Forecast combination with truncation

Daniel Hsiao

November 12, 2018

Contents

1 Introduction

Everyone makes predictions. Drawing a line or a curve through a plot, guessesing base on intuition, or complex mathematical models are all predictions. However, people do not always make the same forecasts. As a matter of facts, most predictions are different. The more people there are, the more variety the predictions will be. The questions are, ”which is better?” and ”can we do better than any forecasts alone?”

Forecast combinations is a study raised from the second question. The idea is simple: if we don't know which forecast is the best, maybe there are a better solution which is a linear combination of all the forecasts. By combining the forecasts, the prediction contains less idiosyncratic risk, and becomes more the general opinion of all forecasters. This has proven to be successful in decreasing forecast error (Clemen (1989), Diebold and Lopez (1995), Chen et al. (2000) , Dunis et al. (2000), Watson and Stock (2004)).

The current literature starts with the seminal article of Bates and Granger (1969) where forecast combination had been introduced to the public. The literature blooms and twenty years later, Clemen (1989) gathered and reviewed the over 200 items in the literature. Despite the extensive research on forecast combinations, Clemen found that some issues are still unsolved. One of the issue is 'What is the explanation for the robustness of the simple average of forecasts?'. Simple average with equal weights often outperforms more complicated weighting schemes when one compares the combination of point forecasts based on mean squared prediction error. This is further supported with further studies, for example, (Watson and Stock, 2004). In 2004, Elliott and Timmermann made another review on recent theoretical contributions. Gibbs and Vasnev (2017) proposed a bias estimation method on the forecast and adjust the weight according to the bias.

This paper looks at the correlation between forecasters. The correlation is a strong factor in the weight determination. While many researchers choose to discard correlation due to uncertainty in the covariance estimation, the effect of the correlation can be in favour of the forecast. If both forecaster always over estimates, the high covariance detects this and results in a combined forecast that is not between the two forecasts. Therefore we want to see the effect of negative weights during estimation and to what extend limiting the negative weights can help in the forecast combination.

We propose to limit the negative weights instead of neglecting the correlation at once. We show with survey data from European Central Bank (ECB) that the limiting negative weight method can improve the prediction. We impose a threshold parameter, where every negative weights below the threshold are set to zero. Not only does it improve the prediction in terms of mean squared prediction error and mean absolute error, it is also able to outperform equal weight in some area.

The result shows that the test statistics[1] decreases in a smooth adjustments and the choice of threshold parameter is not very sensitive to minor changes. This relieves the error during threshold

[1] Lower is better.

selection, as choosing a slightly wrong threshold does not impact the performance largely. We show that even with in-sample threshold parameter selection, the threshold method still outperforms the equal weights.

We look further in bias correction and simulation study. Bias correction aims to decrease the weights of the forecasts with known bias, while in the simulation study we show the effect of noise-to-signal ratio and correlation on the shape of mean squared prediction error. Bias estimation does not improve in most cases, but improve greatly for one timeseries where threshold does not help. The simulation on the other hand shows that when the correlation increases, the end position of the MSPE decreases. The similar effect can be found in the noise-to-signal ratio. When the noise-to-signal ratio increases, the achievable MSPE with negative weights increases.

The paper is structured as following. Section two starts with methodology and the implications on the truncation. Section three explains the data structure and preliminary analysis on the data. Section four provides empirical results and a simulation study on the threshold method in relation with signal-to-noise ratio and correlation.

2 Methodology

2.1 Standard Approach with High Covariance

For the methodology section, we consider the two variable case for illustrative purpose. We have two forecasts, y_1 and y_2, of the true variable y. We want to combine y_1 and y_2 with a weight w such that we have $y_c = wy_1 + (1-w)y_2$. Assuming they follow some distribution with finite first and second moment, e.g. $y_1 \sim D(0, \sigma_1)$, $y_2 \sim D(0, \sigma_2)$, and $corr(y_1, y_2) = \rho$, then the variance of the combined forecast y_c is

$$Var(y_c) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\sigma_1\sigma_2\rho, \tag{1}$$

and the optimal weight with minimal variance is

$$w^* = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}. \tag{2}$$

Equation 2 is the standard benchmark approach in the combination theory, where extensive researches had been done on. We refer this weight as the optimal weight. Equation 2 has a few empirical results that are against this approach. Two common alternative solutions are diagonal covariance matrix and equal weight.

Ignoring the correlation term ρ by setting $\rho = 0$, we get the inverse relation on the variance

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{3}$$

This is a robust way to avoid the estimation of the covariance when the dimension goes up. The amount of parameters to estimate for the covariance with dimension n is $\frac{1}{2}n(n+1)$, which is quadratic in n. When the user only estimates the variances, the amount of parameters to estimate reduces to n, which greatly decreases the estimation error Stock and Watson (2001).

Applying more restriction by setting $\sigma_1 = \sigma_2$, we get equal weight. Equal weight is another common approach that works better empirically. Clemen (1989) The forecast combination is in this case just an arithmetic mean of all forecasts. The reason behind using equal weight is the fact that estimating weight increases or shifts the forecast errors due to additional estimation error in the estimation of w. This estimation error has a negative effect on the forecast.

Looking back to equation 2, we examine the effect of high correlation term. Assuming without loss of generality that $\sigma_1 = \sigma_2(1+\delta)$, where $\delta > 0$, we rewrite the weight as

$$w^* = \frac{(-\rho\delta + (1-\rho))}{2(1-\rho-(1-\rho)\delta) + \delta^2}. \tag{4}$$

The numerator in w consist of a weighted mean between $-\delta$ and 1 with weight ρ and $1-\rho$ respectively. When ρ is small, the weight is close to equation 3, which is $\frac{1}{1+(1+\delta)^2}$. When ρ is large, the negative in numerator, $-\delta$, takes over and results in negative weight. The weight becomes

$$w^* = \frac{-\delta}{\delta^2} \tag{5}$$

when ρ approaches 1. As δ approaches 0, the weight quickly goes toward $-\infty$. This is in contrary with the intuition of equal weights, where the variance are equal. Only when $\delta > 1$, that is $\sigma_1 > 2\sigma_2$, will the weight be above -1. The boundary case is

$$\rho = \frac{\sigma_2}{\sigma_1}, \tag{6}$$

which w decreases to 0 and $y_c = y_2$.

2.2 With estimation error of w

We can also consider the weight as non-deterministic, but related with y, e.g., in a trivariate distribution with finite third and fourth moments (Claeskens et al., 2014). Under trivariate distribution, the variance of the weight influences the expected value and the variance of the combined forecast.

The expected value and the variance of the combined forecast becomes

$$\begin{aligned}
E(y_c) =& \mu + (cov(w, y_1 - y_2))^2 \\
var(y_c) =& E(w)^2\sigma_1^2 + (1 - E(w))^2\sigma_2^2 + 2E(w)(1 - E(w))\rho\sigma_1\sigma_2 \\
&+ E[(w - E(w))(y_1 - y_2)(E(w)y_1 + (1 - E(w))y_2 - \mu)] \\
&+ E[(w - E(w))^2(y_1 - y_2)^2] - cov(w, y_1 - y_2)^2.
\end{aligned} \tag{7}$$

Equation 7 shows the general case of the forecast combination. If the covariance between w and $y_1 - y_2$ is not 0, the forecast is biased when combining, with bias $cov(w, y_1 - y_2)^2$. The variance also increases from equation 1 with $E[(w - E(w))(y_1 - y_2)(E(w)y_1 + (1 - E(w))y_2 - \mu)] + E[(w - E(w))^2(y_1 - y_2)^2] - cov(w, y_1 - y_2)^2$. This case the only requirement is the individual forecast has to be unbiased and that the weight sums up to 1.

Let $d = (d_1, d_2)'$ be the third moment between y_1, y_2, and w, and let $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ be the (co)variance matrix of (y_1, y_2), we have

$$w^\dagger = w^*(1 + \frac{\sigma_{22}d_1 + \sigma_{11}d_2 - \sigma_{12}(d_1 + d_2)}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}) - \frac{\sigma_{22}d_1 - \sigma_{12}d_2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}. \tag{8}$$

The non-deterministic weight selection is a linear combination of the original weight. The non-deterministic weight does not change if the third moments are 0. In real life case, it is not hard to imagine that the weights are correlated with the data used to calculate the weights. However, since the weights are unobserved, there are no straightforward solution for the third moment.

From equation 8, we look in the scaling parameter and the intercept adjustment vector. Under same condition where $\sigma_1 = \sigma_2(1 + \delta)$, we rewrite the determinant to

$$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{22}\sigma_{11}(1 - \rho^2), \tag{9}$$

and the weights become

$$w^\dagger = w^*(1 + \frac{d_1(1 - \rho(1 + \delta)) + d_2((1 + \delta)^2 - \rho(1 + \delta))}{\sigma_{11}(1 - \rho^2)}) - \frac{d_1 - \rho(1 + \delta)d_2}{\sigma_{11}(1 - \rho^2)}. \tag{10}$$

Looking at a special case where $\sigma_{22}d_1 - \sigma_{12}d_2$, we get $d_1 = \rho(1 + \delta)d_2$ and

$$w^\dagger = w^*(1 + \frac{d_2}{\sigma_{22}}). \tag{11}$$

The direction of the weight bias is only dependent on the sign of d_2. If d_2 is large, $w^\dagger$ becomes large, and vice versa. The correlation no longer plays a role in the bias of $w^\dagger$.

The behaviour of $w^\dagger$ when the correlation is close to 0 or 1 is also of interest. When ρ is 0, equation 10 simplifies to

$$\begin{aligned} w^{\dagger} &= w^*(1+\frac{d_1+d_2(1+\delta)^2}{\sigma_{11}})-\frac{d_1}{\sigma_{11}} \\ &= w^*+w^*(\frac{d_2}{\sigma_{22}})+(1-w^*)(-\frac{d_1}{\sigma_{11}}). \end{aligned} \tag{12}$$

The bias of $w^{\dagger}$ is now a weighted average between $\frac{d_1}{\sigma_{11}}$ and $\frac{d_2}{\sigma_{22}}$ with weight w itself. We get the same special case in equation 11 when $d_1 = 0$.

When ρ reaches 1, the determinant becomes 0, and the limit of weight becomes ∞ or $-\infty$ depending on the sign of d_1 and d_2. Due to the fact that the sign of the third moment between w and y is unknown, we cannot infer the sign of the scaling factor. Comparing to equation 5, where the effect is limited as long as δ is big enough, the weight in equation 8 is unrestricted and can take any value. This has a negative effect on the estimation as the increase in ρ increases the estimation uncertainty.

2.3 Truncated weight

To avoid the high correlated forecasts, we use truncation on the variable. The weight estimation is as follows

$$\tilde{w} = \begin{cases} w^* & , \text{ if } c < w^* < 1-c \\ 0 & , \text{ if } w^* < c \\ 1 & , \text{ if } w^* > 1-c. \end{cases} \tag{13}$$

This truncation is effectively searching for the forecasts that are above a certain level of weights. If the weight is below the threshold, the forecast is discarded. Therefore the truncation can also be viewed as a forecast selection method.

Assuming that there is no skewness in the joint distribution, e.g., w^* is an unbiased estimator of the true w, then by law of total expectation, the expected value of the weight $\tilde{w}$ is

$$E(\tilde{w}) = E(w^*|c < w^* < 1-c)P(c < w^* < 1-c) + P(w^* > 1-c). \tag{14}$$

The bias is therefore

$$\begin{aligned} Bias_c(\tilde{w}) &= E(w^*) - E(\tilde{w}) \\ &= E(w^*|w^* < c)P(w^* < c) + E(w^* - 1|w^* > 1-c)P(w^* > 1-c). \end{aligned} \tag{15}$$

In the first term, we have $w^* < c < 0$, which cancels out the negative sign and becomes positive. In the second term we have $w^* > 1-c > 1$, which gives a positive value in $w^* - 1$. In general case where w^* can go above $1-c$ or below c and the distribution of w is not symmetric, then the bias is non-zero.

By law of total variance, the variance of the weight is

$$
\begin{aligned}
Var(\tilde{w}) &= Var(I_{c<w^*<1-c}w^* + I_{w^*>1-c}) \\
&= Var(w^*|c < w^* < 1-c)P(c < w^* < 1-c) + P(w^* > 1-c)(1 - P(w^* > 1-c)) \\
&- 2E(w^*|c < w^* < 1-c)P(c < w^* < 1-c)P(w^* > 1-c),
\end{aligned} \tag{16}
$$

and the difference between $\tilde{w}$ and w^* is

$$
\begin{aligned}
Var(w^*) - Var(\tilde{w}) =& Var(w^*|w^* < c)P(w < c) + Var(w^*|w^* > 1-c)P(w^* > 1-c) \\
&+E(w^*|w^* < c)^2 P(w^* < c)(1 - P(w^* < c)) \\
&+E(w^*|w^* > 1-c)^2 P(w^* > 1-c)(1 - P(w^* > 1-c)) \\
&-2E(w^*|w^* < c)P(w^* < c)E(w^*|w^* > 1-c)P(w^* > 1-c) \\
&-2E(w^*|c < w^* < 1-c)P(c < w^* < 1-c)E(w^*|w^* > 1-c)P(w^* > 1-c) \\
&-2E(w^*|w^* > 1-c)P(w^* > 1-c)E(w^*|c < w^* < 1-c)P(c < w^* < 1-c) \\
&+2P(w^* > 1-c)E(w^*|c < w^* < 1-c)P(c < w^* < 1-c) \\
=& Var(w^*|w^* < c, w^* > 1-c) \\
&-2E(w^*|c < w^* < 1-c)P(c < w^* < 1-c) \\
&\quad [P(w^* > 1-c)(E(w^*|w^* > 1-c) - 1) - E(w^*|w^* < c)P(w^* < c)].
\end{aligned} \tag{17}
$$

where row 1, 2, 3, and 5 collapse to $Var(w^*|w^* < c, w^* > 1-c)$. Since $E(w^*|w^* > 1-c) - 1 > 0$ and $E(w^*|w^* < c) < 0$, the term within the square bracket is positive. The effect of the second term depends solely on $E(w^*|c < w^* < 1-c)$. The value of the expectation can be both positive or negative, and we cannot infer the effect without the distribution of w^*.

The mean squared error (MSE) is then given by

$$
\begin{aligned}
MSE_c &= Bias_c(\tilde{w})^2 + Var(\tilde{w}) \\
&= (E(w^*|w^* < c)P(w^* < c) + E(w^* - 1|w^* > 1-c)P(w^* > 1-c))^2 \\
&+ Var(w^*|c < w^* < 1-c)P(c < w^* < 1-c) + P(w^* > 1-c)(1 - P(w^* > 1-c)) \\
&- 2E(w^*|c < w^* < 1-c)P(c < w^* < 1-c)P(w^* > 1-c)
\end{aligned} \tag{18}
$$

We expect that when c is small enough, e.g. $\tilde{w} \approx w^*$, the increase in bias is small compared to the decrease in variance. When the truncation c gradually increases, the decrease of the variance starts to slow down, and the bias starts to increase more. This is due to the quadratic behaviour of MSE. Thus, somewhere between w^* and $\frac{1}{n}$, there will be an optimal trade-off. This is similar to shrinkage estimator (James and Stein, 1961). However, truncation set the truncated weight to 0 instead of c. This effectively removes the forecast from the list to be combined. By incorporating this truncation style, the weight can be viewed as a variable selection method in the area of forecast combination, where negative weight is empirically inferior to positive weight.

2.4 Bias Correction

In this section we provide the bias correction following Gibbs and Vasnev (2017). They suggest that knowing the conditional bias helps the weight estimation. If the conditional bias is large for a certain forecast, the forecast should be under-weighted, similar to the case where the variance of the forecast is large.

To start with bias estimation, let the prediction error from forecast i be $\epsilon_i = y - y_i$, that we can decompose it into predictable term and unpredictable term:

$$\epsilon_i = b_i + \xi_i. \tag{19}$$

with the expected value

$$E(\epsilon_i) = b_i. \tag{20}$$

Let $b_i \neq 0$, then the expected error is non-zero

$$E(\epsilon_c) = wb_1 + (1 - w)b_2 \tag{21}$$

and the variance

$$Var(\epsilon_c) = w^2\sigma_{\xi,11} + (1 - w)^2\sigma_{\xi,22} + 2w(1 - w)\rho_{\xi_1,\xi_2}\sigma_{\xi,1}\sigma_{\xi,2} \tag{22}$$

where σ_{ξ_i} is the variance of ξ_i.

Since the expectation is non-zero, we minimise the mean squared error to calculate the weight. Write $b_{ij} = b_i * b_j$, then minimising the mean squared error gives

$$\hat{w} = \frac{\sigma_{\xi,22} - \sigma_{\xi,12} + b_{22} - b_{12}}{\sigma_{\xi,11} + \sigma_{\xi,22} - 2\sigma_{\xi,12} + b_{11} + b_{22} - 2b_{12}}. \tag{23}$$

Notice for equation 23, we can use the equation 2 from optimal weight w^*, with a slight modification on the variance matrix. Instead of using the original variance matrix, we can replace it with $\begin{bmatrix} \sigma_{\xi,11} + b_{11} & \sigma_{\xi,12} + b_{12} \\ \sigma_{\xi,12} + b_{12} & \sigma_{\xi,22} + b_{22} \end{bmatrix}$, and the results will be equal.

We study a few special cases here. When bias is equal, $b_1 = b_2$, we have $\hat{w} = w^*$. The bias cancels each other out.

If $\rho_{\xi_1,\xi_2} = 0$ and $b_1 = 0$, we have

$$\hat{w} = \frac{\sigma_{\xi,22} + b_{22}}{\sigma_{\xi,11} + \sigma_{\xi,22} + b_{22}} \tag{24}$$

and converge to 1 if $b_2 \to \infty$.

When $\rho_{\xi_1,\xi_2} = 0$ and $b_2 = 0$, we have

$$\hat{w} = \frac{\sigma_{\xi,22}}{\sigma_{\xi,11} + \sigma_{\xi,22} + b_{11}} \tag{25}$$

and converge to 0 if $b_1 \to \infty$. Here we see that the effect of bias are similar to variance. The higher the bias, the lower the weight for the forecast.

The final special case is when $\rho_{\xi_1,\xi_2} = 0$. We have

$$\hat{w} = \frac{\sigma_{\xi,22} + b_{22} - b_{12}}{\sigma_{\xi,11} + \sigma_{\xi,22} + b_{11} + b_{22} - 2b_{12}}. \tag{26}$$

A similar idea approaches here to equation 4, with $b_1 = b_2(1+\gamma)$. However, b_{22} cannot be cancelled out

$$\hat{w} = \frac{\sigma_{\xi,22} - \gamma b_{22}}{\sigma_{\xi,11} + \sigma_{\xi,22} + b_{22}\gamma^2}. \tag{27}$$

The implication on the optimal weight is also different from equation 4. When $\gamma \to 0$, we have the special case with $b_1 = b_2$, which gives the simple weight w^*. When $\gamma \to \infty$, the denominator becomes huge and $\hat{W} \to 0$. It is worth noting that the boundary of 0 and ∞ does not imply $\hat{w} \in [0, \infty]$. One can think of a case where b_{22} is large enough such that $\sigma_{\xi,ii}$ are negligible and the weight becomes $\frac{-\gamma}{\gamma^2}$.

3 Survey of Professional Forecasters (SPF)

To illustrate the empirical results, we use the data from European Central Bank (ECB) in this paper. The data, Survey of Professional Forecasters (SPF), is a quarterly survey initiated by ECB, with the aim to obtain future estimates on inflation (HICP), Real Gross Domestic Product year-on-year growth (RGDP) and unemployment rate (UNEM) from the private sector. Every quarter, a group of professional forecasters from financial and non-financial institution, such as economic research institutions, respond to the survey with the idea on the future economy. Starting 1999, SPF is the longest survey of macroeconomic expectation in the Euro area. Until the date of this paper, there are 75 quarters of observation available, with 1999 Q4 as the first forecasted value, and 2018 Q2 as the last observed true macroeconomic indices. We take 2016 Q1 as the first quarter to forecast, with a extending window size. The first in-sample time frame is forecasts made on 1999 Q4 to 2015 Q4, and the second in-sample time frame adds 2016 Q1 to the first.

The set up of the survey consists of multiple magnitudes of questions, ranging from different horizon to different distribution. The forecasters are asked to provide their point forecast and the probability of a certain scenario to happen. The survey enables ECB to do a quantitative assessment on the consensus of the market, like the distribution statistics and standard deviations. For this paper, we take the two most answered periods, which is 1 year ahead and 2 years ahead as our dataset for all HICP, RGDP, and UNEM.

To compare the forecasts with the actual macroeconomics, we obtain the true value from ECB database. The data cannot be observed from the economic in 100% accuracy within the first time frame and exhibits changes to the initial estimates after revision. We use the final revision of the macroeconomics where possible, that is, the revision done on 2018 Q2. The use of a final revision is fine because the original forecast is not the real target to forecast.

Within the datasets, not all forecasters did a forecast every time period. To avoid singular outliers, we remove all forecasters with a total forecasted period of less than 24 quarter (6 years). The removal approach is in-line with Matsypura et al. (2018).

Following Matsypura et al. (2018), we calculate the covariance by looking at the intersection between each forecaster. Let T_i be the set of periods which the i-th forecaster has respond to the survey, $T_i \subseteq \{1, \ldots, T\}$, and let e_{it} be the forecast error of i-th forecaster on time t, $e_{it} \in T_i$. Then the variance/covariance can be calculated with

$$\sigma_{i,j} = \frac{1}{|T_i \cap T_j|} \sum_{t \in \{T_i \cap T_j\}} e_{it} * e_{jt} \tag{28}$$

When there is no intersection between 2 forecasters, we set the covariance value to 0. When $i = j$, the intersection of T_i and T_j does not influence the variance calculation, and the variance becomes mean squared error. Additionally, we calculate the correlation by using the covariance divided by the standard deviation. Standard deviation is obtained from the square root of the diagonal.

$$\rho_{i,j} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \tag{29}$$

The cleaned up gives us a preliminary view on the SPF data without the noises.

In figure 1 and table 6 show the plots of the forecasts alongside with the true value in the macroeconomics and the statistics of the covariance of the forecast error. We plot only the minimum, mean, and maximum from the forecasts. We see that there exist a high consistency across all forecasts, with two years ahead stronger than one year. The consistency in the forecast is lower in UNEM than the other two. Furthermore, many true values lie outside of the forecast range, with RGDP the worse of all three. More values outside of the forecast range suggest that restricting

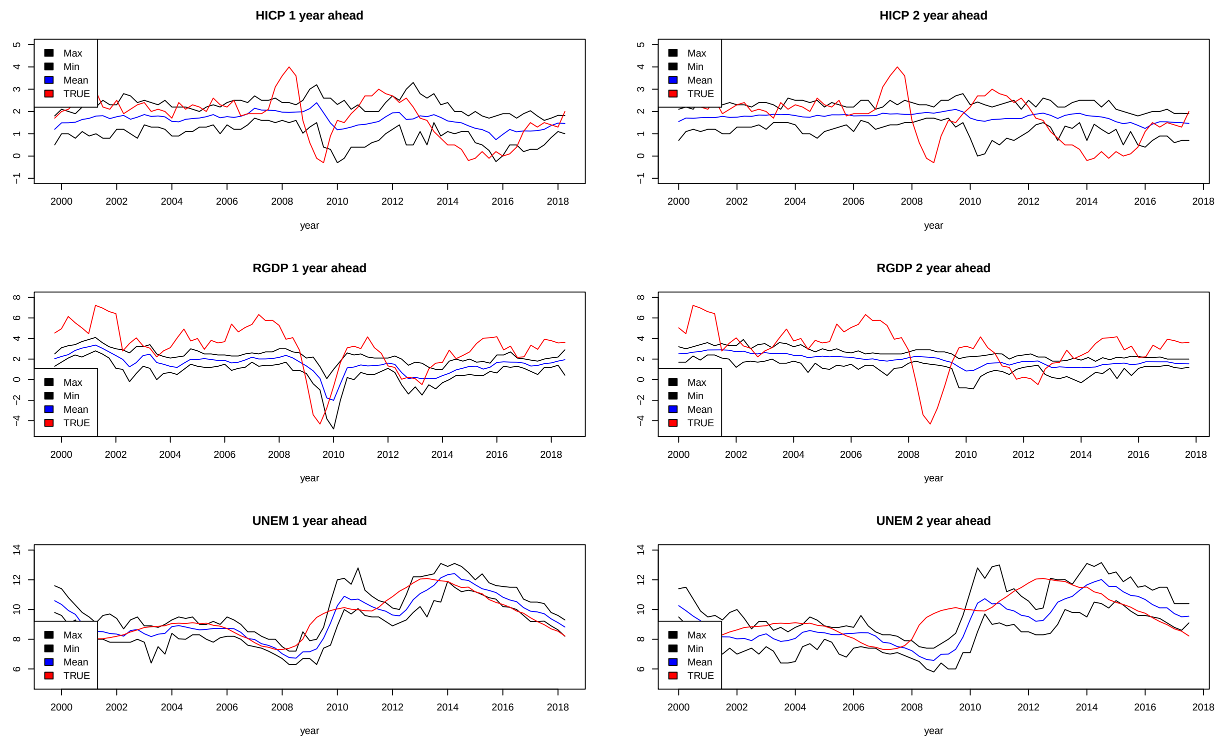

Figure 1: Survey of Professional Forecasters data illustration

positive weight may be a strong limitation in the forecast combination.

We examine the amount of true value outside of forecast range by looking at the summary statistics. Let model space be an indicator with 1 when the true value is outside of the forecast range, and 0 when the true value is inside. Table 1 shows the mean of the indicator. From the three macro topics, RGDP has the highest percentage of time periods to be outside the forecast range, 83%, followed by HICP with an average of 48% of the time periods. UNEM with 38% of the time periods has the lowest percentage to be outside the forecast range. Changing from 1 year to 2 years generally does not influence the mean of the indicator a lot. From the results from figure 1 and table 1, we expect to have large effect using truncation in the forecast of RGDP, while HICP and UNEM do not have a strong effect. We also expect the two years ahead forecast will be better than the one year ahead.

Table 1: Mean of the model space indicator of the forecast. The indicators are split up into different forecast topics and different forecast horizons. From the three macro topic, RGDP has the least chance within the forecast range, followed by HICP. UNEM has the highest chance to be in the forecast range.

	Model Space Indicator					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Mean	0.45	0.51	0.83	0.82	0.37	0.39

As not all forecasters replied to each survey, we show the amount of forecasters per macroeconomic topic in table 2. The amount of forecasters after filtering is around 70 for 1 year ahead forecasts, and 60 for 2 year ahead forecasts. We also show the amount of replies per forecaster in a summary statistic in table 3. The minimum amount of forecast are set to 24, and all forecasters with less than 24 forecasts are discarded. The maximum replies are around 70, which is close to the full amount of 75 observations. With the mean around 50, we have enough observations to do pair-wise covariance, but multivariate covariance remains sceptical.

Table 2: Amount of Forecasters. The amount of the forecaster for 1 year ahead forecast is higher than 2 year ahead.

	Amount of Forecasters					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Amount	72	60	70	64	65	53

Table 3: Summary statistics of the survey reply amounts per forecasters. The average forecasters made around 50 replies for the whole duration of the survey. No forecaster replied to all survey. Means are round down to whole number for readability.

	Summary Statistics of the Amounts Per Forecaster					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Minimum	24	24	26	24	25	24
First Q	40	40	42	35	40	36
Median	55	52	55	51	53	52
Mean	53	50	54	49	52	50
Third Q	67	62	66	63	66	63
Maximum	74	70	75	71	74	70

Table 4 shows the summary statistics of the amount of survey replies per time period. On average there are 40 replies per time period. The deviation to the time is small with more than 50% within 35 to 45 range. The minimum of 22 in UNEM pose a possible problem in the estimation of the covariance. If the estimation of the covariance is bad, limiting negative weights will pose a better result.

Table 5 shows the amount of replies per time period that we use to test the model. The values are lower than the mean for almost all time period except a few. Only RGDP 1 year on 2018 Q1 and 2018 Q2, and UNEM 1 year on 2016Q1 are above the average. This decreases the dimension of the covariance matrix, but on the other hand decreases the aomunt of observations. The trade-off between lower dimension should be more important than the loss in observations.

The table 6 tell us on how the forecast errors are correlated. The correlations are split up into different forecast topics and different forecast horizons. The diagonal elements of the correlation

Table 4: Summary statistics of the survey reply amounts per time period. There are around 35 to 45 responses for half of the time. Means are round down to whole number for readability.

	Summary Statistics of the Amounts Per Time Period					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Minimum	37	26	38	28	32	22
First Q	43	35	42	36	38	30
Median	45	38	45	40	40	34
Mean	46	39	45	40	41	34
Third Q	49	43	48	45	44	38
Maximum	54	49	54	52	51	44

Table 5: Amount of replies to the survey on the testing period starting 2016. All forecasters that replied in the test periods had also replied in previous time periods.

	Survey Reply Amounts					
Macro topic	HICP		RGDP		UNEM	
Test Period	1 year	2 years	1 year	2 years	1 year	2 years
2016 Q1	45	32	38	33	41	30
2016 Q2	40	33	42	37	36	28
2016 Q3	44	35	42	38	39	28
2016 Q4	45	34	41	40	39	30
2017 Q1	39	35	38	30	32	32
2017 Q2	38	27	38	35	33	25
2017 Q3	37	31	43	36	33	29
2017 Q4	43	34	41	36	36	29
2018 Q1	40	27	46	30	34	22
2018 Q2	44	26	45	31	40	23

matrix are not within the table when generating the summary statistics. For all of the series, the correlations are on average above 60%. In HICP and UNEM the correlation increases across all statistics when the forecast horizon increases, while RGDP remains the same. The lowest correlation to be found is -0.02, but this number is not too different from the minimum correlation in the other two topics.

We conclude in for the analysis of SPF that there exhibits characteristics that are not in the standard cases where correlation are close to 0. By incorporating the possibility of negative weight, we expect to see some improvement in the forecast errors.

Table 6: Summary statistics of the correlation of the forecast error. The correlation are split up into different forecast topics and different forecast horizons. For all of the series, the correlation are on average above 60%. In HICP and UNEM the correlation increases across all statistics when the forecast horizon increases, while RGDP remains the same.

	Correlation					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Minimum	0.03	0.02	0.11	0.13	-0.02	0.08
First Q	0.53	0.62	0.63	0.65	0.47	0.56
Mean	0.64	0.70	0.75	0.75	0.59	0.67
Median	0.66	0.72	0.81	0.80	0.62	0.70
Third Q	0.77	0.81	0.89	0.87	0.74	0.80
Maximum	0.96	0.97	0.98	0.98	0.94	0.95

4 Empirical Results

4.1 Optimal Weight

Analysis of the optimal weight given by equation 2 provides us with a preliminary understanding of the variability. Table 7 shows the summary statistics of the optimal weight. We see that for all macroeconomic topics, the mean and median variate around 0.02, which is close to the calculation from equal weight. Looking at the first and the third quantile, we see that half of the weight are between -0.6 and 0.6, which gives us a possible range of effectiveness of the threshold. The minimum and the maximum is extreme considering that 1 is 100%. With negativity up to -11, -21, and -14, and first and third quartile only up to -0.6 and 0.6, the weight indicates a strong thick tail behaviour. Looking further to the difference between 1 year and 2 years ahead, we see that in two out of three macro topics, the extreme values converge to 0, while RGDP becomes worse. On the other hand, the gap between first and the third quantile increases for all topics. This indicates the increase in the choice of negative weight in the general cases for 2 yearss horizon. The mean and median does not change much in relation to the forecast horizon.

4.2 Proceure

We seek to evaluate the effect of truncating the weight with SPF data. The procedure to get the weight can be different between different researchers. Therefore we provide a through order of the steps we take in the weight estimation. We do the weight estimation for the truncation using the following steps:

Given each time to forecast, we take all the know observation before that forecast time period.

1. Find the nearest positive definite covariance matrix. This step is required for the covariance

Table 7: Summary statistics of the weight from equation 2. The mean and median is close to the value from equal weight. However, extreme values exist in the weight, visible in the mimum and maximum. Based on the first and the third quantile, we see that half of the weight are between -0.6 and 0.6.

	Optimal Weight					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Minimum	-11.05	-6.21	-9.00	-21.79	-14.41	-5.64
First Q	-0.30	-0.59	-0.20	-0.63	-0.28	-0.40
Mean	0.04	0.04	0.01	-0.01	0.03	0.05
Median	0.05	0.03	0.04	0.01	0.04	0.03
Third Q	0.41	0.60	0.25	0.82	0.39	0.43
Maximum	12.94	7.56	8.74	16.34	11.26	7.90

to be invertible. We employ the nearPD function from r package *Matrix*. The nearPD function first decompose the covariance into univariate variance and the correlation. The function then uses the algorithm by *higham* on the correlation matrix to compute the nearest positive definite matrix. The final results the the covariance matrix that is combined from the univariate variance and the correlation matrix.

2. Subset the covariance. We take the sub matrix containing only the available forecast on the testing period. That is, if 80 forecasters had made some forecasts before, but out of them, only 40 has a forecast this time, we discard the 40 extra forecasters.

3. Estimate the optimal weight w^* from equation 2.

4. Truncate the optimal weight with equation 13. The weight now has less values and does not sum to one. We normalise the weights back such that the sum of the weights is 1.

5. Combine the forecast using the new weights.

6. Calculate the test statistics of new weights and equal weights.

7. Calculate the ratio of the test statistics of new weights over equal weights.

4.3 Ratio of the Test Statistics

We evaluate the performance of different weights by mean squared prediction error (MSPE) and mean absolute prediciton error (MAPE). The truncated MSPE and MAPE are then compared with the MSPE and from equal weight to produce an test ratio. The results of HICP, RGDP, and UNEM are given in table 8, 9, and 10 respectively. We report the ratio of the MSPE and MAPE. This helps us visualize the effect of truncation instead of comparing per number to the equal weights. A value smaller than 1 means the test statistic of the truncation is smaller than test statistic of equal weight.

We see in table 8 that for the truncation value -2 and -1.5, the minimum of MSPE is attained for 1 year horizon and 2 years horizon respectively. The minimum of MAPE are attended at -2 and 0 respectively. The 1 year selection based on MAPE collides with selection based on MSPE, but the minimum for MAPE is not at -1.5. The optimal weight obtains a ratio of 3.29 and 2.21 for MSPE, significantly larger than the ratio obtained by truncation. The ratio difference of MAPE between optimal weight and truncated weights are smaller compared with MSPE. Truncating all the way to 0 means that no forecast with weight less than 0 are selected. For HICP, selecting too strict threshold does not help in 3 out of 4 cases.

Table 8: Mean Squared Prediction Error (MSPE) and Mean Absolute Prediction Error (MAPE) of inflation with different truncated value. The MSPE and MAPE of the truncated value are given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE or MAPE.

	HICP			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
$-\infty$	3.2902	1.5898	2.2072	1.4347
-5	2.8211	1.4178	2.1148	1.3537
-4.5	2.8211	1.4178	2.1148	1.3537
-4	2.8187	1.414	2.1500	1.3753
-3.5	2.8178	1.4126	1.6747	1.1988
-3	2.8178	1.4126	1.5479	1.1551
-2.5	2.8180	1.4129	1.0181	1.0638
-2	0.8078	0.9025	0.993	1.0653
-1.5	0.8877	0.9619	0.9688	1.0390
-1	0.8835	0.9610	0.9764	1.0170
-0.5	0.9465	0.9912	0.9877	1.0108
0	0.9844	0.9994	0.9834	1.0005

Table 9 shows a similar pattern in RGDP as in HICP. The optimal threshold for 1 year is around -1.5, obtaining 0.83 and 0.88 for MSPE and MAPE respectively. We did not look for threshold larger than 0, and the minimum MSPE of RGDP 2 years is not within the search region. This cut-off gives us 1.01 as the boundary case. The MAPE on 2 year horizon acts differently. The minimum MAPE is attended at -4.5. The optimal weight MSPE ratio in 1 year horizon is similar to equal weight, scoring 1.44 in the ratio. 2 years ahead MSPE ratio in other hand does not perform that well in optimal weight, with a ratio of 6.07. The optimal weight in MAPE ratio for 1 and 2 year show similar values to equal weights, scoring 1.07 and 1.80 respectively.

When we look further in UNEM, the results show promising ratios. UNEM achieved the lowest MSPE ratio in 1 year and 2 years at -1.5 and -0.5. Comparing to HICP and RGDP, the MSPE ratio of 0.70 and 0.82 are the lowest of all three. The MSPE ratio of 0.70 is also an 40% decrease in the MSPE, even when the non-truncated is already at 1.16. MAPE ratio shows that in 1 year ahead, optimal weight already outperforms equal weights with 0.89, but with truncation the MAPE

Table 9: Mean Squared Prediction Error (MSPE) and Mean Absolute Prediction Error (MAPE) of economic growth with different truncated value. The MSPE and MAPE of the truncated value are given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE or MAPE.

	RGDP			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
$-\infty$	1.4387	1.0681	6.0654	1.8047
-5	1.1000	0.9561	1.0684	0.9496
-4.5	1.1462	0.9861	1.0683	0.9496
-4	1.1462	0.9861	1.0885	0.9621
-3.5	0.8738	0.9029	1.0765	0.9539
-3	0.8768	0.9046	1.0824	0.9579
-2.5	0.8611	0.8967	1.102	0.9899
-2	0.8402	0.8824	1.0824	0.9567
-1.5	0.8278	0.8768	1.1712	1.0438
-1	0.8610	0.9102	1.1356	1.0617
-0.5	0.8999	0.9396	1.0198	1.0114
0	0.9721	0.9844	1.0073	1.0056

ratio drops to 0.74. The MAPE ratio in 2 year ahead decreases to 0.91. UNEM seems to have an effect from truncation that is not explained by the correlation. UNEM does not have a difference in correlation to other macroeconomics topics that cannot be explained by estimation noise. One possible explanation is the optimal choice of the truncation is more consistent and does not vary over time. We evaluate the consistency of the truncation in the out-of-sample selection area.

Figure 2 shows the MSPE ratio and MAPE ratio for HICP, RGDP, and UNEM with forecast horizon 1 and 2 years. The figures are plotted with a step size of 0.1. From the figure, HICP shows a sudden decrease in both statistics and minor changes afterwards. RGDP 1 year shows a similar pattern with a larger increase towards the end. For RGDP 2 years, there is a spike in both statistics around 1.3. Given that the value never falls below 1, we suspect that the correlation does not capture the error well enough, and there might be a better measure to counter the forecast error. UNEM 1 year is a bit volatile but exhibits a well-defined U-shape. For UNEM 2 years, the ratios drop in a smooth way to the minimum.

To conclude, we see the possibility to improve the MSPE and MAPE by using the truncation in all cases. In the beginning, there is a drop in the test statistics, but closer to the threshold of 0, test statistics increase. The MSPE and MAPE both have a U-shape characteristic, which attribute to variance and bias. By applying truncation, the drop in variance is larger than the increase in bias, and this contributes to the decrease in the test statistics. It is worth noting that all macroeconomic topics has on average better performance than equal weight, which is considered a benchmark that is hard to beat. Two of those macroeconomic forecasts, HICP 2 years and RGDP 2 year, have a

Table 10: Mean Squared Prediction Error (MSPE) and Mean Absolute Prediction Error (MAPE) of unemployment with different truncated value. The MSPE and MAPE of the truncated value are given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE or MAPE.

	UNEM			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
$-\infty$	1.1639	0.8899	4.3535	1.6517
-5	1.0249	0.8438	2.3189	1.3796
-4.5	0.8926	0.7991	2.3189	1.3796
-4	0.8926	0.7991	2.2467	1.3552
-3.5	0.9447	0.8189	2.0462	1.3042
-3	0.9335	0.8137	1.9501	1.2435
-2.5	0.9097	0.8050	1.9146	1.2427
-2	0.7684	0.7858	1.8718	1.2264
-1.5	0.6988	0.7399	1.6610	1.1328
-1	0.7283	0.7554	1.4222	1.0346
-0.5	0.8232	0.8471	0.8242	0.9066
0	0.9346	0.9543	0.8769	0.9466

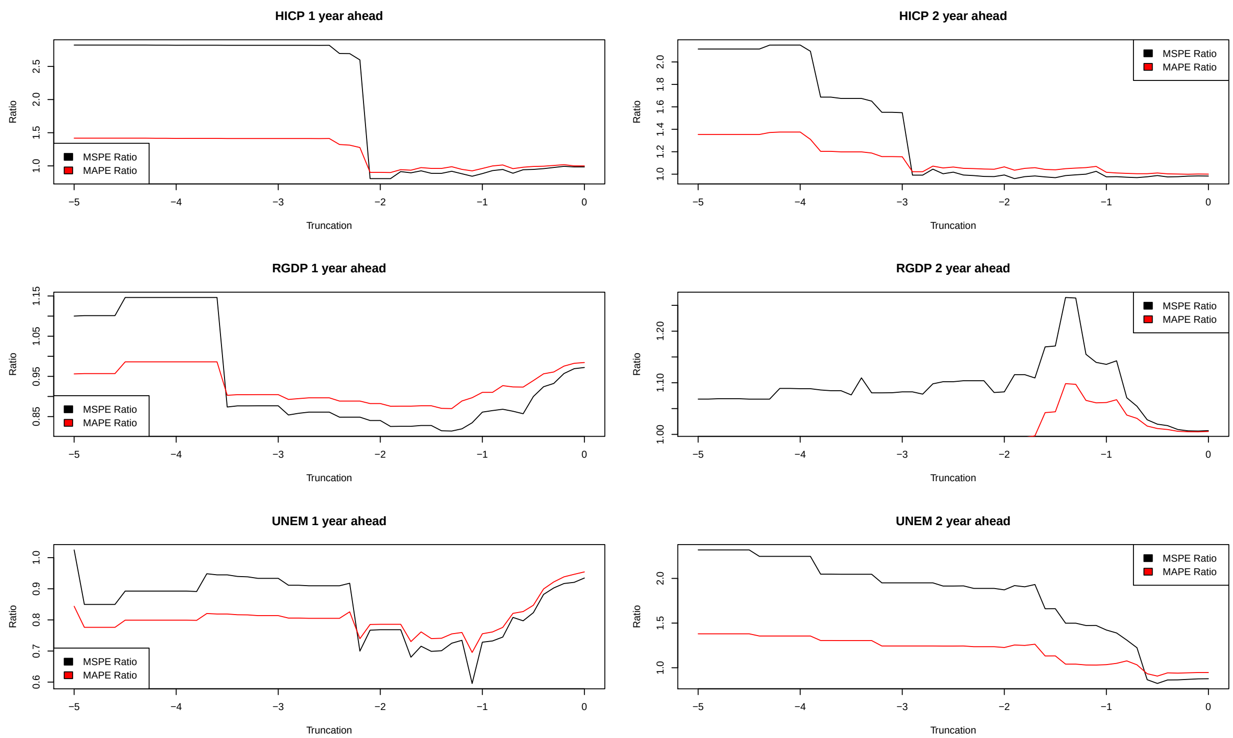

Figure 2: MSPE Ratio for different macroeconomic topics

hard to justify improvement in MSPE or MAPE when estimation error is taken into consideration. Other four obtain from 17% to 30% decrease when compared to equal weight. The results also show that the decrease in MSPE is not just one point, but a graduate movement. This implies the robustness in truncation selection. Even if the user does not select the most optimal case, but deviates a small amount next to it, the increase in the test statisitcs is relatively contained.

4.4 Optimal Truncation Selection

In this section, we show the MSPE and MAPE in weight truncation when the truncation parameter is in-sample optimally selected. To perform the selection of the truncation parameter, we follow the following step:

1. Estimate the optimal weight in equation 2.
2. Calculate the in-sample mean squared error (MSE) with truncated weight, e.g., the first window is 1999 Q4 to 2015 Q4. The truncation parameter ranges from -10 to 0 with a step of 0.1. We also look into other starting values, namely -5, -2, and -1.
3. Select the truncation parameter with the lowest in-sample MSE.
4. Use the selected truncation parameter to combine the predictions.
5. Calculate the test statistics of new weights and equal weights.
6. Calculate the ratio of the test statistics of new weights over equal weights.

If there are multiple truncation points with the same MSE value, we take the largest truncation value within the same MSE. The reason to choose largest value is that the threshold is closer to the next changing point. Assume for example, a weight with a minimum of -1 is optimal, and any higher weight results in higher MSE. Then all truncation values between -10 and -1 gives the same minimal MSE. We therefore choose -1 to record. The choice does not influence the forecast and is pure for the analysis later.

4.5 Optimal Truncation Selection Result

We evaluate the Optimal Truncation (OT) forecast ability using the same way as section 4.3. The MSPE ratio and the MAPE ratio are given in table 11.

As shown in table 11, OT improves in all MSPE cases when we compare with equal weights. In MAPE, OT failed to improve HICP 1 year, but obtained positive results in other 5 topics. The truncation works better in RGDP and UNEM. If we compare the value to the MSPE or MAPE of optimal weights, we see that the truncation improves the forecasts for all topics except UNEM 1 year. RGDP 2 years also shows in MSPE that none of the constant truncations works as good as changing truncation. This means that the optimal truncation selection overweight the selection uncertainty. These results verify the idea that the truncation improves the upon equal weight even if the truncation uncertainty can increase the MSPE.

Table 11 also illustrates the effect of changing the search area. Smaller search area gives less uncertainty, but are more restrictive in the solution space. For HICP 1 year and UNEM 2 year, changing the search area does not influence the test statistics enough to be visible under 4 digits of precision. For HICP 2 year and RGDP 1 year, being more restictive increases the test statistics, effectively means worse predictions. For RGDP 2 year and UNEM 1 year, being restrictive decreases the prediction error, and therefore MSPE and MAPE. Since if being restrictive or not relies on prior information on the forecast set, it is hard to determine which is a good choice.

Table 11: Mean squared prediction error ratio and mean absolute prediction error ratio when the selection of the truncation is optimal. The minimum search to optimal threshold are -10, -5, -2, and -1.

	MSPE with Optimal Truncation					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
-10	0.9670	0.9611	0.9275	0.9558	0.9153	0.8752
-5	0.9670	0.9611	0.9275	0.9558	0.9153	0.8752
-2	0.9670	0.9719	0.9319	0.9518	0.8982	0.8752
-1	0.9670	0.9709	0.9319	0.9949	0.8982	0.8752
	MAPE with Optimal Truncation					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
-10	1.0015	0.9928	0.9532	0.9577	0.9320	0.9502
-5	1.0015	0.9928	0.9532	0.9577	0.9320	0.9502
-2	1.0015	0.9968	0.9562	0.9533	0.9260	0.9502
-1	1.0015	0.9988	0.9562	0.9524	0.9260	0.9502

To better understand the OT, table 12 looks at the selection of the truncation. The results show that on average, a truncation around -0.5 is selected, indicating a preference in the negative weight. HICP 2 years and RGDP 2 years select on average lower truncation, -0.8 and -1.0 respectively. UNEM selects -0.2 on average, highest among all. The minimum across all macro topics are at most -1, with RGDP 1 year, RGDP 2 year, and UNEM 1 year the lowest selected minimum. The first and the third quantiles show that for the most cases, the truncations are not extreme, ranging between -0.5 to -0.2 in 1 year and -1 to 0 in 2 years horizon. We notice an increase in the variation when the horizon increases. This can be attribute to the uncertainty in the horizon adds uncertainty in the truncation parameter selection.

Figure 3 gives a visual inspection on the stability of the selection. For all cases, the fluctuation is limited. Excluding the spikes in some plots, the selection ranges within -1.5 to 0. We suspect that by setting the search range from -1.5 to 0 will improve the OOS truncation selection.

Table 12: Summary statistics of the truncation selected. The available truncation ranges from -10 to 0 with steps of 0.1. Additionally, we add -∞ to the truncation choice, essentially return the non-truncated weight.

	Selected Optimal Truncation					
Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 years	1 year	2 years	1 year	2 years
Minimum	-1.0	-2.1	-3.3	***-4.6***	-3.0	-1.0
First Q	-0.6	-1.0	-0.5	-1.1	-0.4	-0.3
Mean	-0.5	-0.8	-0.5	-1.0	-0.4	-0.2
Median	-0.4	-0.6	-0.2	-0.4	-0.2	-0.2
Third Q	-0.3	-0.3	-0.1	-0.2	-0.2	0.0
Maximum	0.0	0.0	0.0	0.0	0.0	0.0

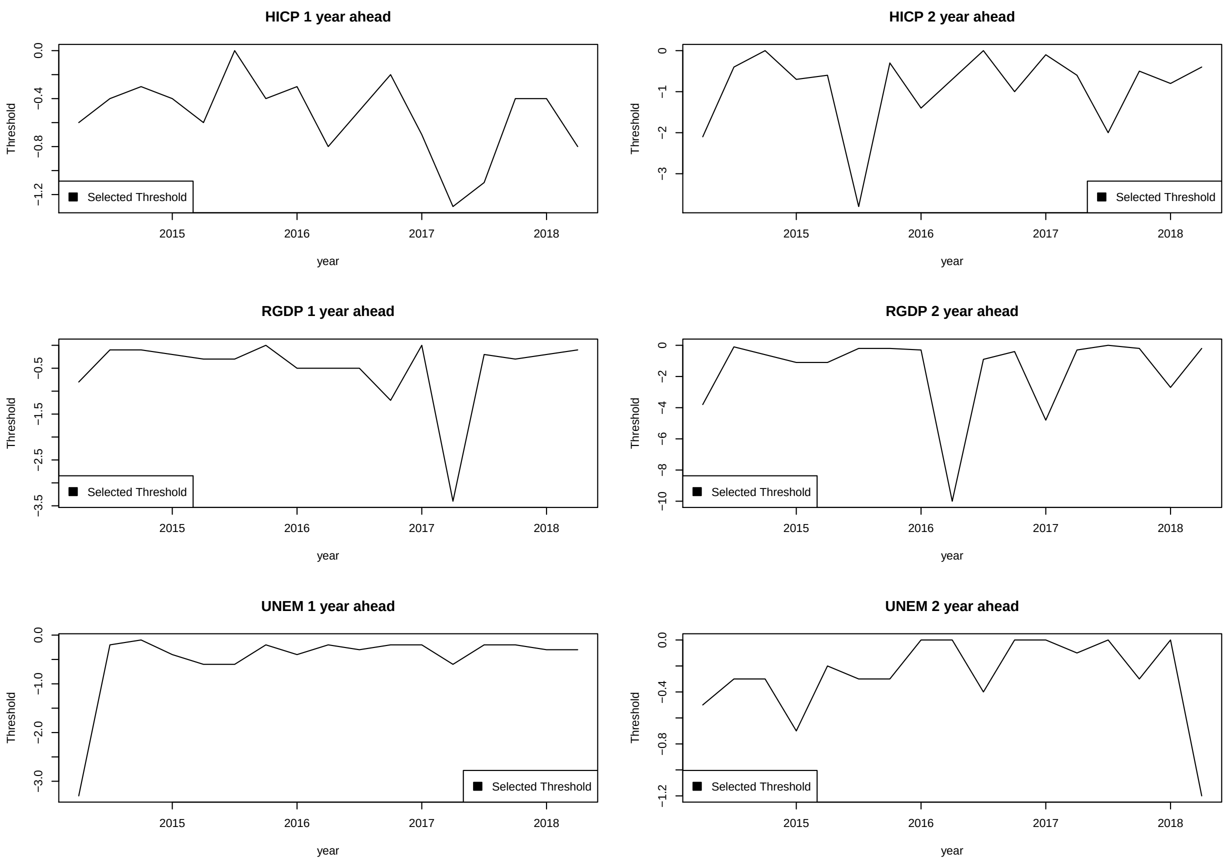

Figure 3: Truncation selection for different OOS period

4.6 Bias weighting

The bias weighting is an interesting case here. If the model in the bias estimation successfully estimates the predictable part and the unpredictable error, this can attribute to a better weight selection with additional information gained. Our approach to bias estimation is similar to Gibbs and Vasnev (2017) but deviates in the variable used. We estimate the bias by

$$\epsilon_{i,t} = \alpha + \gamma_i y_{i,t} + \eta_{i,t}, \tag{30}$$

which we then produce the estimated future forecast bias by taking the expected value

$$E(\epsilon_{i,t+1}) = \alpha + \gamma_i y_{i,t+1}. \tag{31}$$

Gibbs and Vasnev uses the true macroeconomic data, while we uses the forecast itself. Essentially they are interchangeable. Rewrite equation 30 into

$$\begin{aligned}\epsilon_{i,t} &= \frac{\alpha}{1+\gamma_i} + \frac{\gamma_i}{1+\gamma_i} y_t + \frac{1}{1+\gamma_i}\eta_{i,t} \\ &= \alpha^* + \gamma_i^* y_t + \eta_{i,t}^*,\end{aligned} \tag{32}$$

and equation 30 becomes the equation given by Gibbs and Vasnev.

Table 13, 14, and 15 shows the effect of truncation with bias-corrected weight. The estimation of bias increases the estimation error and therefore is another trade-off between estimation error and the bias. Only two out of six topics have a ratio below 1, considerably worse than without bias correction. The two topics are RGDP 2 year and UNEM 1 year, with minimum of 0.83 and 0.56 respectively. For HICP with bias correction, the truncation performs best with the highest value. This is different from no-bias selection.

Table 13: Mean Squared Prediction Error (MSPE) of bias-corrected inflation with different truncated value. The MSPE of the truncated value is given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE.

	HICP			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
-∞	2.7747	1.7760	1.6452	1.3735
-5	2.7747	1.7760	1.6452	1.3735
-4.5	2.7747	1.7760	1.6452	1.3735
-4	2.7747	1.7760	1.6452	1.3735
-3.5	2.7747	1.7760	1.6452	1.3735
-3	2.7747	1.7760	1.5888	1.2976
-2.5	2.7747	1.7760	1.6059	1.3276
-2	2.7747	1.7760	1.5955	1.3064
-1.5	2.7747	1.7760	1.5887	1.2871
-1	2.6528	1.6894	1.1640	1.1141
-0.5	1.7696	1.3260	1.0464	1.0365
0	1.1118	1.0576	1.0221	1.0231

RGDP shows similar results in 1 year, where truncation helps, but selecting 0 is the best performing one. On the other hand, RGDP 2 years shows that it can achieve a MSPE ratio of 0.83 with a truncation value of -1. The selection from MAPE ratio yields the same truncation. Comparing this to RGDP without bias correction, the bias-corrected performs worse in 1 year horizon, while performing better in 2 years horizon. The best truncation also changes to a different value, from -1.5 for 1 year horizon and 0 for 2 years horizon in no correction to 0 and -1 respectively.

Table 14: Mean Squared Prediction Error (MSPE) of bias-corrected economic growth with different truncated value. The MSPE of the truncated value is given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE.

	RGDP			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
$-\infty$	3.3002	1.5457	1.0535	0.9458
-5	1.7581	1.2626	1.1094	0.9846
-4.5	1.7915	1.2894	1.1094	0.9846
-4	1.7910	1.2888	1.1069	0.9833
-3.5	1.8086	1.298	1.1079	0.9838
-3	1.8101	1.2984	1.1076	0.9837
-2.5	1.8118	1.3004	1.0691	0.9716
-2	1.6089	1.3075	0.9950	0.9438
-1.5	1.4604	1.2530	0.8630	0.8931
-1	1.2764	1.1320	0.8275	0.8753
-0.5	1.0758	1.0464	0.9437	0.9508
0	1.0436	1.0264	0.9835	0.9892

With UNEM 1 year as low as 0.56, UNEM achieved is a large improvement to the equal weight case. However, this effect does not show in the 2 years horizon. The 2 years horizon MSPE ratio increased to higher than equal weight.

Figure 4 shows the MSPE ratio and MAPE ratio over the different threshold in higher granularity. HICP does not attend the minimum with threshold up to 0 for both forecast horizon. The same conclusion also holds for RGDP 1 year and UNEM 2 years. On the other hand, RGDP 2 years has a large effect in truncation with bias-correction occurs, while performs badly without bias-correction. One reason for this is that the problem with RGDP is the highly biasses over time, which cannot be corrected using cross-sectional truncation. UNEM 1 year is another macroeconomic topic that perform better with bias correction. We see from the plot that a large area is well below 0.7, relieving the problem of selecting precisely -0.5.

All in all, we observed the possibility to combine bias-correction and truncation. Some macroeconomic performs better with only truncation, while some perform better with the addition of bias-correction. A possible explanation is that the truncation has effect in a different direction than the bias-correction. Truncation relies on the high correlation between different y_i, while bias-correction relies on the consistency and predictability of the forecast error in a univariate set-

Table 15: Mean Squared Prediction Error (MSPE) of bias-corrected unemployment with different truncated value. The MSPE of the truncated value is given in the ratio of truncation to equal weight. Larger than 1 means truncated is worse, whereas smaller than 1 means truncated weight helps in reducing MSPE.

	UNEM			
	1 Year Horizon		2 Years Horizon	
Threshold	MSPE Ratio	MAPE Ratio	MSPE Ratio	MAPE Ratio
-∞	3.7993	1.4829	21.2312	3.0358
-5	3.7993	1.4829	2.5172	1.5149
-4.5	3.7993	1.4829	2.5172	1.5149
-4	3.7993	1.4829	2.5376	1.5293
-3.5	3.7993	1.4829	2.5287	1.5246
-3	3.3185	1.3959	2.3569	1.4439
-2.5	2.3783	1.1662	2.3491	1.4493
-2	0.9828	0.8032	2.3203	1.4382
-1.5	0.7950	0.7275	1.6368	1.2530
-1	0.5564	0.6864	1.3372	1.1226
-0.5	0.7253	0.8155	1.1141	1.0643
0	0.8736	0.9201	1.0327	1.0176

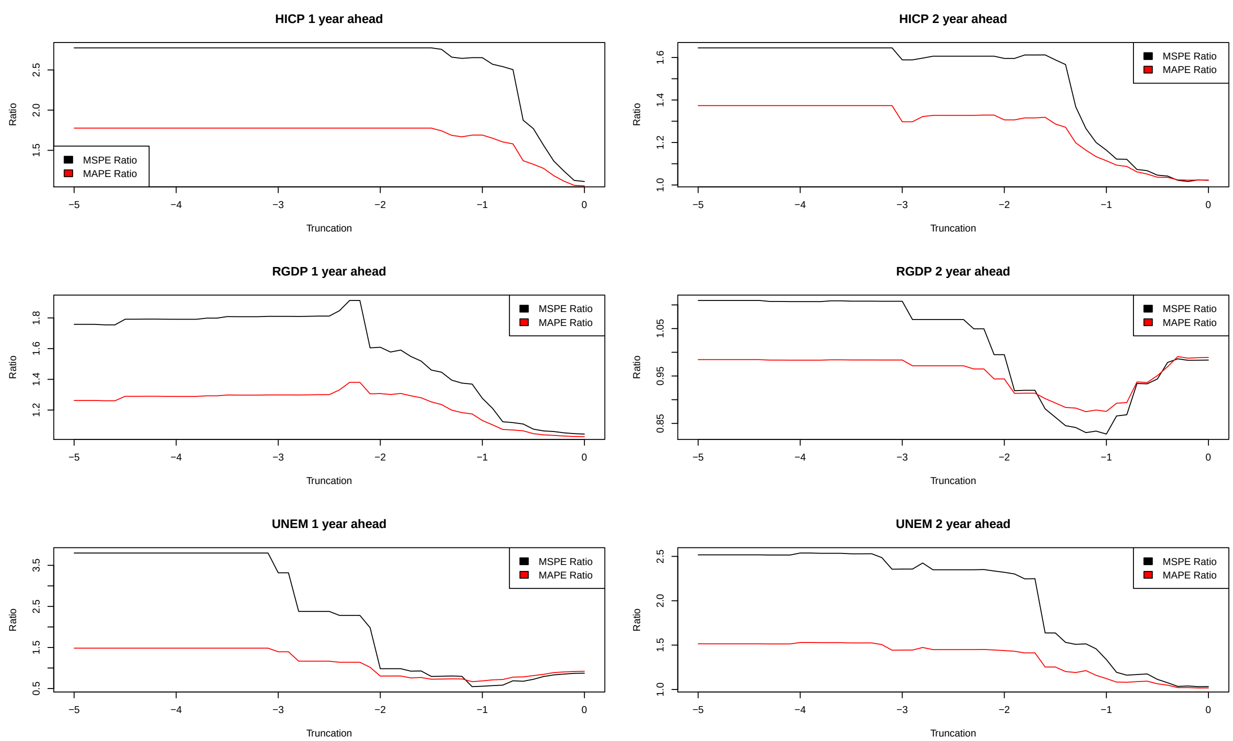

Figure 4: MSPE Ratio for different macroeconomic topics with bias correction

ting. This difference between two approaches explains the different effect between series. If the predictability of the forecast error is small, such that predicting the forecast increases the MSE, adjusting the bias does not help. Of course, the truncation and bias-correction are not mutually independent. An increase in the correlation increases the similarity of the estimation error. In turn, the estimation errors do not cancel out but magnifies.

4.7 Simulation

In this section, we show different effect on the underlying model influences the MSPE. For simplicity, we select four true weights and change the data generating model regarding correlation and error variance. We start with the true weight as $w = (-0.5, 0.3, 0, 1.2)'$, and proceed with correlation matrix where all off-diagonals are the same value. Covariance matrix follows by setting all variance to 1. The error term is generated with univariate random normal distribution, while the forecasts are generated by multivariate normal. We do not use biased forecast in this simulation. The simulation runs 10 observations per time for 2000 times for each correlation and error variance combination. Then the simulation takes the average of the MSPE.

In figure 5 different effects of correlation and error variance are shown. We select four different correlation, from 0.5 to 0.8, covering most of the correlation we observed in SPF. For the error variance, we show a range of selection from 0.4 to 0.9. Since the data variance is 1, the error variance quickly gives an idea on the fit of each forecast.

From figure 5, changes in the correlation and error variance does not influence the position where the MSPE drops or rises. Changes in error variance influence trivially on the level of MSPE. Higher error variance leads to higher MSPE. Changes in correlation show the location where the truncation effect converge to. For low correlation, 0.5 and 0.6, there is a distinct upward movement when truncation goes to 0 for all error variance. This forms the U-shape one expect when the bias increases. For high correlation, the effect of truncation does not increase the bias large enough, making the U-shape observable in only a few cases. Other cases with higher error variance show a continuous decline in MSPE.

Referring to section 4.6, the changes in the shape can be accounted for the additional estimation error caused by the bias term. In figure 5, it would be similar to plots of correlation 0.7 where error variance goes from 0.7 to 0.9.

For a U-shape to exist, e.g. equal weight is not the best, the correlation and error variance both plays an important role. Higher correlation only allows lower error when finding the optimal truncation parameter, while lower correlation also allows higher error to exist.

5 Conclusion and discussion

In this paper, we look into the possibility to use truncation to improve the forecast within forecast combination. Simple weight derived with a deterministic weight assumption often show bad empirical behaviour when compared to equal weight by arithmetic mean. Equal weight has a behaviour

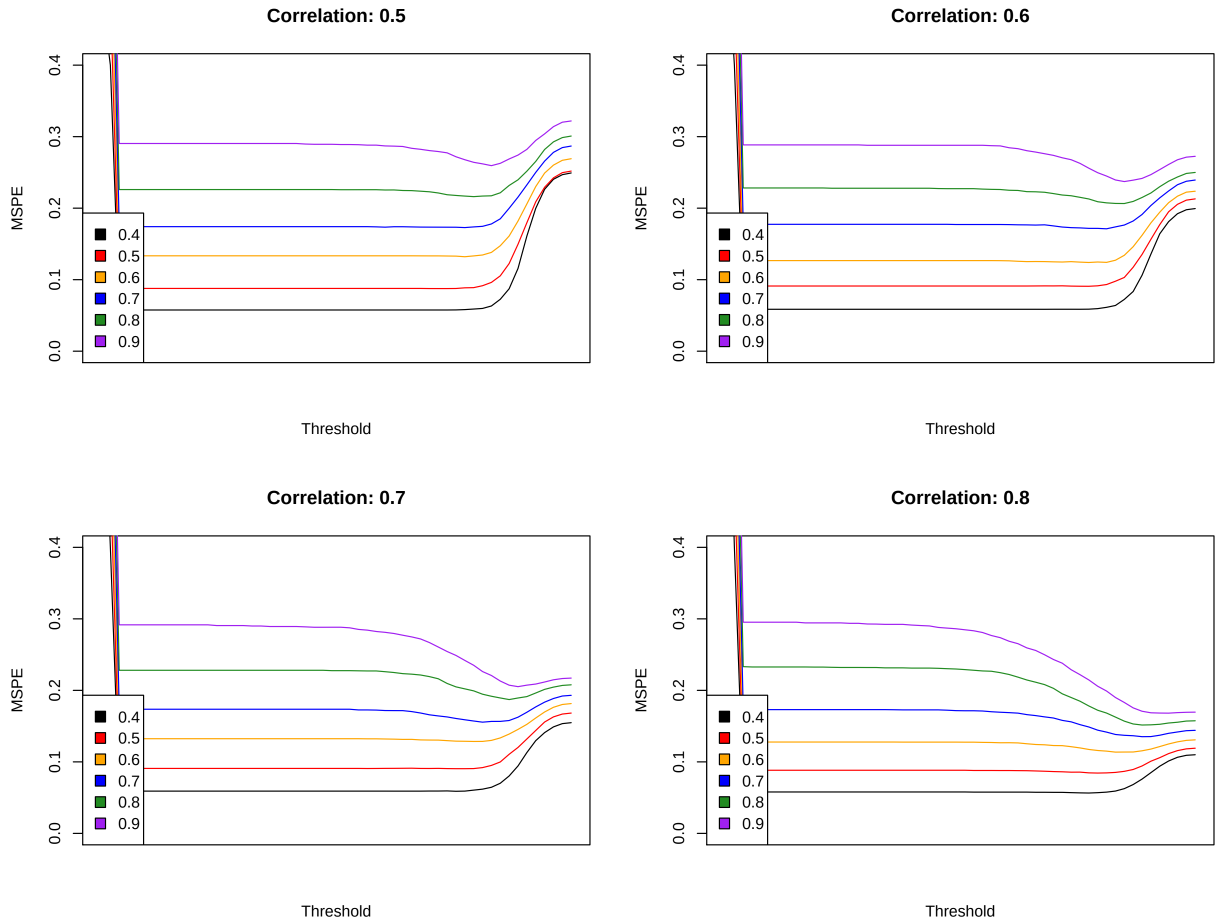

Figure 5: MSPE for different correlation and different error variance.

that the combined forecast cannot exceed the minimum or the maximum of all forecast. This limitation also holds for all weight selection method where the weight cannot be negative.

Using truncation, five out of six macroeconomic topics have a positive effect on the mean squared prediction error (MSPE) and mean absolute prediction error (MAPE) for the European Central Bank's (ECB) survey of professional forecasters (SPF). SPF is a quarterly survey sent to the private sector to understand their view on inflation (HICP), economic growth (RGDP), and unemployment (UNEM) in the next one and two year. The survey forecasts have a strong correlation within the forecasters, which we show that the true value often lies outside of the minimum/maximum range. Additionally, the increase in the correlation leads to an increase in the negative weight. Therefore the relaxation of no negative weight helps in this aspect.

By truncation, we set the weight below a certain threshold to 0, which removes the forecasts from the combination list. The results indicate a positive effect up to 30% decrease in MSPE compared to equal weight. This is positive news considering the amount of forecaster is relatively large compared to the sample size. With up to 120 forecasters and 60 observations, the weight estimation contains a large amount of estimation noise. This gives the estimated weight a wide variation, with a minimum of -21.79 and maximum of 16.34. On the other hand, the weights calculated by equal weight method are around 0.01.

The truncation effect holds for optimal threshold (OT) selection on the truncation parameter. The effect decreases from up to 30% to up to 12% decrease in MSPE. Contrary to the truncation where the parameter stays constant, OT selection allows the parameter to vary over time. This attributes to the increase in performance of RGDP with 2 years forecast horizon. In all six macroeconomic cases, we see an improvement in MSPE and MAPE when compared to equal weight. For the selected truncation, most of the selected parameter are below -1.5, which is well below the original -21.79. Setting the parameter below -1.5 limits the variation, and therefore decreases the variance and the estimation error.

Looking further into the weight selection, it is possible that the characteristic of SPF is dominated by a consistent bias over a certain forecaster. An example would be a case where the forecaster undervalues the growth and the decline, resulting in a forecast closer to 0 than the true value. Therefore we construct the bias correcting forecast weight and applies the truncation analysis on the corrected weight. The result is a mix bag of good and bad news. Four out of six macroeconomic topics are worse than equal weight, while without bias correction, we have only 1 truncation weight that is under-performing. There are still declines in the test statistics compared to the non-truncated ones. On the positive news, bias-correction improves the only macro topic that does not outperform equal weight without bias-correction. The MSPE decreases up to 17% compared with equal weight and MAPE up to 12 %. Another positive news is an additional decrease of 13% in UNEM 1 year ahead. The difference with and without bias correction could be in the direction of the changes of the weight. Truncation tries to extend the forecast upon the range by using negative weight, while bias-correction tries to identify the consistent bias within each forecaster.

We also study the different effect of correlation and estimation uncertainty on the shape of the MSPE. Correlation and estimation uncertainty both play important intertwined roles in the truncation. Higher correlation only allows lower error when finding the optimal truncation parameter, while lower correlation also allows higher error to exist. Since estimation uncertainty and correlation can be known before applying truncation, the selection of best truncation using that information is a possible future research topic.

We list a few other points in this paper that require further research. All analytic derivations are done in two forecasts scenarios, which can be extended to multivariate setting. RGDP 2 years ahead shows an unexpected shape of MSPE, with the reason to this shape unclear. Knowing the reason to spikes in the truncation selection is helpful as this can help in limiting variation. Combination of other techniques like shrinkage on the covariance to improve the estimates. Additionally, use the threshold value instead of 0 to the truncation can also change the effect of the method.

All in all, we demonstrate the ability to improve upon equal weight by truncation. The requirements for the improvement lies in the correlation and forecast uncertainty. When the conditions are met, for example, the SPF data, equal weight is no longer the prime choice in the forecast combination.

References

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *OR*, 20(4):451–468.

Chen, J., Hong, H., and Stein, J. C. (2000). Forecasting crashes: Trading volume, past returns and conditional skewness in stock prices. Working Paper 7687, National Bureau of Economic Research.

Claeskens, G., Magnus, J., Vasnev, A., and Wang, W. (2014). The forecast combination puzzle: A simple theoretical explanation. *SSRN Electronic Journal*.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.

Diebold, F. and Lopez, J. (1995). Forecast evaluation and combination. Research Paper 9525, Federal Reserve Bank of New York.

Dunis, C., Laws, J., and Chauvin, S. (2000). The use of market data and model combinations to improve forecast accuracy.

Elliott, G. and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, 122(1):47–79.

Gibbs, C. G. and Vasnev, A. L. (2017). Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts. Discussion Papers 2017-10, School of Economics, The University of New South Wales.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1.

Matsypura, D., Thompson, R., and Vasnev, A. L. (2018). Optimal selection of expert forecasts with integer programming. *Omega*, 78(C):165–175.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

Watson, M. W. and Stock, J. H. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.