Getting the Most out of Ensemble Forecasts: A Valuation Model Based on User–Forecast Interactions

ANTONY MILLNER

Oxford University Centre for the Environment, Oxford, United Kingdom

(Manuscript received 17 October 2007, in final form 19 February 2008)

ABSTRACT

A flexible theoretical model of perceived forecast value is proposed that explicitly includes the effects of user and ensemble characteristics and their interactions. The model can be applied to arbitrary decision problems and is sensitive to a much wider range of factors than traditional forecast valuation models. A simple illustration of its application to the cost–loss decision problem familiar from the forecast valuation literature is discussed. It is shown that perceived value is highly sensitive to perceived model accuracy and that in most cases a high level of perceived accuracy is required for the forecasts to be thought to have any value at all. Decisions with a cost–benefit ratio that is close to the climatological probability of the adverse event are shown to be less sensitive to perceived accuracy. The model shows that it is possible for perceived value to remain unchanged when perceived accuracy increases, thus suggesting an explanation for why forecast uptake often does not increase after improvements in model performance are made. Last, it is argued that attempts to increase forecast uptake should be targeted at those users whose cost–benefit ratios fall in a restricted range that depends on the climatological probability of the event and an objective measure of the ensemble accuracy.

1. Introduction

It has long been recognized that seasonal forecasts hold tremendous potential value for managing climate risks (Mjelde et al. 1998; Messina et al. 1999; Palmer 2002). Despite this widely accepted assertion, relatively little of that potential value is extracted by actual forecast users (Rayner et al. 2005; Vogel and O'Brien 2006), often despite an increase in forecast skill over the past decade (Saha et al. 2006). There is growing awareness in the forecasting community that forecast skill does not translate linearly into forecast value (Zhu et al. 2002; Richardson 2001; Wilks 2001; Katz and Murphy 1997). Research conducted by social scientists (Patt and Gwata 2002; Ziervogel and Calder 2003; Vogel and O'Brien 2006; Artikov et al. 2006) has confirmed that users' perceptions of forecast value, and hence their rate of forecast use, are contingent on a variety of behavioral, cultural, and economic constraints. In a similar way, economic and applied meteorological work has shown that forecast value is heavily dependent on user wealth levels and attitudes to risk (Mjelde and Cochran 1988; Jones et al. 2000; Letson et al. 2005) and trust in the forecasting product (Adams et al. 1995; Luseno et al. 2003; Bharwani et al. 2005).

The methods used to estimate forecast value fall into two categories: those based on modeling studies (e.g., Jones et al. 2000; Zhu et al. 2002) and those based on direct empirical measurements (e.g., Patt et al. 2005). The former set of methods tends to emphasize the potential value that idealized users can extract from forecast information and suffers from several drawbacks. Some include the effects of user wealth and risk aversion on value but assume perfectly accurate and trusted forecasts, while others incorporate the effects of imperfectly trusted forecasts but neglect the effects of wealth and risk aversion. Neglecting such behavioral and economic factors leads to estimates of forecast value that are consistently greater than the value that users actually believe the forecasts to have. In addition, the economic value scores that are prominent in the meteorological literature (e.g., Katz and Murphy 1997) focus on a somewhat artificial decision problem—a cost–loss scenario in which there is no opportunity to use the forecast to hedge against alternative climatic out-

Corresponding author address: Antony Millner, Oxford University Centre for the Environment, South Parks Road, Oxford, OX1 3QY, United Kingdom.
E-mail: antony.millner@magd.ox.ac.uk

DOI: 10.1175/2008JAMC1899.1

comes.[1] Because the structure of the decision problem the user faces can have a significant effect on perceived forecast value, the restriction to a single case throws doubt on the relevance of these models' conclusions for many real-world decisions. Modeling studies do, however, have the advantage of being based on abstractions of the users and their decision-making process, allowing their conclusions to be generalized to the extent that their assumptions remain approximately valid. On the other hand, empirical studies such as those described in Patt et al. (2007) have the advantage of being rooted in the experience of actual forecast users and thus provide a more complete picture of how much value is realized by users. They are, however, very data intensive and of necessity are based on the local experiences of a specific group of forecast users, thus making it difficult to generalize their findings. If the aim is to understand the factors that determine forecast uptake generally and to tailor forecasting products to specific user groups, then one needs a valuation model that is both flexible (can treat any decision problem) and more comprehensive, in that it incorporates those factors that can contribute in important ways to users' appraisal of the product. This paper will propose such a model in the hope that it will act as a bridge between existing modeling studies and empirical findings.

The valuation model presented in this paper differs from previous attempts to quantify forecast value in an important respect. Most forecast skill scores are heavily focused on ensemble performance (see, e.g., Wilks 2006). As such their level of abstraction of the ensemble output is fairly low while their level of abstraction of the users is high, as in the cost–loss models, or the user is simply not accounted for. Conversely, social research on forecast value and uptake is far more concerned with the users and their response to information than it is with the details of the information that the ensemble provides. Here an approach intermediate between these two has been adopted, in that both user behavior and the ensemble and its output have been abstracted in order to focus on the interaction between them. Thus, although some clarity of description of the components that contribute to perceived forecast value is lost, insight into how they interact is gained. I believe that it is this interaction that is most important in determining forecast value.

The model presented below focuses on ensemble forecasts, because they have the potential to provide the most useful information for risk-management decisions and because they are based on numerical models that are believed to hold the greatest promise for future forecasting development (WCRP 2005). It can, however, be specialized to simple deterministic forecasts as a special case.

In section 2, the forecast valuation model is described in mathematical detail. Section 3 recaps the assumptions that went into the model, lists the parameters to which perceived forecast value is sensitive, and discusses their interpretation. As a simple example of the application of the valuation model, section 4 applies the new framework to the cost–loss decisions familiar from the literature on the economic value of forecasts. Section 5 concludes.

2. Forecast valuation model

Because the problem of forecast valuation is essentially economic in nature, it seems natural to tackle it from the perspective of conventional economic decision theory, based on rational choice. There have, however, been several challenges to the tenets of rational choice theory in recent years, behavioral economics modifications of expected utility theory (Kahneman and Tversky 1979) and bounded rationality (Kahneman 2003) being among the most prominent in the economics community. The model presented below is firmly based in the rational choice tradition, being nothing more than a combination of expected utility theory and Bayesian updating. Real decisions are often complicated by psychosocial factors, but the rational choice assumption seems a good first approximation for understanding forecast uptake. In fact, it may be thought of as providing an upper bound on perceived forecast value, because it has been suggested (Grothmann and Patt 2005) that psychological factors probably never work to increase perceived value.

The valuation model is composed of four components: a specification of the user's decision problem, which is applicable to arbitrary decisions; a specification of the user's utility function, which encodes user attitude to risk; a specification of user response to forecast information through Bayesian updating; and a parameterized abstraction of the ensemble forecasting system.

a. User decision making

It is assumed that users act as if they are expected utility maximizers [see, e.g., Mas-Colell et al. (1995) for

[1] Decisions that admit hedging opportunities allow the decision maker to take an action that trades off the risks of one eventuality against another. For example, farmers may plant a fraction of their land with heat-resistant crops to hedge against the possibility of a hotter-than-usual growing season. The cost–loss decision problem does not admit hedging opportunities because there is no action that is intermediate between protecting and not protecting that provides a degree of insurance against either possible outcome.

the conditions under which this is true and Machina (1987) and Quiggin (2001) for alternative viewpoints]. Suppose the user perceives the set $S = \{s\}$ of "states of the world," which occur with probability $p(s)$, respectively. A Bayesian interpretation of $p(s)$ is adopted, in that instead of representing the frequency of event s, it rather represents the user's beliefs about the likelihood of event s occurring. In general, different users will have different $p(s)$. The user can choose from a set of actions $X = \{x\}$. The set X may be discrete (e.g., insure, or do not insure) or continuous (plant a proportion x of a field with maize). For every pair (x, s), the user anticipates a consequence $c(x, s)$. The expected utility framework can be written as follows: Suppose users are faced with a decision as to which action to pursue. They will choose their actions as if they maximize their expected utility; that is,

$$\max_x EU = \sum_{s \in S} p(s)U[c(x, s)], \tag{1}$$

where $U[c(x, s)]$ is the utility the user derives from consequence $c(x, s)$ and Ef denotes the expectation of the function f. The user's attitude to risk is determined by the concavity of the utility function and is captured by the Arrow–Pratt coefficients of risk aversion:

$$R_{\mathrm{abs}}(W) = -\frac{U''(W)}{U'(W)} \quad \text{and} \tag{2}$$

$$R_{\mathrm{rel}}(W) = -W\frac{U''(W)}{U'(W)}, \tag{3}$$

where a prime on a function denotes a derivative with respect to its argument. See Gollier (2001) for a derivation and discussion of these expressions. In general, they represent the decision maker's tolerance for risk and are related to the absolute and relative amounts the user would be willing to pay to remove all risk from the decision. It is often supposed that $R_{\mathrm{abs}}(W)$ is a decreasing function (the more wealthy one is, the more likely one is to take a bet of fixed size) while $R_{\mathrm{rel}}(W)$ is an increasing or constant function (the more wealthy one is, the less likely one is to bet a fixed fraction of one's wealth; see Gollier 2001, chapter 2). It is assumed here that the user has a constant R_{rel}, which implies that

$$U(W) = \begin{cases} \dfrac{W^{1-r}}{1-r} & r \neq 1 \\ \ln W & r = 1 \end{cases}, \tag{4}$$

where $r \geq 0$ is a constant. It can be shown that $R_{\mathrm{abs}}(W) = 1/W$ and that $R_{\mathrm{rel}}(W) = r$. This makes it a reasonable choice as well as one that is easy to handle. It is a widely used workhorse in the economics literature. The coefficient of relative risk aversion r will be used to parameterize the user's attitude to risk. Low values of r correspond to risky behavior, and high values of r correspond to conservative, risk-averse behavior.

b. *Bayesian updating of beliefs*

Forecast information can be thought of as a message that users use to update their beliefs about the future state of the world. It would be unrealistic to assume that once a user receives a forecast, her beliefs about the future are dictated by the forecast alone. Most forecast users have strongly entrenched beliefs about possible future outcomes, which they arrive at through a combination of localized knowledge and long experience (Roncoli et al. 2002). It is extremely difficult in general to understand the way in which forecast users filter the new information they have received through their consciousness to arrive at some new beliefs. To attempt to represent this process, I will make use of Bayes's theorem, the traditional method of updating beliefs in the economics of information.

Suppose the user receives a forecast message s'. For the moment, assume that s' is a deterministic message, such as "it will rain more than normal next season," rather than an ensemble probability prediction. The user updates her beliefs about the states of the world according to Bayes's rule as follows:

$$p(s|s') = \frac{\mathbf{p}(s'|s)p_c(s)}{p(s')}. \tag{5}$$

Here $p(s|s')$ is the posterior probability of state s once message s' is received, and $p_c(s)$ is the prior probability of state s before the message was received. These two quantities are related through $\mathbf{p}(s'|s)$, the likelihood matrix. The likelihood matrix can be interpreted as the probability of receiving message s' given that the state of the world is s and is thus a natural measure of the user's beliefs of forecast accuracy. Here $p(s')$ is a normalization factor and can be computed from the likelihood matrix and the prior probability:

$$p(s') = \sum_{t \in s} \mathbf{p}(s'|t)p_c(t).$$

Now that the user's response to a deterministic message can be estimated, how can her response to an ensemble forecast message, which may take the form of a probability density, be quantified? To answer this question, some additional assumptions about the ensemble will be needed.

c. The model ensemble

Real multimodel ensembles are composed of several different climate models, each of which is run with several parameter sets to generate a distribution of results. Each of the models performs differently and has different levels of accuracy that may not be homogenous over states of the world. For example, a particular model may be better at predicting extreme events than normal ones. These complexities will be disregarded, given that it is unlikely that the user is aware of them. Each realization of a model (i.e., a particular model, and a particular set of parameters) will be treated as identical and independent from the perspective of the user. The total number of model realizations in the ensemble will be referred to as N, and each model realization will be referred to as an ensemble member. For example, if three different climate models are run, each with nine parameter sets, then there are $N = 3 \times 9 = 27$ ensemble members. The assumptions made about the user's knowledge of the ensemble can be formalized as two levels of democracy in the ensemble, inter-, and intramodel democracy. Intermodel democracy assumes that all ensemble members are created equal. In particular, they are all believed to have the same accuracy, as encoded by their likelihood matrix. Also, in arriving at the final probability distribution, each model's predictions are weighted equally. Intramodel democracy assumes that each ensemble member's chances of making a correct prediction are independent of the state of the world. Moreover, forecast errors are distributed equitably between the incorrect states of the world; that is, if an incorrect forecast is made, it is believed to be equally likely to be a forecast of any of the "incorrect" states of the world.

These assumptions allow us to write down the likelihood matrix for each model. Suppose, for example, that there are three possible states of the world and assume that each model generates a forecast that corresponds to a particular state of the world s'. Furthermore, suppose the forecast is believed to be correct with probability λ. Then the likelihood matrix has the following form:

$$\mathbf{L}(s'|s) = \begin{pmatrix} \lambda & \frac{1-\lambda}{2} & \frac{1-\lambda}{2} \\ \frac{1-\lambda}{2} & \lambda & \frac{1-\lambda}{2} \\ \frac{1-\lambda}{2} & \frac{1-\lambda}{2} & \lambda \end{pmatrix}. \tag{6}$$

These two assumptions allow perceived model accuracy to be described by just one parameter λ, rather than a likelihood matrix with independent entries for each ensemble member. Thus a little complexity is sacrificed for the sake of clarity of interpretation. There is an important subtlety implicit in the definition of λ, in that it represents the user's *beliefs* about forecast accuracy rather than the actual model accuracy. There is no a priori reason why λ should coincide with the model accuracy, and it may be subject to a variety of behavioral effects. As such, it is a measure of the trust the user places in each ensemble member's predictions.

d. Bayesian updating from an ensemble prediction

Here several of the previous assumptions are combined to answer the question: How do users respond to probability density forecasts from ensembles? Suppose the user receives a forecast probability density $\underline{\pi} = \underline{\pi}(s')$ from an ensemble.[2] How should her beliefs be updated? In the case in which the user received only one deterministic message s', one was able to use Bayes's theorem, and the same can be done here. Consider the case in which there are only three states of the world. There is a sense in which, when the user receives the density forecast $\underline{\pi}$, she is receiving all three possible messages, only with different weightings, that is, their forecast probabilities. This suggests the following natural generalization of Bayes's theorem to the case of a probability density message:

$$p(s|\underline{\pi}) = \sum_{s' \in S} \underline{\pi}(s')p(s|s') \tag{7}$$

$$= \sum_{s' \in S} \underline{\pi}(s') \frac{\mathbf{p}(s'|s)p_c(s)}{p(s')}, \tag{8}$$

where it is important to remember that s' is the forecast message (because it is the argument of $\underline{\pi}$), and s is the observation. Moreover, because of the assumptions about model democracy, $\mathbf{p}(s'|s)$ may be replaced by the likelihood matrix $\mathbf{L}(s'|s)$ defined in Eq. (6). Expanding the normalization constant, one has

$$p(s|\underline{\pi}) = \sum_{s' \in S} \underline{\pi}(s') \frac{\mathbf{L}(s'|s)p_c(s)}{\sum_{t \in S} \mathbf{L}(s'|t)p_c(t)}. \tag{9}$$

This equation will be used to update user beliefs once a density forecast is received.

[2] The symbol $\underline{\pi}$ will always be used to refer to density forecasts.

e. Forecast value

Given that the effects of forecast information on user beliefs can now be estimated, how can a value be attached to the forecast information? The value of the forecast is defined as the amount the user would have to be paid to make her indifferent between having forecast information and not having it. This is formalized as follows: Let the decision variable be x, and let x_0 be the value of x that maximizes expected utility given prior beliefs $p_c(s)$ and x_1 be the value that maximizes expected utility given updated beliefs $p(s|\underline{\pi})$. Then the value V of the density forecast $\underline{\pi}$ is defined through

$$\sum_{s\in S} p_c(s)U[c(x_0, s) + V] = \sum_{s\in S} p(s|\underline{\pi})U[c(x_1, s)]. \tag{10}$$

Note that this measure is explicitly dependent on user characteristics as encoded by the utility function.

This is how the value of a specific forecast may be defined; however, the user does not know a priori which forecast will be received. To place a value on the forecasting system as a whole, one needs to incorporate every possible prediction the system can make, weighted by the appropriate probability. It is important to note that every conceivable forecast $\underline{\pi}$ will lead to a different set of user beliefs and hence a different value of x that maximizes expected utility under those beliefs. Let $x_{p(s|\underline{\pi})}$ be this value, and let $q[\underline{\pi}]$ be the probability of receiving the density forecast $\underline{\pi}$ (more on this later). Then the value V of the forecast service as a whole is defined so as to satisfy

$$\sum_{s\in S} p_c(s)U\{c[x_{p(s)}, s] + V\} = \sum_{\underline{\pi}} q[\underline{\pi}]\left\langle \sum_{t\in S} p(t|\underline{\pi})U\{c[x_{p(t|\underline{\pi})}, t]\}\right\rangle, \tag{11}$$

where the outer sum on the right-hand side is over all possible realizations of the density forecast $\underline{\pi}$.

f. The probability of an ensemble prediction

What is the subjective probability of receiving a particular probability density as a forecast? This can be calculated for a deterministic forecast from a single model as follows: Suppose a deterministic forecast s' is received, and the likelihood matrix is $\mathbf{L}(s'|s)$. Then

$$q(s') = \sum_{t\in S} \mathbf{L}(s'|t)p_c(t), \tag{12}$$

where $p_c(t)$ is the prior probability of state t. I proceed in analogy with this simple case for an ensemble prediction:

$$q[\underline{\pi}] = \sum_{t\in S} \mathcal{P}(\underline{\pi}|t)p_c(t), \tag{13}$$

where $\mathcal{P}(\underline{\pi}|t)$ denotes the probability of receiving forecast $\underline{\pi}$ given that the state of the world is t. The probability $\mathcal{P}(\underline{\pi}|t)$ can be calculated by taking advantage of model democracy and considering how the probabilities are arrived at in an ensemble prediction. Let there be N members in the ensemble, and let the likelihood matrix associated with an individual ensemble member be $\mathbf{L}(s'|s)$. Then,

$$\mathcal{P}(\underline{\pi}|t) = N!\prod_{s'\in S}\frac{[\mathbf{L}(s'|t)]^{N\underline{\pi}(s')}}{[N\underline{\pi}(s')]!}. \tag{14}$$

Using this formula,[3] $q[\underline{\pi}]$ can now be calculated. Notice that it only depends on three things: the number of members in the ensemble N, the perceived model accuracy λ, and the prior probability $p_c(t)$.

The forecast valuation model is thus defined by Eq. (11), where

$$q[\underline{\pi}] = N!\sum_{t\in S}\left\{\prod_{s'\in S}\frac{[\mathbf{L}(s'|t)]^{N\underline{\pi}(s')}}{[N\underline{\pi}(s')]!}\right\}p_c(t). \tag{15}$$

3. Model sensitivities and interpretation

The model presented above is an adaptation of standard results in expected utility theory [see Gollier (2001) for a discussion of this method of calculating the value of information] to the case of ensemble forecasts,

[3] To understand why this function has this form, first note that since $\underline{\pi} = \underline{\pi}(s')$ is nothing more than a histogram of the model predictions, $N\underline{\pi}(s')$ is always an integer. Now consider an example in which there are three states of the world and 10 ensemble members. What is the probability that $\underline{\pi} = \{0.3\ 0.5\ 0.2\}$? Think of it as follows: There is a bag of 10 balls, and 3 different colored bins [red (R), green (G), and blue (B)], into which the balls are thrown with probability $L(\mathrm{R})$, $L(\mathrm{G})$, and $L(\mathrm{B})$. The $\underline{\pi}$ corresponds to taking the balls out of the bag and throwing them into the bins in the follow order: RRRGGGGGBB. This occurs with probability $[L(\mathrm{R})]^3[L(\mathrm{G})]^5[L(\mathrm{B})]^2$. However, one could equally well have thrown them in another order, say RGGBRBGGRG, and the result would be the same. In total there are 10!/(3!5!2!) ways the balls could have been thrown to achieve the same result, and so the final probability is $10![L(\mathrm{R})^3/3!][L(\mathrm{G})^5/5!][L(\mathrm{B})^2/2!]$, which is the result the formula yields.

TABLE 1. Functions defining the valuation model.

Symbol	Name	Represents	Assumption
U	Utility function	Risk preferences	Eq. (4)
$c(x, s)$	Consequence function	Decision problem	Is well-defined and known
$\mathbf{L}(s'\vert s)$	Likelihood matrix	Perceived forecast accuracy	E.g., Eq. (6)

as represented by the simplifying assumptions about model democracy.

In specifying the model, assumptions about three key quantities have been made, listed in Table 1. I believe that the specifications of these quantities given above are the simplest and most useful for generating qualitative results on the dependencies of forecast value. This assertion is of course open to debate, and other specifications—in particular, of the likelihood matrix—would no doubt be interesting to pursue.

Once these quantities are defined, the model is closed by specifying the parameters listed in Table 2. There may also be additional parameters, such as an initial wealth level, that arise from the specifics of the consequence function for the decision being considered.

Although it is possible that, when combined with careful econometric work to estimate the parameters r and λ, the model could be used to give quantitative estimates of the value of ensemble forecasting systems for particular user groups, I feel that its true value lies in providing qualitative information about the dependence of value on the above parameters. The manner in which the value function V changes as a function of its parameters gives an indication of which groups of users are likely to respond to forecast information and is useful information for informing policy and forecast uptake strategies. Such qualitative information can also be used to set targets for trustworthiness and hence forecast accuracy. These points are illustrated in a simple example below.

4. Application to cost–loss decisions

To illustrate the application of the model, I focus on the cost–loss decision problem much studied in the literature (e.g., Zhu et al. 2002). As noted above, this is a somewhat artificial problem in that it does not admit any hedging opportunities. It will, however, be sufficient to illustrate some of the insights the model provides.

TABLE 2. Parameters of the model, which interact to determine perceived forecast value.

Parameter	Interpretation
r	Coef of relative risk aversion
λ	Perceived model accuracy
$p_c(s)$	Prior beliefs, e.g., climatology
N	Ensemble size

In the cost–loss decision problem, the user has the choice to protect herself (at a cost C) against an adverse event that occurs with climatological probability p_c and that causes losses L if it occurs and the user is unprotected. If the user is protected, those losses are diminished by an amount A, where $C < A \leq L$. The consequences of the various possible outcomes are represented in Table 3, which specifies the consequence function $c(x, s)$.

Suppose now that one is most interested in how the user's perception of forecast accuracy affects perceived forecast value. To facilitate comparison with existing forecast value scores for this problem, I assume a risk-neutral user ($r = 0$ so that the utility function is linear). The results obtained can be easily extended to include the effects of risk aversion. In this simple case, however, one can show that if the user is an expected utility maximizer she decides to protect if she believes the probability p of the event occurring satisfies

$$p > C/A. \tag{16}$$

Because one is interested in isolating the effect of λ, a forecast value score that allows for direct comparison of perceived forecast value against a forecasting system believed to be perfectly accurate, that is, $\lambda = 1$, is defined. Let the perceived value of a forecasting system that is believed to have accuracy λ be V_λ and define $z \equiv C/A$. The forecast value score S_λ is defined as[4]

$$S_\lambda = V_\lambda/V_1. \tag{17}$$

It can be shown (see the derivation in the appendix) that

[4] Note that, unlike the economic value scores familiar from the literature (e.g., Zhu et al. 2002), there is no need to subtract the economic value of climatological information, because the Bayesian updating procedure incorporates the effects of climatological knowledge automatically by using it as the prior.

TABLE 3. Matrix of consequences for the cost–loss scenario.

	Protect	Do not protect
Event occurs	$-C - (L - A)$	$-L$
No event	$-C$	0

$$S_\lambda = \frac{E_{\pi_i}[\theta(p_i - z)(p_i - z)] - \theta(p_c - z)(p_c - z)}{z(1 - p_c)\theta(p_c - z) + p_c(1 - z)\theta(z - p_c)}, \tag{18}$$

where the expectation is taken over all possible forecasts π_i,[5] p_i is the believed probability of the event once forecast π_i has been received, and $\theta(y)$ is the Heaviside step function that is equal to 1 for positive y and 0 otherwise. Notice that S_λ depends on the parameters of the decision problem only through the cost–benefit ratio z. The score S_λ is plotted as a function of λ for several values of z in Fig. 1, with $p_c = 0.3$.

It is clear from the figure that S_λ is symmetric about $\lambda = 1/2$. At first this seems like a surprising result, but a moment's reflection suggests that this must be so: believing the forecast is not accurate gives the user the same amount of useful information in this binary decision as does believing it is accurate. Let us thus concentrate on the region $\lambda > 1/2$ from now on, because the user is not expected to be making use of the forecast in the hope that she will be able to profit by how poorly it does. Next, observe that for all values of z (except $z = p_c$) there is a range of λ for which $S_\lambda = 0$—the forecast has no value. For a particular value of z, given p_c, there is some critical value of λ, call it $\lambda^*(z)$, at which the forecasts begin to have value.[6] In other words, the user would have to believe that the probability of the forecast being correct is greater than λ^* for the forecasting system to have any value. It is interesting to ask how the critical value λ^* depends on the decision parameter z. This will give an indication of the minimum amount of trust the user will have to have in the forecast for her to believe it to have any value. In Fig. 2, $\lambda^*(z)$ is plotted for several values of p_c.

This figure suggests several interesting and perhaps counterintuitive results. The first thing to notice is that $\lambda^*(z)$ is not a monotonically increasing function. If it is remembered that z represents the cost–benefit ratio of taking protective action against the adverse event, it seems reasonable to expect that as z increases so to will

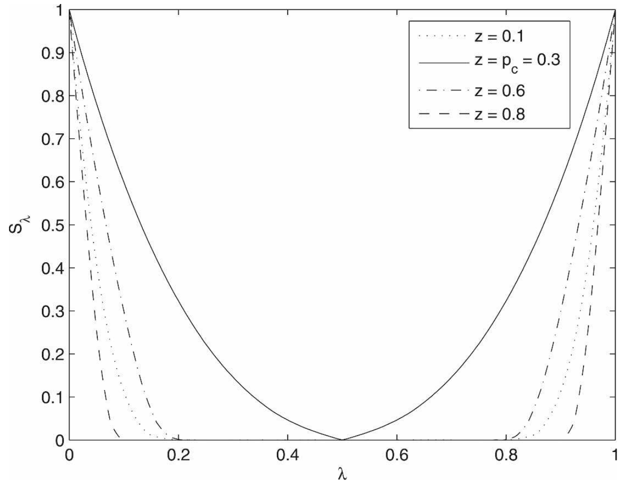

FIG. 1. The dependence of the perceived forecast value score S_λ on λ ($p_c = 0.3$; $N = 20$).

$\lambda^*(z)$. The intuition behind this is as follows: The larger z is, the less likely the user is to choose the protective action. For the forecasting system to have any value, it must recommend at least one course of action that is different from the action the user would have taken without the forecast, that is, based on her knowledge of the climatological probability alone. Now as z increases, the user needs more and more convincing to take protective action, because her margin for error decreases. She will thus need to place increasing amounts of trust in the forecast for it to alter her beliefs sufficiently for her to change her actions. Thus one would expect $\lambda^*(z)$ to increase with z. Why then are all the curves plotted in Fig. 2 increasing for $z > p_c$ but decreasing for $z < p_c$? The answer lies in the same intuition, but in reverse. When z is small, the user will

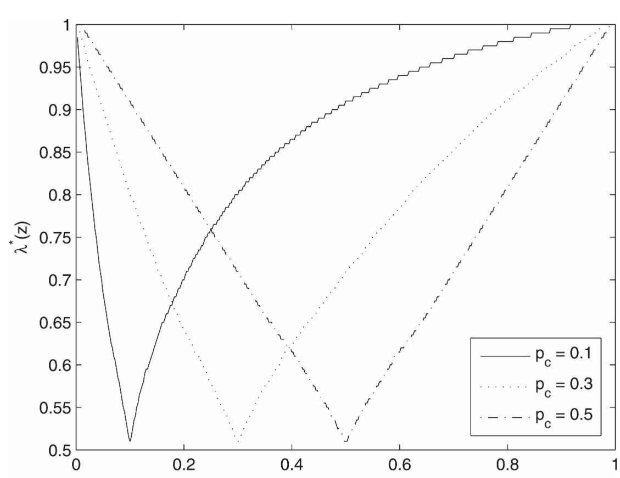

FIG. 2. The dependence of λ^* on z ($N = 20$).

[5] In the case of a binary decision problem the forecast density $\underline{\pi}_i = (\pi_i, 1 - \pi_i)$, and so nothing is lost by referring to it as the number π_i rather than the full density $\underline{\pi}_i$.

[6] It is assumed that $\lambda^* > 0.5$; that is, one is interested in the upper boundary of the region where $S_\lambda = 0$.

almost always take the protective action, because its potential benefits far outweigh its costs. For the forecasting system to be perceived as valuable, the user will have to be convinced that there is at least one forecast for which her actions would differ from what they would have been without the forecast. As z decreases, it becomes increasingly difficult to convince her, and she needs to have increasing amounts of trust in the forecasting system for there to be at least one such forecast. Thus, since $\lambda^*(z)$ decreases for small z and increases for large z, there must be a critical value of z at which its behavior changes and jumps from decreasing to increasing. What Fig. 2 shows is that this critical value is exactly $z = p_c$, the climatological probability of the adverse event. In general, the closer z is to p_c, the less the user needs to trust the forecast for it to have value for her. This suggests that decisions that are in some sense "aligned" with the climatological probability of the event will tend to get more value from forecasts than those that are not.

Figures 1 and 2 encode several insights that may be relevant for understanding patterns of forecast uptake, at least for cost–loss decisions. First, as was illustrated in Fig. 1, regardless of the values of z or p_c (except when $z = p_c$), there is invariably a set of values of λ for which the perceived value of the forecasts is zero. Suppose now that we seek to increase forecast uptake by investing in the development of climate models with increased accuracy. Then the model suggests that even though an objective increase in the accuracy in the forecasting system may lead to an increase in its perceived accuracy this does not necessarily imply a corresponding increase in perceived forecast value. If one assumes, as seems reasonable, that it is perceived value that determines uptake, then it is seen that scientific improvements in forecasting products need not lead to increases in the rate of forecast uptake. Only once a critical value of trust is reached [given by $\lambda^*(z)$] can we be assured that increases in forecast accuracy lead to increases in perceived value. Thus, if we wish scientific improvements in forecasting products to be converted into real gains in forecast usage, we had better make sure that the trust users have in the forecasting system exceeds the required threshold level. How can this be achieved? As has been seen, the required threshold is dependent on the parameters of the user's decision problem, as encoded by z, the cost–benefit ratio. Figure 2 suggests that it may be no easy task to raise a user's beliefs about forecast accuracy above the required threshold if z is much larger or much smaller than p_c. This is not necessarily due to any significant deficiency on the part of the forecasting system, but rather is due to the fact that for these values of z the user requires a very high degree of trust in the system for it to change her actions. It is important to remember that it generally costs users something—time and effort—to learn how to make use of a forecast and to decide how best to apply it to their particular decision problem. They cannot be expected to invest in learning how to apply the product unless it will have a material effect on their actions in a large number of circumstances. Thus users with a moderate degree of trust in the forecasting system, but with values of z far from p_c, will actually lose out by learning to use the forecasts, because there will be few if any circumstances in which having access to forecast information will alter their choices. In fact, encouraging users with high/low z to use the forecasts may require them to hold beliefs about the forecast accuracy that are not supported by objective measures; that is, they would need to believe the forecasts to be more accurate than they actually are. In such cases, users who perceive the forecasts to be valuable put themselves at risk—perceived value may be greater than actual realized value. This suggests that, when attempts are made to increase rates of forecast uptake, they should target those users whose decision parameters would allow them to profit from increased forecast use without putting themselves at risk, that is, those users for whom $\lambda^*(z)$ is less than an objective measure of the accuracy of the ensemble.[7] The range of values of z for which this is achieved is plotted as a function of objective average ensemble accuracy in Fig. 3, for $p_c = 0.1$. The endpoints of this range are given by taking the inverse of $\lambda^*(z)$.

The message of this figure is that interventions aimed at increasing rates of forecast uptake need to be sensitive to the details of an individual user's decision problem. For some, getting them to trust the forecast sufficiently for it to be seen as valuable will require overselling its accuracy. For a given value of the objective forecast accuracy, interventions are best aimed at those users who have values of z that lie inside the shaded area in Fig. 3.

5. Conclusions

This paper aimed to provide a flexible tool for understanding the factors that contribute to users' perception of the value of ensemble forecasts and hence their rate of forecast use. A model that is sensitive to key features of the forecasting product, user characteristics,

[7] Objective measures of ensemble accuracy will, in general, depend on the details of the ensemble. For our purposes, one can simply imagine taking the average of the accuracies of the individual ensemble members.

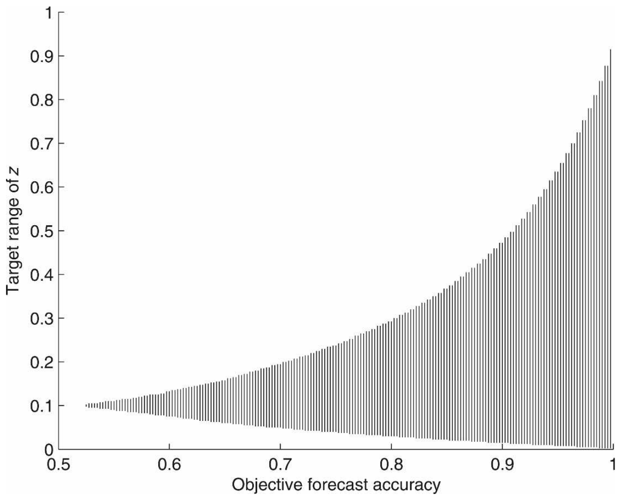

FIG. 3. The range of values of z for which attempts should be made to increase forecast uptake, as a function of objective forecast accuracy ($p_c = 0.1$; $N = 20$).

the decision problem itself, and geographical parameters such as the climatological distribution of the forecast variable was put forward. The model has potentially wide application as a qualitative tool for understanding the dependencies of forecast value on these factors. The additional insights it is capable of providing were illustrated by considering the simple cost–loss decision problem that has formed the basis of most economic value scores in the meteorological literature. It was shown that one of the parameters of the model, the perceived forecast accuracy λ, can have a dramatic, and sometimes counterintuitive, impact on forecast value. In most circumstances, high values of λ are required for the forecast product to be perceived as having any value at all, but for decisions with a cost–benefit ratio close to the climatological probability of the event lower values of λ can be tolerated. The results obtained suggest several insights that may be useful for explaining forecast uptake patterns. The model suggests that it is possible for users' appraisal of the value of a forecast product to remain the same even when they believe that forecast accuracy has increased, but that, once users have a critical amount of trust in the forecast accuracy, increases in perceived accuracy are accompanied by increases in perceived value. It was shown that this critical value of perceived accuracy is very sensitive to the parameters of the user's decision problem and is, in general, high for values of the cost–benefit ratio that are far from the climatological probability of the event and low for values close to the climatological probability. It was suggested that attempts to encourage increased forecast use should be restricted to those users whose decision parameters naturally allow them to perceive the forecasts as valuable without having to attribute to the forecasts more accuracy than they actually have.

The cost–loss scenario considered above is a very restricted application of the model to an extremely simple decision problem that does not exploit its full power. Future work will be directed toward exploring the dependence of forecast value on the other model parameters for more realistic decision problems. There are several interesting applications that immediately suggest themselves. Understanding the influence of the user's wealth and risk preferences on perceived value is of great practical significance for understanding responses to forecast information among different user groups. People in economically depressed situations or managers in conservative institutions are known to be relatively strongly risk averse (Dercon 2002; Rayner et al. 2005), whereas the wealthy can afford to be more risk tolerant as they tend to have access to a large number of economic safety nets. These economic considerations can have important effects on perceived forecast value and go a long way toward explaining forecast uptake patterns. In addition, there are interesting questions relating to the influence of forecast presentation on perceived value. Are there optimal ways of presenting the output from an ensemble to the user, and, if so, do they depend on the underlying climatological distribution of the forecast variable? Also, what is the effect of ensemble size on perceived value? These factors will be explored in the context of continuous decision problems, such as the choice between two assets whose returns are dependent on the forecast variable. These more complex decisions allow hedging behavior and are of more immediate relevance to many real-world scenarios.

Acknowledgments. Many thanks are given to Richard Washington and Diana Liverman for helpful discussions and comments. The financial support of the Commonwealth Scholarship Commission and NRF is gratefully acknowledged.

APPENDIX

Derivation of S_λ

The expression in Eq. (18) for S_λ can be derived as follows:

Recall that the user takes action if she believes that the probability p of the adverse event occurring satisfies $p > C/A$, and that it has been assumed that the user is risk neutral ($r = 0$) so that she has a linear utility function. Using the consequence matrix in Table 3, if the user decides to protect, the expected utility from her action is $p(A - C - L) + (1 - p)(-C)$, whereas, if she decides not to protect, her expected utility is just $p(-L)$. Thus, in general, her expected utility EU can be written as

$$EU = \theta(p - C/A)[p(A - C - L) + (1 - p)(-C)] + [1 - \theta(p - C/A)]p(-L) \tag{A1}$$

$$= \theta(p - C/A)(pA - C) - pL, \tag{A2}$$

where $\theta(x)$ is the Heaviside step function. Using this expression and the fact that the utility function is linear, one can solve for V in the forecast value Eq. (11):

$$V_\lambda = E_{\pi_i}[\theta(p_i - C/A)(p_i A - C) - p_i L] - \theta(p_c - C/A)(p_c A - C) - p_c L \tag{A3}$$

$$= A\{E_{\pi_i}[\theta(p_i - z)(p_i - z)] - \theta(p_c - z)(p_c - z)\} + L(p_c - E_{\pi_i} p_i), \tag{A4}$$

where the expectation is taken over all possible forecasts π_i, p_i are the user's beliefs given forecast π_i as given by Eq. (9), and $z = C/A$. Note that the expression multiplied by A in the first term is the numerator of Eq. (18). It is now shown that the second term is identically zero.

Suppose that the forecast for the adverse event is π_i and that perceived forecast accuracy is λ. Then the ensemble forecast update Eq. (9) tells us that

$$p_i = p(\text{event occurs}|\pi_i) = \frac{\lambda \pi_i p_c}{\lambda p_c + (1 - \lambda)(1 - p_c)} + \frac{(1 - \lambda)(1 - \pi_i)p_c}{(1 - \lambda)p_c + \lambda(1 - p_c)}. \tag{A5}$$

I now want to compute the expected value of this quantity over all possible forecasts π_i, where the probability of receiving forecast π_i is given by Eq. (15). To do this all I need do is compute

$$\sum_{\pi_i} q[\pi_i]\pi_i, \tag{A6}$$

because the expression in Eq. (A5) is linear in π_i. Summing over all forecasts is especially simple in this case because there are only two forecast categories. I begin by computing $q[\pi_i]$. Write the forecast $\pi_i = i/N$, that is, i out of N ensemble members forecast the adverse event. Then

$$q[\pi_i] = \mathcal{P}\left(\pi_i = \frac{i}{N}|\text{event occurs}\right)p_c + \mathcal{P}\left(\pi_i = \frac{i}{N}|\text{no event}\right)(1 - p_c) \tag{A7}$$

$$= \binom{N}{i}\lambda^i(1 - \lambda)^{N-i}p_c + \binom{N}{N-i}(1 - \lambda)^i\lambda^{N-i}(1 - p_c), \tag{A8}$$

where

$$\binom{N}{i} = \frac{N!}{i!(N - i)!}$$

is a binomial coefficient. To take all possible forecasts into account in the expectation of π_i, I must now sum over all possible values of i. Making use of the fact that

$$\binom{N}{i} = \binom{N}{N-i}$$

and that the mean value of the binomial distribution with parameters (N, λ) is $N\lambda$, I find that

$$\sum_{\pi_i} q[\pi_i]\pi_i = \sum_{i=0}^{N} q[\pi_i]\frac{i}{N} = \lambda p_c + (1 - \lambda)(1 - p_c). \tag{A9}$$

Taking the expected value over all π_i in Eq. (A5) and substituting the above result for the expectation of π_i, I find that

$$E_{\pi_i} p_i = p_c. \tag{A10}$$

Hence the term that depends on L in Eq. (A4) is zero.

Last, to compute S_λ, I need to find V_1. Note from Eq. (A8) that, when $\lambda = 1$,

$$q\left[\pi_i = \frac{i}{N}\right] = \begin{cases} 1 - p_c & i = 0 \\ 0 & 0 < i < N. \\ p_c & i = N \end{cases} \tag{A11}$$

This follows from direct substitution into Eq. (A8). Using this result for $q[\pi_i]$, and the fact that $p_i = \pi_i$ when $\lambda = 1$, one can verify that

$$\sum_{\pi_i} q[\pi_i]\theta(p_i - z)(p_i - z) = p_c(1 - z) \tag{A12}$$

and hence

$$V_1 = A[p_c(1 - z) - \theta(p_c - z)(p_c - z)] \tag{A13}$$

$$= A[z(1 - p_c)\theta(p_c - z) + p_c(1 - z)\theta(z - p_c)]. \tag{A14}$$

Thus,

$$S_\lambda = V_\lambda / V_1 \tag{A15}$$

$$= \frac{E_{\pi_i}[\theta(p_i - z)(p_i - z)] - \theta(p_c - z)(p_c - z)}{z(1 - p_c)\theta(p_c - z) + p_c(1 - z)\theta(z - p_c)}, \tag{A16}$$

which is Eq. (18).

REFERENCES

Adams, R., K. Bryant, B. McCarl, D. Legler, J. O'Brien, A. Solow, and R. Weiher, 1995: Value of improved long-range weather information. *Contemp. Econ. Policy,* **13,** 10–19.

Artikov, I., and Coauthors, 2006: Understanding the influence of climate forecasts on farmer decisions as planned behavior. *J. Appl. Meteor. Climatol.,* **45,** 1202–1214.

Bharwani, S., M. Bithell, T. E. Downing, M. New, R. Washington, and G. Ziervogel, 2005: Multi-agent modelling of climate outlooks and food security on a community garden scheme in Limpopo, South Africa. *Philos. Trans. Roy. Soc. London,* **360,** 2183–2194.

Dercon, S., 2002: Income risk, coping strategies, and safety nets. *World Bank Res. Obs.,* **17,** 141–166.

Gollier, C., 2001: *The Economics of Risk and Time.* MIT Press, 444 pp.

Grothmann, T., and A. Patt, 2005: Adaptive capacity and human cognition: The process of individual adaptation to climate change. *Global Environ. Change,* **15A,** 199–213.

Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina, 2000: Potential benefits of climate forecasting to agriculture. *Agric. Ecosyst. Environ.,* **82,** 169–184.

Kahneman, D., 2003: Maps of bounded rationality: Psychology for behavioral economics. *Amer. Econ. Rev.,* **93,** 1449–1475.

——, and A. Tversky, 1979: Prospect theory: An analysis of decision under risk. *Econometrica,* **47,** 263–292.

Katz, R., and A. Murphy, 1997: *Economic Value of Weather and Climate Forecasts.* 1st ed. Cambridge University Press, 222 pp.

Letson, D., G. P. Podesta, C. D. Messina, and R. A. Ferreyra, 2005: The uncertain value of perfect ENSO phase forecasts: Stochastic agricultural prices and intra-phase climatic variations. *Climatic Change,* **69,** 163–196.

Luseno, W. K., J. G. McPeak, C. B. Barrett, P. D. Little, and G. Gebru, 2003: Assessing the value of climate forecast information for pastoralists: Evidence from southern Ethiopia and northern Kenya. *World Dev.,* **31,** 1477–1494.

Machina, M. J., 1987: Choice under uncertainty: Problems solved and unsolved. *J. Econ. Perspect.,* **1,** 121–154.

Mas-Colell, A., M. D. Whinston, and J. R. Green, 1995: *Microeconomic Theory.* Oxford University Press, 981 pp.

Messina, C. D., J. W. Hansen, and A. J. Hall, 1999: Land allocation conditioned on El Niño–Southern Oscillation phases in the Pampas of Argentina. *Agric. Syst.,* **60,** 197–212.

Mjelde, J., and M. Cochran, 1988: Obtaining lower and upper bounds on the value of seasonal climate forecasts as a function of risk preferences. *West. J. Agric. Econ.,* **13,** 285–293.

——, H. S. J. Hill, and J. F. Griffiths, 1998: A review of current evidence on climate forecasts and their economic effects in agriculture. *Amer. J. Agric. Econ.,* **80,** 1089–1095.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.,* **128,** 747–774.

Patt, A., and C. Gwata, 2002: Effective seasonal climate forecast applications: Examining constraints for subsistence farmers in Zimbabwe. *Global Environ. Change,* **12,** 185–195.

——, P. Suarez, and C. Gwata, 2005: Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. *Proc. Natl. Acad. Sci. USA,* **102,** 12 623–12 628.

——, L. Ogallo, and M. Hellmuth, 2007: Learning from 10 years of climate outlook forums in Africa. *Science,* **318,** 49–50.

Quiggin, J., 2001: Production under uncertainty and choice under uncertainty in the emergence of generalized expected utility theory. *Theor. Decis.,* **51,** 125–144.

Rayner, S., D. Lach, and H. Ingram, 2005: Weather forecasts are for wimps: Why water resource managers do not use climate forecasts. *Climatic Change,* **69,** 197–227.

Richardson, D., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.,* **127,** 2473–2489

Roncoli, C., K. Ingram, and P. Kirshen, 2002: Reading the rains: Local knowledge and rainfall forecasting in Burkina Faso. *Soc. Nat. Resour.,* **15,** 409–427.

Saha, S., and Coauthors, 2006: The NCEP climate forecast system. *J. Climate,* **19,** 3483–3517.

Vogel, C., and K. O'Brien, 2006: Who can eat information? Examining the effectiveness of seasonal climate forecasts and regional climate-risk management strategies. *Climate Res.,* **33,** 111–122.

WCRP, 2005: The World Climate Research Programme strategic framework 2005–2015: Coordinated Observation and Prediction of the Earth System (COPES). WCRP-123, WMO/TD 1291, 44 pp.

Wilks, D. S., 2001: A skill score based on economic value for probability forecasts. *Meteor. Appl.,* **8,** 209–219.

——, 2006: *Statistical Methods in the Atmospheric Science.* 2nd ed. Academic Press, 627 pp.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.,* **83,** 73–83.

Ziervogel, G., and R. Calder, 2003: Climate variability and rural livelihoods: Assessing the impact of seasonal climate forecasts in Lesotho. *Area,* **35,** 403–417.