# Combining Density and Interval Forecasts: A Modest Proposal*

**Kenneth F. Wallis**

Department of Economics
University of Warwick
Coventry CV4 7AL, UK
[K.F.Wallis@warwick.ac.uk]

**October 2004**

**Abstract** The finite mixture distribution is proposed as an appropriate statistical model for a combined density forecast. Its implications for measures of uncertainty and for combining interval forecasts are described. Related proposals in the literature and applications to the U.S. Survey of Professional Forecasters are discussed.

***With apologies to Jonathan Swift (1667-1745). Initial work was done during a visit to the program on Econometric Forecasting and High-Frequency Data Analysis at the Institute for Mathematical Sciences, National University of Singapore, jointly organised by the School of Economics and Social Sciences, Singapore Management University, May 2004; the kind hospitality of both institutions is gratefully acknowledged.

## 1. Introduction

Different forecasts of the same event exist whenever forecasters have different information sets or process given information in different ways. The idea that combining such forecasts might be worthwhile rests on the recognition that the combination of forecasts implicitly pools the information sets, and additional information is almost always helpful. The basic idea applies to any kind of forecasts, and has spawned an extensive literature, but this is almost exclusively concerned with point forecasts of random variables. The review and annotated bibliography by Clemen (1989), for example, mostly considers combinations of point forecasts, noting that an issue still deserving attention is the robustness of the simple average of forecasts. Likewise Granger (1989), in his personal commentary twenty years on from the seminal article of Bates and Granger (1969), mostly considers "developments of the basic idea" insofar as they concern point forecasts. Both articles go beyond this central case to some extent: Clemen considers some event probability forecasting applications; Granger suggests approaches to combining interval and density forecasts. His suggestions have had little impact on subsequent practice, however, and possible reasons for this are discussed below. The general forecast combination literature has maintained its focus on point forecasts, as indicated in several contributions to the compendium edited by Clements and Hendry (2002), and this paper aims to widen the focus to include interval and density forecasts.

The relative neglect of combined density forecasts is surprising once it is recalled that they appeared in the original article on the U.S. Survey of Professional Forecasters (SPF), then known as the ASA-NBER Survey (Zarnowitz, 1969), as mean probability distributions of future changes in GNP and prices. Survey respondents are asked not only for their point forecasts of several variables but also to attach a probability to each of a number of preassigned intervals, or bins, into which future GNP growth and inflation might fall. In this way, respondents provide density forecasts of these two variables, in the form of histograms. The probabilities are then averaged over respondents to obtain the mean or combined density forecasts, again in the form of histograms. The reports on the survey results previously published in the *NBER Reporter* and the *American Statistician* did not always refer to the density forecasts, and sometimes combined bins, but mean density forecasts have been included in the press releases of the Federal Reserve Bank of Philadelphia since it assumed responsibility for the survey in 1990 (and changed its name). Initially there was little interest in the individual density forecasts, due to data processing difficulties and variation over time in the number and identity of respondents, although Zarnowitz and Lambros (1987) is a notable exception. More recently, increased interest in density forecasts in general, thanks to publication of such forecasts in several arenas, and the increased accessibility of the SPF in particular, has led to several contributions. Diebold, Tay and Wallis (1999) use the SPF mean density forecast of inflation to illustrate new methods for the evaluation of density forecasts, and Wallis (2003) further extends these methods using the same illustration, among others. The individual SPF responses are used to address several questions of interest by Lahiri, Teigland and Zaporowski (1988), McNees and Fine (1996) and Giordani and Soderlind (2003a). These contributions are further discussed below, following presentation of our preferred statistical framework.

The finite mixture distribution is presented as the appropriate way to think about a combined density forecast in Section 2, and its implications for analyses of consensus and uncertainty are discussed in Section 3. The alternative statistical framework of Giordani and Soderlind (2003a) is considered in Section 4, and a simple example in Section 5 illustrates the difference between the two approaches. To combine interval forecasts it is recommended in Section 6 that the implied density forecasts first be combined, then the combined interval forecast with the required probability be read off from the combined density.

## 2. A finite mixture distribution

We denote $n$ individual density forecasts of a random variable $Y$ at some future time as $f_i(y)$, $i = 1,...,n$. These may come from different forecasters and/or different models and methods, and may be expressed numerically or analytically. For economy of notation time subscripts and references to the information sets on which the forecasts are conditioned are suppressed. The finite mixture distribution is proposed as an appropriate statistical representation for a combined density forecast. It is defined as

$$f_w(y) = \sum_{i=1}^{n} w_i f_i(y),$$

with weights $w_i \geq 0$, $i = 1,...,n$, $\Sigma w_i = 1$. The same expression appears in statistical decision theory as the linear opinion pool, the commonest form of group consensus probability distribution (see French, 1985, and references therein); again the finite mixture distribution is a relevant statistical model. For a general discussion of finite mixture distributions see Everitt and Hand (1981) or, for a briefer introduction, Everitt's entry on mixture distributions in Kotz and Johnson (1985, pp.559-569), or Stuart and Ord (1994, §5.20-5.24).

Much of the literature on finite mixture distributions is concerned with the problem of identifying and estimating the parameters of the component densities and the mixing proportions. For a mixture of two normal distributions this problem was first considered by Pearson (1894). Mixtures of normal distributions remain a leading case, and although the estimation problem is not our present concern, this case is relevant to many applications in interval and density forecast combination. In reporting probabilities associated with interval forecasts a normal distribution is often assumed, and some current density forecasts are constructed as normal distributions with mean equal to an associated point forecast and variance equal to that of past forecast errors. The perspective of a mixture distribution immediately prompts the observation that a combination of such normal density forecasts is not in general a normal distribution, contrary to what is often assumed (by Hendry and Clements, 2004, §9, for example).

Various weighting schemes appear in the literature on the combination of point forecasts and can be carried over to the present context. Given a series of past forecasts weights can be estimated in a forecast evaluation regression or by Bayesian model averaging, for example. However a surprisingly frequent finding is that, in combining point forecasts, a simple average, with equal weights, outperforms more complicated weighting schemes. The use of equal weights is sufficiently general for our present purposes, hence in what follows we restrict attention to the particular combined density forecast

$$f_C(y) = \frac{1}{n}\sum_{i=1}^{n} f_i(y). \tag{1}$$

This also reflects the use of simple averages in constructing the SPF aggregate density forecasts.

The moments about the origin of $f_C(y)$ are given as the corresponding equally-weighted combinations of the moments about the origin of the individual densities. We assume that the individual point forecasts are the means of the individual forecast densities and so denote these means as $\hat{y}_i$; the individual variances are $s_i^2$. Then for the first two moments we have

$$m_1' = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \hat{y}_C \, , \tag{2}$$

namely the combined or average point forecast, and

$$m_2' = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i^2 + s_i^2\right).$$

Hence the variance of $f_C$ is

$$s_C^2 = m_2' - m_1'^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \hat{y}_C\right)^2 . \tag{3}$$

The first term on the right-hand side of (3) is the average individual variance, and the second term is a measure of the dispersion of the individual point forecasts. This decomposition of the variance of the aggregate distribution lies behind several analyses of uncertainty and disagreement, to which we now turn.

## 3. Consensus and uncertainty

Zarnowitz and Lambros (1987) define "consensus" as the degree of agreement among point forecasts of the same variable by different forecasters, and "uncertainty" as the dispersion of the corresponding probability distributions. Their emphasis on the distinction between them was motivated by several previous studies in which high dispersion of point forecasts had been interpreted as indicating high uncertainty. Those studies had not had access to any direct measure of uncertainty, whereas the SPF data provided the opportunity for Zarnowitz and Lambros to check this presumption, among other things. Their definitions are made operational by calculating time series of (a) the mean of the standard deviations calculated from the individual density forecasts, and (b) the standard deviations of the corresponding sets of point forecasts, for two variables and four forecast horizons. These are analogous measures to the two terms on the right-hand side of equation (3), although the use of standard deviations rather than variances breaks the equation; in any event Zarnowitz and Lambros seem unaware of the decomposition. They find that the "uncertainty" series (a) are typically larger and more stable than the (lack of) "consensus" series (b), thus measures of uncertainty based on the forecast distributions "should be more dependable". The two series are positively correlated, however, hence in the absence of direct measures of uncertainty a measure of disagreement among point forecasts may be a useful proxy.

Lahiri *et al*. (1988) calculate the first four moments of the individual SPF density forecasts of inflation, and use time series of their average values to examine the effect of inflation uncertainty on the real rate of interest. In passing, they obtain a version of the decomposition in equation (3) above, although the left-hand side, described as “total variation in the individual probabilistic forecasts”, is not presented as the variance of the aggregate distribution. In their development they work directly with the individual forecast histograms, and there is no consideration of the mean density forecast, nor any underlying statistical model. As with the second moment relationships in equation (3), their average individual skewness and kurtosis coefficients differ from the skewness and kurtosis coefficients of the aggregate distribution.

## 4. Giordani and Soderlind’s statistical framework

An alternative statistical model for the SPF mean density forecast is presented by Giordani and Soderlind (2003a). They summarize the information set of forecaster *i* by a scalar signal $z_i$ and write the probability density function of inflation conditional on receiving the signal of forecaster *i* as $\text{pdf}\left(p\,|i\right)$. With $p$ and $z_i$ random variables, the latter having density function $\text{pdf}\left(i\right)$, they write the aggregate distribution $\text{pdf}_A\left(p\right)$ as

$$\text{pdf}_A\left(p\right) = \int_{-\infty}^{\infty} \text{pdf}\left(p\,|i\right)\,\text{pdf}\left(i\right)\,\text{d}i\,. \tag{1'}$$

This is Giordani and Soderlind’s equation (1), which they say “amounts to calculating the “marginal” distribution of $p$” (quotation marks in original). From the standard relation between the variances of conditional and marginal distributions the variance of the aggregate distribution is then written as

$$\text{Var}_A\left(p\right) = \text{E}\left(s_i^{\,2}\right) + \text{Var}\left(m_i\right), \tag{2'}$$

which is equation (2) in their article. These equations fulfil similar functions to equations (1) and (3) above, although their statistical foundations are different.

One approach to a mixture distribution is, given a density function dependent on a parameter $q$, $f(x|q)$, and a weighting distribution for $q$, $p(q)$ say, then integrating with respect to $q$ we obtain

$$f(x)=\int_{-\infty}^{\infty} f(x|q)p(q)\,\mathrm{d}q \tag{4}$$

(Stuart and Ord, 1994, p.181). The marginal distribution $f(x)$ is sometimes known as a compound distribution. One consequence of this approach is that "any density function $f(x)$ can be viewed as a mixture density simply by imagining extra variables which have been integrated over" (Everitt and Hand, 1981, p.4). Although (4) is at first sight of the form of (1'), it is an inappropriate model for a combined density forecast, because combination in effect pools information sets, rather than integrates them out to obtain a marginal density, which Giordani and Soderlind may be acknowledging when they place marginal in quotation marks. In practice the representation of diverse, overlapping information sets as random variables with well-defined distributions, as required in this approach, presents conceptual difficulties, and we recall Granger's (1989) remark that "aggregating forecasts is not the same as aggregating information sets". The statistical framework in Section 2 is preferred because it directly captures the SPF averaging of individual densities without introducing a conditioning random variable. Moreover the sample average notation on the right-hand side of equation (3) is statistically more accurate than the use of E and Var on the right-hand side of (2').

In the light of their simple model of many forecasters, Giordani and Soderlind find that "it is not obvious what the aggregate distribution represents". The point of view of the present paper is that it is a combined forecast in the tradition of the point forecast pooling literature. In a later paper (Giordani and Soderlind, 2003b), however, they show that it is the correct approach to aggregation in an asset-pricing problem in which individual agents have logarithmic utility functions. This last assumption makes aggregation straightforward, and the correspondence does not hold otherwise.

The analysis of the SPF inflation forecasts by Giordani and Soderlind (2003a) extends the line of research initiated by Zarnowitz and Lambros (1987) and strengthens their conclusion – disagreement is a better proxy for inflation uncertainty than previously thought. Like Zarnowitz and Lambros, Giordani and Soderlind calculate standard deviations, not variances, so identity (3) cannot be checked; unlike them, the standard deviation calculations are based on normal approximations to the forecast histograms. And the time series are now longer, of course. From their evaluation of the individual density forecasts Giordani and Soderlind conclude that the forecasters underestimated uncertainty. They contrast this finding with that of Diebold *et al*. (1999) who, treating the mean density forecast as that of a representative forecaster, found that it overestimated uncertainty. That such disagreement is possible is obvious from equation (3), and which is the better measure of collective uncertainty – the variance of the mean density forecast or the average individual variance – remains an open question.

## 5. Example

A simple univariate example illustrates our statistical framework. Consider the Gaussian AR(2) data generating process

$$Y_t = f_1 Y_{t-1} + f_2 Y_{t-2} + e_t,\ \ e_t \sim N\left(0, s_e^2\right),$$

for which the true forecast density of $Y_t$ given observations $y_{t-1}$ and $y_{t-2}$ is

$$Y_t \left| y_{t-1}, y_{t-2} \sim N\left(f_1 y_{t-1} + f_2 y_{t-2}, s_e^2\right).\right.$$

However two competing forecasters use only a single past observation, lagged one and two periods respectively, thus their density forecasts are

$$Y_t \left| y_{t-i} \sim N\left(r_i y_{t-i}, s_i^2\right),\ \ i = 1,2\right.$$

where $r_i$ are autocorrelation coefficients and $s_i^2 = \left(1 - r_i^2\right) s_y^2$. The combined density forecast is not the marginal distribution of $Y_t$, as suggested by (1') above, but the mixture of normals

$$Y_{Ct} \sim \frac{1}{2} N\left(r_1 y_{t-1}, s_1^2\right) + \frac{1}{2} N\left(r_2 y_{t-2}, s_2^2\right).$$

The composite information set for the combined density forecast is identical to the information set of the true forecast density: both contain the same two observations. However the combined forecast uses the information inefficiently, relative to the true forecast density. If these two observations are sufficiently different from one another, the condition depending on the parameter values of the data generating process, then the combined forecast density is bimodal.

## 6. Combining interval forecasts

The representation of a combined density forecast as a finite mixture distribution leads to the proposal that interval forecasts be combined via the same route. A density forecast is implicit in an interval forecast, since the calculation of the probability to be attached to an interval requires a distributional assumption, often normality. A combined interval forecast for any required probability can then be obtained from the relevant combined density forecast, whereas combining intervals directly will not in general give an interval with the correct probability.

The proposal by Granger (1989), applied by Granger *et al*. (1989), attempts to overcome this difficulty by estimating combining weights from data on past forecasts that in effect recalibrate the forecast quantiles. For the forecast cumulative distribution function $F(y)$ define the corresponding quantile function $Q(q)$ as

$$Q(q) = F^{-1}(q), \quad 0 \le q \le 1.$$

A central interval with forecast coverage $p$, that is an interval with equal tail probabilities $q/2$, where $q = 1 - p$, is then $\left(Q(q/2), Q(1-q/2)\right)$. Given a pair of quantile estimates $Q^A(q)$ and $Q^B(q)$, Granger suggests the combined estimate

$$Q^C(q) = \hat{a}_1(q)Q^A(q) + \hat{a}_2(q)Q^B(q).$$

The notation makes it clear that the weights vary with $q$, moreover this recalibration requires a historical record of similar forecasts and their realizations, which may not always be available. His final proposal is to base a combined density forecast on the distribution function corresponding to $Q^C(q)$, presumably evaluated over a grid of values of $q$ and so requiring several forecast intervals with different coverages – almost a complete density. This seems much more cumbersome than combining the density forecasts directly, as proposed here.

A further example illustrates these issues. To simplify matters we consider two forecast densities differing only in scale, not location, and assume normal distributions. Then the combined density is a mixture of normals with the same mean, which we take to be zero without loss of generality. Both forecasters report 50% and 90% intervals, and we weight the forecasters equally. The density of forecaster 1 is $N(0,1)$, so the respective intervals are $\pm 0.67$ and $\pm 1.64$, whereas forecaster 2 is much more uncertain, with forecast density $N(0,4)$ and associated intervals $\pm 1.35$ and $\pm 3.29$. With equal weights, combined intervals are then $\pm 1.01$ and $\pm 2.47$, and in the combined density these have coverage 54% and 88% respectively, so the directly combined intervals are respectively too wide and too narrow. Intervals with the correct 50% and 90% coverages in the combined density are $\pm 0.92$ and $\pm 2.62$, respectively narrower and wider than the directly combined intervals. On seeking unequal weights such that a combination of the individual intervals yields the correct combined interval we find $(0.64,\ 0.36)$ for the 50% intervals and $(0.41,\ 0.59)$ for the 90% intervals: the equally-weighted combination of the 50% intervals is too wide, so more weight is needed on the narrower interval of forecaster 1, and vice versa for the 90% interval. These weights vary with the required coverage, as anticipated above, and deriving intervals from the combined density, rather than seeking to combine individual intervals, is recommended.

**References**


Bates, J.M. and Granger, C.W.J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20, 451-468.

Clemen, R.T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5, 559-583.

Clements, M.P. and Hendry, D.F. (eds) (2002). *A Companion to Economic Forecasting*. Oxford: Blackwell.

Diebold, F.X., Tay, A.S. and Wallis, K.F. (1999). Evaluating density forecasts of inflation: the Survey of Professional Forecasters. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (R.F. Engle and H. White, eds), pp.76-90. Oxford: Oxford University Press.

Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.

French, S. (1985). Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds), pp.183-201. Amsterdam: North-Holland.

Giordani, P. and Soderlind, P. (2003a). Inflation forecast uncertainty. *European Economic Review*, 47, 1037-1059.

Giordani, P. and Soderlind, P. (2003b). Is there evidence of pessimism and doubt in subjective distributions? A comment on Abel. Unpublished paper, available at http://home.tiscalinet.ch/paulsoderlind/

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

Granger, C.W.J., White, H. and Kamstra, M. (1989). Interval forecasting: an analysis based on ARCH-quantile estimators. *Journal of Econometrics*, 40, 87-96.

Hendry, D.F. and Clements, M.P. (2004). Pooling of forecasts. *Econometrics Journal*, 7, 1-31.

Kotz, S. and Johnson, N.L. (eds) (1985). *Encyclopedia of Statistical Sciences*, vol. 5. New York: Wiley.

Lahiri, K., Teigland, C. and Zaporowski, M. (1988). Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit, and Banking*, 20, 233-248.

McNees, S.K. and Fine, L.K. (1996). Forecast uncertainty: can it be measured? Paper presented at the Conference on Expectations in Economics, Federal Reserve Bank of Philadelphia, October 1996.

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

Stuart, A. and Ord, J.K. (1994). *Kendall's Advanced Theory of Statistics*, 6th ed., vol. 1. London: Edward Arnold.

Wallis, K.F. (2003). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting*, 19, 165-175.

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

Zarnowitz, V. and Lambros, L.A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, 95, 591-621.