Untitled

Daniel Hsiao

November 12, 2018

Contents

1 Methodology

1.1 Standard Approach

Consider the two variable case here for illustration 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 that we have $y_c = wy_1 + (1-w)y_2$. Assume they follow some distribution, 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 research had been done on. **add research**. Equation 2 has a few empirical results that are against this approach. Two common alternatives are diagonal covariance matrix and equal weights.

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 parameter 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 parameter to estimate reduces to n, which greatly decreases the estimation error. **add citation**

Equal weights is another common approach that works better empirically. **add citation** The forecast combination is in this case just an arithmetic mean of all forecasts. The reason behind this is the fact that estimating weights increases or shifts the forecast errors due to additional estimation error in the estimation of w. We laborate on the estimation error of w more later.

add more on equal weight

1.2 With estimation error of w

We can also consider the weight as non-deterministic, but related with y, e.g., in a trivairate distribution with finite third and fourth moments. Under trivariate distribution, the variance of the weights influences the expected value and the variance of the combined forecast. The expected value and the variance of the combined forecast becomes

$$
\begin{aligned}
E(y_c) =& \mu + (cov(w, y_1 - y_2))^2 \\
var(y_c) =& E(w)^2\sigma_1^2 + (1-E(w))^2\sigma_2^2 + 2E(w)(1-E(w))\rho\sigma_1\sigma_2 \\
& + E[(w-E(w))(y_1-y_2)(E(w)y_1 + (1-E(w))y_2 - \mu)] \\
& + E[(w-E(w))^2(y_1-y_2)^2] - cov(w, y_1-y_2)^2.
\end{aligned} \tag{4}
$$

Equation **??** 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 euqation 1 with $E[(w-E(w))(y_1-y_2)(E(w)y_1 + (1-E(w))y_2 - \mu)] + E[(w-E(w))^2(y_1-y_2)^2] - cov(w, y_1-y_2)^2$. **can we prove that the change in var(yc) is positive? I tried to prove it but got stuck at** $E[(w-E(w))(y_1-y_2)(E(w)y_1 + (1-E(w))y_2 - \mu)] + Var((w-Ew)(y_1-y_2)) > 0$. This case the only requirements are that the individual forecast to be unbaised and that the weights sum up to 1.

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

$$w^\dagger = w^*(1 + \frac{\sigma_{22}d_1 + \sigma_{11}d_2 - \sigma_{12}(d_1 + d_2)}{\sigma_{11}\sigma_{22} - 2\sigma_{12}}) - \frac{\sigma_{22}d_1 - \sigma{12}d_2}{\sigma_{11}\sigma_{22} - 2\sigma_{12}}. \tag{5}$$

The non-deterministic weight selection is a linear combination of the original weight. The non-deterministic weight does not change if the third moment is 0.

1.3 Negative weights

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

$$w = \frac{\sigma_2^2(-\rho\delta + (1 - \rho))}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}. \tag{6}$$

The numerator in w consist of σ_2^2 scaled with a weighted mean between $-\delta$ and 1 with weight ρ. When ρ is small, the weights is close to equation 3. When ρ is large, the negative difference in variance $-\delta$ takes over and results in negative weights. The level of negativity accounts for both the negtaive difference and correlation. The boundary case is

$$\rho = \frac{\sigma_2}{\sigma_1}, \tag{7}$$

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

From equation 5, 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

$$\sigma_{11}\sigma_{22} - 2\sigma_{12} = \sigma_2 2((1 + \delta)^2 - 2(1 + \delta)), \tag{8}$$

and the scaling factors

$$\begin{aligned} \frac{\sigma_{22}d_1 - \sigma{12}d_2}{\sigma_{11}\sigma_{22} - 2\sigma_{12}} &= \frac{d_1 - \rho(1 + \delta)d_2}{(1 + \delta)^2 - 2(1 + \delta)} \\ \frac{\sigma_{22}d_1 + \sigma_{11}d_2 - \sigma_{12}(d_1 + d_2)}{\sigma_{11}\sigma_{22} - 2\sigma_{12}} &= \frac{d_1(1 - \rho(1 + \delta)) + d_2((1 + \delta)^2 - \rho(1 + \delta))}{(1 + \delta)^2 - 2(1 + \delta))} \end{aligned} \tag{9}$$

and I'm lost in what I want to say

1.4 Truncated weights

To avoid the high correlated forecasts, we use truncation on the variable. The weight estimation is as follows ** insert latex equation of truncation** where the sum of the weights are rescaled to 1.

Assume that there is no skewness in the joint distribution, e.g., w^* is unbiased estimator of the true w. The expected value of the weights $\tilde{w}$ is

$$\begin{aligned} E(\tilde{w}) &= E(I_{w>c, w<1-c}w + I_{w>1-c}) \\ &= \int_c^{1-c} wf(w)\,dw + \int_c^{1-c} f(w)\,dw. \end{aligned} \tag{10}$$

The bias is therefore

$$E(\tilde{w}) - E(w^*) = \int_{-\infty}^c -wf(w)\,dw + \int_{1-c}^\infty (1 - w)f(w)\,dw, \tag{11}$$

3

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 negative value in $1 - w$. In general case where w can go above $1 - c$ or below c and the skewness of w is not 0, the bias is non-zero. This increases MSE to the estimated y_c.

The variance is

$$
\begin{aligned}
Var(\tilde{w}) &= E(\tilde{w}^2) - E(\tilde{w})^2 \\
&= \int_c^{1-c} w^2 f(w)\,\mathrm{d}w + \int_c^{1-c} f(w)\,\mathrm{d}w - (\int_c^{1-c} w f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} f(w)\,\mathrm{d}w)^2,
\end{aligned}
\tag{12}
$$

And the change in variance is

$$
\begin{aligned}
Var(\tilde{w}) - Var(w) &= \int_{-\infty}^{c} -w^2 f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} (1 - w^2) f(w)\,\mathrm{d}w \\
&\quad - (\int_c^{1-c} w f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} f(w)\,\mathrm{d}w)^2 + (\int_{-\infty}^{\infty} w f(w)\,\mathrm{d}w)^2.
\end{aligned}
\tag{13}
$$

By Cauchy–Schwarz inequality, and take $g(w)$ as the weight generating function of $\tilde{w}$, we have

$$
\begin{aligned}
Var(\tilde{w}) - Var(w) &\leq \int_{-\infty}^{c} -w^2 f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} (1 - w^2) f(w)\,\mathrm{d}w \\
&\quad - (\int_c^{\infty} g(w) f(w)\,\mathrm{d}w)^2 + \int_{-\infty}^{\infty} w^2 f^2(w)\,\mathrm{d}w.
\end{aligned}
\tag{14}
$$

Since $|g(w)| \leq |w|$ for all w, thus $g^2(w) \leq w^2$ for all w. Equation 14 becomes **rewrite starting here**

$$
\begin{aligned}
Var(\tilde{w}) - Var(w) &\geq \int_{-\infty}^{c} -w^2 f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} (1 - w^2) f(w)\,\mathrm{d}w \\
&\quad - \int_c^{\infty} w^2 f^2(w)\,\mathrm{d}w + (\int_{-\infty}^{\infty} w f(w)\,\mathrm{d}w)^2.
\end{aligned}
\tag{15}
$$

Since $f(w) \leq 1$ for all w, we have $f(w) \geq f^2(w)$ for all w. Rewrite into

$$
\begin{aligned}
Var(\tilde{w}) - Var(w) &\geq \int_{-\infty}^{\infty} -w^2 f(w)\,\mathrm{d}w + \int_{1-c}^{\infty} (1 - w^2) f(w)\,\mathrm{d}w \\
&\quad + (\int_{-\infty}^{\infty} w f(w)\,\mathrm{d}w)^2.
\end{aligned}
\tag{16}
$$

1.5   Bias Correction

2   Survey of Professional Forecasters (SPF)

To illustrate the empirical results, we use the data from ECB **(footnote link to data)** in this paper. The data, SPF, is a quarterly survey initiated by ECB, with the aim to obtain future estimates on inflation (HICP), RGDP and unemployment rate (UNEM) from the private sector. Every quarter, a gourp of professional forecasters from financial and non-financial instutition, such as economic research institutions, respond to the survey with the idea on the future economic. 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 indice.

The set up of the survey consist of multiple magnitudes of questions, ranging from different horizon to different distribution. The forecaststers are asked to provide their point forecast and the probability of a certain scenario to happen. The survey enables ECB to do quantitative assessment on the consensus of the

market, like the distribution statistics and standard deviations. For this paper, we take the 2 most answered time periods, which is 1 year ahead and 2 year ahead as our data set for all HICP, RGDP, and UNEM.

To compare the forecasts with the actual macroeconomics, we obtain the true value from ECB data base **(footnote link to data)**. The data cannot be observed from the economic in 100% accuracy within the first time frame, and exihibits changes to the initial estimates after revision. We use the final estimate of the macroeconomics where possible. The use of final estimate is fine is due to the fact that the original forecast is not the real target to be forecasts.

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 inline with **(ref paper)**.

Following **(ref paper)**, we calculate the covariance by looking at the intersection between each forecasters.

insert equation

When there are no intersection between 2 forecasters, we set the covariance value to 0. 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.

$$\rho_{i,j} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \tag{17}$$

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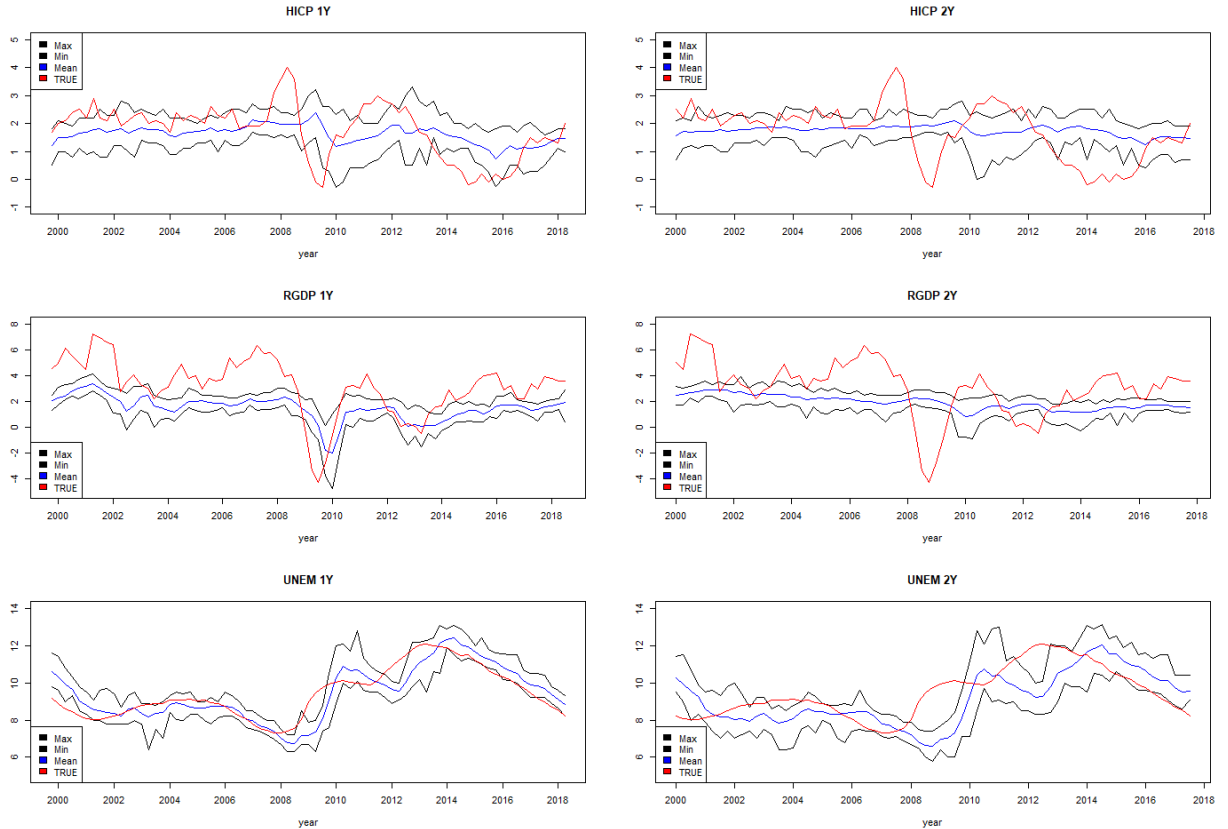

Figure 1: Survey of Professional Forecasters data illustration

\begin{table} \caption{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.}

Macro topic	HICP		RGDP		UNEM	
Horizon	1 year	2 year	1 year	2 year	1 year	2 year
min	0.03	0.02	0.11	0.13	-0.02	0.08
1st Q	0.53	0.62	0.63	0.65	0.47	0.56
median	0.66	0.72	0.81	0.8	0.62	0.7
mean	0.64	0.7	0.75	0.75	0.59	0.67
3rd Q	0.77	0.81	0.89	0.87	0.74	0.8
max	0.96	0.97	0.98	0.98	0.94	0.95

\end{table}

In figure 1 and table 2 we show the plots of the forecasts along side with the true value in the macroeconomics and the statistics of the covariance of the forecast error. To avoid too many lines on the figure by plotting all forecasts, 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 lies outside of the forecast range, with RGDP the worse of all three. More values outside of the forecast range suggest that restricting positive weights may be a strong limitation in the forecast combination. We expect to have large effect using truncation in the forecast of RGDP, while HICP and UNEM does not have too strong effect. We also expect the two year ahead forecast will be better than the one year ahead.

The table 2 tell us on how the forecast error are correlated. The correlation are split up into different forecast topics and different forecast horizons. Diagonal element of the correlation matrix is not within the correlation 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 than the minimum correlation in the other two topics.

add more explanation

3   Empirical Results

3.1   Proceure

The procedure to get the weights can be different between different researcher. Therefore we provide a through order of the steps we take in the weight estimation. We do the weights estimation for the truncation using the following steps:

1. Estimate the simple weight w^* from equation 2.