Forecast combinations in R using the ForecastCombinations package

Applied Finance with R

Eran Raviv
May 21, 2016
Talk overview

♦ Introduction
♦ Implemented combination schemes
♦ Some practical examples
♦ Discussion and takeaways
♦ Credits and references
Some targets are easy to forecast

Some targets are easy to forecast


Next eclipse: **August 12, 2026**
Some evidence is mixed
Introduction

♦ What is it?

\[ f_{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \] (1)
Introduction

♦ What is it?

\[ f_{combined} = \frac{\sum_{i=1}^{P} f_i}{P} \]  \hspace{1cm} (1)

♦ Why is it?

*Because it works*
Introduction

♦ What is it?

\[ f_{\text{combined}} = \frac{\sum_{i=1}^{P} f_i}{P} \]  

(1)

♦ Why is it?

Because it works

♦ And why is that?

More research is needed
Introduction

♦ What is it?

\[ f_{\text{combined}} = \sum_{i=1}^{P} f_i \]  

(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

\[ f^c = \hat{\alpha} + \sum_{i=1}^{P} \hat{\beta}_i f_i, \tag{3} \]

+’s:
- OLS (optimality)
- Flexibility, unconstrained

-’s:
- Flexibility, unconstrained
- Interpretation
Regression based (LAD)

Train the individual forecasts using:

\[ y_t = \alpha + \sum_{i=1}^{P} \beta_i f_{i,t} + \varepsilon_t, \] (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:

\[ y_t = \alpha + \sum_{i=1}^{P} \beta_i f_{i,t} + \varepsilon_t. \]  

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):

$$\text{MSE}_i = \frac{1}{T} \sum_{t=1}^{T} (f_{i,t} - y_t)^2,$$

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

$$f^c = \frac{\left( \frac{\text{MSE}_i}{\sum_{i=1}^{P} \text{MSE}_i} \right)^{-1}}{\sum_{i=1}^{P} \left( \frac{\text{MSE}_i}{\sum_{i=1}^{P} \text{MSE}_i} \right)^{-1}} f_i = \frac{1}{\sum_{i=1}^{P} \frac{1}{\text{MSE}_i}} f_i.$$  \hspace{1cm} (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):

$$\text{MSE}_i = \frac{1}{T} \sum_{t=1}^{T} (f_{i,t} - y_t)^2,$$

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

$$f_c = \frac{1}{\sum_{i=1}^{P} \left( \frac{\text{MSE}_i}{\sum_{i=1}^{P} \text{MSE}_i} \right)^{-1}} \left( \frac{\text{MSE}_i}{\sum_{i=1}^{P} \text{MSE}_i} \right)^{-1} f_{i} = \frac{1}{\sum_{i=1}^{P} \frac{1}{\text{MSE}_i}} f_{i}.$$  

\[ (6) \]

♦ 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

\[ f^c = w_i f_i, \quad \text{where} \quad w_i = 1 \quad \text{if} \quad MSE_i < MSE_{-i} \quad \forall i \in \{1, \ldots, P\} \]

\[ w_i = 0 \quad \text{otherwise} \]

♦ Very restrictive
Best individual (BI)

Basically (ex-post) model selection

\[ f^c = w_if_i, \quad \text{where} \]
\[ w_i = 1 \quad \text{if } MSE_i < MSE_{-i} \quad \forall i \in \{1, \ldots, P\} \]
\[ w_i = 0 \quad \text{otherwise} \]

- Very restrictive
- Easy to explain
Best individual (BI)

Basically (ex-post) model selection

\[ f^c = w_i f_i, \quad \text{where} \quad w_i = \begin{cases} 1 & \text{if } MSE_i < MSE_{-i} \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in \{1, \ldots, P\} \quad (7) \]

♦ Very restrictive
♦ Easy to explain
♦ Don’t dismiss it beforehand
PPP estimation

PPP FX rates against the euro (%, based on CPI)

Source: Bloomberg, author’s calculation

PPP FX rates against the euro (%, export prices)

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]
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
<table>
<thead>
<tr>
<th>Background</th>
<th>How</th>
<th>Practical Examples</th>
<th>Discussion</th>
<th>Some selective references</th>
</tr>
</thead>
</table>

**Why not use it?**

- Interpretation is often lost
- Does not always add value (garbage in ⇒ garbage out)
- Especially when you have a strong dominant model to begin with
Why not use it?

♦ Interpretation is often lost
Why not use it?

♦ Interpretation is often lost
♦ Does not always add value (garbage in $\rightarrow$ 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)

DISCLOSURE
The content present is solely the responsibility of the presenter, and should not be interpreted as related to the views of APG or APG-AM.
<table>
<thead>
<tr>
<th>Background</th>
<th>How Practical Examples</th>
<th>Discussion</th>
<th>Some selective references</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background</strong></td>
<td><strong>How</strong></td>
<td><strong>Practical Examples</strong></td>
<td><strong>Discussion</strong></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
<td>-----------------------</td>
<td>----------------</td>
</tr>
</tbody>
</table>