Journal of Forecasting Vol. 4, 329–348 1985,

Extraneous Expert Information

ROBERT T. CLEMEN
College of Business Administration, University of Oregon, U.S.A.

ABSTRACT

When a Bayesian decision maker has to choose among information sources, he should consider the anticipated impact that the information will have on his posterior distribution. In some cases he may determine at the outset that an information source will have no effect on his posterior beliefs, no matter what that source says. Such an information source is called extraneous. In this paper we discuss Bayesian conditions for extraneous information sources, and show a hypothetical example involving experts with overlapping information. Analysis of U.S. weather forecasts demonstrates how this concept can be operationalized to test hypotheses concerning the use of information by forecasters.

KEY WORDS Bayesian aggregation of information Dependence Calibration Weather forecasting

When a decision maker (DM) is interested in the outcome of some chance event, he may solicit information from one or more sources. These sources of information may be human experts, mathematical models or computer simulations. The DM's task of combining the information to form his personal beliefs about the event is complicated if he believes that the information is statistically dependent. The appropriate Bayesian procedure for combining the information has been outlined by Morris (1974), with models based on Morris' scheme provided by Winkler (1981), Lindley (1983, 1985) and Clemen (1984). The key step in the procedure is the assessment of a likelihood function that jointly relates the information and the chance event.

One of the main motivations for modelling information sources (experts, for simplicity) as being statistically dependent is to capture the DM's belief that the experts may share information and hence may provide reports that are redundant to some extent. In this regard, there may be occasions when the DM believes that one expert's opinion is perfectly redundant in the light of another expert's opinion. Clemen (1984) showed that, depending on the nature of the statistical model that jointly relates the experts' information, the combination formula could put zero weight on an expert's opinion, in which case the expert could be ignored. Clemen's motivation was in terms of experts with overlapping information on which they based their opinions, but this concept can be extended to any collection of experts.

These ideas are not without some significance to real world decision makers who face choices

0277-6693/85/040329-20$10.00 *Received September 1985*

involving substantial stakes. In virtually any such decision one may consult any number of consultants or subscribe to a variety of information services. There is public information to gather as well. However, if one hires a consultant, it seems reasonable to presume that the consultant has been able to incorporate the public information into his own opinion; if so, there is no point in gathering the public information. If two consultants are thought to obtain their information from the same sources, but one has a better 'track record', it seems reasonable to presume that the better expert is simply a better processor of the information. In such a case, one may hire only the better expert. In the construction of an expert system (in the artificial intelligence sense), one component of the system might process information, performing the role of the DM in aggregating information. A proper Bayesian information processor would identify and ignore those information sources that add nothing to the system.

Clemen and Winkler (1985) discuss the idea of selecting among experts whose information is dependent. The optimal selection depends upon a careful analysis of the marginal value of information any one expert brings to the combination and hence to the DM's posterior distribution. The value of information analysis depends on the DM's cost function and can be quite complicated. On the other hand, if an expert's information can be ignored, this means that his opinion has no effect on the DM's posterior beliefs. Such an expert can be excluded from the analysis regardless of the specific cost function.

Another motivation for this study is that decision makers sometimes implicitly assume that an information source may be ignored. The hypothetical example in Section 2 and the weather forecasting analysis in Section 4 show that our intuition can be incorrect. A DM may be able to make use of information sources currently taken for granted, thus improving the quality of both his aggregated information and subsequent decisions.

We will refer to an expert whose information may be safely ignored as an extraneous expert. This term is chosen specifically to contrast with the weaker notion of dominance, which is often associated with concepts such as stochastic dominance (e.g. Whitmore and Findlay, 1978), refinement (DeGroot and Fienberg, 1982, 1983) or sufficiency (Blackwell, 1953). DeGroot and Fienberg have shown that if one expert (A) is sufficient for a pair of experts A and B (i.e. B is extraneous) then A is sufficient for B. The reverse, however, is not necessarily true.

Another ground rule to state at the outset is that throughout the article the DM is considered to have no information that the experts do not have. This is not to say that the DM knows nothing, but that the experts are truly experts relative to the DM; there is nothing the DM knows that the experts have not already incorporated into their opinions. Although this seems like a minor point, it avoids the problem of the DM having to assess the dependence between his own information and that of the experts (e.g. see French, 1980). Alternatively, the DM can think of himself as an extraneous expert since he can ignore his own information regarding the chance event in the process of aggregating the experts' information. (A knowledgeable DM can consider himself as another expert, proceeding with the analysis presented below for $k + 1$ rather than k experts.)

In the next section we formalize the concept of extraneous experts and specify the Bayesian procedure that the DM must follow in order to be assured that he is ignoring only information that should be ignored. In Section 2 we study an hypothetical example involving experts with overlapping information. Section 3 discusses the connection between extraneous experts and DeGroot's and Fienberg's sufficiency and refinement concepts. Section 4 shows how the methodology developed in Sections 1 and 3 can be used to study weather forecasters' use of information in the context of probability of precipitation forecasts. Section 5 concludes with a brief discussion of expert information and potential uses for the methodology we develop.

1. CONDITIONS FOR EXTRANEOUS EXPERTS

Suppose that the DM is interested in the random variable θ and has assessed a subjective prior distribution $f(\theta)$. He has access to k experts, each of whom provide him with information about θ. Denote expert j's information by φ_j and let $\Phi^T = (\varphi_1, \ldots, \varphi_k)$, where T denotes transposition. Note that φ_j could be an estimate of θ, a probability distribution, or even a vague opinion. Following Morris (1974) the DM must assess a joint likelihood function for Φ, $\mathscr{L}(\Phi \mid \theta)$. Applying Bayes' rule gives the DM's posterior distribution $f(\theta \mid \Phi)$:

$$f(\theta \mid \Phi) \propto \mathscr{L}(\Phi \mid \theta) f(\theta) \tag{1}$$

Partition Φ so that $\Phi^T = (\Phi_1, \Phi_2)$. Let E_j denote the set of experts corresponding to Φ_j. We define the notion of extraneous experts by the following condition: E_1 is extraneous with respect to E_2 if and only if

$$f(\theta \mid \Phi_1, \Phi_2) = f(\theta \mid \Phi_2) \text{ for all values of } \Phi \tag{2a}$$

By expanding both sides using Bayes' rule and rearranging, this condition is equivalent to

$$\mathscr{L}(\Phi_1, \Phi_2 \mid \theta) = \mathscr{L}(\Phi_2 \mid \theta) f(\Phi_2 \mid \Phi_1) \propto \mathscr{L}(\Phi_2 \mid \theta) \text{ for all values of } \theta \text{ and } \Phi \tag{2b}$$

which also implies that

$$\mathscr{L}(\Phi_1 \mid \Phi_2, \theta) \propto f(\Phi_1 \mid \Phi_2) \text{ for all values of } \theta \text{ and } \Phi \tag{2c}$$

or that Φ_1 is conditionally independent of θ, given Φ_2. In other words, once E_2 is consulted, E_1 adds no new information. In principle this condition is easy to apply in assessing the likelihood. The DM must ask whether, given Φ_2, the relative likelihoods of the possible values of Φ_1 are the same no matter what value θ takes.

The identification of extraneous experts can be thought of as a preposterior operation performed by the DM as he contemplates the acquisition of information and the selection of appropriate experts. This is in contrast to the situation where the particular information an expert provides has no impact, although if he had said something else, his information would have made a difference. In such a case, we might say that the expert is extraneous *ex post*. In such a case, however, the DM could not coherently ignore the expert *a priori*. We can also imagine a situation where for some values of Φ_2 knowledge of Φ_1 will not affect the posterior distribution of θ. In such a case we would say that E_1 is conditionally extraneous with respect to E_2 for those values of Φ_2. Clearly, for E_1 to be completely extraneous (in the preposterior sense), then E_1 must be conditionally extraneous for all values of Φ_2.

We can extend the notion of extraneous experts somewhat. Partition Φ so that $\Phi^T = (\Phi_1, \Phi_2, \Phi_3)$. Note that the DM's likelihood function can be written in the form

$$\mathscr{L}(\Phi \mid \theta) \propto \mathscr{L}(\Phi_3 \mid \theta, \Phi_1, \Phi_2) \mathscr{L}(\Phi_1, \Phi_2 \mid \theta) \tag{3}$$

If E_1 is extraneous with respect to E_2, then (2b) holds. However, for E_1 to be extraneous with respect to both E_2 and E_3 we must have an expanded version of (2b) that includes Φ_3. A sufficient condition for this to occur is for Φ_3 to be conditionally independent of Φ_1:

$$\mathscr{L}(\Phi_3 \mid \theta, \Phi_1, \Phi_2) \propto \mathscr{L}(\Phi_3 \mid \theta, \Phi_2) \text{ for all values of } \Phi \text{ and } \theta \tag{4}$$

Multiplying both sides of (4) by $\mathscr{L}(\Phi_1, \Phi_2 \mid \theta) \propto \mathscr{L}(\Phi_2 \mid \theta)$ shows that under these conditions E_1 is extraneous with respect to the remaining experts E_2 and E_3.

We can realistically suppose that the experts form their opinions φ_j on the basis of data that they observe. Now Φ_3 can be conditionally independent of Φ_1 if there is no overlap between the underlying information sets that E_1 and E_3 use to form their opinions Φ_1 and Φ_3, and if one group's observations are independent of the other group's observations. For example, consider the case where each of three experts forms an opinion after observing a number of independent and identically distributed data. Suppose that Expert 1 is extraneous when Experts 1 and 2 are considered alone. If there is no information overlap between Experts 1 and 2 on one hand and Expert 3 on the other hand, φ_3 is conditionally independent of φ_1, and Expert 1 is extraneous with respect to the combination of Experts 2 and 3.

This suggests an approach for exploiting the notion of extraneous experts. Suppose that the DM has access to a number of experts. First he can decide whether there are groups of experts who use data sets that do not overlap among the groups, and for whom the information is conditionally independent among the groups. We might think of these groups as being 'informationally isolated' from one another. The DM can then look for extraneous experts in each group. Any expert that is extraneous with respect to the other members of his group will be extraneous in the entire set of experts by conditional independence among the groups.

An example of this might be the plant supervisor who has two separate assembly lines that use some of the same kinds of components in their assembly. He may be interested in the proportion p of defectives of component r which is used on both lines. Suppose he hears from a number of workers on each assembly line, and each of these workers has formed an estimate of p based on their observations during the day. In this case it might make sense to assume that the workers on the two lines are informationally isolated. There should be no information overlap; workers on the two lines observed different samples of components. Further, we might assume that defective items occur on one assembly line independently of the occurrence of defective items on the other assembly line. If these two assumptions hold, the supervisor can decide whether to ignore information from a worker solely on the basis of the individual's status with respect to the other workers on his assembly line.

2. A BERNOULLI EXAMPLE: EXPERTS WITH OVERLAPPING INFORMATION

In this section we look at an example of a DM using expert information, in which one expert's information completely subsumes another's. We would typically presume that this indicates the presence of an extraneous expert. However, as we shall see, whether an expert is extraneous depends on how many experts are available and the exact nature of the overlap.

In this example, the DM is interested in the proportion p of a Bernoulli process, from which a sequence of n independent trials may be observed. We begin with two experts; both experts observe n_{12} trials from this process, and Expert 1 sees an additional n_1 trials privately (thus $n = n_{12} + n_1$). Let s_j^* denote the total number of successes observed by expert j and n_j^* the total number of trials observed by expert j. Assume that each expert j reports to the DM both n_j^* and s_j^*. The DM, knowing that Expert 2's information is completely subsumed by Expert 1's, realizes that there were s_1^* successes in n_1^* independent Bernoulli trials. The DM's posterior distribution will depend only on what Expert 1 tells him; Expert 2 is extraneous since his entire information set is observed by Expert 1.

Suppose we include a third expert. Denote a sequence of eight Bernoulli trials by $(x_1, \ldots, x_8)$, and suppose that Expert 1 observes $(x_1, \ldots, x_6)$, Expert 2 observes $(x_1, \ldots, x_4)$ and Expert 3 observes (x_1, x_2, x_7, x_8). If each expert reports n_j^* and s_j^*, then the DM can calculate his posterior distribution given this information, provided he knows the nature of the information overlap among the experts. If the DM knows how the data sets overlap and has an improper beta prior for p (that is, $f(p) \propto p^{-1}(1-p)^{-1}$) then his posterior distribution $f''(p)$ is a mixture of beta distributions:

$$f''(p) = \sum_{q=0}^{n} v_q^* f_\beta(p \mid q, n) \tag{5}$$

where v^* is the posterior probability, given the expert's reports, that there actually were q successe. .n the n trials, and $f_\beta(p \mid q, n)$ is a beta distribution with parameters q and n. (The derivation of this posterior distribution is included in the Appendix.) If the DM knew the total number of successes (s) in the entire sequence of n trials, his posterior distribution would be a single beta distribution with parameters s and n. However, the fact that the information overlaps means that he may not be certain of the value of s, even though he knows n.

Assume that $s_1^* = 3$ and $s_3^* = 2$ (or each of these two experts would estimate $p = 0.5$). If the DM ignores s_2^*, his posterior distribution would be

$$f''(p) = 0.154 f_\beta(p \mid 3, 8) + 0.692 f_\beta(p \mid 4, 8) + 0.154 f_\beta(p \mid 5, 8)$$

Given the values of s_1^* and s_3^*, the value of s_2^* can be 1, 2 or 3. If $s_2^* = 1$ and is not ignored, then the DM's posterior distribution is

$$f''(p) = 0.6 f_\beta(p \mid 4, 8) + 0.4 f_\beta(p \mid 5, 8)$$

Similarly, for $s_2^* = 2$ we have

$$f''(p) = 0.125 f_\beta(p \mid 3, 8) + 0.750 f_\beta(p \mid 4, 8) + 0.125 f_\beta(p \mid 5, 8)$$

and for $s_2^* = 3$

$$f''(p) = 0.4 f_\beta(p \mid 3, 8) + 0.6 f_\beta(p \mid 4, 8)$$

Since Expert 2's information affects the posterior distribution, Expert 2 is not extraneous in this case.

Note also the relationships among the posterior expectations $E''(p)$. Exhibit 1 shows the values of s_2^*, Expert 2's estimate of p (s_2^*/n_2^*) and the corresponding $E''(p)$. When Expert 2's estimate of p is less than 0.5, $E''(p)$ is greater than 0.5, and vice versa. The reason for this effect is apparent. A low (high) value for s_2^* means that there are few (many) successes seen in common by the experts, and Experts 2 and 3 must have seen relatively high (low) numbers of successes in their private data.

If Expert 3's information had not overlapped with that of the other experts (say Expert 3 observed $(x_7, \ldots, x_{10})$), then Expert 3's information would be conditionally independent of the information from Experts 1 and 2. Thus Expert 2 is extraneous with respect to the combination of Experts 1 and 3. The DM would know that there were $s = s_1^* + s_3^*$ successes in $n = n_1^* + n_3^*$ independent Bernoulli trials and would calculate his posterior distribution on the basis of this information, ignoring the information from Expert 2.

s_2^*	s_2^*/n_2^*	$E''(p)$
1	0.25	0.55
2	0.50	0.50
3	0.75	0.45

Exhibit 1. Posterior expectations of p for various values of s_2^*. In all cases, Experts 1 and 3 have estimates of p that are $s_1^*/n_1^* = s_3^*/n_3^* = 0.50$

3. EXTRANEOUS EXPERTS, SUFFICIENCY AND REFINEMENT

In a recent series of articles DeGroot and Fienberg (1982, 1983) have developed a theory of calibration and refinement in the context of weather forecasting. They focus on forecasters who provide a probability for a single dichotomous event. In this paper we discuss the notion of information in a more general context; nevertheless, DeGroot's and Fienberg's results are useful in discussing extraneous weather forecasters.

A weather forecaster's performance can be characterized by two functions: (1) $\rho(x)$, the calibration function, and (2) $v(x)$, the refinement function. The calibration function gives the probability of rain (usually based on historical frequencies) when the forecaster states that the probability of rain is x. The refinement function provides the relative frequency with which a forecaster says that the probability of rain is x. The importance of these two aspects of forecast performance are readily apparent. We typically desire a forecaster to be calibrated; it should rain x proportion of the days when he says x. In terms of refinement, the calibrated forecaster who gives forecasts of 0.1 and 0.9 is more useful to us than the calibrated forecaster who provides probabilities of 0.4 or 0.6. The better forecaster is the one who is more certain of the outcome.

These concepts are consistent with a Bayesian interpretation. The calibration function represents the DM's posterior probability of rain, having heard the forecast. In this regard, we note than an expert can be calibrated purely on the basis of historical data, on the basis of subjective assessment by the DM, or through a combination of the two procedures. The refinement function can be interpreted as the DM's subjective marginal distribution for the forecaster's information.

By definition, one forecaster is 'at least as refined' as a second if the second forecaster's information can be represented as a noisy version of the first's. This notion is based on the idea of sufficient experiments in the sense of Blackwell (1953). Thus, if one has a choice among calibrated forecasters, one should choose the forecaster who is most refined.

Consider two forecasters A and B. A provides forecast $x \in X$ with calibration and refinement functions $\rho_A(x)$ and $v_A(x)$, and B provides forecast $y \in Y$ with corresponding calibration and refinement functions. If the forecasters are 'well calibrated', then $\rho_A(x) = x$ and $\rho_B(y) = y$. We can also talk about the two forecasters' joint forecasting performance characterized by a joint calibration function $\rho(x, y)$ and joint refinement function $v(x, y)$. These functions are the bivariate counterparts of the calibration and refinement functions discussed above. DeGroot and Fienberg show that forecaster A is sufficient for the pair A and B if and only if

$$\rho(x, y) = \rho_A(x) \text{ for all } x \in X \text{ and } y \in Y \tag{6}$$

We can also consider the relative frequency with which a particular calibrated forecast is found on the basis of the two forecasts. Call this relative frequency $h(z)$, where z takes values in the range of $\rho(x, y)$. If A is well calibrated and A is sufficient for the pair, then $\rho(x, y) = x$ and $h(\rho(x, y)) = v_A(x)$. Thus, the jointly calibrated forecast would have exactly the same characteristics as A's calibrated forecasts, and forecaster B is extraneous. In the next section we show how these ideas can be operationalized in the context of U.S. weather forecasts using simple statistical procedures.

4. INFORMATION USE IN WEATHER FORECASTING

In 1966 the National Weather Service (NWS) of the U.S. began to issue precipitation forecasts in terms of the probability of precipitation (PoP). Thus, the official PoP forecast for an area is given in terms of the local weather forecaster's subjective probability that more than a trace of precipitation will occur at his weather station during the specified period of time. Meteorologists use inputs from a variety of sources to develop their subjective PoP forecasts. Since 1972, one of the sources of information available to them has been a guidance forecast, also a PoP forecast, prepared at the NWS headquarters in Washington, D.C. These forecasts are 'objective' in the sense that they are generated by a mathematical/statistical model of the global atmospheric system. Murphy and Winkler (1984) provide an excellent overview of the forecasting process and the meteorological literature concerning both subjective and objective forecasts.

Since two precipitation forecasts are made in any given instance, we may ask whether one of the forecasts is extraneous. The first question is whether the local forecast is extraneous once we know the guidance forecast. This is the same as asking whether the local forecaster adds any new information that is not already contained in the guidance forecast. *A priori* we might expect to conclude that the local forecaster does indeed add information; he has the advantages of being able to observe local conditions and of generating his forecasts slightly later than the guidance forecast. (However, the opposite conclusion could have important implications for the use of taxpayers' money to forecast the weather!)

The second question we can ask is whether the guidance forecast is extraneous once the local forecast is known. We can rephrase this question to ask whether the local forecaster is making full use (in a statistical sense) of the information contained in the guidance forecast. We would hope that the local forecaster is adept at extracting this information and incorporating it into his subjective forecast, although the exact process by which he should do this is not clearly understood and probably varies from one forecaster to the next. On the other hand, if the guidance forecast is not extraneous with respect to the local forecast, it should be possible to combine the two forecasts and improve forecasting performance.

In the remainder of this section, we describe an empirical analysis of U.S. weather forecasters. The next subsection describes the data and the calibration characteristics of the forecasters. In subsection 4.2 we develop a statistical procedure for identifying extraneous weather forecasters, and we find that for both forecasters we can reject the null hypothesis that the forecaster is extraneous. Subsection 4.3 presents a performance analysis of the individual forecasts as well as of a variety of techniques for combining the forecasts. The results indicate that forecast performance can be improved slightly by combining the guidance and local forecasts.

4.1. Data description and forecast calibration functions

To answer the questions posed above we analysed local and guidance PoP forecasts from weather stations around the United States, spanning the period from April 1972 to September 1983. We aggregated the data into four regions, and each region included data from four separate weather stations. Exhibit 2 shows the four regions and the stations that comprised each one. The data were aggregated in this way to obtain sample sizes large enough to permit a reasonable statistical analysis. Since data from individual stations are actually made up of forecasts made by several individuals over the years, aggregating data over regions with relatively homogeneous weather patterns would not appear to compromise the analysis in any material way.

We analysed forecasts made during the warm season (April–September). All forecasts were generated in the evening and specified the probability of rain during a period of time beginning twelve hours later and ending twenty-four hours later. For example, in the North-east (NE) and South-east (SE) regions the forecasts were made at 7.00 p.m. for the period of time from 7.00 a.m. to 7.00 p.m. the following day. For the Southern Plains (SP) and Rocky Mountain (RM) regions the corresponding hours were 6.00 and 5.00, respectively.

Exhibits 3–6 show estimates of both individual and joint calibration functions as well as

Region	Stations	Call letters
North-east (NE)	Albany, NY	ALB
	Boston, MA	BOS
	New York City, NY	LGA
	Philadelphia, PA	PHL
South-east (SE)	Atlanta, GA	ATL
	Asheville, NC	AVL
	Birmingham, AL	BHM
	Columbia, SC	CAE
Southern Plains (SP)	Amarillo, TX	AMA
	Dallas/Ft. Worth, TX	DFW
	Oklahoma City, OK	OKC
	Wichita, KS	ICT
Rocky Mountain (RM)	Boise, ID	BOI
	Denver, CO	DEN
	Great Falls, MT	GTF
	Salt Lake City, UT	SLC

Exhibit 2. Weather stations and regions included in the empirical analysis

frequencies for the forecasts from each region. (The refinement functions are not of use to us here but can be calculated from the frequencies.) Even though the observations totalled well over 5000 in each case, the majority occur when the local and guidance forecasts are close to each other. In each table there are many cells far from the main diagonal that have few (if any) observations. The individual calibration functions indicate that, in general, both guidance and local forecasters overestimate the probability of precipitation. This result is consistent with previous calibration studies (e.g. Murphy and Winkler, 1977).

To use the joint calibration figures, we can read from Exhibit 3, for example, that in the NE it rained on 43.88 per cent of the 98 days when the local forecast was 40 per cent and the guidance forecast was 30 per cent. Of course, for those joint forecasts where there are few historical cases, this empirical calibration procedure has virtually no meaning. However, when there are many observations, it may be useful to consider the forecasters' joint performance. Indeed, the joint calibration functions would appear to indicate that neither of the forecasters is extraneous; for any given forecast value from either forecaster, the joint calibration function appears to increase with increasing forecast values from the other forecaster. Exhibits 7 and 8 demonstrate this effect graphically for the SE region. The curves in Exhibit 7 can be thought of as the local forecaster's conditional calibration curves. If the local forecast were extraneous, the curves would be flat. On the other hand, Exhibit 8 shows conditional calibration curves for the guidance forecast, given the local forecast. The upward trend is much less pronounced and is only obvious in the curves when the local forecast is 0.4 or 0.6. This provides graphical evidence that the guidance forecast may be 'almost' extraneous even though it may not be perfectly so.

4.2. Statistical analysis to identify extraneous experts

Although the above discussion provides some insight as to whether one of the forecasters is extraneous, we can develop a more complete statistical procedure on the basis of the definition of extraneous information. In this case the definition is equivalent to the statement that, given the forecast from one forecaster, θ (rain or no rain) is independent of the information from the

Local forecast: probability of precipitation (%)	Guidance forecast: probability of precipitation (%)													Local forecast calibration and frequency
	0	2	5	10	20	30	40	50	60	70	80	90	100	
0	0.30	0.00	1.62	1.16	4.55	0.00	100.00	—	—	—	0.00	—	—	0.86
	677	305	371	344	44	9	1	—	—	—	1	—	—	1752
2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5	—	0.00	—	—	—	—	—	—	—	—	—	—	—	0.00
	—	1	—	—	—	—	—	—	—	—	—	—	—	1
10	1.53	2.13	6.04	6.61	7.43	7.84	28.57	0.00	0.00	—	—	—	—	5.94
	131	94	182	454	148	51	14	3	1	—	—	—	—	1078
20	13.33	0.00	10.53	12.25	11.55	16.67	15.38	20.00	30.00	0.00	0.00	—	—	13.07
	30	15	76	204	251	150	52	20	10	2	1	—	—	811
30	35.29	0.00	24.14	15.66	22.37	32.82	23.33	24.00	50.00	40.00	40.00	0.00	—	26.06
	17	5	29	83	152	195	60	25	10	5	5	1	—	587
40	0.00	0.00	40.00	9.52	28.85	43.88	38.53	27.91	37.04	44.44	50.00	100.00	—	35.85
	1	1	5	21	52	98	109	43	27	9	4	1	—	371
50	20.00	25.00	0.00	37.50	28.13	49.18	50.85	54.88	35.00	40.00	83.33	100.00	—	45.78
	5	8	3	8	32	61	59	82	40	25	6	3	—	332
60	0.00	—	0.00	20.00	44.44	57.14	48.78	49.02	59.09	57.14	56.25	33.33	—	52.17
	2	—	1	5	9	28	41	51	66	28	16	6	—	253
70	—	—	0.00	—	83.33	64.29	61.54	62.96	75.76	76.09	72.73	63.64	100.00	70.29
	—	—	2	—	6	14	13	27	33	46	22	11	1	175
80	100.00	100.00	—	—	80.00	46.15	28.57	54.55	65.63	78.38	80.00	69.57	80.00	70.27
	1	1	—	—	5	13	7	11	32	37	50	23	5	185
90	—	—	100.00	—	0.00	100.00	85.71	75.00	85.71	83.33	81.82	85.71	90.91	83.85
	—	—	1	—	1	1	7	8	14	12	33	42	11	130
100	—	100.00	100.00	100.00	50.00	100.00	75.00	77.78	93.75	100.00	84.62	94.12	100.00	93.07
	—	1	1	1	2	6	4	9	16	20	39	51	52	202
Guidance forecast calibration and frequency	1.85	1.39	5.37	7.05	16.24	32.59	37.60	45.88	57.83	68.48	75.71	81.88	97.10	22.21
	864	431	671	1120	702	626	367	279	249	184	177	138	69	5877

Exhibit 3. Individual and joint calibration functions for local and guidance PoP forecasts in the NE region

Local forecast: probability of precipitation (%)	Guidance forecast: probability of precipitation (%)													Local forecast calibration and frequency
	0	2	5	10	20	30	40	50	60	70	80	90	100	
0	0.46	0.47	0.72	1.79	1.45	13.33	12.50	0.00	—	—	0.00	—	—	0.99
	655	211	279	279	69	15	8	4	—	—	1	—	—	1521
2	—	0.00	0.00	0.00	0.00	—	—	—	—	—	—	—	—	0.00
	—	4	4	4	1	—	—	—	—	—	—	—	—	13
5	0.00	0.00	0.00	4.55	16.67	0.00	—	—	0.00	—	—	—	—	3.19
	8	3	29	44	6	3	—	—	1	—	—	—	—	94
10	0.00	3.51	5.26	6.08	5.88	5.33	0.00	0.00	50.00	—	—	—	—	5.25
	65	57	133	395	238	75	23	2	2	—	—	—	—	990
20	11.54	7.14	8.77	14.62	13.11	16.67	17.39	38.89	42.86	—	—	—	—	14.62
	26	14	57	212	389	168	46	18	7	—	—	—	—	937
30	0.00	14.29	9.09	15.96	22.01	29.94	29.31	21.54	38.10	57.14	50.00	—	—	25.14
	7	7	22	94	268	314	116	65	21	7	2	—	—	923
40	0.00	—	0.00	23.33	28.38	37.33	36.61	33.78	31.43	40.00	25.00	100.00	—	34.04
	1	—	6	30	74	150	183	74	35	5	4	2	—	564
50	—	—	33.33	22.22	26.67	39.19	46.99	45.36	48.57	62.96	50.00	—	100.00	44.61
	—	—	3	9	30	74	83	183	70	27	2	—	1	482
60	—	—	—	66.67	21.43	50.00	51.06	62.26	58.46	67.57	50.00	0.00	100.00	57.25
	—	—	—	6	14	28	47	106	130	37	16	1	1	386
70	—	—	—	—	75.00	66.67	77.27	72.41	69.77	76.32	84.21	33.33	100.00	74.38
	—	—	—	—	4	6	22	29	43	76	19	3	1	203
80	—	—	—	100.00	100.00	50.00	50.00	82.35	64.00	84.38	75.68	100.00	66.67	75.56
	—	—	—	1	1	2	8	17	25	32	37	6	6	135
90	—	—	—	—	100.00	50.00	—	80.00	70.00	100.00	95.45	91.67	100.00	90.41
	—	—	—	—	2	2	—	5	10	11	22	12	9	73
100	—	—	—	0.00	100.00	50.00	100.00	66.67	90.91	100.00	95.00	81.25	90.91	87.78
	—	—	—	1	1	4	2	3	11	10	20	16	22	90
Guidance forecast calibration and frequency	0.79	1.69	3.19	8.47	15.04	27.94	36.43	46.64	55.21	75.12	77.24	82.50	90.00	22.85
	762	296	533	1075	1097	841	538	506	355	205	123	40	40	6411

Exhibit 4. **Individual and joint calibration functions for local and guidance PoP forecasts in the SE region**

Local forecast: probability of precipitation (%)	Guidance forecast: probability of precipitation (%)													Local forecast calibration and frequency
	0	2	5	10	20	30	40	50	60	70	80	90	100	
0	0.35	0.35	0.96	2.13	5.88	0.00	0.00	0.00	—	—	—	—	—	1.02
	855	286	417	422	68	13	4	1	—	—	—	—	—	2066
2	0.00	0.00	0.00	0.00	0.00	—	—	—	—	—	—	—	—	0.00
	13	6	7	3	1	—	—	—	—	—	—	—	—	30
5	2.94	0.00	4.00	1.05	3.13	14.29	—	—	—	—	—	—	—	2.63
	34	23	75	95	32	7	—	—	—	—	—	—	—	266
10	2.50	4.30	2.89	5.10	6.45	12.61	9.09	16.67	50.00	33.33	—	—	—	5.51
	160	93	242	726	372	111	44	6	2	3	—	—	—	1759
20	6.67	2.70	6.80	10.26	15.68	24.55	10.53	17.65	11.11	0.00	0.00	—	—	13.49
	45	37	103	380	440	167	57	17	9	2	3	—	—	1260
30	12.50	0.00	6.67	24.42	29.41	24.87	33.33	45.45	53.85	50.00	0.00	—	—	28.08
	8	5	30	86	170	189	84	44	13	4	1	—	—	634
40	33.33	40.00	25.00	33.33	28.33	33.33	33.65	46.67	35.71	40.00	50.00	—	—	34.66
	3	5	8	12	60	66	104	45	14	5	4	—	—	326
50	0.00	0.00	40.00	37.50	19.23	42.86	55.00	46.34	28.57	50.00	50.00	100.00	—	41.05
	2	1	5	8	26	35	40	41	21	8	2	1	—	190
60	—	0.00	0.00	50.00	37.50	43.75	48.65	66.67	66.67	45.45	28.57	100.00	0.00	50.77
	—	2	2	2	8	16	37	21	21	11	7	2	1	130
70	—	—	100.00	—	40.00	33.33	70.00	43.75	61.54	66.67	50.00	100.00	—	57.95
	—	—	1	—	5	6	10	16	13	27	8	2	—	88
80	—	—	—	0.00	50.00	100.00	100.00	75.00	92.31	85.71	80.00	66.67	—	81.43
	—	—	—	2	4	5	5	4	13	14	20	3	—	70
90	—	—	—	100.00	50.00	0.00	—	33.33	71.43	66.67	66.67	25.00	100.00	58.62
	—	—	—	2	2	1	—	3	7	6	3	4	1	29
100	0.00	0.00	—	—	100.00	—	100.00	100.00	66.67	0.00	85.71	100.00	100.00	77.27
	1	1	—	—	2	—	1	2	3	1	7	3	1	22
Guidance forecast calibration and frequency	1.16	1.74	3.15	6.73	15.13	25.00	32.64	45.50	52.59	59.26	60.00	73.33	66.67	12.69
	1121	459	890	1738	1190	616	386	200	116	81	55	15	3	6870

Exhibit 5. Individual and joint calibration functions for local and guidance PoP forecasts in the SP region

Local forecast: probability of precipitation (%)	Guidance forecast: probability of precipitation (%)													Local forecast calibration and frequency
	0	2	5	10	20	30	40	50	60	70	80	90	100	
0	0.35 859	0.00 139	0.36 276	0.71 280	2.17 46	0.00 13	0.00 1	0.00 1	50.00 2	100.00 1	— —	— —	— —	0.56 1618
2	0.00 49	7.41 27	0.00 19	0.00 14	0.00 2	0.00 1	— —	— —	— —	— —	— —	— —	— —	1.79 112
5	2.17 138	1.56 64	2.82 177	1.23 244	0.00 64	16.67 6	0.00 3	0.00 1	— —	— —	— —	— —	— —	1.87 697
10	4.69 128	5.41 74	4.71 191	4.36 597	5.32 301	14.86 74	20.83 24	0.00 7	0.00 1	— —	— —	— —	— —	5.51 1397
20	8.51 47	8.82 34	13.40 97	10.03 319	14.15 417	16.54 133	19.64 56	29.41 17	40.00 5	100.00 1	0.00 2	— —	— —	13.48 1128
30	5.56 18	25.00 12	18.75 32	23.89 113	26.34 205	30.20 255	30.69 101	44.44 36	36.84 19	50.00 4	— —	— —	— —	28.18 795
40	50.00 2	100.00 1	20.00 5	35.29 17	33.33 60	57.45 94	44.44 108	46.94 49	50.00 22	62.50 8	0.00 3	— —	100.00 1	46.22 370
50	25.00 4	100.00 2	100.00 1	25.00 8	29.63 27	44.19 43	47.27 55	57.32 82	47.50 40	52.94 17	71.43 7	33.33 3	— —	48.44 289
60	— —	0.00 1	100.00 1	0.00 3	60.00 10	57.14 35	69.70 33	66.67 54	71.43 70	54.29 35	66.67 6	100.00 2	— —	64.40 250
70	— —	— —	100.00 1	— —	66.67 3	44.44 9	70.59 17	62.50 8	83.33 18	75.00 36	83.33 18	80.00 5	100.00 3	74.58 118
80	— —	— —	— —	100.00 1	— —	66.67 6	40.00 5	80.00 5	86.67 15	88.89 18	87.10 31	71.43 7	100.00 2	82.22 90
90	— —	— —	— —	— —	0.00 1	— —	50.00 2	75.00 4	100.00 3	100.00 2	92.31 13	94.12 17	100.00 4	89.13 46
100	— —	— —	— —	— —	— —	— —	66.67 3	75.00 4	100.00 4	90.91 11	100.00 12	100.00 10	100.00 9	94.34 53
Guidance forecast calibration and frequency	1.53 1245	4.52 354	4.75 800	6.20 1596	14.61 1136	31.69 669	39.46 408	52.99 268	62.81 199	69.17 133	81.52 92	86.36 44	100.00 19	17.26 6963

Exhibit 6. Individual and joint calibration functions for local and guidance PoP forecasts in the RM region

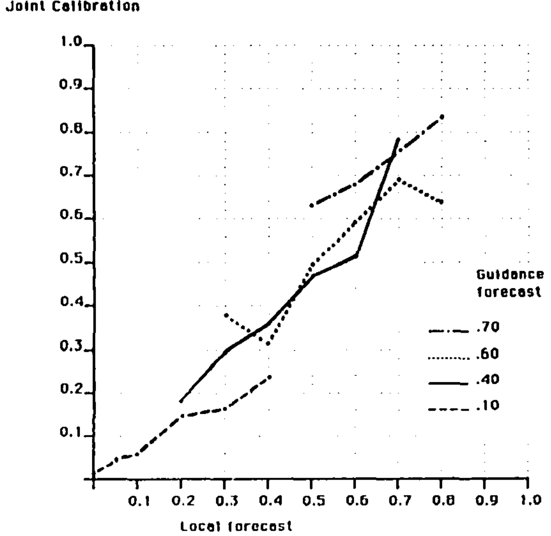

Exhibit 7. Some conditional calibration curves for the local forecast, given the guidance forecast, in the SE region. Calibration figures based on fewer than twenty observations have been excluded

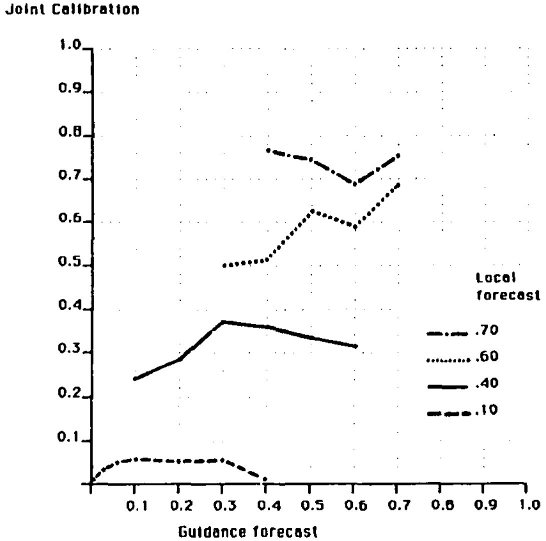

Exhibit 8. Some conditional calibration curves for the guidance forecast, given the local forecast, in the SE region. Calibration figures based on fewer than twenty observations have been excluded

	Guidance forecast	0	2	5	10	20	30	40	50	60	70	80	90	100	Overall chi-square	Degrees of freedom‡
NE region	Chi-square (d.o.f. = 1)	41.69*	14.73*	22.96	26.28	17.42	50.73	22.11	25.62	25.72	24.98	8.37	13.69	6.30*	237.87	10
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.012		
	n	864	431	671	1120	702	626	367	279	249	184	177	138	69		
SE region	Chi-square (d.o.f. = 1)	6.49*	6.53*	12.05	24.26	34.05	41.82	26.02	32.78	15.77	9.25	9.82	0.67*	0.05*	205.85	9
	p-value	0.011	0.011	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.414	0.832		
	n	762	296	533	1075	1097	841	538	506	355	205	123	40	40		
SP region	Chi-square (d.o.f. = 1)	20.56*	8.60*	12.89	27.33	49.48	11.87	26.47	3.38	16.82	6.54	11.00	1.76*	†	165.76	9
	p-value	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.066	0.000	0.011	0.001	0.185	†		
	n	1121	459	890	1738	1190	616	386	200	116	81	55	15	3		
RM region	Chi-square (d.o.f. = 1)	25.53	10.83	20.10	38.79	57.30	44.32	27.99	13.55	22.01	10.56	8.66	5.85*	†	279.66	11
	p-value	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.016	†		
	n	1245	354	800	1596	1136	669	408	268	199	133	92	44	19		

* One or more of the cells had an expected frequency less than 5. Chi-square may not be an appropriate test.
† Statistics cannot be calculated.
‡ Overall chi-square excludes values with *.

Exhibit 9. Chi-square analysis to determine whether local forecasts are extraneous, given the guidance forecast

other forecaster. For example, suppose we ask whether the local forecast (Π_L) is extraneous. Once we know the guidance forecast (Π_G), the probability of rain should be independent of whether $\Pi_L \geqslant \Pi_G$ or $\Pi_L < \Pi_G$. We can use the data to create a series of 2×2 contingency tables, one for each value of the guidance forecast; the horizontal dimension in each table is whether rain occurred, whereas the vertical dimension is whether $\Pi_L < \Pi_G$ or not. (For $\Pi_G = 0$ the dichotomy is whether $\Pi_L = \Pi_G$ or $\Pi_L > \Pi_G$.) The χ^2 test for independence is appropriate for testing the null hypothesis that the local forecast and the occurrence of rain are independent for the given value of the guidance forecast. Since the contingency table is 2×2, the χ^2 random variable has one degree of freedom. To test the hypothesis that the local forecaster is extraneous for all values of the guidance forecast, we can sum the individual χ^2 values. If there are k_G different values of Π_G, there are k_G individual χ^2 statistics, each having one degree of freedom, and the sum of these statistics is also χ^2 with k_G degrees of freedom. A significant χ^2 value allows us to reject the null hypothesis that a forecaster is extraneous.

Some comments about this procedure are in order. First, the use of the χ^2 statistic with k_G degrees of freedom to test the hypothesis of extraneousness is very powerful; it will detect very slight deviations from independence and can be substantially affected by a significant result in any one of the individual χ^2 terms. Offsetting the power of the test is the fact that we have lost a good deal of information in creating the dichotomy $\Pi_L < \Pi_G$ versus $\Pi_L \geqslant \Pi_G$. If there were k_L different values for Π_L, we could have created a $k_L \times 2$ contingency table and performed the χ^2 test of independence on that table with k_L degrees of freedom. The motivation for compressing the data into a 2×2 configuration is that Π_L and Π_G are usually similar. Using the $k_L \times 2$ contingency table would in most cases result in a substantial number of cells with very small expected values, thus rendering the χ^2 approximation inappropriate.

Exhibit 9 shows the results of the analysis to determine whether the local forecaster is extraneous. These results demonstrate without a doubt that none of the local forecasters in the regions we analysed are extraneous. All of the overall χ^2 values are highly significant, as are all of the valid individual χ^2 terms. Thus, the conclusion must be that the local forecaster does add information to the weather forecasting system and should not be ignored in favour of using the guidance forecast alone.

The second question is whether the guidance forecast is extraneous; are the local forecasters successful in extracting the information from the guidance forecast and incorporating it into their official forecasts? The results of the χ^2 analysis to answer this question are shown in Exhibit 10. All of the overall χ^2 values are significant at the 0.01 level, indicating that the guidance forecast indeed is not extraneous. However, these overall χ^2 values are an order of magnitude smaller than those in Exhibit 9. Also, it is apparent that the high significance levels of the overall χ^2 values in Exhibit 10 are due primarily to a few high individual values. For example, in the SP region three individual terms account for nearly 84 per cent of the overall χ^2. In the SE region one individual term accounts for over half of the overall statistic; excluding this one term would yield an overall $\chi^2 = 11.87$ with 8 degrees of freedom. This is not significant at the 0.05 level.

Is the guidance forecast extraneous, given the local forecast? Strictly speaking, the statistics say no, but our intuition might suggest that it is nearly extraneous. If the guidance forecast is nearly extraneous, then combining local and guidance forecasts should result in only a slight improvement in the forecast performance of the local forecast.

4.3. Measuring forecast performance

It is possible to measure forecast performance using standard evaluation techniques. The first step in the evaluation is to calculate a score for each forecast. Let the variable δ_i indicate whether

	Local forecast	0	2	5	10	20	30	40	50	60	70	80	90	100	Overall chi-square	Degrees of freedom‡
NE region	Chi-square (d.o.f. = 1)	4.09	†	†	5.94	1.36	7.49	0.15	1.52	1.92	0.85	2.86	0.57	5.22*	26.73	10
	p-value	0.043	†	†	0.015	0.244	0.006	0.695	0.218	0.166	0.357	0.091	0.449	0.022		
	n	1752	0	1	1078	811	587	371	332	253	175	185	130	202		
SE region	Chi-square (d.o.f. = 1)	3.29	†	0.41*	2.05	1.04	12.46	0.75	3.30	0.71	0.58	0.17	0.79*	0.27*	24.33	9
	p-value	0.070	†	0.522	0.152	0.308	0.000	0.387	0.069	0.401	0.448	0.684	0.373	0.606		
	n	1521	13	94	990	937	923	564	482	386	203	135	73	90		
SP region	Chi-square (d.o.f. = 1)	6.42	†	0.22*	8.16	18.92	3.10	1.57	0.10	0.40	1.25	0.23*	0.86*	0.31*	39.92	8
	p-value	0.011	†	0.641	0.004	0.000	0.078	0.210	0.755	0.529	0.263	0.634	0.353	0.579		
	n	2066	30	266	1759	1260	634	326	190	130	88	70	29	22		
RM region	Chi-square (d.o.f. = 1)	1.42*	1.58*	0.02*	0.48	6.92	6.43	0.00	4.32	0.35	1.37	0.38	1.49*	0.65*	20.25	8
	p-value	0.234	0.208	0.886	0.488	0.009	0.011	0.955	0.038	0.554	0.242	0.538	0.223	0.420		
	n	1618	112	697	1397	1128	795	370	289	250	118	90	46	53		

* One or more of the cells had an expected frequency less than 5. Chi-square may not be an appropriate test.
† Statistics cannot be calculated.
‡ Overall chi-square excludes values with *.

Exhibit 10. Chi-square analysis to determine whether guidance forecasts are extraneous, given the local forecast

it rained on day i ($\delta_i = 1$ if rain occurred on day i, and $\delta_i = 0$ otherwise). The score for forecaster j on day i when he assesses probability Π_{ji} is given by

$$b_{ji} = (\Pi_{ji} - \delta_i)^2$$

This is known as the Brier score (Brier, 1950) and can be thought of simply as a squared error. Smaller Brier scores indicate better performance. The individual scores for n forecast occasions are then averaged for each individual forecaster or combination technique to find the average Brier score b_j (mean squared error) for forecaster or technique j. To facilitate comparison among the forecasts from different regions, skill scores can also be computed. A forecaster's skill score is his average Brier score's percentage improvement over the average Brier score b^* obtained by using the climatological probability of rain as the forecast. Thus the skill score for forecaster j is $100(b^* - b_j)/b^*$.

Both calibrated and uncalibiated forecasts were evaluated. We used the joint calibration procedure to combine local and guidance forecasts, a reasonable approach in the absence of any subjective information. Two other simple combination techniques that we included were (1) the average of the uncalibrated individual forecasts and (2) the average of the calibrated forecasts. The data were partitioned into two sections: data from April 1972 to September 1980 were used to fit both individual and joint calibration functions, whereas the data from April 1980 to September 1983 were used to evaluate the forecasts. In fitting the joint calibration function, if fewer than 25 observations were available for a particular joint forecast, the calibrated local forecast was taken as the jointly calibrated forecast. In turn, in the few instances where there were fewer than 25 observations on which to base an individual calibration estimate, the raw forecast was taken as the calibrated forecast. This procedure has the effect of narrowing the difference in the measured performance between the jointly and individually calibrated forecasts.

Skill scores and average Brier scores are displayed in Exhibit 11. In all but the NE region, the uncalibrated local forecasts had higher skill scores than the uncalibrated guidance forecasts, and the calibrated local forecasts performed uniformly better than the calibrated guidance forecasts. As we anticipated, the jointly calibrated forecast performed better than the guidance forecast (calibrated or not), and slightly better than the uncalibrated local forecast. However, only in the SP region did the jointly calibrated forecast show a performance improvement over the calibrated local forecast.

In contrast to the joint calibration procedure, averaging the forecasts performed well. The average of the calibrated forecasts was the best performer in each region, whereas the simple average took second place in each region except the NE. This robustness of the averaging

	Local forecast	Local forecast (calibrated)	Guidance forecast	Guidance forecast (calibrated)	Jointly calibrated forecast	Simple average	Average of calibrated forecasts	Climatology
NE region	26.80 0.0984	29.71 0.0945	28.74 0.0958	28.74 0.0958	29.48 0.0948	28.66 0.0959	30.23 0.0938	0.00 0.1344
SE region	27.26 0.1231	28.09 0.1217	24.49 0.1278	24.60 0.1276	27.68 0.1224	28.50 0.1210	29.27 0.1197	0.00 0.1692
SP region	25.27 0.0788	27.45 0.0765	23.94 0.0802	22.80 0.0814	28.58 0.0753	28.87 0.0750	28.87 0.0750	0.00 0.1054
RM region	31.74 0.0931	32.77 0.0917	30.43 0.0949	30.28 0.0951	31.82 0.0930	32.92 0.0915	33.87 0.0902	0.00 0.1364

Exhibit 11. Skill scores and Brier scores for individual and combined weather forecasts. In each cell, the top figure is the skill score (%)

techniques is consistent with empirical results in the combination of economic forecasts, where averaging appears to perform better in many cases than more complex combination schemes (e.g. see Makridakis and Winkler, 1983; Clemen and Winkler, 1986).

The relatively poor performance of the joint calibration procedure may be attributed in part to the relatively small number of data points available for estimating most of the individual points on the joint calibration curve. However, we suspect that it results from the guidance forecast being almost extraneous; combining the two forecasts does not substantially improve the overall forecast performance. This idea is supported also by the fact that even the averaging techniques resulted in a fairly limited improvement in performance over the calibrated local forecast. Thus, we conclude that although the local forecasters are unable to extract all of the statistical information that is available in the guidance forecast, they are nevertheless able to incorporate most of that information into their official forecasts.

5. CONCLUSION

In contemplating the acquisition of information from a number of sources, a DM must evaluate the sources in order to choose among them. We have discussed a relatively simple criterion that he may be able to use to screen potential information suppliers. If the DM's likelihood for an expert's information satisfies the necessary conditions, then the expert can be ignored. It is important to realize that this is a much stronger condition that the typical condition of sufficiency; the DM can eliminate an expert from consideration on the basis of extraneousness only if that expert brings no additional information to the aggregated information. This is in contrast to the notion of sufficiency, by which a dominated expert simply has 'worse' information than another expert. If the DM has to make a choice between the two, he should choose the dominant one. If he has the opportunity to choose both, he may be able to drop the dominated one if that expert satisfies the stronger condition of being extraneous.

The example in Section 2 and the analysis of weather forecasters in Section 4 seem to suggest that extraneous experts are uncommon. A tentative conclusion we might reach is that too often decision makers ignore pertinent information in the mistaken belief that such information has already been incorporated into the expert information they receive. However, this conclusion misses an important point. Using stochastically dependent information is difficult at best even within a sophisticated probabilistic framework. Many real world experts, such as weather forecasters and economists, must use inputs from a variety of sources to form their opinions. Typically they use an informal process to aggregate these inputs, a process based largely on intuition, and could not feasibly perform a careful analysis of complex interrelationships. Under these conditions, it is not surprising that experts use information imperfectly. On the contrary, it is surprising that some experts are able to use the available information as well as they do. Our weather forecasters are a case in point in the extent to which they are able to incorporate the guidance forecast into the subjective local forecast. While it may be theoretically possible to extract more information through the joint calibration procedure or some other combining technique, the gains that may be achieved in practice may be small.

The development of operational models for combining dependent expert information is still in its infancy; the models developed so far are mathematically complex and difficult to use in practice. In contrast, the concept of extraneous expert information is a relatively straightforward way to think about and evaluate expert information. The concept proved useful in the study of weather forecasts, and further research can determine whether our results extend to PoP forecasts with longer lead times, or even to other kinds of weather forecasts. Another area in which this notion might be useful is in the combination of economic and financial forecasts. The idea of extraneous information that we have developed here should be valuable in modelling and evaluating dependent information sources in many real world situations.

ACKNOWLEDGEMENTS

This article has benefited greatly from discussions with Allan Murphy, Robert Winkler and Larry Richards. Gary Carter of the National Weather Service Techniques Development Laboratory graciously provided the data analysed in Section 4.

APPENDIX

To derive the DM's posterior distribution for the Bernoulli process model, first assume that the DM knows how the experts' information overlaps. Let $n_{i\ldots j}$ represent the number of Bernoulli trials observed in common by experts $i,\ldots,j$. The fact that the DM knows the overlap structure means that he knows the vector $\mathbf{N}=(n_1,n_2,\ldots,n_{123\ldots k})$. Note that each n_j represents the sample size for a set of non-overlapping Bernoulli trials. Let $\mathbf{S}^*=(s_1^*,\ldots,s_k^*)$ and $\mathbf{S}=(s_1,s_2,\ldots,s_{123\ldots k})$, where s_i is the number of successes in the jth sample with sample size n_i. Then

$$\mathscr{L}(\mathbf{S}^*\mid p,\mathbf{N})\propto\sum_{s_1=0}^{n_1}\cdots\sum_{s_{1\ldots k}=0}^{n_{1\ldots k}}\mathscr{L}(\mathbf{S}^*\mid p,\mathbf{N},\mathbf{S})P(\mathbf{S}\mid p,\mathbf{N})$$

The conditional likelihood in the RHS of this equation is 1 if $\mathbf{S}$ is consistent with $\mathbf{S}^*$ (that is, if the s_is add up correctly for each s_j^*), and 0 otherwise. The probability $P(\mathbf{S}\mid p,\mathbf{N})$ is simply the joint probability of a number of non-overlapping and hence independent binomial samples. Let

$$\delta(Q)=\begin{cases}1 & \text{if expression } Q \text{ is true}\\ 0 & \text{otherwise}\end{cases}\qquad s=\sum_{i=1}^{1\ldots k}s_i\qquad n=\sum_{i=1}^{1\ldots k}n_i$$

and $G\{\mathbf{R}\}$ be a transformation that maps a vector $\mathbf{R}$ into a vector $(\mathbf{R}^*)$ of successes reported by experts $1,\ldots,k$, according to the way the experts' observations overlap. Then

$$\mathscr{L}(\mathbf{S}^*\mid p,\mathbf{N})\propto\sum_{s_1=0}^{n_1}\cdots\sum_{s_{1\ldots k}=0}^{n_{1\ldots k}}\delta(\mathbf{S}^*=G\{\mathbf{S}\})\prod_{i=1}^{1\ldots k}\binom{n_i}{s_i}p^s(1-p)^{n-s}$$

By assumption, the DM's prior is $f(p)\propto p^{-1}(1-p)^{-1}$. As long as $s_j^*>0$ for some j and $s_i^*<n_i^*$ for some i, the DM's posterior distribution is proper and is

$$f(p\mid\mathbf{S}^*,\mathbf{N})\propto\sum_{s_1=0}^{n_1}\cdots\sum_{s_{1\ldots k}=0}^{n_{1\ldots k}}u_{\mathbf{S}}f_\beta(p\mid s,n)$$

where

$$u_{\mathbf{S}}=u_{s_1,\ldots,s_{1\ldots k}}=\delta(\mathbf{S}^*=G\{\mathbf{S}\})\prod_{i=1}^{1\ldots k}\binom{n_i}{s_i}B(s,n)$$

$$f_\beta(p\mid s,n)=B(s,n)^{-1}p^{s-1}(1-p)^{n-s-1}$$

and $B(s, n)$ is the beta function. If we let

$$v_q^* = v_q / \sum_{Y=0}^{n} v_Y \qquad v_q = \sum_{s_1=0}^{n_1} \cdots \sum_{s_{1\ldots k}=0}^{n_{1\ldots k}} \delta(q = s) u s$$

and

$$q \in \{0, \ldots, n\}$$

then

$$f''(p) = f(p \mid \mathbf{S}^*, \mathbf{N}) = \sum_{q=0}^{n} v_q^* f_\beta(p \mid q, n)$$

REFERENCES

Blackwell, D., 'Equivalent comparisons of experiments', *Annals of Mathematical Statistics*, **24** (1953), 265–272.

Brier, G., 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review*, **78** (1950), 1–3.

Clemen, R., 'Modelling dependent information: a Bayesian approach', *Ph.D. Dissertation* (Indiana University), Ann Arbor: University Microfilms, 1984.

Clemen, R. and Winkler, R., 'Limits for the precision and value of information from dependent sources', *Operations Research*, **33** (1985), 427–442.

Clemen, R. and Winkler, R., 'Combining economic forecasts', *Journal of Business and Economic Statistics* (1986), in press.

DeGroot, M. and Fienberg, S., 'Assessing probability assessors: calibration and refinement', in Gupta, S. and Berger, J. (eds), *Statistical Decision Theory and Related Topics III, Vol. 1*, New York: Academic Press, 1982, pp. 291–314.

DeGroot, M. and Fienberg, S., 'The comparison and evaluation of forecasters', *The Statistician*, **32** (1983), 12–22.

French, S., 'Updating of belief in the light of someone else's opinion', *Journal of the Royal Statistical Society, Series A*, **143** (1980), 43–48.

Lindley, D., 'Reconciliation of probability distributions', *Operations Research*, **31** (1983), 866–880.

Lindley, D., 'Reconciliation of discrete probability distributions', in Bernardo, J. *et al.* (eds), *Bayesian Statistics 2*, Amsterdam: North-Holland, 1985, pp. 375–390.

Makridakis, S. and Winkler, R., 'Averages of forecasts: some empirical results', *Management Science*, **29** (1983), 987–996.

Morris, P., 'Decision analysis expert use', *Management Science*, **20** (1974), 987–996.

Murphy, A. and Winkler, R., 'Reliability of subjective probability forecasts of precipitation and temperature', *Journal of the Royal Statistical Society, Series C*, **26** (1977), 41–47.

Murphy, A. and Winkler, R., 'Probability forecasting in meteorology', *Journal of the American Statistical Association*, **79** (1984), 489–500.

Whitmore, G. and Findlay, M., *Stochastic Dominance*, Lexington, MA: D. C. Heath, 1978.

Winkler, R., 'Combining probability distributions from dependent information sources', *Management Science*, **27** (1981), 479–488.

Author's biography:
Robert T. Clemen is an Assistant Professor of Business at the University of Oregon. He has a Ph.D. in quantitative business analysis from Indiana University. His research interests are in combining forecasts, decision theory and Bayesian statistics, and he has published articles in *Operations Research, Management Science* and *Journal of Business Economics and Statistics.*

Author's address:
Robert T. Clemen, College of Business Administration, University of Oregon, Eugene, OR 97403, U.S.A.