Forecast combinations in R using the *ForecastCombinations* package

Applied Finance with R

Eran Raviv

May 21, 2016

apg
tomorrow is today

Talk overview

- Introduction
- Implemented combination schemes
- Some practical examples
- Discussion and takeaways
- Credits and references

Some targets are easy to forecast

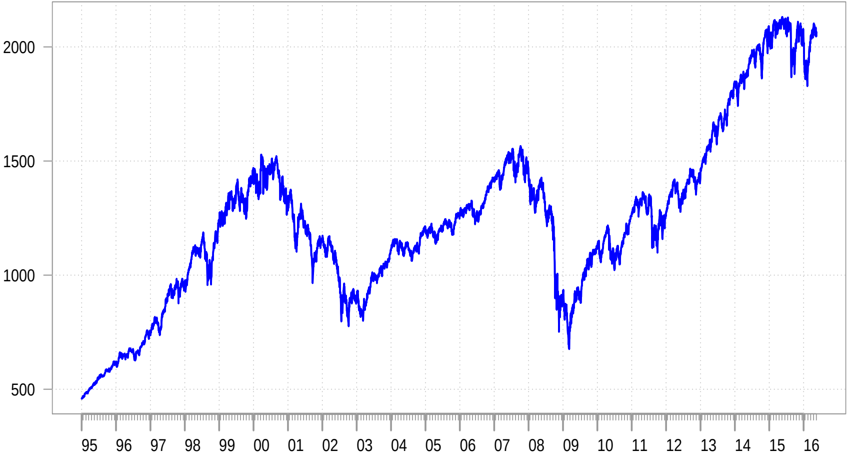
Solar eclipse of March 20, 2015. Source: Wikipedia

Some targets are easy to forecast

Solar eclipse of March 20, 2015. Source: Wikipedia

Next eclipse: **August 12, 2026**

Some.. evidence is mixed



S&P 500. Source: Bloomberg

Introduction

- What is it?

$$f^{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \quad (1)$$

Introduction

- What is it?

$$f^{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \quad (1)$$

- Why is it?

Because it works

Introduction

- What is it?

$$f^{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \quad (1)$$

- Why is it?

  Because it works

- And why is that?

  More research is needed

Introduction

- What is it?

$$f^{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \quad (1)$$

- Why is it?

  Because it works

- And why is that?

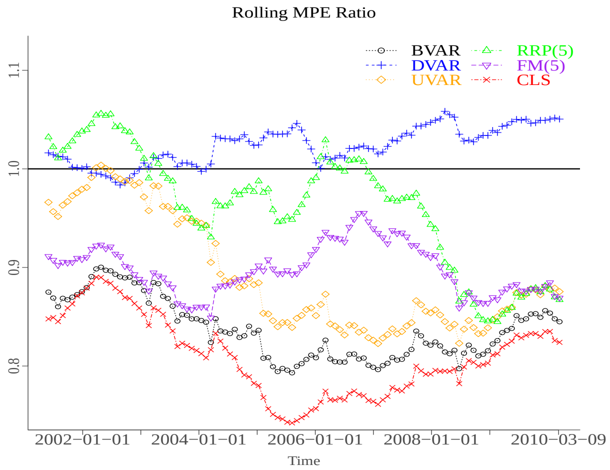
  More research is needed

- What is the intuition?

  Biases
  Model risk
  Structural breaks

Different models perform differently

under different conditions, and/or in different points in time:



Electricity price forecasting, accuracy of different models. Source: Author’s calculation

The thing is, going forward, we don't know which forecasting model will outperform

- So, as we don't bet on the **one** horse in investments, we don't bet on the **one** horse here neither

That is the idea, but how to combine?

Regression based (OLS)

Train the individual forecasts using:

$$y_t = \alpha + \sum_{i=1}^{P} \beta_i f_{i,t} + \varepsilon_t, \tag{2}$$

The combined forecast is then given by

$$f^c = \widehat{\alpha} + \sum_{i=1}^{P} \widehat{\beta}_i f_i, \tag{3}$$

+'s:

- OLS (optimality)
- Flexibility, unconstrained

-'s:

- Flexibility, unconstrained
- Interpretation

Regression based (LAD)

Train the individual forecasts using:

$$y_t = \alpha + \sum_{i=1}^{P} \beta_i f_{i,t} + \varepsilon_t, \tag{4}$$

But minimise the absolute loss function

$$\sum_t |\varepsilon_t|$$

instead of the squared loss function

$$\sum_t {\varepsilon_t}^2$$

- If the cost of missing the target is not very high, this combination scheme may be preferred

Regression based (CLS)

Train the individual forecasts using:

$$y_t = \alpha + \sum_{i=1}^{P} \beta_i f_{i,t} + \varepsilon_t. \tag{5}$$

Minimise the squared loss function:

$$\sum_i {\varepsilon_t}^2,$$

but under additional constraints:

- $\beta_i \geq 0, \; \forall i,$ or
- $\sum_{i=1}^{P} \beta_i = 1,$ or both
- Lacks optimality properties
- Works very well, especially when correlation between individual forecasts is high
- Better interpretability

Accuracy-based (Inverse MSE)

Use some accuracy measure, for example mean squared error (MSE):

$$\mathrm{MSE_i} = \frac{1}{T}\sum_{t=1}^{T}(f_{i,t} - y_t)^2,$$

and combine the forecasts based on how well each individual is doing:

$$f^c = \frac{\left(\frac{MSE_i}{\sum_{i=1}^{P} MSE_i}\right)^{-1}}{\sum_{i=1}^{P}\left(\frac{MSE_i}{\sum_{i=1}^{P} MSE_i}\right)^{-1}} f_i = \frac{\frac{1}{MSE_i}}{\sum_{i=1}^{P}\frac{1}{MSE_i}} f_i. \tag{6}$$

- When individual forecasts are highly correlated this is not much different than the simple average
- You can tailor the accuracy measure

Accuracy-based (Inverse MSE)

Use some accuracy measure, for example mean squared error (MSE):

$$\mathrm{MSE_i} = \frac{1}{T}\sum_{t=1}^{T}(f_{i,t} - y_t)^2,$$

and combine the forecasts based on how well each individual is doing:

$$f^c = \frac{\left(\frac{MSE_i}{\sum_{i=1}^{P} MSE_i}\right)^{-1}}{\sum_{i=1}^{P}\left(\frac{MSE_i}{\sum_{i=1}^{P} MSE_i}\right)^{-1}} f_i = \frac{\frac{1}{MSE_i}}{\sum_{i=1}^{P}\frac{1}{MSE_i}} f_i. \quad (6)$$

- When individual forecasts are highly correlated this is not much different than the simple average
- You can tailor the accuracy measure
- Since 2004, aggregated forecast through exponential re-weighting method (AFTER)

Best individual (BI)

Basically (ex-post) model selection

$$f^c = w_i f_i, \quad \text{where} \quad \begin{array}{ll} w_i = 1 & \text{if } MSE_i < MSE_{-i} \quad \forall i \in \{1, \ldots, P\} \\ w_i = 0 & \text{otherwise} \end{array} \tag{7}$$

- Very restrictive

Best individual (BI)

Basically (ex-post) model selection

$$f^c = w_i f_i, \quad \text{where} \quad \begin{array}{ll} w_i = 1 & \text{if } MSE_i < MSE_{-i} \quad \forall i \in \{1, \ldots, P\} \\ w_i = 0 & \text{otherwise} \end{array} \tag{7}$$

- Very restrictive
- Easy to explain

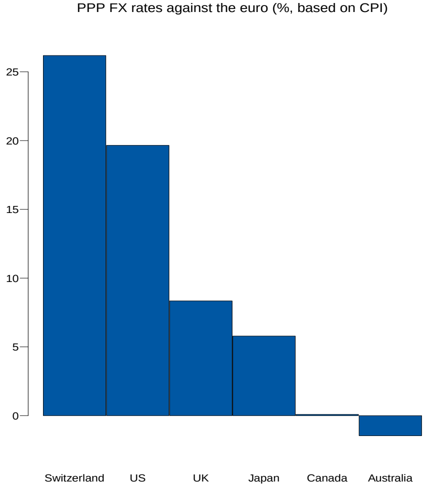
Best individual (BI)

Basically (ex-post) model selection

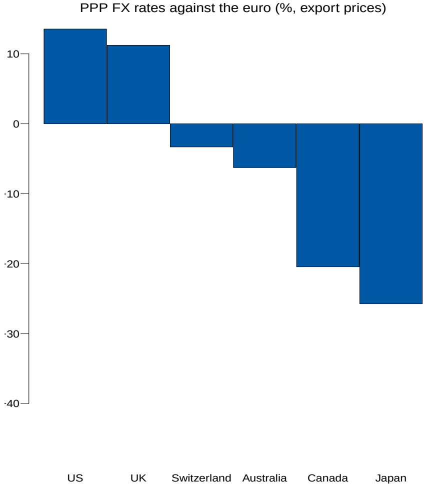
$$f^c = w_i f_i, \quad \text{where} \quad \begin{array}{ll} w_i = 1 & \text{if } MSE_i < MSE_{-i} \quad \forall i \in \{1, \ldots, P\} \\ w_i = 0 & \text{otherwise} \end{array} \tag{7}$$

- Very restrictive
- Easy to explain
- Don't dismiss it beforehand

PPP estimation



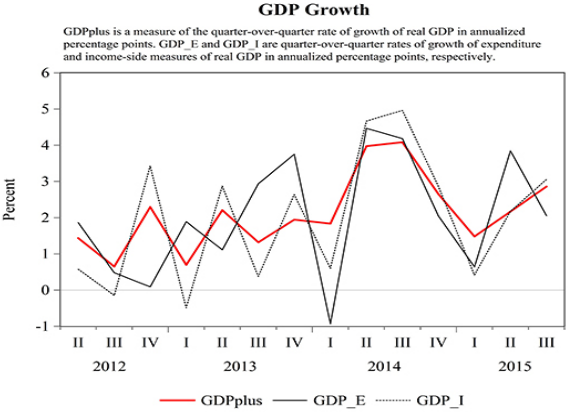
Source: Bloomberg, author's calculation



Source: Bloomberg, author's calculation

GDP measurements

“The current system emphasizes data on spending, but the bureau also collects data on income. In theory the two should match perfectly - a penny spent is a penny earned by someone else. But estimates of the two measures can diverge widely” [Aruoba et al., 2015]



Source: Fed of Philadelphia

Discussion

- Paper here
- Popular (across estimation window; bootstrapping; rolling vs expanding and more). Research is still going strong
- You, as well, are using it already:

$$D_t = (1-\lambda)\sum_{t=1}^{\infty}\lambda^{t-1}(\varepsilon_{t-1}\varepsilon'_{t-1}) = (1-\lambda)(\varepsilon_{t-1}\varepsilon'_{t-1}) + \lambda D_{t-1}, \quad (8)$$

Discussion

- Paper here
- Popular (across estimation window; bootstrapping; rolling vs expanding and more). Research is still going strong
- You, as well, are using it already:

$$D_t = (1-\lambda)\sum_{t=1}^{\infty}\lambda^{t-1}(\varepsilon_{t-1}\varepsilon'_{t-1}) = (1-\lambda)(\varepsilon_{t-1}\varepsilon'_{t-1}) + \lambda D_{t-1}, \quad (8)$$

- Some research ideas I did not get around to:
  - Different regimes
  - Dynamic model averaging

Why not use it?

Why not use it?

- Interpretation is often lost

Why not use it?

- Interpretation is often lost
- Does not always add value (garbage in $\Longrightarrow$ garbage out)
- Especially when you have a one strong dominant model to begin with

Why use it?

- Good “hedge” against wrong modelling choices

Why use it?

- Good “hedge” against wrong modelling choices
- No consensus on a ’best’ approach.
- Simple average is very **robust**
- Combining models eliminate the need to choose, which can be a very good thing

Why use it?

- Good “hedge” against wrong modelling choices
- No consensus on a ’best’ approach.
- Simple average is very **robust**
- Combining models eliminate the need to choose, which can be a very good thing
- Useful in changing environment where structural breaks are likely

References

[De Pooter et al., 2010]: Forecast combination with application to yield curve forecasting

[Cheng and Yang, 2015]: Forecast combination with outlier protection

[Pesaran and Pick, 2011]: Across estimation windows (PPP style)

[Fawcett et al., 2013]: Generalised density forecast combinations (the word “Generalised” here can be replaced by “time-varying”)

[Hyndman et al., 2011]: Reconciling disaggregated forecasts with aggregated forecasts

[Casarin et al., 2015]: Predictive density combination

[Götz et al., 2016]: Forecasts based on different vintages of macro data

[Taieb and Hyndman, 2012]: Combining recursive and direct forecasts

[Jore et al., 2010]: Combining forecast densities from VARs

[Opschoor et al., 2014]: For Value-at-Risk estimates

[Liu, 2015]: Taking on the theory of the least squares averaging estimator

[Yang, 2004]: Exponential re- weighting method (AFTER)

Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., and Song, D. (2015).
Improving measurement: A measurement-error perspective.
Journal of Econometrics.

Casarin, R., Grassi, S., Ravazzolo, F., and van Dijk, H. K. (2015).
Dynamic predictive density combinations for large data sets in economics and finance.

Cheng, G. and Yang, Y. (2015).
Forecast combination with outlier protection.
International Journal of Forecasting, 31(2):223–237.

De Pooter, M., Ravazzolo, F., and Van Dijk, D. J. (2010).
Term structure forecasting using macro factors and forecast combination.
FRB International Finance Discussion Paper, (993).

Fawcett, N., Kapetanios, G., Mitchell, J., and Price, S. (2013).
Generalised density forecast combinations.

Götz, T. B., Hecq, A., and Urbain, J.-P. (2016).
Combining forecasts from successive data vintages: An application to us growth.
International Journal of Forecasting, 32(1):61–74.

Hyndman, R., Ahmed, R., Athanasopoulos, G., and Shang, H. (2011).
Optimal combination forecasts for hierarchical time series.
Computational Statistics & Data Analysis, 55(9):2579–2589.

Jore, A. S., Mitchell, J., and Vahey, S. P. (2010).
Combining forecast densities from vars with uncertain instabilities.
Journal of Applied Econometrics, 25(4):621–634.

Liu, C.-A. (2015).
Distribution theory of the least squares averaging estimator.
Journal of Econometrics, 186(1):142–159.

Opschoor, A., Van Dijk, D. J., and Van der Wel, M. (2014).
Improving density forecasts and value-at-risk estimates by combining densities.

Pesaran, M. and Pick, A. (2011).
Forecast combination across estimation windows.
Journal of business and economic statistics, 29(2):307–318.

Taieb, S. and Hyndman, R. (2012).
Recursive and direct multi-step forecasting: the best of both worlds.

Yang, Y. (2004).
Combining forecasting procedures: some theoretical results.
Econometric Theory, 20(01):176–222.