J. R. Statist. Soc. B (1992)
54, *No.* 2, *pp.* 553–567

Modelling Agent Forecast Distributions

By MIKE WEST†

Duke University, Durham, USA

[Received November 1989. Revised November 1990]

SUMMARY

Problems in the analysis and interpretation of uncertain inferences obtained from individuals, or *agents*, are considered in the context of forecasting a scalar random quantity. Based on new theoretical results presented by West and Crosse, Bayesian updating is developed for cases in which agents' opinions are provided in terms of full forecast distributions (discrete or continuous), or partial summaries such as forecast means and variances. A particular focus is on problems in which partial agent information is communicated in terms of a few summary probabilities or, alternatively, quantiles of the agent's distribution. Example models highlight the key features of the theory and illustrate applications.

Keywords: AGENT OPINION; COMBINATION OF FORECASTS; EXPERTISE; FORECAST DISTRIBUTIONS; QUANTILES

1. INTRODUCTION

A variety of problems of information synthesis and aggregation can be usefully considered within a framework of agent opinion analysis. In the context of inference about uncertain quantities, a decision maker may have available relevant information from several sources. These include standard observational or experimental information and non-standard sources such as uncertain inferences from statistical models and stated opinions of individuals, possibly expert in the domain of interest. Such non-standard sources are referred to as agents, the inferences or forecasts made available to the decision maker from such sources being termed agent opinion. Recent works, discussing basic issues and previous modelling approaches, include Lindley (1983, 1988), Lindley and Singpurwalla (1986), Pratt and Schlaiffer (1985), West (1988) and West and Crosse (1992). The last reference provides discussion of basic issues and modelling approaches in opinion analysis and develops new theoretical results which generalize and extend the work of Genest and Schervish (1985). In the current paper, the results of West and Crosse (1992) are reformulated, interpreted and illustrated in contexts of practical interest.

A decision maker ($\mathscr{D}$) is interested in a scalar random quantity Y, discrete or continuous, which has an initial prior distribution $P(y)$. Often, though not always, Y is to be observed at some future time, as in time series forecasting. An agent ($\mathscr{A}$), who is an informed individual or expert, a group of individuals acting collectively in their inferences or a model producing uncertain forecast information about Y, provides some form of summary inferences of potential relevance to $\mathscr{D}$ in forecasting Y. All uncertain inferences from such an agent are referred to as agent opinion. The

†*Address for correspondence*: Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27706, USA.

© 1992 Royal Statistical Society 0035–9246/92/54553 $2.00

terminology used throughout the paper is that appropriate to the canonical case in which $\mathscr{A}$ is an informed individual, stated inferences representing features of his subjective beliefs about Y.

As an example, let Y be the future value of a financial time series, such as exchange rates or money market indicators, with $P(y)$ $\mathscr{D}$'s forecast distribution based on a time series model combined with historical data. Agent information takes the form of external forecasts from other forecasters, forecasting agencies or governmental groups, or econometric models, and may come in the form of point forecasts of Y, forecast means or medians, for example, with associated uncertainty measures such as forecast variances or interquartile ranges. More refined forecasts may be phrased in terms of collections of forecast probabilities or percentage points for Y, or, rather less often, completely specified forecast distributions. Other forms of agent information include forecasts of predictor variables, such as 'official' money supply forecasts, that $\mathscr{D}$ may use to revise beliefs about Y indirectly.

The central problem of Bayesian agent opinion analysis is calculation of $\mathscr{D}$'s posterior distribution $P(y|H)$ where H denote the information provided by $\mathscr{A}$, whatever form this may take.

2. GENERAL UPDATING RESULTS

The key results in West and Crosse (1992), their theorems 4 and 5 that generalize basic results of Genest and Schervish (1985), may be reinterpreted and summarized in the following framework.

For a scalar random quantity Y, $\mathscr{D}$ has a prior distribution $P(y)$ $(-\infty < y < \infty)$ assumed to have a density $p(y)$ with respect to a suitable dominating measure. Agent $\mathscr{A}$ is to provide $\mathscr{D}$ with a distribution for a scalar random quantity X, possibly related to Y. The agent distribution function is denoted by $F(\ )$, so that

$$H = \{F(x): (-\infty < x < \infty)\}.$$

Before learning H, the values of $F(\ )$ are uncertain quantities about which $\mathscr{D}$ has a prior view, his prior information summarized in principle by a complete joint distribution for $\{Y, H\}$. The framework of West and Crosse (1992) follows Genest and Schervish (1985) in assuming that $\mathscr{D}$ only partially specifies his joint prior through the margin $P(y)$ for Y, and his prior expectation of the function $F(\ )$, namely

$$M(x) = E[F(x)] \qquad (-\infty < x < \infty).$$

With these specifications, theorems 4 and 5 of West and Crosse (1992) may be reformulated and summarized as follows.

Theorem 1. If $\mathscr{D}$ specifies only $P(y)$ and $M(x)$, and these specifications are consistent with a complete joint prior over $\{Y, H\}$, then his posterior distribution for $(Y|H)$ has the form

$$P(y|H) = P(y) + \int \Pi(y|x)\{\mathrm{d}F(x) - \mathrm{d}M(x)\} \qquad (-\infty < y < \infty), \quad (1)$$

for some conditional distributions $\Pi(y|x)$ that satisfy $P(y) = \int \Pi(y|x)\, \mathrm{d}M(x)$. An alternative representation is

$$P(y|H) = \int \Pi(y|x)\, dF(x) \qquad (-\infty < y < \infty). \tag{2}$$

In terms of density functions, with respect to a suitable dominating measure over Y, equations (1) and (2) become

$$p(y|H) = p(y) + \int \pi(y|x)\{dF(x) - dM(x)\} = \int \pi(y|x)\, dF(x)\, dx, \tag{3}$$

where $\pi(\ |x)$ is the density of $\Pi(\ |x)$, for each x.

By way of interpretation, note that $P(y|H_x) = \Pi(y|x)$ where $H_x = \{F(x) \doteq \delta_x\}$, δ_x being the distribution degenerate at $X=x$. Thus $\Pi(y|x)$ is $\mathscr{D}$'s posterior were he to learn that $\mathscr{A}$ believed X to take the particular value x with probability 1. Though the theory starts out assuming $P(y)$ and $M(x)$, the reformulation (2) shows that $P(y|H)$ may be identified by identifying suitable functions $\Pi(y|x)$. The way to think about these is as follows. Consider the hypothetical information H_x resulting from $\mathscr{A}$ stating that he believes $X=x$ for sure. Usually, $\mathscr{D}$ would be unwilling to accept this directly and so would view the stated value x as a simple estimate of X based on $\mathscr{A}$'s information and experience. Specification of the ingredients required for theorem 1 involve $\mathscr{D}$ in directly modelling the effect of H_x by assessing $\Pi(y|x)$, his response to such extreme agent information, and in performing this assessment across all possible x. This is generally a challenging task, but, at the least, provides a structure within which to embed the agent opinion issue and sharpens up the problem of model assessment by appeal to the extreme forms of information H_x, temporarily removing from consideration issues associated with $\mathscr{A}$'s own uncertainties about X.

Before proceeding to examples, we make two further comments. Firstly, there is a rather nice feature of the general result: if $F(x) = M(x)$ for all x, the identity

$$P(y|H) = \int \Pi(y|x)\, dM(x)$$

stated in the theorem implies that $P(y|H) = P(y)$. Thus, if $\mathscr{A}$'s views are exactly as expected, they do not influence $\mathscr{D}$ at all. The second comment concerns calibration issues. The development in theorem 1 assumes that $\mathscr{D}$ judges $\mathscr{A}$'s forecasts to be calibrated in that they suffer no systematic biases, due either to inexperience or inaccuracies in assessment, or due to $\mathscr{A}$ misrepresenting his views. In some problems, historical records of $\mathscr{A}$'s past forecasts and actual outcomes may be used to assess and correct for any miscalibration, as West and Crosse (1992) discuss in the context of event forecasting. The issue of calibration clearly applies more generally but will not be developed further here. We simply assume, throughout the following discussion, that $\mathscr{D}$ views $\mathscr{A}$ to be calibrated in the following sense: $\mathscr{A}$'s stated views are accepted as unbiased and accurate representations of his opinions, or have been adjusted (recalibrated) by $\mathscr{D}$ in the light of relevant historical data.

Example 1. The simplest example, event forecasting, is developed in detail in West and Crosse (1992). Here $X = Y = 0$ or 1, an event indicator, and we have the canonical agent opinion problem of Genest and Schervish (1985) in which $\mathscr{D}$ has a prior probability $p = \Pr[Y=1]$ for an event, and $\mathscr{A}$ is to state his own probability f for the event. The identity

$$p(y|H) = \int \pi(y|x)\,\mathrm{d}F(x)\,\mathrm{d}x$$

of equation (3) implies $\mathscr{D}$'s posterior probability

$$\Pr[Y=1|H] = p(1|H) = \pi(1|1)f + \pi(1|0)(1-f) = \alpha f + (1-\beta)(1-f),$$

where $\alpha = \pi(1|1)$ and $\beta = \pi(0|0)$. As in the discussion following theorem 1, probabilities α and β, that completely determine the updating rule, are defined as $\alpha = \Pr[Y=1|f=1]$ and $\beta = \Pr[Y=0|f=0]$, and thus they determine limits on the extent to which $\mathscr{A}$'s stated probability influences $\mathscr{D}$. See Genest and Schervish (1985) and West and Crosse (1992) for further discussion of this case.

Example 2. Suppose now that both Y and X are continuous, and all distributions involved have continuous density functions; thus $P(y)$ has density $p(y)$, $\Pi(y|x)$ has density $\pi(y|x)$, $M(x)$ has density $m(x)$, and $F(x)$ has density $f(x)$. Then equation (3) becomes

$$p(y|H) = p(y) + \int \pi(y|x)\{f(x) - m(x)\}\,\mathrm{d}x = \int \pi(y|x)f(x)\,\mathrm{d}x, \qquad (4)$$

with the identity $p(y) = \int \pi(y|x)\,m(x)\,\mathrm{d}x$. An interesting special case is that of $X=Y$, when the latter identity in equation (4) provides a very simple though general rule for Bayesian updating when $\mathscr{D}$ has a prior density $p(\ )$ for Y and $\mathscr{A}$ states his own density $f(\ )$ for the same quantity. If $f(\ )$ were degenerate at any point x, $\mathscr{D}$'s posterior would be simply $\pi(y|x)$; $\mathscr{A}$'s uncertainty comes into the picture by averaging $\pi(y|x)$ over possible points of degeneracy x with respect to $\mathscr{A}$'s prior $f(\ )$.

For illustration suppose Y to be a future value of a time series, and that $X=Y$. Suppose also that $P(y)$ is normal with known mean t and variance T, denoted by $Y \sim N[t, T]$. A general normal structure for opinion analysis is provided by supposing that the distributions $M(x)$ and $\Pi(y|x)$ determine a bivariate normal distribution, with margin $M(x)$ as $N[a, A]$, for some mean a and variance A, and with correlation r. Then $\Pi(y|x)$ is given by $(Y|X) \sim N[t+b(X-a), (1-r^2)T]$ with regression coefficient $b = r(T/A)^{1/2}$. Take $r > 0$ and suppose that $\mathscr{A}$ has $f(\ )$ degenerate at x. $\mathscr{D}$'s posterior $\Pi(y|x)$ has a shift in location from his prior mean t towards the agent's 'estimate' x of Y, and the magnitude of the shift increases with r. The conditional variance $(1-r^2)T$ decreases as r increases. Thus r is a direct measure of $\mathscr{D}$'s views about $\mathscr{A}$'s subject matter expertise. As a specific example, consider the case $a=t$ and $A=T$ so that $\mathscr{D}$ expects $\mathscr{A}$ to hold views in accord with his own. In the limit as $r \to 1$, $\Pi(y|x)$ degenerates at $y=x$ reflecting a belief that $\mathscr{A}$ is a subject matter expert; this results in $p(y|H) = f(y)$, so that $\mathscr{D}$ simply adopts $\mathscr{A}$'s opinion as his own. With $r=0$, $\mathscr{D}$ judges $\mathscr{A}$'s beliefs to be totally irrelevant, and $P(y|x) = P(y)$ is unchanged whatever the value of x, and, from theorem 1, $P(y|H) = P(y)$ for all H.

Suppose now that $\mathscr{A}$ provides an information set H. Some features of $P(y|H)$ in particular cases are as follows.

(a) If $F(\ )$ has a finite mean s, and H includes the value of s, then $\mathscr{D}$'s posterior mean is

$$E[Y|H] = E[t+b(X-a)|H] = t + b(s-a),$$

a linear function of t and s with coefficient $b = r(T/A)^{1/2}$ on s increasing with the relevance measure r. This form of the posterior mean is common to alternative approaches using normal theory models (Lindley, 1988; West, 1988; West and Harrison, 1989) and is also related to *ad hoc* approaches to the combination of point forecasts (Bordley, 1982; Granger and Ramanathan, 1984). In the symmetric case that $a = t$ and $A = T$, then $b = r$ and $E[Y|H]$ is a convex linear combination of t and s in which the weight r on s is a direct measure of the relevance of $\mathscr{A}$'s opinion. Note finally that $E[Y|H] = E[Y|s]$ here, the point forecast alone being sufficient to determine the posterior mean.

(b) If, in addition to the mean s, $F(\ )$ has variance S, then, regardless of other features of H,

$$V[Y|H] = E[(1-r^2)T|H] + V[t+b(X-a)|H] = (1-r^2)T + b^2S.$$

This is a linear combination of T and S with weight b^2 on S increasing with r. In the symmetric case $A = T$, the posterior variance is a convex linear combination of T and S. An alternative expression is $V[Y|H] = T + r^2(S-A)T/A$, a rather natural form given by the prior variance plus a correction based on the difference between the observed agent variance S and that A of the expected agent distribution $M(\ )$. If $\mathscr{A}$'s distribution is more diffuse than expected in that $S > A$, then this is recognized by the fact that $\mathscr{D}$'s variance increases, and vice versa.

(c) If $F(\ )$ is normal, so is $P(Y|H)$.

(d) The linear updating rule for the posterior mean in (a) is obtained from any model of the form discussed here in which the joint distribution specified for (Y, X) leads to a linear conditional mean $E[Y|X]$, whether the distribution is normal or not. Other possibilities include bivariate elliptical distributions, such as the bivariate t-distribution, with $k > 0$ degrees of freedom, in which the margins are Student t, $P(y) = T_k[t, T]$, with mode t and scale factor $T^{1/2}$, and $M(x) = T_k[a, A]$, with mode a and scale factor $A^{1/2}$. Assuming that $k > 2$, the modes are also means, the marginal variances are $Tk/(k-2)$ and $Ak/(k-2)$, and the correlation is r. It follows that

$$\Pi(y|x) = T_{k+1}[t+b(x-a), (1-r^2)T\,w(x)]$$

where $w(x) = \{k+(x-a)^2/A\}/(k+1)$. In terms of the dependence of r, this is qualitatively similar to the normal model (which is a limiting case as $k \to \infty$). Now consider updating using equation (4). The linear updating rule (a) still applies; if $F(\ )$ has mean s then $E[Y|H] = E[Y|s] = t + b(s-a)$. However, the effect on spread is somewhat different for small and moderate degrees of freedom k. If the agent distribution has mean s and variance S then

$$V[Y|H] = V[Y|s, S] = b^2S + (1-r^2)T\,w(k+1)/(k-1)$$

where

$$w = E[w(X)|H] = \frac{k + (s-a)^2/A + S/A}{k+1}.$$

If k is large, $w \approx 1$ and the normal theory result is approached. Otherwise,

$\mathscr{D}$'s posterior variance increases with both S and, now, the absolute value of the agent 'forecast error' $s-a$.

3. SUMMARY PROBABILITY AND QUANTILE INFORMATION

3.1. *Form of Updating with Summary Probability Information*

Agent opinion is often provided in terms of only a partial specification through collections of quantiles. Pratt and Schlaiffer (1985) provide a case study of collections of quantiles, and West (1988) was particularly concerned with such cases. As in example 2 earlier, suppose that Y is continuous with $P(y)$ having a continuous density $p(y)$, and that $\mathscr{A}$ provides a partial summary of his distribution for a related continuous quantity X. For example, X may be an independent predictor variable to be related to Y, or, as a key special case developed further below, we may have $X=Y$.

Suppose that $\mathscr{A}$ has a continuous distribution $F(x)$ for X, and that $\mathscr{D}$ specifies his expectation $M(x) = E[F(x)]$ for this function. In terms of densities, $E[f(x)] = m(x)$. Suppose now that $\mathscr{A}$ only partially communicates his distribution to $\mathscr{D}$ by stating probabilities

$$H_n = \{f_1, \ldots, f_{n+1}: f_i = F(q_i) - F(q_{i-1}), i=1, \ldots, n+1\} \tag{5}$$

where $q_0 = -\infty < q_1 < \ldots < q_n < q_{n+1} = \infty$ are specified in advance. The global form of $F(x)$ remains uncertain to $\mathscr{D}$. Introduce a discrete random quantity X_n defined as

$$X_n = j, \qquad \text{if and only if } q_{j-1} < X \leqslant q_j \qquad (j=1, \ldots, n+1).$$

Clearly X_n is just a discretized version of X. In $\mathscr{A}$'s opinion, X_n has a discrete distribution $F_n(x_n)$ giving probabilities f_j to the points $x_n=j$; for $\mathscr{D}$, f_j is uncertain, though his prior expectation $M(\ )$ for $F(\ )$ implies that $E[f_j] = \mu_j = M(q_j) - M(q_{j-1})$, for each j. Write $M_n(x_n)$ for the discrete distribution defined by these probabilities μ_j at point $x_n=j$, so that $M_n(\ )$ is $\mathscr{D}$'s expectation for $F_n(\ )$.

Theorem 1 now applies to the restricted information set H_n with the discrete quantity X_n replacing X throughout, and the (discrete) distributions F_n and M_n replacing their continuous counterparts F and M. Equations (3) become

$$p(y|H_n) = p(y) + \int \pi_n(y|x_n)\{\mathrm{d}F_n(x_n) - \mathrm{d}M_n(x_n)\} = \int \pi_n(y|x)\,\mathrm{d}F_n(x_n)\,\mathrm{d}x$$

which reduces to

$$p(y|H_n) = p(y) + \sum_{j=1}^{n+1} \pi_{n,j}(y)(f_j - \mu_j) = \sum_{j=1}^{n+1} \pi_{n,j}(y) f_j, \tag{6}$$

where $\pi_{n,j}(y) = \pi_n(y|j) = p(y|f_j=1)$. Thus $\pi_{n,j}(y)$ is the posterior density of $\mathscr{D}$ were he to learn that $\mathscr{A}$ believed X to lie in $(q_{j-1}, q_j]$ with probability 1.

Equation (6) may be rewritten, in a form more suited to calculation, as follows. Recall the identity $p(y) = \int \pi(y|x)\, m(x)\, \mathrm{d}x$ in the statement of theorem 1 as it applies to learning the full information set $H = \{f(x): -\infty < x < \infty\}$. Then, using Bayes theorem, define the conditional densities

$$m(x|y) = m(x)\,\pi(y|x)/p(y),$$

and let $M(x|y)$ denote the corresponding distribution functions. Define conditional probabilities $\mu_j(y) = M(q_j|y) - M(q_{j-1}|y)$, for each j and all y. Then, trivially, the posterior density $p(y|H_n)$ may be rewritten, in a form more suited to calculation than equation (6), as

$$p(y|H_n) = p(y)\sum_{j=1}^{n+1}\frac{\mu_j(y)}{\mu_j}f_j. \tag{7}$$

This form is used later in illustrations of the development and use of this model.

3.2. *Normal Theory Examples*

Consider the normal framework of example 2 earlier. Recall that $\mathscr{D}$'s prior $p(y)$ is $N[t, T]$, and his expectation $m(x)$ for $f(\ )$ is $N[a, A]$. Also, $\pi(y|x)$ is the conditional distribution of $(Y|X)$ from the bivariate normal having margins $p(y)$ and $m(x)$ and correlation r. It follows that $m(x|y) = m(x)\,\pi(y|x)/p(y)$ is the density of the normal distribution $M(x|y)$ with mean $a + c(y-t)$ and variance $(1-r^2)A$ with $c = r(A/T)^{1/2}$.

Suppose, as a first example, that $\mathscr{A}$ is to state only a single value from his distribution. This is the simplest case of the quantile specification model, with $n=1, f_1 = F(q_1)$ and $f_2 = 1 - f_1$. Writing $\Phi[\ ]$ for the standard normal distribution function, it follows that $\mu_1 = M(q_1) = \Phi[(q_1-a)/A^{1/2}]$ and

$$\mu_1(y) = M(q_1|y) = \Phi\left[\frac{q_1-\{a+c(y-m)\}}{\{(1-r^2)A\}^{1/2}}\right],$$

with $\mu_2(y) = 1 - \mu_1(y)$, for all y.

From equation (7) therefore, $p(y|H_1) = p(y)\,L(y|H_1)$ where

$$L(y|H_1) = \frac{\mu_1(y)}{\mu_1}f_1 + \frac{1-\mu_1(y)}{1-\mu_1}(1-f_1). \tag{8}$$

To illustrate the form of this result, consider the particular case in which $X = Y$. Suppose also that $t = a = 0$ and $T = A = 1$; thus $\mathscr{D}$'s prior for Y is standard normal and he expects the same of $\mathscr{A}$. Suppose also that the two are in positive accord (in $\mathscr{D}$'s opinion) with $r = 0.8$. Then $\mu_1 = \Phi[q_1]$ and $\mu_1(y) = \Phi[5(q_1 - 0.8y)/3]$. Applying the results of example 2 in Section 2, note that, were $\mathscr{D}$ to learn the mean s of $F(\ )$, his own posterior mean would be $0.8s$; were he to learn the variance S, his posterior variance would be $0.36 + 0.64S$. Two illustrations, summarized in Figs 1 and 2, are as follows.

Firstly, suppose that $q_1 = 1.5$, so that $\mu_1 = \Phi[1.5] = 0.9332$ and $\mu_1(y) = \Phi[5(1.5 - 0.8y)/3]$. Fig. 1 plots $\mathscr{D}$'s posterior $p(y|H_1)$ for four possible values of the stated agent probability f_1, as follows.

(a) Curve 1, $f_1 = 0.9332$, its prior expectation; here $L(y|H_1) = 1$ for all y, and $p(y|H_1) = p(y)$ is unchanged, the data H_1 simply confirming $\mathscr{D}$'s initial expectations.

(b) Curve 2, $f_1 = 0.5$ suggests that Y is greater than initially expected by $\mathscr{D}$; his

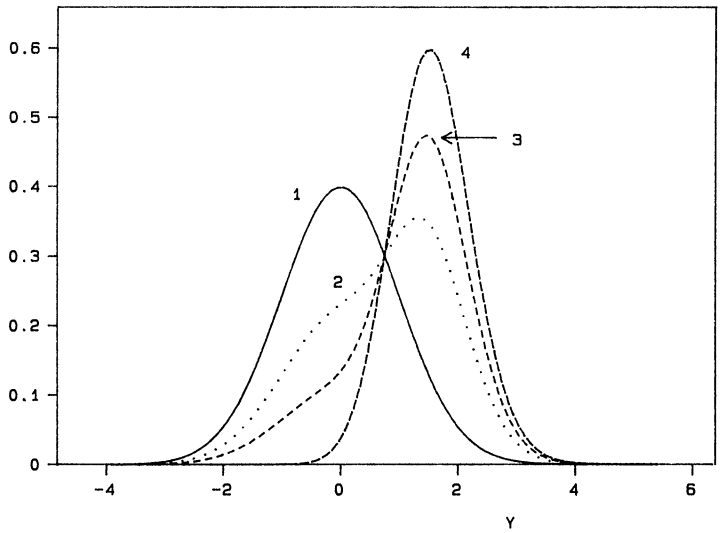

Fig. 1. Posteriors $p(Y|H_1)$ when $q_1=1.5$: curve 1, $f_1 = E[f_1] = 0.9332$; curve 2, $f_1 = 0.5$; curve 3, $f_1 = 0.25$; curve 4, $f_1 = 0.0$

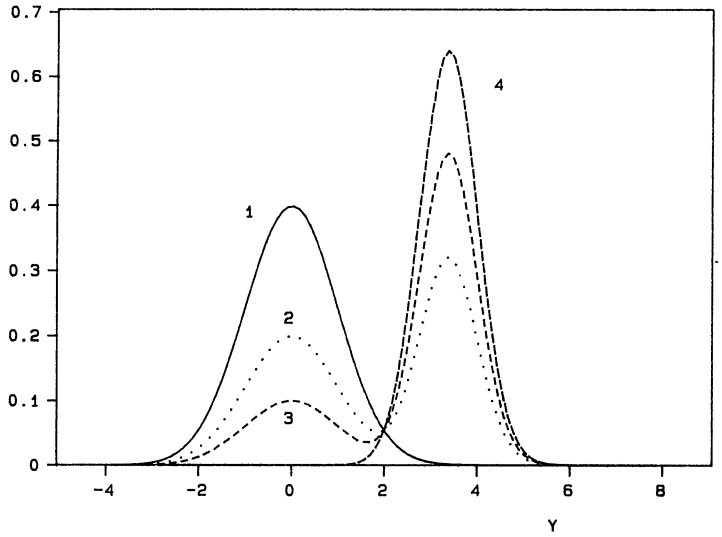

Fig. 2. Posteriors $p(Y|H_1)$ when $q_1=4.0$: curve 1, $f_1 = E[f_1] = 0.9999$; curve 2, $f_1 = 0.5$; curve 3, $f_1 = 0.25$; curve 4, $f_1 = 0.0$

posterior mode is near 1.5, but the posterior is skewed with much of the mass remaining below $Y=0$.

(c) Curve 3, $f_1 = 0.25$, is similar though more extreme data than case (b).

(d) Curve 4, $f_1 = 0$, $\mathscr{A}$ believes that $Y > 0$; $\mathscr{D}$'s posterior strongly reflects this since the agent is viewed as highly relevant ($r=0.8$), though $p(y|H_1)$ retains positive density over the whole real line.

Fig. 2 provides similar plots for $q_1 = 4.0$, far into the upper tail $M(x)$ with $\mu_1 = E[f_1] = 0.9999$. Now cases (b) and (c) lead to bimodal posteriors, heavily influenced by $\mathscr{A}$'s information but with secondary modes near zero, the region favoured initially under $\mathscr{D}$'s prior. This bimodality reflects conflict due to the extremeness of the data H_1 relative to prior expectation.

3.3. *Random Probabilities of Random Quantiles?*

The results and analyses above explore features of $\mathscr{D}$'s prior distribution over Y and $F(\ )$ *conditional* on the values of the q_j. Formally including $\mathbf{q} = (q_0, \ldots, q_{n+1})$ in the conditioning, $\mathscr{D}$'s prior for Y and agent probabilities H_n in equation (5) is written $p(y, H_n \mid \mathbf{q})$, or just $p(y, \mathbf{f} \mid \mathbf{q})$ where $\mathbf{f} = (f_1, \ldots, f_{n+1})$. Then, as a result of theorem 1, equation (7) is just the corresponding conditional density $p(y \mid \mathbf{f}, \mathbf{q})$. In many problems, however, it is the quantiles rather than the probabilities that are more naturally viewed as uncertain. Examples are given by Pratt and Schlaiffer (1985) and West (1988). Eliciting an agent's median and quartiles, for example, corresponds to fixing $n = 3$ and $f_1 = f_2 = f_3 = 0.25$, and then asking the agent to state q_1, q_2 and q_3. Thus the question arises about the relevance of the theory here in such problems, which is answered as follows.

The random probability development above effectively deduces $p(y \mid \mathbf{f}, \mathbf{q})$ from Bayes theorem in which the prior is $p(y)$, and the sampling density $p(\mathbf{f} \mid y, \mathbf{q})$. This formally is equivalent to $p(y \mid \mathbf{f}, \mathbf{q})$ (viewed as a function of all three arguments) if, and only if, $p(y) = p(y \mid \mathbf{q})$ for all y and $\mathbf{q}$. This says that, in the absence of knowledge of $\mathbf{f}$, $\mathscr{D}$ learns nothing about Y from $\mathbf{q}$ alone. This is generally reasonable. Similar comments apply to the alternative approach that updates the prior $p(y)$ to a posterior by using the sampling model $p(\mathbf{q} \mid y, \mathbf{f})$ for random agent quantiles $\mathbf{q}$; this is coherent if, and only if, $p(y \mid \mathbf{f}) = p(y)$, so that $\mathbf{f}$ provides no information about Y in the absence of knowledge of $\mathbf{q}$, which is similarly sensible in practice. Hence, under the reasonable assumptions $p(y) = p(y \mid \mathbf{f}) = p(y \mid \mathbf{q})$, for all y, $\mathbf{f}$ and $\mathbf{q}$, the analysis and updating results of the random probability model of Section 3.2 apply without modification to problems in which it is actually the quantiles that are random. The next section concerns an application in which this is the case.

4. ILLUSTRATION WITH NORMAL AGENT OPINION MODELS

Pratt and Schlaiffer (1985) report a case study in which the agent forecasts the gross receipts for a series of programmes at two cinemas over a period of time. The forecasts were made sequentially over time in terms of forecast median, quartiles and two (upper and lower) extreme points of $\mathscr{A}$'s forecast distributions. To illustrate the models here, a small subset of the data in that study has been selected, namely that relating to 88 of the programmes at the Harvard Square Theater provided in Pratt and Schlaiffer (1985). The first 50 (numbers 1–50 inclusive) actual receipts per show were logged, their sample mean calculated as 6.82, the sample standard deviation as 0.35. For illustration, assume that $\mathscr{D}$ uses these numbers as the basis for his normal prior for the logged per show receipts for future programmes, such outcomes being assumed exchangeable. If Y is the (logged per show) outcome of a particular future programme, $\mathscr{D}$'s view is that $Y \sim N[t, T]$ with $t = 6.82$ and $T = 0.35^2 = 0.122$. $\mathscr{D}$ is to receive forecast information from $\mathscr{A}$; take $X = Y$ and suppose that $\mathscr{A}$ has similar

information and can be expected to hold similar views, reflected in the model of example 2 by taking $M(x)$ as $N[t, T]$ too; thus $a = t = 6.82$ and $A = T = 0.122$. The correlation r appearing in the model is positive reflecting the expected concordance between $\mathscr{D}$ and $\mathscr{A}$.

To estimate r, we use further outcomes from the original data set, namely the 38 programmes numbered 51–88 in Pratt and Schlaiffer (1985). Logged per show receipts y_i ($i = 1, \ldots, 38$) appear in Table 1. The corresponding logged values of the per show lower quartile, median and upper quartile of $\mathscr{A}$'s forecast distributions, namely q_{1i}, q_{2i} and q_{3i} ($i = 1, \ldots, 38$), appear in Table 1. We consider two separate analyses, differing in the type of forecast information assumed provided by $\mathscr{A}$.

TABLE 1
Data and forecasts for illustration in Section 4

Number i	*Outcome y_i*	*Lower quartile q_{1i}*	*Median $t_i = q_{2i}$*	*Upper quartile q_{3i}*	*Variance S_i*
1	7.33	6.40	6.50	6.91	0.1430
2	7.16	6.55	6.75	6.91	0.0699
3	6.46	6.40	6.55	6.80	0.0903
4	7.34	6.91	7.04	7.16	0.0347
5	7.09	6.91	7.20	7.42	0.1430
6	6.59	6.31	6.45	6.60	0.0455
7	7.18	6.75	6.91	7.16	0.0903
8	7.22	6.80	6.91	7.07	0.0391
9	6.81	6.73	6.91	7.20	0.1210
10	7.20	6.77	6.91	7.13	0.0699
11	6.60	6.45	6.68	6.91	0.1150
12	7.48	6.62	6.80	7.21	0.1900
13	6.92	6.70	6.91	7.13	0.1020
14	6.40	6.34	6.45	6.55	0.0245
15	7.14	6.72	6.93	7.19	0.1230
16	6.72	6.40	6.50	6.64	0.0330
17	6.88	6.72	7.00	7.13	0.0949
18	7.04	6.50	6.73	6.91	0.0903
19	6.72	6.62	6.86	6.93	0.0536
20	6.66	6.73	6.91	7.06	0.0622
21	6.98	6.91	7.04	7.16	0.0347
22	6.67	6.85	6.91	7.04	0.0203
23	6.92	6.86	7.13	7.31	0.1150
24	6.63	6.75	6.91	7.10	0.0667
25	6.70	6.75	6.91	7.00	0.0365
26	7.77	6.91	7.09	7.31	0.0903
27	7.29	7.31	7.42	7.68	0.0743
28	6.33	6.40	6.55	6.80	0.0903
29	7.16	6.66	6.91	7.20	0.1600
30	6.64	6.40	6.60	6.76	0.0743
31	7.36	6.75	6.98	7.16	0.0903
32	7.09	6.91	7.23	7.39	0.1300
33	6.71	6.73	6.91	7.09	0.0731
34	6.83	6.98	7.16	7.36	0.0806
35	7.30	6.48	6.62	6.86	0.0791
36	7.19	6.91	7.04	7.16	0.0347
37	6.93	7.04	7.18	7.31	0.0406
38	7.13	6.98	7.10	7.26	0.0455

4.1. *Analysis 1: Normal Agent Forecasts*

To illustrate analysis with full distributional information, suppose that all agent forecast distributions are normal, that outcomes are assumed independent by both $\mathscr{D}$ and $\mathscr{A}$, and that $\mathscr{D}$ assumes $\mathscr{A}$ to be calibrated with no need for correction of his stated forecasts for anticipated biases. (The last assumption could be assessed, and recalibration performed, on the basis of his forecasts over the first 50 programmes, though this is not done here.) For each i, $\mathscr{D}$'s prior is $Y_i \sim N[t, T]$, with $t = 6.82$ and $T = 0.122$, and that of the agent is $N[s_i, S_i]$ for some mean and variance s_i and S_i. On the basis of the available data in Table 1, the mean s_i is taken as the stated median $s_i = q_{2i}$, the variance taken as $S_i = 0.55(q_{3i} - q_{1i})^2$ (the true variances were q_{3i} and q_{1i}, the upper and lower quartiles of a normal distribution). The S_i also appear in Table 1. The discussion in Section 3.2 now applies for each i; conditional on r, and with H^i denoting the fully specified agent distribution for case i, theorem 1 implies that

$$(Y_i | r, H^i) \sim N[6.82 + r(s_i - 6.82), 0.122 + r^2(S_i - 0.122)], \tag{9}$$

independently over i ($i = 1, \ldots, 38$). On observing the actual values $y_1, \ldots, y_{38}$, the product over i of the terms in distribution (9) gives a likelihood function for the correlation parameter r; specifying a prior $p(r)$ allows $\mathscr{D}$ to update to his posterior $p(r|D)$, where $D = \{y_i, H^i: i = 1, \ldots, 38\}$. The posterior chosen here takes $\ln\{r/(1-r)\} \sim N[0, 1.5]$, which gives zero mass to $r \leqslant 0$, is rather flat and uninformative for $r > 0$ and symmetric about the mode at 0.5. Fig. 3, curve a, is a graph of the resulting posterior $p(r|D)$, whose mean and variance are approximately $E[r|D] = 0.75$ and $V[r|D] = 0.193$, and whose mode is approximately 0.84.

For a future programme, with outcome Y, $\mathscr{A}$ will provide information H specifying a normal forecast distribution $N[s, S]$ for some moments s and S. As in distribution (9), $\mathscr{D}$ would forecast Y by using $(Y|r, H) \sim N[6.82 + r(s - 6.82), 0.122$

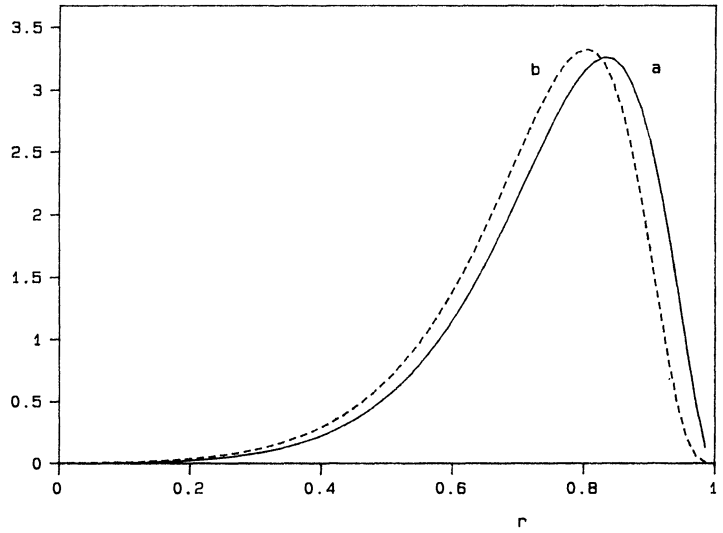

Fig. 3. Posteriors $p(r|D)$ in the normal model: curve a, full distribution; curve b, median plus quartiles

$+ r^2(S-0.122)]$ if r were known; integrating with respect to r gives his unconditional forecast density

$$p(y|H, D) = \int_0^1 p(y|r, H)p(r|D)\,dr \tag{10}$$

(implicitly assuming that, given r and H, Y is conditionally independent of D, and that, given D, r is conditionally independent of H). The mean of density (10) is therefore $E[6.82 + r(s-6.82)|D] = 6.82 + 0.75(s-6.82)$, and the variance may be similarly calculated as $V[Y|H, D] = 0.03 + 0.76S + 0.193(s-6.82)^2$. For any s and S, density (10) may be calculated by simple numerical integration over r. (The results reported here are based on the use of Gaussian quadrature; full details may be obtained on request from the author.) Fig. 4 displays density (10) for the following three cases.

(a) Curve 1, $s=6.82$ and $S=0.122$—since $F(\ )$ is expected to be $N[6.82, 0.122]$, this corresponds to $\mathscr{A}$'s information coinciding exactly with expectation and the posterior is just $\mathscr{D}$'s initial prior.

(b) Curve 2, $s=6$ and $S=0.01$—this represents a precise agent distribution favouring values somewhat lower than expected, though not surprisingly so; $\mathscr{D}$'s posterior reflects both this precision and location.

(c) Curve 3, $s=4.5$ and $S=0.01$—the location $s=4.50$ is well below expectation so that, even though $\mathscr{A}$'s distribution is very precise with $S=0.01$, $\mathscr{D}$'s posterior flattens out, the increased spread reflecting the surprisingly low agent mean.

Qualitatively similar results are obtained with larger values of the agent variance S, though $\mathscr{D}$'s posteriors are naturally more diffuse in such cases.

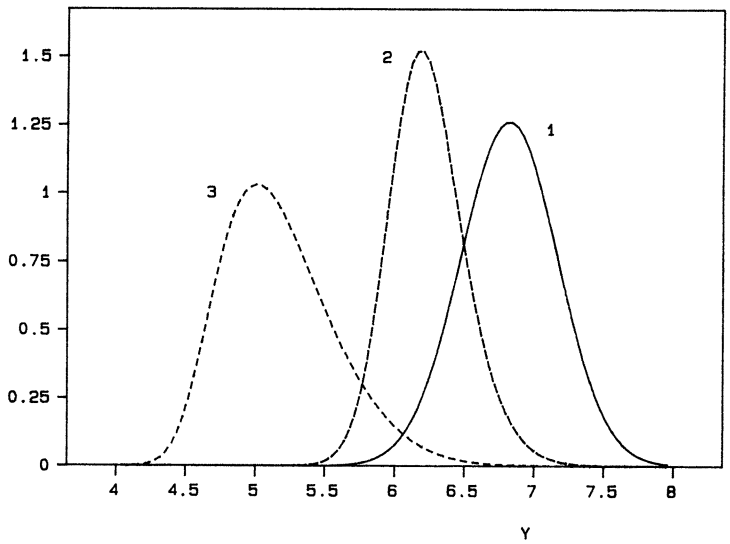

Fig. 4. Posteriors $p(Y|H, D)$ in the normal model with $H = \{s, S\}$: curve 1, {6.82, 0.122}; curve 2, {6, 0.01}; curve 3, {4.5, 0.01}

4.2. *Analysis 2: Agent Forecasts in Terms of Median and Quartiles*

A similar analysis is discussed here, now with $\mathscr{A}$'s information restricted to only a partial specification in terms of forecast median and quartiles, the values for the 38 cases of interest given in Table 1. In dealing with uncertainty about r, the details follow analysis 1 and so are omitted. The major difference is in the structure of densities (9); if Y is any one of the 38 cases of interest, or any future case, and $\mathscr{A}$'s stated information is $H = \{q_1, q_2, q_3\}$, the lower quartile, median and upper quartile respectively of $F(\ )$ for this particular Y, then the discussion of Section 3.2 applies with $n=3$ and, for $j=1, \ldots, 4$, $f_j=0.25$. Recalling that $q_0 = -\infty$ and $q_4=\infty$, equation (7) implies

$$p(y|r, H) = 0.25\,p(y) \sum_{j=1}^{4} \mu_j(y)/\mu_j \tag{11}$$

where, for $j=1, \ldots, 4$, $\mu_j = \Phi[(q_j-6.82)/0.35] - \Phi[(q_{j-1}-6.82)/0.35]$, and $\mu_j(y) = \Phi[b_j] - \Phi[b_{j-1}]$ with $b_0=-\infty$, $b_4=\infty$, and otherwise $b_j = \{q_j-6.82-r(y-6.82)\}/0.35\sqrt{(1-r^2)}$. For each of the 38 training programmes equation (11) gives $\mathscr{D}$'s density $p(y_i|r, H^i)$ with $H^i = \{q_{1i}, q_{2i}, q_{3i}\}$, so that, as in analysis 1, $\mathscr{D}$ deduces his posterior $p(r|D)$. With the same prior, the posterior for this analysis appears as Fig. 3, curve b. This has approximate moments $E[r|D] = 0.73$ and $V[r|D] = 0.192$, mode 0.81, and is similar to that in analysis 1.

For a future outcome Y, density (11) inserted in the integral equation (10) leads to $\mathscr{D}$'s forecast density unconditional on r. Fig. 5 displays this density for three cases in which $\mathscr{A}$'s quartiles are symmetrically located about the median.

(a) Curve 1, $q_1=6.59$, $q_2=6.82$ and $q_3=7.05$—these values coincide with those of the $N[6.82, 0.122]$ prior expected, and as a result $\mathscr{D}$'s posterior coincides with the prior.

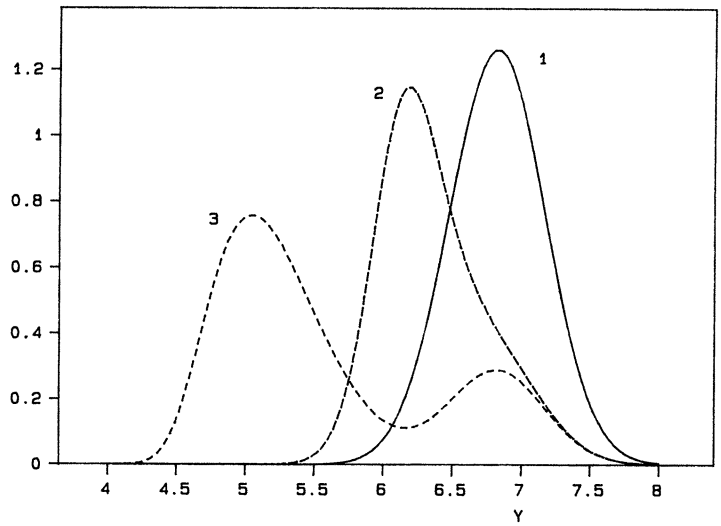

Fig. 5. Posteriors $p(Y|H, D)$ in the normal quantile model $\{q_1, q_2, q_3\}$: curve 1, {6.59, 6.82, 7.05}; curve 2, {5.93, 6, 6.07}; curve 3, {4.43, 4.5, 4.57}

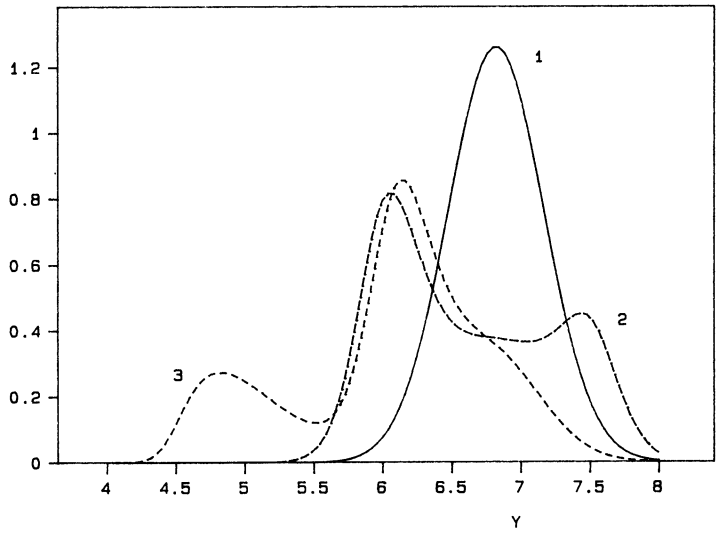

Fig. 6. Posteriors $p(Y|H, D)$ in the normal quantile model $\{q_1, q_2, q_3\}$: curve 1, {6.59, 6.82, 7.05}; curve 2, {5.9, 6, 7.5}; curve 3, {4.5, 6, 6.1}

(b) Curve 2, $q_1 = 5.93$, $q_2 = 6$ and $q_3 = 6.07$—these values agree with those of the $N[6, 0.01]$ forecast distribution of case (b) in analysis 1, so that Fig. 5, curve 2, may be compared with Fig. 4, curve 2. In comparison with the latter, the former is more skewed towards the region favoured by $\mathscr{D}$'s initial expectation of $F(\ )$, since the information here leaves greater uncertainty about the distribution of mass under $\mathscr{A}$'s distribution $F(\ )$.

(c) Curve 3, $q_1 = 4.33$, $q_2 = 4.5$ and $q_3 = 4.57$—as in (b), these values are consistent with case (c) of analysis 1, the resulting density to be compared with Fig. 4, curve 3. Now, since the location of $\mathscr{A}$'s distribution (in terms of the median) is far below expectation, there is much greater uncertainty about Y and $\mathscr{D}$'s posterior is bimodal.

Similar calculations are illustrated in Fig. 6, though (b) and (c) are cases in which $\mathscr{A}$'s information indicates heavily skewed forecast distributions. The data here are

(a) curve 1, as in Fig. 5, curve 1, just $\mathscr{D}$'s prior;

(b) curve 2, $q_1 = 5.9$, $q_2 = 6$ and $q_3 = 7.5$—$F(\ )$ is located near 6 with a long right-hand tail; $\mathscr{D}$'s posterior is bimodal, the major mode just above 6 and a secondary mode near 7.5;

(c) curve 3, $q_1 = 4.5$, $q_2 = 6$ and $q_3 = 6.1$—$F(\ )$ is located near 6 with a long left-hand tail; the posterior is similarly bimodal, reflecting the conflict and increased uncertainty about Y in the light of $\mathscr{A}$'s information that differs quite wildly from expectation.

These analyses illustrate applications of a basic normal model with various types of agent information, some according with expectation and some conflicting. Variants on these analyses are obviously possibly. In particular, calibration adjustments for expected biases in the location and/or scale of $\mathscr{A}$'s distribution may be incorporated,

and other distributions, such as bivariate t-distributions, may be used in place of normality. Models involving exponential distributions, for example, might be used in problems of reliability analysis, such as explored in Lindley and Singpurwalla (1986).

ACKNOWLEDGEMENT

Partial financial support was provided by National Science Foundation grant DMS-8903842.

REFERENCES

Bordley, R. F. (1982) The combination of forecasts: a Bayesian approach. *J. Ops Res. Soc.*, **33**, 171–174.

Genest, C. and Schervish, M. J. (1985) Modelling expert judgements for Bayesian updating. *Ann. Statist.*, **13**, 1198–1212.

Granger, C. W. J. and Ramanathan, R. (1984) Improved methods of combining forecasts. *J. Forecast.*, **3**, 197–204.

Lindley, D. V. (1983) Reconciliation of probability distributions. *Ops Res.*, **31**, 806–886.

——(1988) The use of probability statements. In *Accelerated Life Tests and Experts' Opinion in Reliability* (eds C. A. Clarotti and D. V. Lindley). Amsterdam: North-Holland.

Lindley, D. V. and Singpurwalla, N. S. (1986) Reliability (and fault tree) analysis using expert opinions. *J. Am. Statist. Ass.*, **81**, 87–91.

Pratt, J. W. and Schlaiffer, R. (1985) Repetitive assessment of judgemental probability distributions: a case study. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

West, M. (1988) Modelling expert opinion. In *Bayesian Statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Oxford: Oxford University Press.

West, M. and Crosse, J. (1992) Modelling probabilistic agent opinion. *J. R. Statist. Soc.* B, **54**, 285–299.

West, M. and Harrison, P. J. (1989) *Bayesian Forecasting and Dynamic Models*, sect. 15.3. New York: Springer.