - L(F_1)\right\}W(\mathcal{D}(L(F_1),F_1))f(L(F_1),u_2)du_2 + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_{3i} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\eta_{3i} &= W(\tau)\phi(m'e_{i,t};U(F_1))\int u_2^* f(U(F_1),u_2)du_2 \\
&\quad -W(\tau)\phi(m'e_{i,t};L(F_1))\int u_2^* f(L(F_1),u_2)du_2
\end{aligned}$$

since $\mathcal{D}(L(F_1),F_1) = \mathcal{D}(U(F_1),F_1) = \tau$. Note that the influence functions of $U(F_1)$ and $L(F_1)$

are defined as

$$\begin{aligned}\phi(x_1; U(F_1)) &= ((1-\tau)/\tau)\phi(x_1; s(F_1)),\\ \phi(x_1; L(F_1)) &= -((1-\tau)/\tau)\phi(x_1; s(F_1)).\end{aligned}$$

We can similarly verify $\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)} W(\mathcal{D}(u_1, \widehat{F}_1))d\widehat{F}_1(u_1) = \int_{L(F_1)}^{U(F_1)} W(\mathcal{D}(u_1, F_1))dF_1(u_1) + o_p(1)$ in the denominator, and the desired result follows by combining the expressions η_{1i}, η_{2i}, and η_{3i} above. □

From Theorem 1, we can conclude that the depth-weighted forecast combination $\widehat{y}_{t+1}$ satisfies $\widehat{y}_{t+1} \to_p y^0_{t+1} + \theta(F)$ and $\sqrt{n}(\widehat{y}_{t+1} - \{y^0_{t+1} + \theta(F)\}) \to_d \mathcal{N}(0, \sigma^2_{t+1})$ as $n \to \infty$. Apparently, when $\theta(F) = 0$, $\widehat{y}_{t+1}$ becomes a consistent forecast.

The specific form of the asymptotic variance $\sigma^2_{t+1} = \mathbb{E}\left[\phi(z_{i,t+1}; \theta(F))^2\right]$ depends on the the influence function of $s_t = s(F_1)$. For instance, for the square root of the FMSE s_t in (2), we can derive

$$\phi(x_1; s(F_1)) = \frac{x_1^2 - \mathbb{E}[(m'e_{i,t})^2|\mathcal{I}_t]}{2(\mathbb{E}[(m'e_{i,t})^2|\mathcal{I}_t])^{1/2}}$$

when $0 < \mathbb{E}[(m'e_{i,t})^2|\mathcal{I}_t] < \infty$ using influence function of the mean. For the FMAD s_t in (3), we let $\text{sgn}(c) = 1$ if $c > 0$; 0 if $c = 0$; -1 if $c < 0$. Then, using the influence functions of the median, we can derive

$$\phi(x_1; s(F_1)) = \frac{\text{sgn}(x_1 - \lambda(m'e_{i,t}))}{2f_1(\lambda(m'e_{i,t}))},$$

where $\lambda(m'e_{i,t}) = \inf\{v : \mathbb{P}(|m'e_{i,t}| \le v|\mathcal{I}_t) \ge 1/2\}$ is the conditional median of $m'e_{i,t}$, provided that the marginal density function $f_1(\cdot)$ of $m'e_{i,t}$ at t satisfies $0 < f_1(\lambda(m'e_{i,t})) < \infty$. Therefore, the choice if s_t determines the robustness of the forecast combination. When the support of F_1 is not bounded, the influence function of the FMSE is not necessarily bounded, whereas that of the FMAD is bounded. Hence the latter is more robust towards outliers or very under-performed forecasts.

Though the analytical form of σ^2_{t+1} is complicated, it can be simply estimated as

$$\widehat{\sigma}^2_{t+1} = \sum_{i=1}^{n}\left[\sqrt{n}\pi_{i,t}(y_{i,t+1} - \widehat{y}_{t+1})\right]^2 = \frac{\sum_{i=1}^{n}\left[\sqrt{n}1\{\widehat{\mathcal{D}}_{i,t} \ge \tau\}W(\widehat{\mathcal{D}}_{i,t})(y_{i,t+1} - \widehat{y}_{t+1})\right]^2}{\left[\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \ge \tau\}W(\widehat{\mathcal{D}}_{i,t})\right]^2}$$

because $\widehat{y}_{t+1}$ is in the form of a weighted average. Hence, the $100(1-\alpha)\%$ pointwise confidence

interval of the combined forecast $\widehat{y}_{t+1}$can be readily obtained as

$$\left[\widehat{y}_{t+1} \pm z_{\alpha/2}\frac{\widehat{\sigma}_{t+1}}{\sqrt{n}}\right], \tag{9}$$

where $z_{\alpha/2}$ is the $(1-(\alpha/2))$th quantile of the standard normal distribution.

5 Concluding Remarks

In this paper, we develop the notion of forecast depth and a depth-weighted forecast combination with trimming. Since the weights are not obtained by minimizing some loss function, we do not discuss any optimality properties. However, the weights can be calculated even when we have many forecasts but the training period is as short as just one, and hence it has much potential in practice as complementing other forecast combinations. In comparison, when long training period is available, a recent work by Diebold and Shin (2019) uses LASSO in estimating the weights by minimizing a L_2 loss function with a L_1 penalty term, from which we can also obtain weights with some trimming (or selection).

Though we focus on a specific form of the forecast depth, we can also consider other types of depth and extend them to this context. When we have a balanced longitudinal data structure (i.e., all forecasts are available over the same period), we can construct the forecast depth based on the Mahalanobis distance. It uses the k-dimensional vector of the forecast error $e_{i,t}$ as $e_{i,t}'\Sigma_t^{-1}e_{i,t}$ without any normalization, where Σ_t is the $k \times k$ variance matrix of $e_{i,t}$, and hence it counts the forecast performance of a specific time during the training period more heavily if the cross-sectional FMSE is small. The projection depth is another option, provided the length of the training period is short and the number of forecasts is very large, though its computation becomes almost infeasible if the length of training period is larger than two. For more discussions of these depths, see Lee and Sul (2019).

We can use this forecast combination idea for multivariate forecasting. Since we can construct depth-based contour (i.e., multivariate quantile), it can provide a ranking among different models based on their forecast performance for multiple economic variables together. In addition, depending on the choice of the dispersion term s_t, the depth-weighted forecast combination does not necessarily require existence of the moments of the forecasts. Therefore, it can be applied for financial data with fat-tailed distributions.

References

Aiolfi, M. and A. Timmermann (2006). Persistence in Forecasting Performance and Conditional Combination Strategies, *Journal of Econometrics*, 135(1–2), 31–53.

Bates, J.M. and C.W.J. Granger (1969). The Combination of Forecasts, *Operational Research Quarterly*, 20(4), 451–468.

Busetti, F. (2017). Quantile Aggregation of Density Forecasts, *Oxford Bulletin of Economics and Statistics*, 79(4), 495-512.

Chang, Y., R.K. Kaufmann, C.S. Kim, J.I. Miller, J.Y. Park, and S. Park (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214(1), 274-294

Chang, Y., C.S. Kim, and J.Y. Park (2016). Nonstationarity in time series of state densities, *Journal of Econometrics*, 192(1), 152-167.

Clemen, R.T. (1989). Combining Forecasts: A Review and Annotated Bibliography, *International Journal of Forecasting*, 5(4), 559–583.

Diebold, F.X. and P. Pauly (1990). The Use of Prior Information in Forecast Combination, *International Journal of Forecasting*, 6(4), 503–508.

Diebold, F.X. and M. Shin (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives, *International Journal of Forecasting*, 35(4), 1679-1691.

Granger, C.W.J. and Y. Jeon (2004). Thick Modelling, *Economic Modelling*, 21(2), 323–343.

Hu, B., J.Y. Park, and J. Qian (2017). Analysis of Distributional Dynamics for Repeated Cross-Sectional and Intra-Period Observations, *working paper*, Indiana University.

Lee, Y. and D. Sul (2019). Depth-weigthed estimation of panel data models, *working paper*, Syracuse University.

Stock, J.H., Watson, M. (2001). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, In *Festschrift in Honour of Clive Granger*, R.F. Engle and H. White (eds.), Cambridge: Cambridge University Press, pp. 1-44.

Timmermann, A. (2006). Forecast Combinations, In *Handbook of Economic Forecasting, Vol. 1*, G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), Amsterdam: North-Holland, pp. 135-196.

Wu, M. and Y. Zuo (2009). Trimmed and Winsorized means based on a scaled deviation, *Journal of Statistical Planning and Inference*, 139(2), 350-365.

Zuo, Y. (2006). Multidimensional trimming based on projection depth, *Annals of Statistics*, 34(5), 2211-2251.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function, *Annals of Statistics*, 28(2), 461-482.