Untitled

Daniel Hsiao

November 12, 2018

Contents

1 Methodology

1.1 Standard Approach

Consider the two variable case. 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

1.2 Truncation Approach

1.3 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.

$$\rho_{i,j} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \tag{3}$$

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

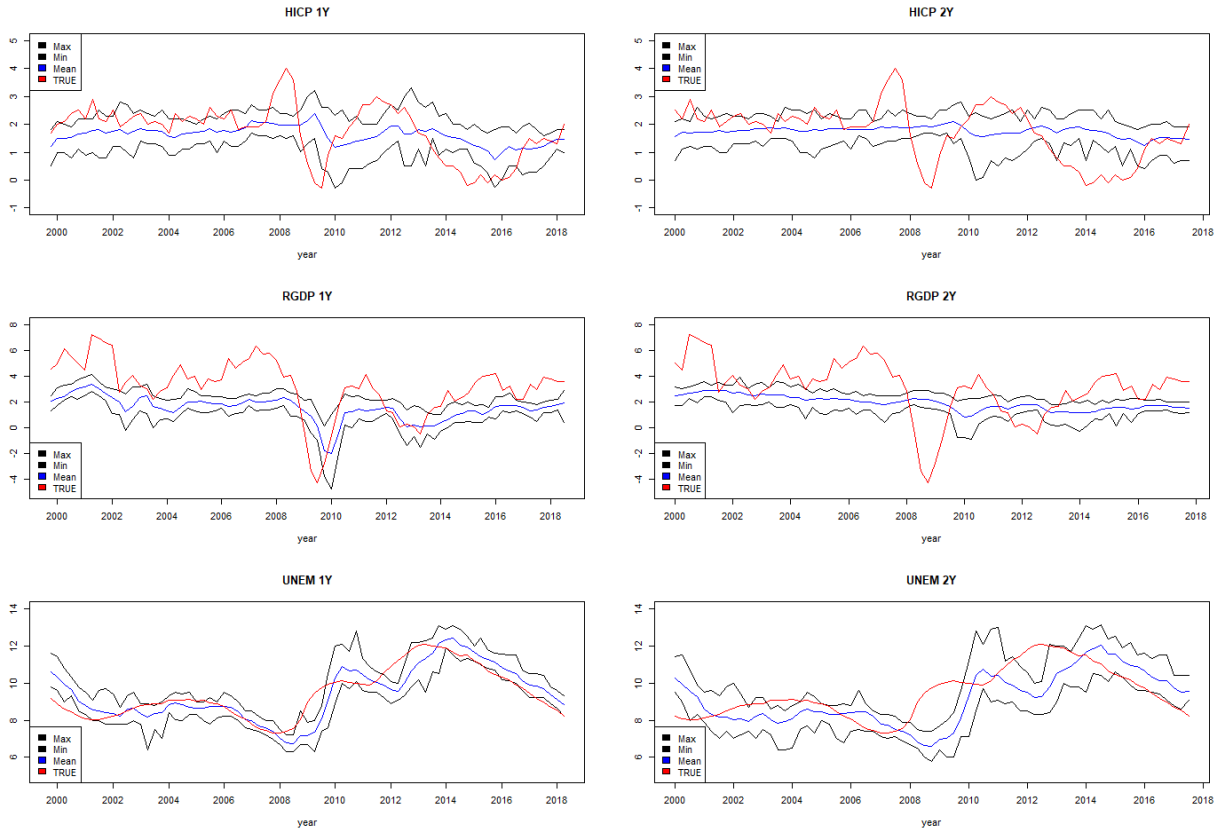

Figure 1: Survey of Professional Forecasters data illustration

\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 Figure 1 and table Figure 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 2 year ahead stronger than 1 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 2 year ahead forecast will be better than the 1 year ahead.

The table Figure 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