Workshop on Central Bank Forecasting

Discussion of "Nowcasting Norwegian GDP in Real-Time: A Density Combination Approach" by K.A. Aastveit, K.R. Gerdrup, A.S. Jore and L.A. Thorsrud

Kenneth F. Wallis

Emeritus Professor of Econometrics, University of Warwick
http://go.warwick.ac.uk/kfwallis

Federal Reserve Bank of Kansas City *14-15 October 2010*

1. Introduction

This paper is part of an impressive programme of work at Norges Bank (NB), which can be seen as an extension to density forecasts of the large-scale point forecast combination studies by Stock and Watson and others over the last decade.

It introduces a two-stage or two-step combination method, in a practical forecasting environment, which raises associated issues of real-time forecasting, nowcasting, data revisions and the "ragged edge" (Wallis, 1986), all of which are dealt with very well.

Forecast combination in general is subject to the "inefficiency of mean forecasts" noted by Granger (1989) – "aggregating forecasts is not the same as aggregating information sets" – but the latter is infeasible given the relative size of k and T.

My comments relate to the authors' terminology, density forecast combination methods, weighting schemes, and calibration tests.

2. Terminology

(**Added note**. In the presentation this slide was left blank. My discussion had been prepared from the preliminary version of the paper, and included my objections to the authors' use of the word "ensemble" to describe their two-stage combination procedure, on the grounds that "ensemble" is already in use in the (mostly meteorological) forecasting literature, where it means something quite different from density forecast combination. To use it here would add confusion, and distract attention from the originality of the authors' work. I had put these points to the authors, and when the revised version of the paper arrived I saw that they had been accepted, since the word in question no longer appears in the paper. However this was at very short notice, and in the time available the simplest solution was to remove this part of the discussion, and offer a brief word of explanation on the day.)

3. Three ways of obtaining a combined density forecast

Given: N individual density forecasts $f_j(y),\ j=1,\ldots,N,$ at some time, horizon, etc.

3.1 Linear combination (Wallis, 2005; Zarnowitz, 1969)

Finite mixture distribution: $f_C(y) = \sum_{j=1}^{N} w_j f_j(y)$ where $w_j \geq 0,\ j=1,\ldots,N,\ \Sigma w_j = 1.$

If $f_j(y) = \mathcal{N}\left(\mu_j, \sigma_j^2\right),\ j=1,\ldots,N,$ then $f_C(y)$ is a mixture of normals (Pearson, 1894).

For any distributional forms, the mean and variance of the combined density are

$$\mu_C = \sum_{j=1}^{N} w_j \mu_j \quad \text{and} \quad \sigma_C^2 = \sum_{j=1}^{N} w_j \sigma_j^2 + \sum_{j=1}^{N} w_j \left(\mu_j - \mu_C\right)^2 .$$

In the case of equal weights, $w_j = 1/N$, the variance equation can be interpreted as:

aggregate uncertainty = average individual uncertainty + disagreement.

3.2 Logarithmic combination (Bacharach, 1972)

This is usually written in geometric form: $f_G(y) = \dfrac{\prod f_j^{w_j}(y)}{\int \prod f_j^{w_j}(y)dy}$

If $f_j(y) = \mathcal{N}\left(\mu_j, \sigma_j^2\right),\ j = 1,\ldots,N,$ then $f_G(y) = \mathcal{N}\left(\mu_G, \sigma_G^2\right),$ where

$$\frac{\mu_G}{\sigma_G^2} = \sum_{j=1}^{N} w_j \frac{\mu_j}{\sigma_j^2} \text{ and } \frac{1}{\sigma_G^2} = \sum_{j=1}^{N} w_j \frac{1}{\sigma_j^2}\ .$$

In the case of equal weights:

- σ_G^2 is the harmonic mean of the individual variances; this is less than their arithmetic mean, which is less than the finite mixture variance σ_C^2;

  (other NB work finds log combinations better probabilistically calibrated than linear)

- μ_G is the linear combination of the μ_j with inverse variance weights.

3.3 A more direct approach

Each component density forecast is obtained by centering a normal distribution on the point forecast and setting its variance equal to the MSE of past point forecasts.

The same method could be adopted for the final combined density forecast: center a normal distribution on the combined point forecast and set its variance as above.

If normality is “too good not to be true” at the level of the individual forecasting models, why abandon it at the last?

Log combination preserves normality, but this direct approach saves much computer time!

Note that the estimated weights used for combining density forecasts are comparable to those used for combining point forecasts ...

4. Weighting schemes

For a normal distribution, the expected log score is $E\left\{\ln f(y)\right\} = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}$.

For a normal density forecast $f_j(y)$ with DGP $f_o(y)$ we have

$$E_o\left\{\ln f_j(y)\right\} = -\frac{1}{2}\ln\left(2\pi\sigma_j^2\right) - \frac{1}{2}\frac{\sigma_o^2}{\sigma_j^2} - \frac{\left(\mu_o - \mu_j\right)^2}{2\sigma_j^2}.$$

The log score weights in the paper are based on the relative values of

$$\exp\sum_t \ln f_{jt}\left(y_t\right), \quad j = 1, \ldots, N,$$

whose ingredients are the same as the sample point forecast mean square error.

So it would be expected that the estimated combination weights for (a) point forecasts, and (b) density forecasts centered on those point forecasts, are rather similar.

But estimated weights have large sampling errors.

This explains the “forecast combination puzzle” (Smith and Wallis, 2009): in many empirical examples the optimal weights are close enough to equality that the estimation error dominates the effect of the bias that results from assuming equal weights.

Intermediate methods between simple averages and estimated weights have appeared:

- shrinkage of estimated weights towards equality
- trim the worst forecasts then use equal weights (Granger and Jeon, 2004)
- clustering of forecasts (Aiolfi and Timmermann, 2006)
- two-stage combining (NB, in the present paper and other work)

To date, the first three of these methods relate to point forecasting, where efficiency gains have been shown to be available from different approaches to combination.

Investigation of the same question for density forecasting would be worthwhile.

5. Calibration tests

An informal check of probabilistic calibration is to look for uniformity of PIT histograms.

Most of the histograms shown have a hump-shaped appearance, suggesting that the forecast densities are too dispersed, although apparently not significantly so.

Table 2 reports chi-square tests for histograms with 8 classes, which is possibly too many with only 33 observations (although it allows a decomposition which might be pursued).

Figure 7 uses 5 classes, and my chi-square tests of these histograms give more favorable results than those in Table 2, with $p>0.10$. However the chi-square distribution is a large-sample approximation, and exact finite-sample p-values could be obtained via *StatXact.*

But Figure 7 also shows a tendency towards skewness in the data, with not enough outcomes falling in the upper tail of the symmetric forecast densities.

Complete calibration in addition requires independence of the PITs of one-step-ahead forecasts, or of the inverse normal transforms of the PITs as used in the Berkowitz test.

Two observations on the Berkowitz test:

- for a normal density forecast, the double transformation returns the standardized point forecast error $z_t = \left(y_t - \hat{y}_t\right)/\hat{\sigma}_t$, say, so a test of the significance of the lagged dependent variable in the Berkowitz regression is equivalent to a test of the first-order autocorrelation coefficient of the point forecast errors;

- the test uses a normal likelihood, so the maintained hypothesis is $z_t \sim \mathcal{N}\left(\mu, \sigma^2, \rho_1\right)$ and the null hypothesis is $z_t \sim \mathcal{N}\left(0,1,0\right)$. To test the distributional form, it is possible to specify an alternative distribution which nests the normal distribution, and include the additional restrictions that reduce it to normality among the hypotheses under test.

Hence complications arising from both the autocorrelation structure and the distributional assumptions may be affecting the performance of the test and delivering the conflicting results in Table 2.

Additional references in the discussion

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *J. Econometrics*, 135, 31-53.

Bacharach, M. (1972). Scientific disagreement. Unpublished ms., Christ Church, Oxford.

Granger, C.W.J. (1989). Combining forecasts – twenty years later. *J. Forecasting*, 8, 167-173.

Granger, C.W.J. and Jeon, Y. (2004). Thick modeling. *Economic Modelling*, 21, 323-343.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, A, 185, 71-110.

Smith, J. and Wallis, K.F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71, 331-355.

Wallis, K.F. (1986). Forecasting with an econometric model: the 'ragged edge' problem. *J. Forecasting*, 5, 1-13.

Wallis, K.F. (2005). Combining interval and density forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, 67, 983-994.

Zarnowitz, V. (1969). The new ASA-NBER survey of forecasts by economic statisticians. *American Statistician*, 23(1), 12-16.