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

Yoonseok Lee*
Syracuse University

Donggyu Sul†
University of Texas at Dallas

July 2022

Abstract

We develop a novel forecast combination approach based on the order statistics of individual predictability from panel data forecasts. To this end, we define the notion of forecast depth, which provides a ranking among different forecasts based on their normalized forecast errors during the training period. The forecast combination is in the form of a depth-weighted trimmed mean. We derive the limiting distribution of the depth-weighted forecast combination, based on which we can readily construct prediction intervals. Using this novel forecast combination, we predict 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 and robust predictions compared with other popular forecast combinations and reports much narrower prediction intervals.

Keywords: Forecast depth, Forecast combination, Panel forecast, Prediction interval, Robust forecast, COVID-19

JEL Classifications: C32, C33, C53

Address: Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244. *E-mail*: ylee41@syr.edu

†*Address*: Department of Economics, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080. *E-mail*: 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. It is also known that combined forecasts often produce better forecasts than the ex ante best single forecasting model. If we know the individual forecasting models and their information sets (i.e., the predictors) that produce multiple forecast values, we take the model averaging approach and obtain the optimal weights by minimizing the forecast mean squared error loss. When the number of candidate models and the information sets are large, however, it can be very costly to pool the information sets particularly in real-time forecast (e.g., Diebold and Pauly, 1990). In practice, it is often the case that the individual forecasting models are not fully known and only the forecast reports are available from the forecasting agencies. In such cases, we can apply ensemble methods such as bagging and boosting to combine forecasts. Such approaches, however, typically require a long training period but a small number of forecasting models, so that the weights can be properly estimated. For this reason, it is common to use the equally weighted average of forecasts or the weighted average based on the inversed forecast mean squared errors (e.g., Stock and Watson, 2001). The equal weight approach is the most popular because it is simple but typically outperforms the estimated optimal weights, which is often called the "forecast combination puzzle". See Clemen (1989), Stock and Watson (2001, 2006), and Timmermann (2006), for instance, for survey of this literature.

In this paper, we propose a forecast combination based on the order statistics of individual predictability when many forecasts are available. We assume (cross-sectional) data rich environment of the forecasts but we do not know of each forecasting model nor its information set. In other words, we only have a set of forecast values reported from many different forecasting agencies, where we do not know their forecasting models, data sets, or predictors. The weights can be obtained using the cross-sectional information and hence we do not need a long training period to estimate the weights. For this reason, we can apply this novel method to a very short panel forecast data set, which is not typically the case for the inversed forecast mean squared error combination. This feature is very useful in practice especially when the time series of interest highly fluctuates and hence the forecasting accuracy from each forecasting agency is not consistent over time.

More precisely, we develop the *forecast depth*, by modifying the notion of data depth (e.g., Zuo and Serfling, 2000; Lee and Sul, 2022a) in the context of forecast combination, which measures the nearness of each vector of forecasts toward the vector of observed values over the training period. The weights for forecast combination are proportional to the forecast depth. Rather than deriving the optimal forecasting weights, we seek for a robust forecast combination against erroneous forecasts (i.e., outliers). Note that the forecast depth naturally provides a ranking among the forecasting agencies. To design a more robust forecast combination toward

extremely poor forecasts during the training period, we trim out the forecasts by the agencies who belong to the lowest ranking group. 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 combinations. 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 popular combination methods, including the equal weight combination and the inversed forecast mean squared error combination.

The main contribution of this paper can be summarized in three folds. First, we develop the forecast depth, based on which we 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 prediction interval of forecast combination without relying on subsampling or bootstrap. Both the weight and the trimming scheme depend on the forecast depths in the sample, and hence they are treated stochastic in deriving the limiting distribution. Since the proposed form of forecast combination encompasses popular ones as its special cases, such as the equally-weighted forecast combination, the trimmed forecast combination, and the median forecast, this result also provides prediction intervals 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. The new forecast combination yields lower forecast mean squared errors than the ensemble forecast reported by the Centers for Disease Control and Prevention (CDC) as well as other popular forecast combination approaches including the equally-weighted combination and inversed forecast mean squared error combination. It provides forecast that is very robust to erroneous or extremely bad predictions in the pool. In addition, it reports much narrower prediction intervals.

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 to reinforce the forecast combination especially in long horizon forecasting when a long panel forecast data set is available. 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 forecast combination. Section 3 compares the forecast depth with the projection depth and the inversed forecast mean squared errors. Section 4 derives the

limiting distribution of the depth-weighted forecast combination and provides the prediction interval. Using the panel of forecasts on new COVID-19 cases in the U.S., Section 5 examines the performance of the proposed forecast combination and compares it with other popular forecast combination approaches. Section 6 concludes with some remarks. The proof of the main theorem is in the Appendix.

2 Forecast Depth and Forecast Combination

We denote y_t^0 as the observed true value of interest at time t and $\{y_{i,t}\}$ be the multiple competing forecasts for y_t^0 from different forecasting agencies $i = 1, \ldots, n$ using different forecasting models or predictors. For the simplicity, we let y_t^0 and $y_{i,t}$ be scalar values, though the idea below can be extended to the vector case when we jointly predict multiple variables.

We suppose there are n different h-step ahead forecasts $y_{i,t+h}$ for y_{t+h}^0. We do not know how each agency produces her own forecast value $y_{i,t+h}$; in other words, we do not know her forecasting model nor predictors. We consider the h-step ahead forecast combination in the form of a linear combination of the forecasts:

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

for some potentially time-varying weights $\pi_{i,t}$, which are obtained using the information available at time t. Popular choices of the weights $\pi_{i,t}$ include the equal weight (i.e., $\pi_{i,t} = 1/n$) and the inversed forecast mean squared error (e.g., $\pi_{i,t} = \widehat{mse}_{i,t}^{-1} / \sum_{j=1}^{n} \widehat{mse}_{j,t}^{-1}$, where $\widehat{mse}_{i,t}$ is the sample forecast mean squared error of agency i using information at time t). In this paper, we define the weights $\pi_{i,t}$ based on a novel idea, the forecast depth.

To define the forecast depth, we first let the $k \times 1$ vector of forecasts during the training period from $t - k + 1$ to t,[1]

$$Y_{i,t} = (y_{i,t-k+1}, \ldots, y_{i,t-1}, y_{i,t})' \text{ for } i = 1, \ldots, n,$$

which is the most recent k observations at time t, and the $k \times 1$ vector of the observed true values during this period,

$$Y_t^0 = (y_{t-k+1}^0, \ldots, y_{t-1}^0, y_t^0)'.$$

[1] When the data set consists of h-step ahead forecasts for multiple values of h, we can use the h-step ahead forecasts $Y_{i,t}$ during the training period for the h-step ahead forecast of y_{t+h}. In this cases, the weights $\pi_{i,t}$ can be different across h, say $\pi_{i,t}^h$. But the weights still use the information available at t and we omit the index h in $\pi_{i,t}$ for the sake of notational simplicity.

We denote the $k \times 1$ forecast error vector of the agency i during the training period as

$$e_{i,t} = Y_{i,t} - Y_t^0 \text{ for } i = 1, \ldots, n.$$

We let $m = (m_1, \ldots, m_k)'$ be a non-random $k \times 1$ discount vector with $\sum_{j=1}^k m_j = 1$, whose examples are to be discussed later. Given m, we define the normalized forecast error of $Y_{i,t}$ (or the forecast outlyingness) as

$$\mathcal{O}_{i,t} = \frac{|m'e_{i,t}|}{s_t} = \frac{\left|m'(Y_{i,t} - Y_t^0)\right|}{s_t}, \tag{2}$$

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, say $\mathcal{I}_t = \sigma(\cup_{i=1}^n \{e_{i,r}\}_{r \leq t})$.[2] For instance, if $\{e_{i,t}\}$ is a random sample from a common distribution across i, we can consider the conditional root mean squared error (RMSE),

$$s_t = (\mathbb{E}[(m'e_{i,t})^2 | \mathcal{I}_t])^{1/2} \tag{3}$$

when the conditional mean of $m'e_{i,t}$ is assumed to be zero; or the conditional median absolute deviation (MAD),

$$s_t = \inf \left\{ v : \mathbb{P}\left(\left|m'e_{i,t}\right| \leq v | \mathcal{I}_t\right) \geq 1/2 \right\} \tag{4}$$

when the conditional median of $m'e_{i,t}$ is assumed to be zero.[3] We define the *forecast depth* of agency i at time t as

$$\mathcal{D}_{i,t} = \frac{1}{1 + \mathcal{O}_{i,t}}, \tag{5}$$

where it is obtained using the data set in the k training period from $t - k + 1$ to t and hence $\mathcal{D}_{i,t} \in \mathcal{I}_t$. By construction, the forecast depth $\mathcal{D}_{i,t}$ takes values between zero and one; it is one when the agency i yields perfect forecasts during the training period and hence $e_{i,t} = 0$.

Several examples of the discount vector m can be considered. In the forecast error vector

$$e_{i,t} = Y_{i,t} - Y_t^0 = ((y_{i,t-k+1} - y_{t-k+1}^0), \ldots, (y_{i,t-1} - y_{t-1}^0), (y_{i,t} - y_t^0))',$$

since we typically consider that the forecast performance for the most recent observations are

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

[3] If the distribution of $\{e_{i,t}\}$ is heterogeneous across i, we can 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 $\sum_{j=1}^k m_{ij} = 1$ for each i. In such cases, s_t can be alternatively defined as $(\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathbb{E}\left[(m_i'e_{i,t})^2 | \mathcal{I}_t\right])^{1/2}$ and $\lim_{n\to\infty} \text{med}_{1 \leq i \leq n} \inf \{v_i : \mathbb{P}\left(|m_i'e_{i,t}| \leq v_i | \mathcal{I}_t\right) \geq 1/2\}$, respectively.

more important than the distant ones, we can let the jth element of $m = (m_1, \ldots, m_k)'$ as

$$m_j = \frac{K(j/k)}{\sum_{\ell=1}^{k} K(\ell/k)} \quad \text{for } j = 1, \ldots, k, \tag{6}$$

where $K(\cdot)$ is some non-decreasing one-side kernel function. Examples include:

- the polynomial kernel, $K(j/k) = (j/k)^q$ for some $q \geq 1$;
- the discount factor approach by Bates and Granger (1969), $K(j/k) = \gamma^{k-j}$ for some $\gamma < 1$;
- the Box–Cox transform weights by Diebold and Pauly (1987).

If we treat all the forecast errors during the training period equally important, then we simply set $K(j/k) = 1$ or

$$m_j = \frac{1}{k} \quad \text{for all } j = 1, \ldots, k.$$

At given t, we can also define m_j as the normalized inverse of cross-sectional MSE, by letting $K(j/k) = [n^{-1} \sum_{i=1}^{n} (y_{i,t-j+1} - y^0_{t-j+1})^2]^{-1}$. Note that all such choices of m do not require a balanced panel data structure in $Y_{i,t}$; that is all forecasts are not necessarily available over the given training period. Therefore, for each i, we can even define a $k_i \times 1$ heterogeneous discount vector $m_i = (m_{i1}, \ldots, m_{ik_i})'$ by letting $m_{ij} = K(j/k_i) / \sum_{\ell=1}^{k_i} K(\ell/k_i)$ for $j = 1, \ldots, k_i$ with $k_i \leq k$.

We propose to define the weights $\pi_{i,t}$ for the forecast combination in (1) that is proportional to the individual forecast depth (5). The forecast depth $\mathcal{D}_{i,t}$ naturally provides a ranking of predictability through the entire training period among the (unrevealed) forecasting models from different agencies, since the better performing ones have higher levels of forecast depth. Therefore, we can also use the forecast depth as a tool to detect the under-performed forecasting agencies. To design a more robust forecast combination toward extremely poor forecasts, we trim out the forecasts by the agencies who belong to the low ranking group. More precisely, we set some trimming parameter $\tau \in (0, 1)$ and let

$$\pi_{i,t} = \frac{1\{\widehat{\mathcal{D}}_{i,t} \geq \tau\} W(\widehat{\mathcal{D}}_{i,t})}{\sum_{j=1}^{n} 1\{\widehat{\mathcal{D}}_{j,t} \geq \tau\} W(\widehat{\mathcal{D}}_{j,t})}, \tag{7}$$

where $1\{\cdot\}$ is the binary indicator[4] and $W(\cdot)$ is some scalar weight function $W : [0, 1] \to [0, 1]$.

[4]Instead of the hard-threshold trimming, we could consider a smooth transition function such as

$$\begin{cases} \dfrac{\exp\left(-c\left(1 - (d/\tau)\right)^2 - \exp\left(c\right)\right)}{1 - \exp\left(c\right)} & \text{if } d < \tau \\ 1 & \text{if } d \geq \tau \end{cases}$$

for some smoothing parameter $c > 0$. As $c \to \infty$, it approaches to a binary indicator function $1\{d \geq \tau\}$.

For most of the cases, we can simply choose $W(d) = d$. $\widehat{\mathcal{D}}_{i,t}$ is the forecast depth estimator that is defined as

$$\widehat{\mathcal{D}}_{i,t} = \frac{1}{1+\widehat{\mathcal{O}}_{i,t}} \quad \text{with} \quad \widehat{\mathcal{O}}_{i,t} = \frac{\left|m_i'(Y_{i,t} - Y_t^0)\right|}{\widehat{s}_t} \tag{8}$$

for some consistent estimator $\widehat{s}_t$. For each of the aforementioned examples in (3) and (4), we can use

$$\widehat{s}_t = \left(\frac{1}{n}\sum_{i=1}^{n} m'(Y_{i,t} - Y_t^0)(Y_{i,t} - Y_t^0)'m\right)^{1/2} \tag{9}$$

and

$$\widehat{s}_t = \text{med}_{1\le i\le n}\left|m'(Y_{i,t} - Y_t^0)\right|, \tag{10}$$

respectively. Since we define the training period over a rolling window of the most recent k periods, the weight in (7) is naturally time-varying.

Using the weight (7) in the forecast combination (1), we define *the depth-weighted forecast combination* as a form of the trimmed depth-weighted mean given as

$$\widehat{y}_{t+h} = \sum_{i=1}^{n} \pi_{i,t} y_{i,t+h} = \frac{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \ge \tau\} W(\widehat{\mathcal{D}}_{i,t}) y_{i,t+h}}{\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} \ge \tau\} W(\widehat{\mathcal{D}}_{i,t})}, \tag{11}$$

where $\{y_{i,t+h}\}_{i=1}^n$ are the individual h-step ahead forecasts for y_{t+h}^0 at the current time t.

The forecast combination in (11) assigns the weight on $y_{i,t+h}$ based on its forecast depth during the k training period. For this reason, we choose a small k when the time series of interest y_t^0 is very volatile and hence the forecasting performance of each agency highly fluctuates. If the ith agency's forecast error $m'e_{i,t}$ is near zero and hence the forecast depth estimator $\widehat{\mathcal{D}}_{i,t}$ is near maximum (i.e., near unity), then its forecast $y_{i,t+h}$ gets a high weight; if its forecast error is too large, on the other hand, it gets a low weight or even a zero weight by being trimmed. In this sense, unlike the forecast combination based on the equal weights or forecast mean squared errors, the depth-weighted forecast combination $\widehat{y}_{t+h}$ in (11) 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 $\widehat{y}_{t+h}$ is a randomly trimmed forecast combination.

Based on the choice of τ and $W(\cdot)$, $\widehat{y}_{t+h}$ in (11) becomes other popular forecast combinations. For instance, when $W(\cdot) = 1$, $\widehat{y}_{t+h}$ is the trimmed equally-weighed combined forecast, which converges to the trimmed mean of $y_{i,t+h}$ as $n \to \infty$ if it exists; when $\tau = 0$ in addition, it is simply the (untrimmed) equally-weighed combined forecast. When $\tau = \max_{1\le i\le n} \widehat{\mathcal{D}}_{i,t}$, $\widehat{y}_{t+h}$ is the same as $y_{i,t+h}$ whose forecast depth is the maximal; if $m_j = 1/k$ in addition, this maximal depth forecast corresponds to the forecast of agency i whose forecasts has been the most accurate during the entire training period (i.e., ex ante the best single forecast). When $y_{i,t+h}$ has a density function that is elliptically symmetric about its mode, this maximal depth forecast becomes the

median combination forecast.

We finish this section by summarizing the steps of obtaining the forecast combination $\widehat{y}_{t+h}$.

1. Determine the length of training period k and obtain the k-period forecast error $e_{i,t} = (Y_{i,t} - Y_t^0) = ((y_{i,t-k+1} - y_{t-k+1}^0), \ldots, (y_{i,t-1} - y_{t-1}^0), (y_{i,t} - y_t^0))'$ for each i.
2. Determine the form of $k \times 1$ discount vector m in (6) and estimate the dispersion measure s_t of $m'e_{i,t}$ as in (9) or (10).
3. Estimate the forecast depth $\widehat{\mathcal{D}}_{i,t}$ as in (8).
4. Determine the trimming level $\tau \in (0, 1)$ and some scalar weight function $W(\cdot)$. (One can simply let $W(d) = d$.) Then, obtain the weights $\pi_{i,t}$ as in (7).
5. Obtain the forecast combination $\widehat{y}_{t+h} = \sum_{i=1}^{n} \pi_{i,t} y_{i,t+h}$ as in (11).
6. The prediction interval can be obtained as (20) in Section 4.

3 Comparisons and Discussions

3.1 Projection Depth

The forecast depth $\mathcal{D}_{i,t}$ developed in the previous section is in a similar form as one of the popular data depths: the projection depth (e.g., Liu, 1992; Zuo and Serfling, 2000). The data depth measures the outlyingness of a given multivariate sample point with respect to its underlying joint distribution, which is formulated as an index between 0 and 1. If the data point is at the center of the distribution, then the depth value of the data point becomes unity. If a data point locates very far from the center, then the depth value of the point becomes near zero. Recall that, for a k-dimensional random sample $\{Z_i\}$, the projection depth of Z_i is given as $PD_i = 1/[1 + PO_i]$, where the projection-based outlyingness is defined as

$$PO_i = \sup_{\xi \in \mathbb{R}^k : \|\xi\| = 1} \frac{|\xi' Z_i - \mu(\xi' Z_i)|}{\sigma(\xi' Z_i)} \tag{12}$$

for some univariate location and dispersion parameters, $\mu(\xi' Z_i)$ and $\sigma(\xi' Z_i)$, of the distribution of $\xi' Z_i$. PO_i is defined to be zero when $\xi' Z_i - \mu(\xi' Z_i) = \sigma(\xi' Z_i) = 0$.

Comparing the forecast depth $\mathcal{D}_{i,t}$ (or forecast outlyingness in (2)) and the projection depth PD_i (or projection-based outlyingness in (12)), we can point two important differences. First, the forecast depth considers the distance from each forecast $m'Y_{i,t}$ toward the observed true $m'Y_t^0$ (i.e., the forecast error $m'e_{i,t} = m'(Y_{i,t} - Y_t^0)$), whereas the projection depth considers the distance toward a central location parameter $\mu(\xi' Z_i)$ of the distribution of $\xi' Z_i$, such as the

mean or the median. Since the true $m'Y_t^0$ is not necessarily the center of the distribution of the forecasts $m'Y_{i,t}$, the forecast depth has different implications from the projection depth. It should be emphasized that the original notion of depth is mainly motivated to define a robust central location of multi-dimensional variables. On the other hand, for the forecasting problem, the target location is not the centrality of the distribution of the forecasts $Y_{i,t}$; instead, the target is already given as the observed true value vector Y_t^0. The forecast depth provides a normalized distance from a vector of forecasts $Y_{i,t}$ toward the vector of observed values Y_t^0.

Second, we preset the discount vector m in defining the forecast depth that can be potentially heterogeneous, whereas the projection depth needs to search for the ξ vector as defined in (12) so that the outlyingness is maximized to the particular direction. This is possible in the forecasting problem because the researcher often has ordering of the importance among the forecast errors during the training period as we see several examples in the previous section. Though it seems unnecessary, one could find m in defining the forecast depth as ξ in the projection depth. It should be noted that, however, searching for such an m vector is computationally very costly when the dimension of m (i.e., the length of training period k) exceeds 2, which is also a well-known limitation of the projection depth.

Though the forecast depth is different from the standard data depths, it still 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 rescaling of the forecast error vector $e_{i,t}$ (*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 forecasting agencies or, in other words, outliers.

The depth-weighted forecast combination in (11) is in the form of the depth-weighted trimmed mean (e.g., Zuo, 2006), and hence it shares the same robustness properties toward outliers (i.e., the very poor forecasts during the training period). In particular, unlike the equal weight combination or inversed forecast mean squared error combination, the depth-weighted forecast combination does not swing much or even stays unchanged when some agencies in the sample yield extremely poor performance during the training period. For instance, when we use the MAD for s_t as in (4), it can be shown that the breakdown point (i.e., the smallest fraction of contaminants in a sample that causes the forecast combination to break down) can reach to the maximal level. For more discussions of the depths and applications in panel data, see Lee and Sul (2022a, 2022b).

3.2 Inversed MSE

Forecast combination based on the inversed forecast mean squared errors (iMSE hereafter; e.g., Stock and Watson, 2001) is a popular approach in practice. For the $k \times 1$ forecast error vector $e_{i,t}$, the sample iMSE is obtained as $[e'_{i,t}e_{i,t}/k]^{-1}$ and the iMSE combination defines the weight $\pi_{i,t}$ in (1) as

$$\frac{[e'_{i,t}e_{i,t}/k]^{-1}}{\sum_{j=1}^{n}[e'_{j,t}e_{j,t}/k]^{-1}} = \frac{\left[(1/k)\sum_{\ell=1}^{k}\left(y_{i,t-\ell+1} - y^0_{t-\ell+1}\right)^2\right]^{-1}}{\sum_{j=1}^{n}\left[(1/k)\sum_{\ell=1}^{k}\left(y_{j,t-\ell+1} - y^0_{t-\ell+1}\right)^2\right]^{-1}}. \tag{13}$$

However, one limitation of this weight is that it could overly praise the perfect forecast during the training period, especially when the training period is short. As an extreme example, when we only have one training period (i.e., $k = 1$), if one agency i yields the perfect forecast and hence $e_{i,t} = 0$ or $(e^2_{i,t})^{-1}$ is unbounded, then her weight $\pi_{i,t}$ in (13) for the forecast combination becomes 1 even when there are other agencies who produce reasonable or even near perfect forecasts. In comparison, the forecast depth is always bounded by unity by construction and it will distribute proper weights both to the agency with perfect forecast and to the other agencies who produce near perfect forecasts. The forecast-depth weight is a more sensible choice in this case, because the single best performer now is not necessarily the best performer in the next periods.

It is also worthy to note that iMSE $[e'_{i,t}e_{i,t}/k]^{-1}$ can be compared with the forecast depth based on the Mahalanobis distance of $e_{i,t}$:

$$\frac{1}{1 + e'_{i,t}\widehat{\Sigma}_t^{-1}e_{i,t}}, \tag{14}$$

where $0 < \widehat{\Sigma}_t < \infty$ is the $k \times k$ sample variance matrix of $e_{i,t}$. If we ignore the variance and simply let $\widehat{\Sigma}_t$ be the identify matrix multiplied by k, then it becomes $(1 + [e'_{i,t}e_{i,t}/k])^{-1}$. It uses the k-dimensional vector of the forecast error $e_{i,t}$ without any discount vector m, but it counts the forecast performance of a specific time during the training period more heavily if the cross-sectional variance of $e_{i,t}$ (i.e., MSE at the specific time) is small.

As an illustration, we compare iMSE with the following three forecast depths with $k = 1$, using 150 simulated forecasts generated from the standard normal when the true value is zero: FD_{mse}, the sample forecast depth in (8) using RMSE $\widehat{s}_t$ in (9); FD_{mad}, the sample forecast depth in (8) using MAD $\widehat{s}_t$ in (10); MD, the sample forecast depth based on the Mahalanobis forecast distance in (14). The graph on the left in Figure 1 shows the weights $\pi_{i,t}$ based on iMSE given in (13). We can see that the weight based on the iMSE assigns a huge weight only on a particular agency, whose forecast error during the training period is near zero. In comparison,

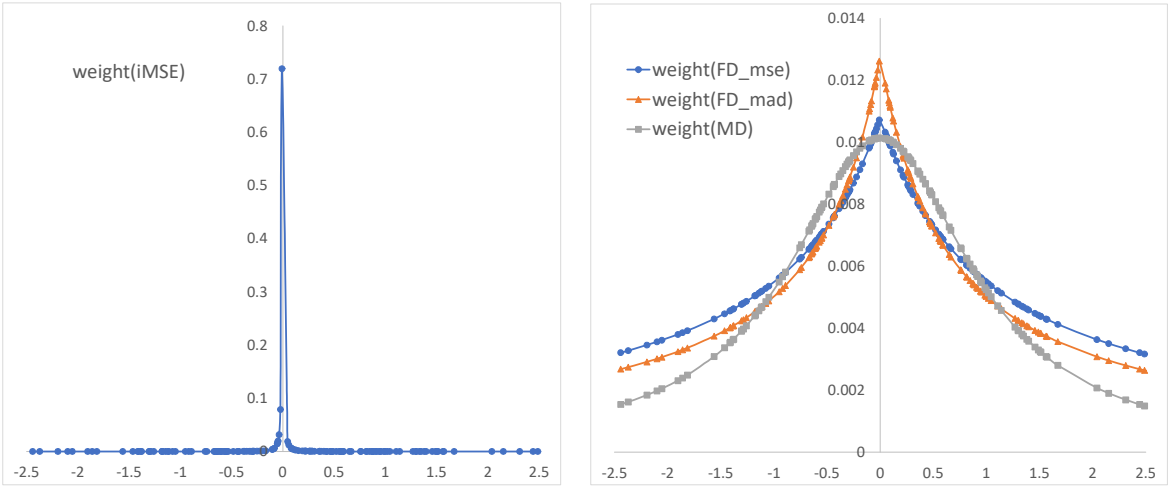


Figure 1: Comparison of weights

the graph on the right in Figure 1 shows the weights based on three forecast depths. Unlike iMSE, we can see that the forecast depths well distribute the weights. This difference can be understood from the fact that the forecast depth in (8) considers the historic absolute deviations of each forecasting agency, whereas the iMSE considers each historic squared errors. Compared with MD, the weights based on the forecast depths FD_{mse} and FD_{mad} are sharply concentrated at zero but levy less penalties on outliers. It hence advocates the usefulness of trimming on the forecast-depth based weights.

4 Limiting Distribution of Combined Forecast

We now derive limiting distribution of the depth-weighted forecast combination (11), from which we can better understand the factors that affect the robustness of the depth-weighted forecast combination. We can also conduct further inferences using the limiting distribution result, such as constructing the prediction intervals, which is presented at the end of this section.

At time t, for the h-step ahead forecast,[5] we define the bivariate forecast error vector as

$$x_{i,t:h} = \begin{pmatrix} m' e_{i,t} \\ y_{i,t+h} - y^0_{t+h} \end{pmatrix} \in \mathbb{R}^2, \tag{15}$$

which is assumed to be a random sample from an underlying joint distribution F_t for all i. We denote the marginal distributions as F_{1t} and F_{2t}.[6] The depth estimator $\widehat{\mathcal{D}}_{i,t}$ is only based on

[5] Note that we consider "direct" forecast because we do not observe each forecasting models nor impose any (autoregressive) dynamic structures.

[6] If we further suppose that, for each i, the forecast error $y_{i,t} - y^0_t$ is stationary over t, we can drop the subscript

the forecast 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 $\widehat{\mathcal{D}}_{i,t} = \mathcal{D}(m'e_{i,t}, \widehat{F}_{1t})$ and $\mathcal{D}_{i,t} = \mathcal{D}(m'e_{i,t}, F_{1t})$, where $\widehat{F}_t$ and $(\widehat{F}_{1t}, \widehat{F}_{2t})$ respectively denote the joint and marginal empirical distributions of $x_{i,t:h}$.[7] Similarly, we denote $\widehat{\mathcal{O}}_{i,t} = \mathcal{O}(m'e_{i,t}, \widehat{F}_{1t})$, $\mathcal{O}_{i,t} = \mathcal{O}(m'e_{i,t}, F_{1t})$, $s_t = s(F_{1t})$, and $\widehat{s}_t = s(\widehat{F}_{1t})$.

Using these notations, we can rewrite the sample forecast error of the depth-weighted forecast combination $\widehat{y}_{t+h}$ in (11) as

$$\begin{aligned}
\theta^h(\widehat{F}_t) &\equiv \widehat{y}_{t+h} - y_{t+h}^0 \\
&= \frac{(1/n)\sum_{i=1}^n \left(y_{i,t+h} - y_{t+h}^0\right) 1\{\mathcal{D}(m'e_{i,t}, \widehat{F}_{1t}) \geq \tau\} W(\mathcal{D}(m'e_{i,t}, \widehat{F}_{1t}))}{(1/n)\sum_{i=1}^n 1\{\mathcal{D}(m'e_{i,t}, \widehat{F}_{1t}) \geq \tau\} W(\mathcal{D}(m'e_{i,t}, \widehat{F}_{1t}))} \\
&= \frac{\int u_2 1\{\mathcal{D}(u_1, \widehat{F}_{1t}) \geq \tau\} W(\mathcal{D}(u_1, \widehat{F}_{1t})) d\widehat{F}_t(u)}{\int 1\{\mathcal{D}(u_1, \widehat{F}_{1t}) \geq \tau\} W(\mathcal{D}(u_1, \widehat{F}_{1t})) d\widehat{F}_{1t}(u_1)}
\end{aligned} \tag{16}$$

for given t, where $u = (u_1, u_2)' \in \mathbb{R}^2$. As the number of forecasts n increases, $\theta^h(\widehat{F}_t)$ will converge to a depth-weighted mean forecast error given by

$$\theta^h(F_t) = \frac{\int u_2 1\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_t(u)}{\int 1\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1)} \tag{17}$$

provided $\sup_{m \in \mathbb{R}^k} |\widehat{s}_t - s_t| = o_p(1)$ and $\sup_{u \in \mathbb{R}^2} |\widehat{F}_t(u) - F_t(u)| = o_p(1)$, which holds in general from the standard results. We can rewrite the numerator of $\theta^h(F_t)$ in (17) as

$$\int \left\{ \int u_2 F_{2|1,t}(du_2) \right\} 1\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1), \tag{18}$$

where $\int u_2 dF_{2|1,t}(u_2) = \mathbb{E}\left[y_{i,t+h} - y_{t+h}^0 | m'e_{i,t}\right]$. This expression implies that the depth-weighted mean forecast error $\theta^h(F_t)$ in (17) is a weighted average of the projection of $y_{i,t+h} - y_{t+h}^0$ on a linear combination of the past forecast errors $e_{i,t}$ during the training period, where the weights are given by $1\{\mathcal{D}(\cdot, F_{1t}) \geq \tau\} W(\mathcal{D}(\cdot, F_{1t}))$.

To obtain the asymptotic representation of the depth-weighted forecast combination, we define the influence function of $\theta^h(F_t)$ in (17). We let δ_x be the point-mass distribution at $x \in \mathbb{R}^2$ and $F_t(\varepsilon, \delta_x) = (1 - \varepsilon)F_t + \varepsilon\delta_x$ be a version of F_t 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

t in the distribution notations. However, it is not required to derived the main results. It is important to note that, however, imposing stationarity of the forecast error does not exclude the potential nonstationarity of the observed series y_t^0 itself and the forecast series $y_{i,t}$ for each i.

[7] By the affine invariance property of the forecast depth, we have $\mathcal{D}(m'e_{i,t}, F_{1t}) = \mathcal{D}(e_{i,t}, \underline{F}_{1t})$, where $\underline{F}_{1t}$ is the joint distribution of $e_{i,t} \in \mathbb{R}^k$.

$\theta^h(F_t)$ is defined as

$$\phi(x;\theta^h(F_t)) = \lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\left\{\theta^h(F_t(\varepsilon,\delta_x)) - \theta^h(F_t)\right\}$$

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

$$\sqrt{n}\left(\theta^h(\widehat{F}_t) - \theta^h(F_t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(x_{i,t:h};\theta^h(F_t)) + o_p(1) \tag{19}$$

for given t. To this end, we first assume the following conditions, similarly as Wu and Zuo (2009) and Lee and Sul (2022a). For $\tau\in(0,1)$, we 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 $\dot{W}(\cdot)$. For each t, it holds that (ii) $\int 1\{\mathcal{D}(u_1,F_{1t})\geq\tau\}W(\mathcal{D}(u_1,F_{1t}))dF_{1t}(u_1) > 0$ and $\int|u_2|1\{\mathcal{D}(u_1,F_{1t})\geq\tau\}W(\mathcal{D}(u_1,F_{1t}))dF_t(u) < \infty$; (iii) $s(\widehat{F}_{1t}), s(F_{1t})\in(0,\infty)$ with satisfying $\sup_{m\in\mathbb{R}^k}|s(\widehat{F}_{1t}) - s(F_{1t})| = o_p(1)$; (iv) $\sup_{u\in\mathbb{R}^2}|\widehat{F}_t(u) - F_t(u)| = o_p(1)$; (v) the joint density function $f_t(u_1,u_2)$ of $x_{i,t:h}$ exists and satisfies $\int(u_2-\theta(F_t))f_t(u_1,u_2)du_2 < \infty$ at $u_1 = U(F_{1t})$, $L(F_{1t})$.*

The following theorem summarizes the asymptotic properties of $\theta^h(\widehat{F}_t)$ in (16). The proof is in the Appendix.

Theorem 1 *Suppose Assumption 1 holds. Then, for given m, k, h, and t, $\theta^h(\widehat{F}_t)\to_p\theta^h(F_t)$ as $n\to\infty$. Furthermore, (19) holds, where*

$$\phi(x_{i,t:h};\theta^h(F_t)) = \frac{\phi_1^h(F_t)+\phi_2^h(F_t)+\phi_3^h(F_t)}{\int 1\{L(F_{1t})\leq u_1\leq U(F_{1t})\}W(\mathcal{D}(u_1,F_{1t}))dF_{1t}(u_1)}$$

with

$$\begin{aligned}
\phi_1^h(F_t) &= (y_{i,t+h} - y_{t+h}^0 - \theta^h(F_t))1\left\{L(F_{1t})\leq m'e_{i,t}\leq U(F_{1t})\right\}W(\mathcal{D}(m'e_{i,t},F_{1t})),\\
\phi_2^h(F_t) &= \int(u_2-\theta^h(F_t))1\{L(F_{1t})\leq u_1\leq U(F_{1t})\}\dot{W}(\mathcal{D}(u_1,F_{1t}))\phi_D(m'e_{i,t};\mathcal{D}(u_1,F_{1t}))dF_t(u),\\
\phi_3^h(F_t) &= \frac{1-\tau}{\tau}W(\tau)\phi_s(m'e_{i,t};s(F_{1t}))\int(u_2-\theta^h(F_t))\{f_t(U(F_{1t}),u_2) - f_t(L(F_{1t}),u_2)\}du_2,
\end{aligned}$$

$$\phi_D(\cdot;\mathcal{D}(u_1,F_{1t})) = \frac{\mathcal{O}(u_1,F_{1t})\phi_s(\cdot;s(F_{1t}))}{s(F_{1t})(1+\mathcal{O}(u_1,F_{1t}))^2},$$

and $\phi_s(\cdot; s(F_{1t}))$ *is the influence function of* $s(F_{1t})$*. Consequently,*

$$\sqrt{n}(\theta^h(\widehat{F}_t) - \theta^h(F_t)) \to_d \mathcal{N}(0, \sigma^2_{t:h})$$

as $n \to \infty$*, where* $\sigma^2_{t:h} = \mathbb{E}\left[\phi(x_{i,t:h}; \theta^h(F_t))^2\right]$*.*

One important finding is that the specific form of the asymptotic variance $\sigma^2_{t:h}$ depends on the the influence function of $s_t = s(F_{1t})$. Therefore, the choice of s_t heavily affect the robustness property of the depth-weighted forecast combination $\widehat{y}_{t+h}$. For instance, for the RMSE s_t in (3), we can derive the influence function of s_t as

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

when $0 < \mathbb{E}[(m'e_{i,t})^2|\mathcal{I}_t] < \infty$ using the influence function of the mean. For the MAD s_t in (4), 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

$$\phi_s(x_1; s(F_{1t})) = \frac{\text{sgn}(x_1 - \lambda(m'e_{i,t}))}{2f_{1t}(\lambda(m'e_{i,t}))},$$

where $\lambda(m'e_{i,t}) = \inf\{v : \mathbb{P}(|m'e_{i,t}| \leq v|\mathcal{I}_t) \geq 1/2\}$ is the conditional median of $m'e_{i,t}$, provided that the marginal density function $f_{1t}(\cdot)$ of $m'e_{i,t}$ at t satisfies $0 < f_{1t}(\lambda(m'e_{i,t})) < \infty$. Recall that the mean has an unbounded influence function when the support of F_{1t} is not bounded, whereas the median has a bounded influence function. It yields that the influence function of the RMSE is not necessarily bounded, whereas that of the MAD is bounded. For this reason, the forecast combination using the MAD is more robust towards outliers or extremely underperformed forecasts.

From Theorem 1, we can conclude that the depth-weighted forecast combination $\widehat{y}_{t+h}$ satisfies $\widehat{y}_{t+h} \to_p y^0_{t+h} + \theta^h(F_t)$ and $\sqrt{n}(\widehat{y}_{t+h} - \{y^0_{t+h} + \theta^h(F_t)\}) \to_d \mathcal{N}(0, \sigma^2_{t:h})$ as $n \to \infty$. Apparently, when $\theta^h(F_t) = 0$, $\widehat{y}_{t+h}$ becomes a consistent forecast. A sufficient condition for $\theta^h(F_t) = 0$ is that the forecasts $\{y_{i,t+h}\}$ are distributed symmetrically about the true value y^0_{t+h}, which could be obtained once extreme forecasts are trimmed out. In this case, the $100(1-\alpha)\%$ prediction interval of y^0_{t+h} can be obtained as

$$\left[\widehat{y}_{t+h} \pm z_{\alpha/2}\frac{\widehat{\sigma}_{t:h}}{\sqrt{n}}\right], \tag{20}$$

where $z_{\alpha/2}$ is the $(1-(\alpha/2))$th quantile of the standard normal distribution and $\widehat{\sigma}_{t:h}$ is a consistent estimator of $\sigma^2_{t:h}$. Knowing that the weighted average of the individual interval forecasts does

not necessarily provide a correct coverage rate (e.g., Timmermann, 2006), the prediction interval in (20) can be alternatively used, which does not rely on subsampling or bootstrap. $\sigma^2_{t:h}$ can be estimated as $\widehat{\sigma}^2_{t:h} = n^{-1}\sum_{i=1}^n \widehat{\phi}(x_{i,t:h};0)^2$, where we replace y^0_{t+h} by $\widehat{y}_{t+h}$ and use kernel density estimators and sample analogues of the tems in $\widehat{\phi}(x_{i,t:h};0)$.[8]

5 Forecasting New COVID-19 Cases

We apply the depth-weighted forecast combination (11) to predict weekly COVID-19 cases in the United States. The data set 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 August 7, 2021, which is updated on August 11, 2021. The data set consists of h-step ahead weekly forecast history from 40 individual modeling groups for $h = 1, 2, 3, 4$.[9] It also includes the h-step ahead ensemble forecast that is reported in the weekly forecast digest by the CDC.[10] We use the past 50 weekly forecasts from the week ending on August 29, 2020 (when all the h-step ahead forecasts became available) to the week ending on August 7, 2021.

We compare different forecast combination approaches, including the equally-weighted average of all the available forecasts at each t, ensemble forecast reported by the CDC ("Ensem"); inversed MSE based forecast combination ("iMSE") as in (13); and the forecast-depth based combination (FD) developed in this paper. For the forecast-depth based combination, for a given training period size $k = 2, 3, 4$, we consider two types of the discount vector $m = (m_1, \ldots, m_k)'$:

- Type 1 (FD1): $m_j = (0.2)^{k-j}/\sum_{\ell=1}^k (0.2)^{k-\ell}$ for $j = 1, \ldots, k$;
- Type 2 (FD2): $m_j = (j/k)^4/\sum_{\ell=1}^k (\ell/k)^4$ for $j = 1, \ldots, k$.

Both discount vectors have similar magnitudes. But Type 1 discounts the past information more heavily with large k; Type 2 discounts the past information more heavily with small k. For each type, two forms of $\widehat{s}_t$ are considered:

- Sample RMSE in (9): "FD1$_{mse}$" and "FD2$_{mse}$";
- Sample MAE in (10): "FD1$_{mad}$" and "FD2$_{mad}$".

[8] Because $\widehat{y}_{t+h}$ is a cross-sectional weighted average, one could instead consider $\widehat{\sigma}^2_{t:h} = n\sum_{i=1}^n \pi^2_{i,t}\left(y_{i,t+h} - \widehat{y}_{t+h}\right)^2$ in this case by ignoring the randomness in $\pi_{i,t}$, when serial dependence between $y_{i,t+h}$ and $Y_{i,t} = (y_{i,t-k+1}, \ldots, y_{i,t-1}, y_{i,t})'$ is not too strong.

[9] The list of the forecasting agencies can be found at https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html.

[10] From April 13 to July 21, 2020, the ensemble was created by the arithmetic average of each prediction quantile for all eligible models for a given location. The confidence interval at each prediction point is also calculated from the corresponding quantile ensembles. (Busetti, 2017). However, starting on the week of July 28, 2020, which is the sample period of our analysis here, the median prediction was instead used across all eligible models at each quantile level. As of August 7, 2021, 23 models are included for the ensemble forecast. For further details about the ensemble forecast, see https://covid19forecasthub.org/doc/ensemble/ for further details.

For $W(\cdot)$, we simply let $W(d) = d$. For the trimming parameter τ, we consider the thresholds at the 0% (no trimming), 10%, 20%, 30%, 40%, and 50% levels of the lowest estimated forecast depth in the sample. For instance, at the 20% threshold, we set τ such that $(1/n)\sum_{i=1}^{n} 1\{\widehat{\mathcal{D}}_{i,t} < \tau\} = 0.2$. We also report the trimmed iMSE forecast combinations, where the trimming thresholds are determined similarly as the forecast depth: 0% (no trimming), 10%, 20%, 30%, 40%, and 50% levels of the lowest estimated iMSE in the sample. For each horizons h, we obtain the weights $\pi_{i,t}$ based on the performance of the h-step ahead forecasts during the training period.

Tables 1 to 4 report the out-of-sample forecast mean squared error (FMSE) comparisons among different forecast combinations for h-step ahead forecasting for $h = 1, \ldots, 4$, respectively. For each k, the values in the tables are the ratio of the FMSE of each combined forecast to that of the equally-weighted forecast combination over the past $(50 - k)$ weeks of forecasting, where the weights for the forecast combinations of the iMSE and the forecast-depth at t are obtained using the training period of $(t - 1, \ldots, t - k)$. So, values less than 1 implies that the FMSE is smaller than that of the equally-weighted forecast combination; smaller values implies better performance.[11] Values marked with "*" indicates the best performer in each case. In each table, k is the training period size and "trim" is the trimming proportion as we described above. The relative performance of different agencies can change over time, and we compute the weights over short rolling windows. For a given target forecasting week t, any agency i is dropped if it does not have at least k most recent forecasts to form the training period. We hence do not need a balanced panel over the entire period; we only use a short balanced panel of length $k + 1$ at each target forecasting week t, whose individual members can be different over t. In each table of the h-step ahead forecast, $\overline{n}$ reports the average of the cross-sectional sample size n_t over the past $(50 - k)$ weeks (i.e., $\overline{n} = (50 - k)^{-1}\sum_{t=k+1}^{50} n_t$), where n_t is the number of individuals who report the h-step ahead forecast at t with k training periods and hence used for forecast combination.

The main point of interest is how the forecast-depth based method performs even with a very short training period. It is evident that the forecast combinations based on the forecast-depth and the iMSE stand out; the forecast-depth based combination shows the best performance in general. It does not show the typical "forecast combination puzzle" – the equal weight outperforms the estimated weights in forecast combination. For some cases, the iMSE combination outperforms the forecast-depth based method, but it is mostly for the cases when we do not employ (enough) trimming. The trimming generally improves the FMSE though the change is

[11] Note that we drop one agency in the forecast pool, who reported very erroneous forecasts between 3/13/2021 and 4/17/2021. If we compare the FMSE of the equally-weighted forecast combinations with ("EQ_{-1}") and without ("EQ_{all}") dropping this agency, we have

h	1	2	3	4
EQ_{all}/EQ_{-1}	173.19	194.49	229.15	245.05

which shows that the equally-weighted forecast combination is prone to swing much by only one outlier.

Table 1: 1-step ahead FMSE ratio to the equal-weight forecast combination

k	trim	$\overline{n}$	Ensem	iMSE	FD1$_{mse}$	FD1$_{mad}$	FD2$_{mse}$	FD2$_{mad}$
2	0.0	24.3	0.931	0.863	0.941	0.938	0.941	0.938
	0.1			0.866	0.871	0.873	0.873	0.873
	0.2			0.857	0.881	0.882	0.855	0.857
	0.3			0.880	0.902	0.901	0.884	0.885
	0.4			0.886	0.892	0.888	0.877	0.879
	0.5			0.899	0.855	*0.854	0.861	0.864
3	0.0	23.9	0.931	0.895	0.945	0.942	0.950	0.946
	0.1			0.900	0.872	0.873	0.895	0.894
	0.2			0.887	0.908	0.906	*0.870	0.872
	0.3			0.912	0.897	0.895	0.894	0.894
	0.4			0.926	0.882	0.877	0.887	0.888
	0.5			0.945	0.889	0.884	0.895	0.896
4	0.0	23.5	0.931	0.872	0.945	0.942	0.952	0.949
	0.1			*0.863	0.875	0.876	0.900	0.900
	0.2			0.869	0.913	0.911	0.907	0.908
	0.3			0.894	0.896	0.895	0.902	0.902
	0.4			0.913	0.889	0.885	0.903	0.903
	0.5			0.953	0.897	0.893	0.915	0.916

Note: The table presents the ratio of the forecast mean square error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 - 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; FD1$_{mse}$ is the forecast-depth combination with type-1 m and RMSE $\widehat{s}_t$; FD1$_{mad}$ is the forecast-depth combination with type-1 m and MAD $\widehat{s}_t$; FD2$_{mse}$ is the forecast-depth combination with type-2 m and RMSE $\widehat{s}_t$; FD2$_{mad}$ is the forecast-depth combination with type-2 m and MAE $\widehat{s}_t$. The best performers are marked with * in each case.

not strictly monotonic to the trimming proportion. However, the benefit from trimming is much larger in FD than iMSE, and all the FD approaches eventually perform better than iMSE with proper trimming for all the cases. Since we compare the FMSE, we expected that the iMSE should perform better than FD as k increases, but the current results do not support it. This seems because the weekly COVID-19 cases fluctuates much and the forecasting agencies can hardly provide good predictions over an extended period consistently. For the choice of $\widehat{s}_t$ in FD, it does not make big difference, but MAE outperforms RMSE in general.

Table 5 compares the forecast combinations based on the forecast-depth and the iMSE in more details. In particular, it summarizes the bias ratios and the variance ratios between FD2$_{mad}$ and iMSE for 2-step ahead forecasts in Table 2 above. It shows that the FMSE improvement of FD2$_{mad}$ mostly comes from the bias reduction.

Table 2: 2-step ahead FMSE ratio to the equal-weight forecast combination

k	trim	$\overline{n}$	Ensem	iMSE	FD1$_{mse}$	FD1$_{mad}$	FD2$_{mse}$	FD2$_{mad}$
2	0.0	22.8	0.943	1.120	0.962	0.963	0.940	0.933
	0.1			1.110	0.916	0.916	0.888	0.881
	0.2			1.100	0.891	0.893	0.875	0.865
	0.3			1.100	0.874	0.877	0.831	0.826
	0.4			1.090	0.866	0.869	0.834	0.829
	0.5			1.080	0.824	0.828	0.761	*0.759
3	0.0	22.4	0.943	0.949	0.971	0.973	0.950	0.943
	0.1			0.934	0.923	0.925	0.896	0.889
	0.2			0.944	0.897	0.900	0.884	0.875
	0.3			0.953	0.882	0.887	0.845	0.841
	0.4			0.971	0.867	0.872	0.848	0.844
	0.5			0.976	0.861	0.866	0.787	*0.785
4	0.0	22.0	0.943	0.934	0.978	0.981	0.959	0.954
	0.1			0.911	0.934	0.935	0.910	0.904
	0.2			0.905	0.908	0.911	0.903	0.896
	0.3			0.917	0.893	0.898	0.856	0.852
	0.4			0.959	0.882	0.887	0.855	0.852
	0.5			0.952	0.862	0.868	0.826	*0.823

Note: The table presents the ratio of the forecast mean square error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 - 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; FD1$_{mse}$ is the forecast-depth combination with type-1 m and RMSE $\widehat{s}_t$; FD1$_{mad}$ is the forecast-depth combination with type-1 m and MAD $\widehat{s}_t$; FD2$_{mse}$ is the forecast-depth combination with type-2 m and RMSE $\widehat{s}_t$; FD2$_{mad}$ is the forecast-depth combination with type-2 m and MAE $\widehat{s}_t$. The best performers are marked with * in each case.

Table 3: 3-step ahead FMSE ratio to the equal-weight forecast combination

k	trim	$\overline{n}$	Ensem	iMSE	FD1$_{mse}$	FD1$_{mad}$	FD2$_{mse}$	FD2$_{mad}$
2	0.0	21.0	0.989	0.832	0.924	0.915	0.920	0.911
	0.1			0.819	0.853	0.848	0.825	0.823
	0.2			0.792	0.791	0.789	0.796	0.794
	0.3			0.789	0.794	0.791	0.773	0.771
	0.4			0.769	0.764	0.761	0.731	0.730
	0.5			0.789	0.744	0.739	0.710	*0.709
3	0.0	20.6	0.989	0.869	0.932	0.923	0.925	0.916
	0.1			0.855	0.863	0.858	0.827	0.825
	0.2			0.835	0.788	0.787	0.801	0.800
	0.3			0.831	0.802	0.799	0.784	0.782
	0.4			0.826	0.782	0.777	0.734	0.734
	0.5			0.805	0.748	0.744	*0.695	0.696
4	0.0	20.1	0.989	0.853	0.940	0.931	0.934	0.924
	0.1			0.832	0.867	0.862	0.832	0.830
	0.2			0.815	0.790	0.789	0.809	0.807
	0.3			0.799	0.801	0.798	0.785	0.783
	0.4			0.795	0.791	0.786	0.744	0.744
	0.5			0.787	0.747	0.744	0.723	*0.722

Note: The table presents the ratio of the forecast mean square error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 - 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; FD1$_{mse}$ is the forecast-depth combination with type-1 m and RMSE $\widehat{s}_t$; FD1$_{mad}$ is the forecast-depth combination with type-1 m and MAD $\widehat{s}_t$; FD2$_{mse}$ is the forecast-depth combination with type-2 m and RMSE $\widehat{s}_t$; FD2$_{mad}$ is the forecast-depth combination with type-2 m and MAE $\widehat{s}_t$. The best performers are marked with * in each case.

Table 4: 4-step ahead FMSE ratio to the equal-weight forecast combination

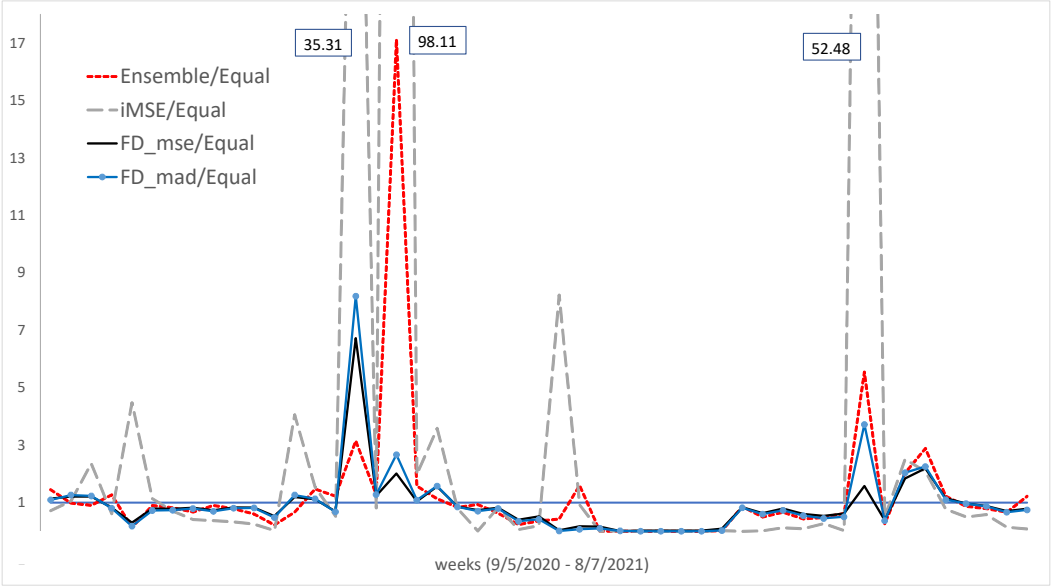
k	trim	$\overline{m}$	Ensem	iMSE	FD1$_{mse}$	FD1$_{mad}$	FD2$_{mse}$	FD2$_{mad}$
2	0.0	17.6	0.949	0.923	0.948	0.937	0.943	0.930
	0.1			0.900	0.876	0.871	0.876	0.870
	0.2			0.894	0.822	0.820	0.819	0.817
	0.3			0.893	0.793	0.789	0.780	0.776
	0.4			0.873	0.765	0.765	0.744	*0.741
	0.5			0.888	0.784	0.783	0.772	0.770
3	0.0	17.0	0.949	0.911	0.959	0.947	0.951	0.937
	0.1			0.897	0.880	0.875	0.879	0.873
	0.2			0.873	0.843	0.839	0.832	0.828
	0.3			0.875	0.796	0.792	0.772	*0.768
	0.4			0.845	0.778	0.778	0.775	0.771
	0.5			0.886	0.779	0.778	0.792	0.789
4	0.0	16.4	0.949	0.928	0.966	0.954	0.959	0.943
	0.1			0.905	0.883	0.878	0.881	0.875
	0.2			0.912	0.843	0.839	0.827	0.823
	0.3			0.908	0.792	0.789	0.761	0.758
	0.4			0.853	0.776	0.776	0.787	0.781
	0.5			0.858	0.732	0.733	*0.727	*0.727

Note: The table presents the ratio of the forecast mean square error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 - 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; FD1$_{mse}$ is the forecast-depth combination with type-1 m and RMSE $\widehat{s}_t$; FD1$_{mad}$ is the forecast-depth combination with type-1 m and MAD $\widehat{s}_t$; FD2$_{mse}$ is the forecast-depth combination with type-2 m and RMSE $\widehat{s}_t$; FD2$_{mad}$ is the forecast-depth combination with type-2 m and MAE $\widehat{s}_t$. The best performers are marked with * in each case.

Table 5: Comparison between FD and iMSE

k	trim	FD2$_{mad}$		iMSE		Ratio (FD/iMSE)		
		bias	stdv	bias	stdv	bias	stdv	mse
2	0.0	-2497	204259	7520	223608	-0.332	0.913	0.834
	0.1	-1230	198483	8547	222433	-0.144	0.892	0.795
	0.2	-310	196746	8123	221738	-0.038	0.887	0.786
	0.3	714	192212	9046	221677	0.079	0.867	0.751
	0.4	4313	192573	10507	220445	0.410	0.874	0.762
	0.5	3492	184243	11998	219330	0.291	0.840	0.704
3	0.0	-2759	207473	3970	208199	-0.695	0.997	0.993
	0.1	-1538	201465	4979	206508	-0.309	0.976	0.951
	0.2	-902	199823	5240	207571	-0.172	0.963	0.926
	0.3	391	195373	6891	208545	0.057	0.937	0.877
	0.4	3754	196203	8156	210470	0.460	0.932	0.868
	0.5	3162	189271	11974	210770	0.264	0.898	0.804
4	0.0	-2675	210510	6363	208627	-0.420	1.009	1.020
	0.1	-1658	204900	6790	206073	-0.244	0.994	0.988
	0.2	-1165	203155	6180	205332	-0.189	0.989	0.978
	0.3	-12	198436	8219	206671	-0.001	0.960	0.920
	0.4	2985	198852	11271	211167	0.265	0.942	0.884
	0.5	4292	191383	16111	210139	0.266	0.911	0.825

Note: "bias" is the average forecast bias; "stdv" is the average standard deviation of the forecast error; and "mse" is the average forecast mean squared error. Values in the table are based on the 2-step ahead forecasts using the same data set as in Table 2.



(Note: The dates on three peaks are 12/19/2020 (35.31), 1/2/2021 (98.11), and 6/12/2021 (52.48).)

Figure 2: Forecast error ratio to the equal-weight forecast combination

In Figure 2, instead of averaging over the entire period as in the Tables 1 to 4, we depict the weekly 1-step ahead forecast error ratio paths of Ensem, iMSE, FD1_{mse}, and FD1_{mae} for $k = 1$ and with 30% trim. Since we consider the case with $k = 1$ period of training, $\text{FD1}_{mse} = \text{FD2}_{mse}$, and $\text{FD1}_{mae} = \text{FD2}_{mae}$. We can see that iMSE is quite volatile and some error ratios (to the error of equally-weighted forecast combination) are even off the boundary. Ensem is still volatile though at a much smaller scale than iMSE. In comparison, the forecast error ratio paths of FD1_{mse} and FD1_{mae} show very little fluctuations and they mostly lie below 1, which implies that their forecast errors are smaller than that of the equally-weighted forecast combination.

Figure 3 depicts the weights based on FD1_{mse}, FD1_{mad}, and iMSE on a random date, October 10, 2020, with $k = 1$ and the 1-step ahead forecasts. Unlike the forecast-depth based forecast combinations, it shows that the weight based on iMSE assigns 99% of the total weight only on three agencies, which are 0.072, 0.171, 0.745, where the weight 0.745 is out of the chart.

Finally, based on the available forecasts up to the week ending September 4, 2021, we report predictions for the next 4 weeks (ending on 8/14/2021, 8/21/2021, 8/28/2021, 9/4/2021) as of August 7, 2021. For each h-week ahead prediction ($h = 1, 2, 3, 4$), we consider 25 different cases of $k = 1, \ldots, 5$ and $trim = 0.1, \ldots, 0.5$ for each of the four forecast combinations, FD1_{mse}, FD1_{mad}, FD2_{mse}, and FD2_{mad}. (Hence, 100 different forecast combinations for each h.) Figure 4 reports the predictions of the average of these depth-weighted forecast combinations in black circle line (FD) and the ensemble forecast reported by the CDC in red triangle line (Ensemble). The predictions between these two methods are quite similar. However, the noticeable difference is in

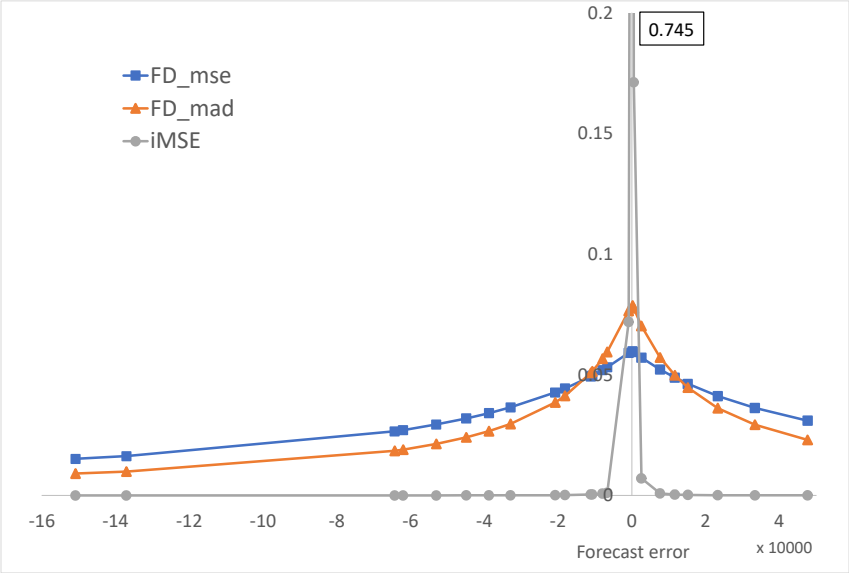


Figure 3: Weights of forecastg combinations

their prediction intervals. The pointwise 95% prediction interval of the FD forecast combination point is depicted by the shaded area, which is obtained as the maximum of the upper bound and the minimum of the lower bound points of the 95% prediction intervals of all the 100 depth-weighted forecast combinations using the normal approximation based on (20). Compared with the prediction interval of the ensemble forecast reported by CDC (in the dashed vertical lines), it shows much narrower bounds.

6 Concluding Remarks

In this paper, we develop the forecast depth and a depth-weighted forecast combination with trimming. Since the weights are not obtained by minimizing a 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 can be practically very useful as complementing other forecast combinations. In comparison, when long training period is available, we can apply LASSO in estimating time-invariant weights by minimizing a L_2 loss function with L_1 penalty terms (e.g., Diebold and Shin, 2019), from which we can obtain weights on each forecasting individual with selection (i.e., trimming).

It should be noted that our approach ignores any estimation error in the forecast values, which is because we do not know the underlying model that yields each forecast value. However, we deal with random weights with endogenous trimming when deriving the limiting distribution of the forecast combination, which is the main analytical contribution of this paper. In this

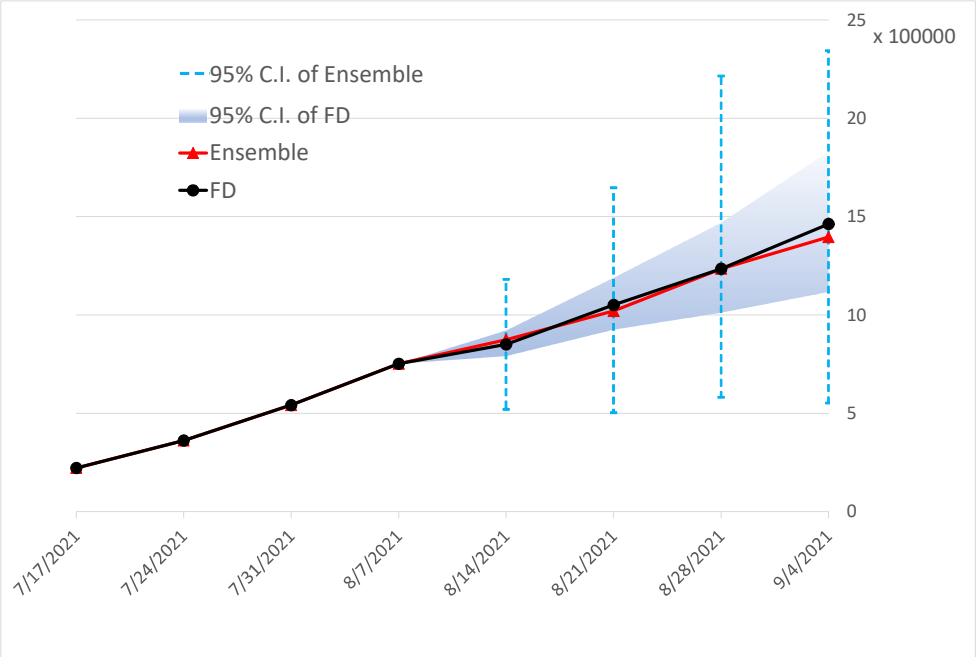


Figure 4: COVID-19 New Cases Forecast (US National)

regards, our approach should be distinguished from the model averaging approaches.

We can extend the forecast combination idea to 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.

Appendix: Proof of Theorem 1

To simplify the notations, we drop the subscript "t" in the distribution notations in the proof. So, we simply denote the joint distribution and density of $x_{i,t:h}$ in (15) as F and f respectively; the marginal distributions as F_1 and F_2. The corresponding empirical distributions are denoted as $\widehat{F}$, $\widehat{F}_1$, and $\widehat{F}_2$, respectively. Over the proof, $u = (u_1, u_2)' \in \mathbb{R}^2$ are used as generic variables of functions and in the integrals. We also define

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

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

$$\begin{aligned}\sup_{u_1\in\mathbb{R}}\left|\mathcal{D}(u_1,\widehat{F}_1)-\mathcal{D}(u_1,F_1)\right| &= \sup_{m\in\mathbb{R}^k}\left|\frac{1}{1+\left|u_1\right|/s(\widehat{F}_1)}-\frac{1}{1+\left|u_1\right|/s(F_1)}\right| \\ &\leq \sup_{m\in\mathbb{R}^k}\left\{\frac{\left|s(\widehat{F}_1)-s(F_1)\right|}{s(\widehat{F}_1)}\frac{\mathcal{O}(u_1,F_1)}{1+\mathcal{O}(u_1,F_1)}\right\}=o_p(1)\end{aligned}$$

as we assume $\sup_{m\in\mathbb{R}^k}\left|s(\widehat{F}_1)-s(F_1)\right|=o_p(1)$ and $s(\cdot)\in(0,\infty)$. Note that $\mathcal{O}(u_1,F_1)=\left|u_1\right|/s(F_1)$ and $\mathcal{O}(u_1,F_1)/(1+\mathcal{O}(u_1,F_1))\in[0,1]$ by construction. The asymptotic normality follows similarly as the proof of Theorem 4.1 in Wu and Zuo (2009), so we sketch the proof here. Recall that we define $L(\cdot)=-((1-\tau)/\tau)s(\cdot)$ and $U(\cdot)=((1-\tau)/\tau)s(\cdot)$, where $\tau\in(0,1)$ is the trimming parameter introduced in (7). Since $L(\cdot)<U(\cdot)$ by construction, we write

$$\begin{aligned}\sqrt{n}\left(\theta^h(\widehat{F})-\theta^h(F)\right) &= \frac{\sqrt{n}\int u_2^*1\{\mathcal{D}(u_1,\widehat{F}_1)\geq\tau\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}(u)}{\int 1\{\mathcal{D}(u_1,\widehat{F}_1)\geq\tau\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)} \\ &= \frac{\sqrt{n}\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)}\left\{\int u_2^*d\widehat{F}_{2|1}(u_2|u_1)\right\}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)}{\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)}W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1)},\end{aligned}$$

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

$$\begin{aligned}N_{1n} &= \sqrt{n}\int_{L(F_1)}^{U(F_1)}\widehat{g}(u_1)W(\mathcal{D}(u_1,F_1))d\widehat{F}_1(u_1), \\ N_{2n} &= \int_{L(F_1)}^{U(F_1)}\widehat{g}(u_1)\sqrt{n}\left\{W(\mathcal{D}(u_1,\widehat{F}_1))-W(\mathcal{D}(u_1,F_1))\right\}d\widehat{F}_1(u_1), \\ N_{3n} &= \sqrt{n}\left\{\int_{L(\widehat{F}_1)}^{U(\widehat{F}_1)}-\int_{L(F_1)}^{U(F_1)}\right\}\widehat{g}(u_1)W(\mathcal{D}(u_1,\widehat{F}_1))d\widehat{F}_1(u_1),\end{aligned}$$

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

$$\eta_{1i}=(y_{i,t+h}-y_{t+h}^0-\theta^h(F))1\left\{L(F_1)\leq m'e_{i,t}\leq U(F_1)\right\}W(\mathcal{D}(m'e_{i,t},F_1)).$$

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 $\phi_D(x_1;\mathcal{D}(u_1,F_1))$ for $x_1\in\mathbb{R}$ such that $\widehat{H}(u_1)=\int\phi_D(x_1;\mathcal{D}(u_1,F_1))d\widehat{\nu}(x_1)+o_p(1)$ uniformly over $x_1\in[L(F_1),U(F_1)]$. Similarly as the proof of Theorem 2 in Lee and Sul (2022a), therefore, we can verify that

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

where

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

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

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

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

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

$$
\begin{aligned}
& -\sqrt{n}\iint_{L(F_1)}^{L(\widehat{F}_1)} u_2^* W(\mathcal{D}(u_1,F_1)) f(u_1,u_2) du_1 du_2 + o_p(1) \\
= & \int u_2^* \sqrt{n}\left\{U(\widehat{F}_1) - U(F_1)\right\} W(\mathcal{D}(U(F_1),F_1)) f(U(F_1),u_2) du_2 \\
& - \int u_2^* \sqrt{n}\left\{L(\widehat{F}_1) - L(F_1)\right\} W(\mathcal{D}(L(F_1),F_1)) f(L(F_1),u_2) du_2 + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{3i} + o_p(1),
\end{aligned}
$$

where

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

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 given as

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

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. *Q.E.D.*

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 (2022a). Depth-Weighted Means of Noisy Data: An Application to Estimating the Average Effect in Heterogeneous Panels, *working paper*, Syracuse University.

Lee, Y. and D. Sul (2022b). Trimmed mean group estimation, *Advances in Econometrics*, Vol. 43B, 177-202.

Liu, R.Y. (1992). Data depth and multivariate rank tests, in (ed.) Y. Dodge, *L1-Statistical Analysis and Related Methods*, North-Holland, 279–294.

Stock, J.H. and M. Watson (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.

Stock, J.H. and M. Watson (2006). Forecasting with Many Predictors, In *Handbook of Economic Forecasting, Vol. 1*, G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), Amsterdam: North-Holland, pp. 515–554.

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.