On the Relationship between the Accuracy and Value of Forecasts in the Cost–Loss Ratio Situation

ALLAN H. MURPHY AND MARTIN EHRENDORFER

Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331

(Manuscript received 23 January 1987, in final form 22 June 1987)

ABSTRACT

This paper explores the relationship between the quality and value of imperfect forecasts. It is assumed that these forecasts are produced by a primitive probabilistic forecasting system and that the decision-making problem of concern is the cost-loss ratio situation. In this context, two parameters describing basic characteristics of the forecasts must be specified in order to determine forecast quality uniquely. As a result, a scalar measure of accuracy such as the Brier score cannot completely and unambiguously describe the quality of the imperfect forecasts. The relationship between forecast accuracy and forecast value is represented by a multivalued function—an accuracy/value envelope. Existence of this envelope implies that the Brier score is an imprecise measure of value and that forecast value can even decrease as forecast accuracy increases (and vice versa). The generality of these results and their implications for verification procedures and practices are discussed.

1. Introduction

Meteorologists frequently use measures such as the mean square error and mean absolute error to evaluate the performance of forecasters or forecasting systems. For example, the Brier score (Brier, 1950)—the mean square error of probabilistic forecasts—has been employed in many operational and experimental programs to determine the accuracy of such forecasts. In addition, forecasting systems that produce probabilistic forecasts are often compared in terms of their respective Brier scores (or corresponding skill scores). In these studies, it is implicitly assumed that smaller Brier scores are indicative of forecasts of higher quality. Moreover, the Brier score has frequently been used as a surrogate for a measure of the economic value of forecasts, in the sense that smaller Brier scores are also assumed to correspond to forecasts of greater value.

The primary purposes of this paper are to demonstrate that, in general, no single measure of performance such as the Brier score can completely and unambiguously describe forecast quality and to show that each numerical value of the Brier score is associated with a range of economic payoffs even in simple decision-making problems. This latter result implies that the Brier score is an imprecise measure of value and, moreover, that forecast value can actually decrease as the Brier score decreases (i.e., as accuracy increases). These results appear to have important implications for the practice of forecast verification and for the use of traditional measures of performance as indicators of forecast value.

To analyze the relationships between measures of the performance of forecasting systems and measures of the value of such systems it is necessary to identify a specific decision-making problem within which to perform the analysis. Here we consider the familiar prototype problem usually referred to as the cost–loss ratio situation (Thompson, 1962; Murphy, 1977). In section 2 we briefly describe this problem, the types of forecasts of interest, and the expected expenses associated with the use of these forecasts. The quality and value of the forecasts are discussed in section 3. This section identifies the basic determinants of forecast quality, specifies a particular measure of forecast accuracy—the Brier score—and discusses its inherent deficiencies as a measure of quality, and defines measures of value. In section 4 we explore the relationship between these measures of accuracy and value and illustrate this relationship by examining specific numerical examples. Section 5 contains a brief summary of the principal results, as well as a discussion of their generality and their implications for verification procedures and practices.

2. Basic considerations: Cost–loss ratio situation, types of forecasts, and expected expenses

The cost–loss ratio situation is a decision-making problem involving two possible actions—protect ($a = 1$) and do not protect ($a = 0$)—and two possible events—adverse weather ($w = 1$) and no adverse weather ($w = 0$). The decision maker is assumed to incur a cost C (>0) if protective action is taken, a loss L (>0) if protective action is not taken and adverse weather occurs, and no cost or loss otherwise. It is convenient to transform these expenses into expenses per unit loss, by dividing each expense by L. The transformed expenses associated with the four possible combinations of actions and events are then defined as follows: C/L for $a = 1$ and $w = 1$ or $w = 0$, one for $a = 0$ and $w = 1$, and zero for $a = 0$ and $w = 0$. To avoid trivial cases we assume that $C < L$. Thus, $0 < C/L < 1$.

Three types of weather forecasts are considered here: imperfect forecasts, climatological forecasts, and perfect forecasts. Imperfect forecasts are assumed to consist of categorical forecasts of the events, where $f = 1$ is a categorical forecast of adverse weather and $f = 0$ is a categorical forecast of no adverse weather. These forecasts can be characterized in terms of two sets of probabilities: (i) four conditional probabilities and (ii) two marginal or predictive probabilities. The conditional probabilities specify the likelihood of occurrence of the events given the forecasts; thus, $p_{11} = \Pr(w = 1|f = 1)$, $p_{01} = \Pr(w = 0|f = 1)$, $p_{10} = \Pr(w = 1|f = 0)$, and $p_{00} = \Pr(w = 0|f = 0)$, where $p_{11} + p_{01} = 1$ and $p_{10} + p_{00} = 1$. The predictive probabilities specify the relative frequency of use of the respective forecasts; thus, $\pi_1 = \Pr(f = 1)$ and $\pi_0 = \Pr(f = 0)$, where $\pi_1 + \pi_0 = 1$. For the purposes of this paper, it is convenient to assume that $p_{11} \geqslant p_{10}$ (or, equivalently, that $p_{01} \leqslant p_{00}$). This assumption implies that adverse weather ($w = 1$) is more likely to follow a forecast of adverse weather ($f = 1$) than to follow a forecast of no adverse weather ($f = 0$). No loss of generality occurs by making this assumption.

Climatological and perfect forecasts represent limiting cases of the imperfect forecasts. Specifically, climatological forecasts correspond to the limiting case of imperfect forecasts in which $p_{11} = p_{10} = p_1$ and $p_{01} = p_{00} = p_0$, where $p_1 = \Pr(w = 1)$ and $p_0 = \Pr(w = 0)$ are the climatological probabilities of adverse weather and no adverse weather, respectively ($p_1 + p_0 = 1$). In this case, the probability of adverse weather (no adverse weather) given the forecast does not depend on the forecast and is equal to the climatological probability of adverse weather (no adverse weather). Perfect forecasts correspond to the limiting case of imperfect forecasts in which $p_{11} = p_{00} = 1$, $p_{01} = p_{10} = 0$, $\pi_1 = p_1$, and $\pi_0 = p_0$. In this case, adverse weather (no adverse weather) always follows a forecast of adverse weather (no adverse weather), and the frequencies of use of the two possible forecasts are necessarily equal to the respective climatological probabilities.

The decision maker is assumed to choose the action that minimizes expected expense, where expected expense is the probability-weighted average of the relevant expenses (costs and/or losses). In the case of climatological forecasts, for example, the expected expense associated with action $a = 1$ (protect) is $e_1 = p_1(C/L) + p_0(C/L) = C/L$ and the expected expense associated with action $a = 0$ (do not protect) is $e_0 = p_1(1) + p_0(0) = p_1$. Thus, the decision criterion implies that the decision maker will take action $a = 1$ if $p_1 > C/L$ and action $a = 0$ if $p_1 < C/L$ (hence the name "cost–loss ratio situation").

Under the assumption that the decision maker adopts the relevant information as the sole basis for choosing the optimal action, expected expense expressions can be readily derived for the three types of forecasts. These expressions, which are denoted by EF, EC, and EP for imperfect, climatological, and perfect forecasts, respectively, are reproduced in appendix A. As indicated in appendix A, the expected expense associated with perfect forecasts is less than or equal to the expected expense associated with imperfect forecasts and the latter is less than or equal to the expected expense associated with climatological forecasts.

3. Quality and value: Definitions and measures

a. *Forecast quality and a measure of accuracy*

In order to realize the objectives of this paper, the (minimum) number of parameters needed to describe the quality of the imperfect forecasts must be specified. As indicated in section 2, these forecasts consist of four conditional probabilities—p_{11}, p_{01}, p_{10}, and p_{00}—and two predictive probabilities—π_1 and π_0, for a total of six parameters. However, since $p_{11} + p_{01} = 1$, $p_{10} + p_{00} = 1$, and $\pi_1 + \pi_0 = 1$, at most three of these six parameters are required to characterize the imperfect forecasts. Moreover, the climatological probabilities—p_1 and p_0 ($p_1 + p_0 = 1$)—are assumed to be known. The relationships among the conditional, predictive, and climatological probabilities, described in appendix B, can then be used to achieve a further reduction in the number of "independent" parameters. As a result, only two parameters must be specified to describe, completely and unambiguously, the quality of the imperfect forecasts. Thus, p_{11} and π_1, or p_{11} and p_{10}, represent basic determinants of forecast quality in this context (assuming that p_1, the climatological probability of adverse weather, is also known).

A natural measure of the overall accuracy of probabilistic forecasting systems that produce two-event forecasts is the Brier score. As noted in section 1, the Brier score is the mean square error of such forecasts and possesses several desirable properties (see Murphy and Daan, 1985). Here we use the expected half Brier score (BS) as the measure of accuracy; this measure is defined and briefly described in appendix C. In particular, appendix C shows that BS can be expressed in terms of the parameters p_{11} and π_1, the basic determinants of forecast quality employed in this paper.

Values of BS are plotted as a function of p_{11} and π_1 in Fig. 1a for the case in which $p_1 = 0.3$ and $C/L = 0.2$ (hereafter referred to as example 1). The ranges of possible values of p_{11} and π_1 are determined by the conditions associated with this case, and these conditions will be described in section 3b. Perfect forecasts arerepresented by the point $p_{11} = 1$ and $\pi_1 = 0.3$ (BS $= 0$), whereas climatological forecasts are represented

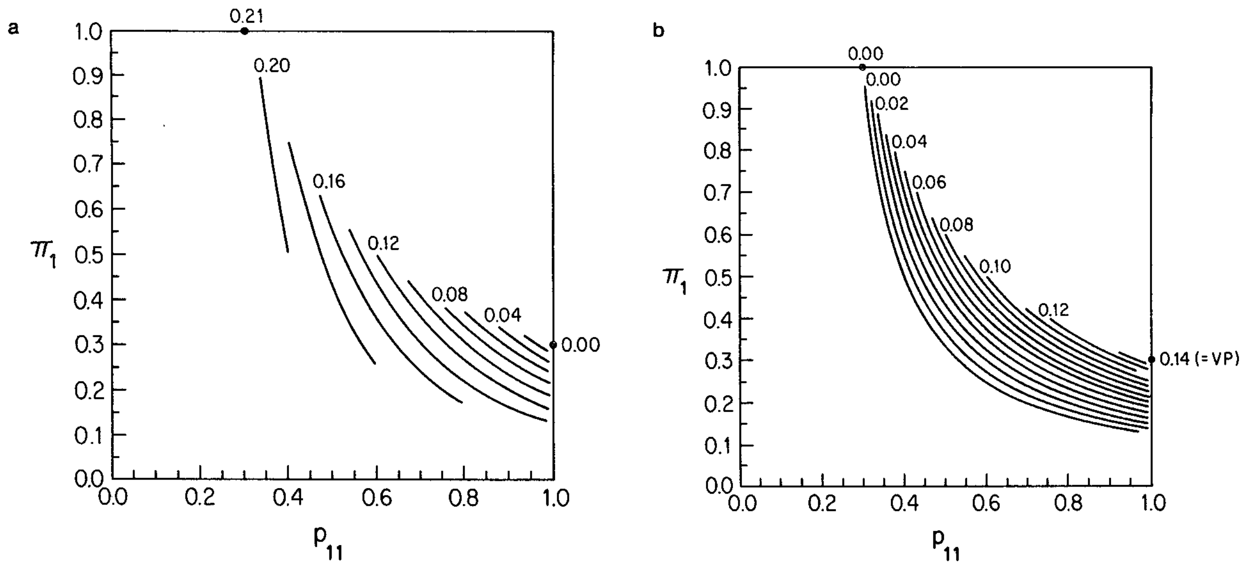

FIG. 1. Isopleths of (a) BS and (b) VF, both as functions of p_{11} and π_1, for example 1.

by the point $p_{11} = 0.3$ and $\pi_1 = 1$ (BS = 0.21). The isopleths of BS in Fig. 1a reveal that different combinations of values of the basic determinants of forecast quality—p_{11} and π_1—can produce the same Brier score. Since these distinct pairs of values of p_{11} and π_1 represent, by definition, different levels of quality, it is evident that BS is not a unique measure of the quality of the imperfect forecasts. It is important to make a clear distinction between forecast quality, as characterized by the basic parameters of the forecasting system (i.e., p_{11} and π_1), and forecast accuracy, as measured by the Brier score.

b. Forecast value and measures of value

The value of the imperfect forecasts depends on the expected expense associated with the use of these forecasts and the expected expense associated with the use of the information consulted by the decision maker in the absence of these forecasts (Winkler and Murphy, 1985). Since it is reasonable to assume that climatological forecasts would always be available to the decision maker, the value of imperfect forecasts can be defined as the difference in expected expenses between the situation involving climatological forecasts and the situation involving imperfect forecasts. If VF denotes the expected value of imperfect forecasts, then

$$\mathrm{VF} = \mathrm{EC} - \mathrm{EF}. \quad (1)$$

Analogously, if VP denotes the expected value of perfect forecasts, then

$$\mathrm{VP} = \mathrm{EC} - \mathrm{EP}. \quad (2)$$

Expressions for VF and VP in terms of p_{11}, π_1, p_1, and C/L are presented in appendix D. As noted in this appendix, the expected value of imperfect forecasts is nonnegative and is less than or equal to the expected value of perfect forecasts.

To facilitate this discussion and the investigation of the relationship between measures of accuracy and value in section 4, it will be convenient to distinguish among several different situations, depending on the values of particular parameters. First, recall that it has been assumed, without loss of generality, that $0 \leq p_{10} \leq p_1 \leq p_{11} \leq 1$. Four cases can be identified, depending on the values of the conditional and climatological probabilities relative to the numerical value of the cost–loss ratio, and these cases are described in Table 1. In addition, this table contains expressions for VF and VP in each case. Note that VF and VP are both equal to zero in cases I and IV. Imperfect and perfect forecasts are of no value in these cases because all of the probabilities, conditional and climatological, lead to the same optimal action [$a = 1$ (protect) in case I and $a = 0$ (do not protect) in case IV]. Forecasts are of positive economic value only if they can lead to different actions than those that the decision maker would have taken in the absence of the forecasts (Winkler and Murphy,

TABLE 1. Expected value expressions for imperfect forecasts (VF) and perfect forecasts (VP) for various cases defined in terms of conditions on the conditional probabilities (p_{11} and p_{10}), climatological probability (p_1), and cost-loss ratio (C/L).

Case: Conditions	VF	VP
I. $0 \leq C/L \leq p_{10} \leq p_1 \leq p_{11} \leq 1$	0	0
II: $0 \leq p_{10} \leq C/L \leq p_1 \leq p_{11} \leq 1$	$\pi_0[(C/L) - p_{10}]$	$(1 - p_1)(C/L)$
III: $0 \leq p_{10} \leq p_1 \leq C/L \leq p_{11} \leq 1$	$\pi_1[p_{11} - (C/L)]$	$p_1[1 - (C/L)]$
IV: $0 \leq p_{10} \leq p_1 \leq p_{11} \leq C/L \leq 1$	0	0

1985). Obviously, we are primarily concerned here with situations associated with cases II and III, in which VF and VP are generally greater than zero.

Values of VF are plotted and isoplethed as a function of p_{11} and π_1 in Fig. 1b for example 1 ($p_1 = 0.3$ and $C/L = 0.2$), a specific instance of case II in which $p_{10} \leq C/L$ (=0.2). The ranges of values of p_{11} and π_1 are determined by the conditions associated with this case. Specifically, this region is bounded above by the curve $\pi_1 p_{11} = p_1 = 0.3$ and below by the curve VF = 0, where VF $= \pi_0[(C/L) - p_{10}] = \pi_1[p_{11} - (C/L)] - p_1 + C/L = \pi_1(p_{11} - 0.2) - 0.1$ in this case (see Table 1). Both sets of isopleths are asymptotic to the vertical line $p_{11} = p_1 = 0.3$ on the left and to the horizontal line $\pi_1 = (p_1 - p_{10})/(p_{11} - p_{10}) = 0.125$ below. Perfect forecasts are represented by the point $p_{11} = 1$ and $\pi_1 = 0.3$ (VF = VP = 0.14), whereas climatological forecasts are represented by the point $p_{11} = 0.3$ and $\pi_1 = 1$ (VF = 0). As indicated in Fig. 1b, different pairs of values of p_{11} and π_1 can yield the same value of VF.

4. Relationships between measures of accuracy and value

To illustrate the relationship between the measure of accuracy and the measure of value, we have overlayed the isopleths of BS (solid curves) and VF (dashed curves) for example 1 in Fig. 2. These two sets of isopleths exhibit similar behavior: BS decreases and VF increases as p_{11} increases *and* π_1 decreases or remains constant over the ranges of permissible values. However, the slopes of the isopleths of BS are steeper than

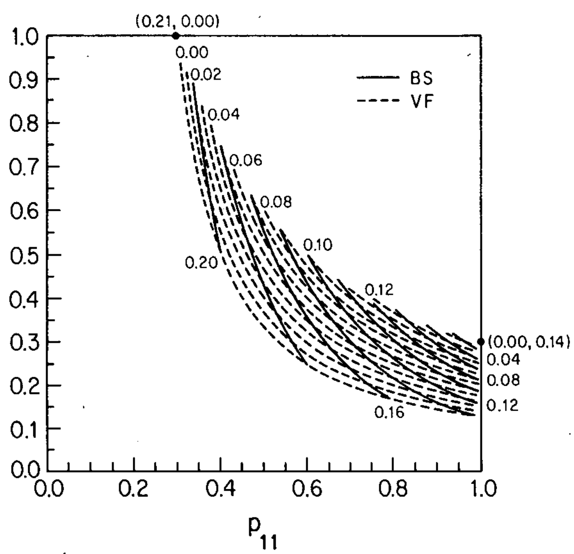

FIG. 2. Overlay of isopleths of BS (solid curves) and VF (dashed curves) for example 1.

the slopes of the isopleths of VF. Since several isopleths of VF generally intersect each isopleth of BS, a range of values of VF is associated with each value of BS. Analogously, several isopleths of BS usually intersect each isopleth of VF, and thus a range of values of the former is associated with each value of the latter. The existence of such ranges of values indicates that the relationship between BS and VF cannot be described by a single-valued function.

Before examining the accuracy/value relationship directly, we briefly consider another example—namely, case III ($0 \leq p_{10} \leq p_1 \leq C/L \leq p_{11} \leq 1$), with $p_1 = 0.15$ and $C/L = 0.20$ (hereafter referred to as example 2). Isopleths of BS and VF, as a function of p_{11} and π_1, are both depicted in Fig. 3 (BS, solid curves; VF, dashed curves). The region of permissible values of p_{11} and π_1 in example 2 is bounded above by the curve $\pi_1 p_{11} = p_1 = 0.15$, below by the horizontal line $\pi_1 = 0$, and on the left by the vertical line $p_{11} = C/L = 0.20$. As in example 1, the slopes of the isopleths of BS are steeper than the slopes of the isopleths of VF, with the intersection of these isopleths implying that VF is not a single-valued function of BS (or vice versa). In example 2, perfect forecasts are represented by the point $p_{11} = 1$ and $\pi_1 = 0.15$ (BS = 0 and VF = VP = 0.12), whereas climatological forecasts are represented by the point $p_{11} = 0.20$ and $\pi_1 = 0$ (BS = 0.1275 and VF = 0).

The relationship between BS and VF in example 1 is described by the "envelope" depicted in Fig. 4. This diagram was produced by computing both BS and VF for all possible values of p_{11} and π_1 (on a two-dimensional grid with a step size equal to 0.005 in each dimension) that satisfy the conditions for this case (see Table 1). (The "points" that constitute the stippling used to define the envelope in Fig. 4 and the other accuracy/value envelopes presented in this paper do *not* correspond to the grid points associated with the computational procedure.) The multivalued nature of the accuracy/value relationship is vividly illustrated by this envelope. Note that the envelope is quite wide for intermediate values of BS for which isopleths of BS intersect many isopleths of VF, whereas it is rather narrow for very small and very large values of BS for which isopleths of BS intersect only a few isopleths of VF (see Fig. 2). Specifically, when BS = 0.15, VF ranges from 0.00 to approximately 0.08, more than one-half of the total range of VF (recall that VP = 0.14 in this example). Moreover, the range of values of VF exceeds 0.03 for all values of BS between 0.050 and 0.019, with narrower ranges of VF only for smaller and larger values of BS. Clearly, knowledge of BS alone generally provides only a rough estimate of VF (and vice versa).

Figure 5 depicts the accuracy/value envelope for example 2. This envelope was obtained in the same manner as the envelope presented in Fig. 4. Although the range of values of VF for specific values of BS is gen-

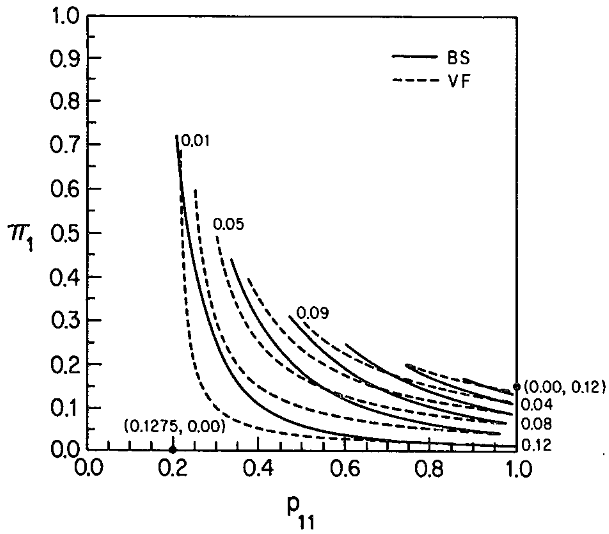

FIG. 3. Overlay of isopleths of BS (solid curves) and VF (dashed curves) for example 2.

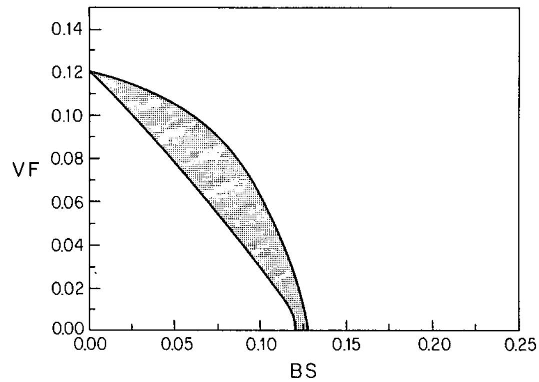

FIG. 5. Accuracy/value envelope for example 2.

erally narrower in example 2 than in example 1, this range (of VF) still exceeds 0.02 for a wide range of values of BS. Once again, knowledge of BS provides an imprecise estimate of VF.

In computing the pairs of values of BS and VF used to describe the accuracy/value envelopes in examples 1 and 2, we have considered all possible values of p_{11} and π_1, subject only to the conditions imposed on these (and other) parameters by the case in question (see Table 1). However, it might be of interest to investigate the effect on the accuracy/value envelopes of imposing restrictions on the values of one or more parameters. In this regard, Katz and Murphy (1987) assumed that $\pi_1 = p_1$ in their study of the quality/value relationship in the cost–loss ratio situation. This assumption implies that, when the forecasts are interpreted as categorical forecasts, the relative frequency of use of the forecast of adverse weather ($f = 1$) is equal to the relative frequency of occurrence of adverse weather ($w = 1$). Alternatively, if the forecasts are interpreted as primitive probabilistic forecasts (see Murphy, 1986), then this assumption implies that forecasts specifying a relatively high probability of adverse weather are issued on the same number of occasions as that on which adverse weather actually occurs (recall that $p_{11} \geq p_{10}$). (Obviously, this assumption does not necessarily imply that $f = 1$ when $w = 1$ or that p_{11} is used only when $w = 1$.) Although such an assumption is not unreasonable, it appears quite restrictive in a general study of quality/value relationships. A less restrictive assumption might involve limiting the range of values of π_1 with respect to p_1. In example 1 ($p_1 = 0.3$), π_1 ranges from 0.125 when $p_{11} = 1$ to one when $p_{11} = 0.3$ (see Fig. 2). Thus, the ratio of π_1 to p_1 ranges from 0.427 to 3.333 in this example. Suppose that we restrict this ratio, r, to the interval from 0.667 to 1.333, in which case $0.2 < \pi_1 < 0.4$. How will the accuracy/value relationship in this example be affected by restricting the values of π_1 in this manner?

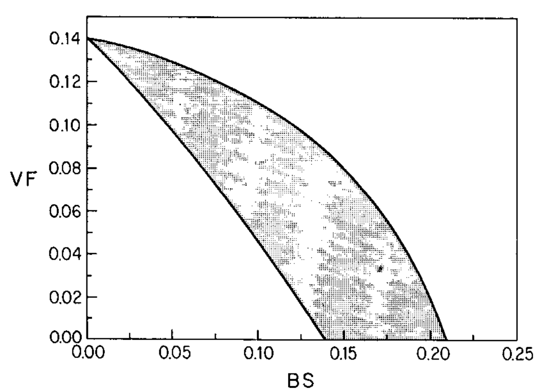

FIG. 4. Accuracy/value envelope for example 1.

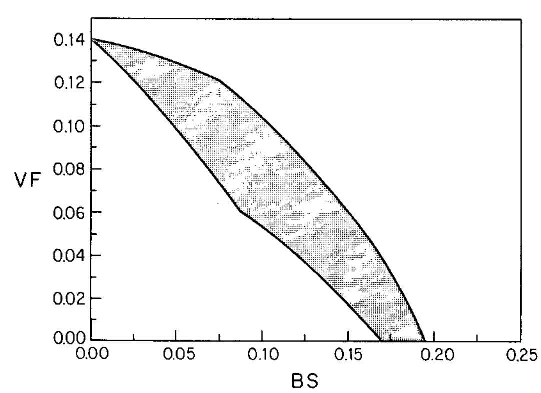

FIG. 6. Accuracy/value envelope for example 1, with $0.2 < \pi_1 < 0.4$.

The accuracy/value envelope under this modified set of conditions is depicted in Fig. 6. Note that the range of values of VF for a specific value of BS (and vice versa) is generally narrower under these modified conditions than under the original conditions. The reduction in the width of the range of values of VF is especially noticeable for larger values of BS. An analogous reduction in the range of values of VF occurs in example 2 when the values of r (and thus π_1) are restricted in a similar manner (this diagram is omitted to conserve space). Thus, if it is known that the ratio of π_1 to p_1 falls in a restricted range of values, this knowledge generally will reduce the range of values of VF for a specific value of BS (and vice versa). Similar results would be obtained if restrictions were placed on the values of p_{11}. Nevertheless, the accuracy/value relationship would still be characterized by an envelope of values of BS and VF, unless π_1 (or p_{11}) is limited to a specific numerical value.

In this regard, it may be of interest to compare the accuracy/value relationships in the cases considered here with the quality/value relationship obtained by Katz and Murphy (1987). As noted previously, the latter assumed that $\pi_1 = p_1$. They also assumed that the parameters p_1 and C/L are specified, as we have done in this paper. Katz and Murphy (1987) then showed that VF is a single-valued linearly increasing function of p_{11} in situations such as those represented by examples 1 and 2. If we assume that $\pi_1 = p_1$ in the context of example 1 or example 2, what will be the nature of the resulting accuracy/value relationship? In effect, we would be taking a "slice" along the horizontal line $\pi_1 = p_1$ through the isopleths of BS and VF in diagrams such as Fig. 2 or Fig. 3. Note that BS decreases (i.e., accuracy increases) and VF increases from left to right along this line.

Figure 7 depicts the accuracy/value relationships, where accuracy is measured by BS and value is measured by VF, for example 1 (solid curve) and example 2 (dashed curve), when $\pi_1 = p_1$. These relationships are characterized by single-valued nonlinear functions, with VF increasing as BS decreases. (For these single-valued functions, the points for which VF = 0 do not represent climatological forecasts since such forecasts generally do not correspond to the case in which $\pi_1 = p_1$.) The relationships are nonlinear because BS is the measure of accuracy employed here and BS is a nonlinear function of p_{11} [see (C2)]. Katz and Murphy (1987) used p_{11} itself as the measure of quality and, as previously noted (see also appendix D), VF is a linear function of this parameter (assuming that π_1, p_1, and C/L are specified). Figure 7 illustrates that complete knowledge of one of the basic parameters that characterize forecast quality reduces the accuracy/value envelope to a single-valued functional relationship between BS and VF. Only in such cases is BS a complete measure of forecast quality. Moreover, the cost–loss ratio must be specified in these cases (as is assumed here) before BS is also a suitable surrogate for forecast value.

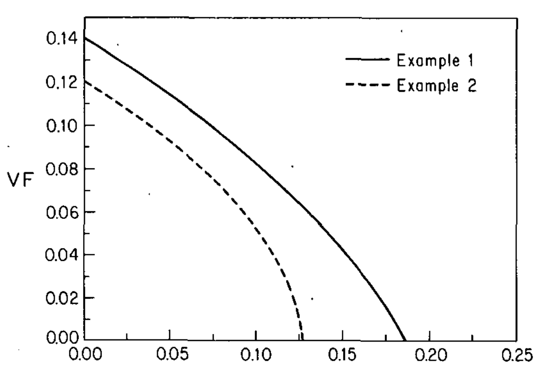

FIG. 7. Accuracy/value curves for example 1 (solid curve) and example 2 (dashed curve), both with $\pi_1 = p_1$.

5. Discussion and conclusion

In this paper we have (i) described the quality and value of forecasts for a simple forecasting system within the context of a prototype decision-making problem; (ii) defined measures of accuracy and value for imperfect forecasts in this context; and (iii) investigated accuracy/value relationships for these forecasts. It has been demonstrated that, for the imperfect forecasts considered here, two parameters (probabilities) must be specified to describe the quality of the forecasts completely and unambiguously. Thus, in general, a one-dimensional measure of accuracy such as the Brier score cannot uniquely characterize forecast quality; different pairs of values of the parameters (i.e., different levels of quality) will lead to the same numerical value of the Brier score. Since the value of the forecasts depends on their quality (i.e., the basic parameters), it should not be surprising that the relationship between accuracy, as measured by the Brier score, and value, as measured by the expected economic value, is found to be a multivalued function. Specifically, this accuracy/value relationship is characterized by an *envelope* of values, with a relatively wide range of forecast value associated with each level of forecast accuracy (and vice versa). Some knowledge of the ranges of possible values of the basic parameters that characterize forecast quality can reduce the width of this envelope. Nevertheless, forecast accuracy when measured solely by a scalar (i.e., one-dimensional) summary measure such as the Brier score generally provides only a rough estimate of forecast value. Moreover, the existence of these accuracy/value envelopes implies that value can actually decrease as accuracy increases (i.e., as the Brier score decreases).

Very simple forecasting systems and decision-making problems are considered in this paper. However, the issues of primary concern here—the distinction between forecast quality and forecast accuracy and the multivalued nature of the relationship between accuracy and value—also arise in situations involving more complex forecasting systems and/or decision-making problems. For example, many more than two parameters would be required, in general, to characterize the quality of a *probabilistic* forecasting system, and no single measure of accuracy such as the Brier score could uniquely describe forecast quality in this context. Moreover, the value of such forecasts would also depend on the specification of a large number of these parameters (i.e., conditional and predictive probabilities), as well as on the specification of the cost–loss ratio or other payoff structure.

At this point, it should be noted that most previous studies of quality/value relationships (e.g., Brown et al., 1986; Katz et al., 1982; Murphy et al., 1985), including an investigation involving the cost–loss ratio situation (Katz and Murphy, 1987), have obtained single-valued relationships between forecast quality and forecast value. These studies yielded single-valued quality/value relationships because of simplifying assumptions, according to which both quality and value depended on only a single parameter such as the variance of the forecasts. For example, Katz and Murphy (1987) assumed that $\pi_1 = p_1$, in which case forecast quality and forecast value are completely described by p_{11} (or p_{10}). In general, however, the quality and value of forecasts both depend on two or more parameters. In this regard, Chen et al. (1987) recently demonstrated the existence of accuracy/value envelopes in the *generalized* cost–loss ratio situation.

What are the implications of the results presented here for verification procedures and practices? First, these results reveal an important deficiency in one-dimensional verification measures in situations in which quality is necessarily multidimensional in character. For example, when quality is inherently two-dimensional, as in the cases considered here, a one-dimensional measure of accuracy such as the Brier score cannot uniquely determine forecast quality. In this sense, then, it is not necessarily true that smaller Brier scores are indicative of forecasts of higher quality. These and other related deficiencies in traditional verification measures have been noted in a recent paper by Murphy and Winkler (1987), in which they described a general framework for forecast verification. This framework is based on the joint distribution of forecasts and observations and on factorizations of this distribution into conditional and marginal distributions. From the perspective provided by this framework, and from the results presented in this paper, it is evident that information regarding basic parameters of forecasting systems is needed to describe forecast quality completely.

Second, we have demonstrated the existence of accuracy/value *envelopes* in the cost–loss ratio situation. In effect, the measure of accuracy (the Brier score in this case) is an interval measure of value. Since this interval is generally quite wide, the Brier score provides only a rough estimate of forecast value. Moreover, the existence of accuracy/value envelopes implies that increases in accuracy (i.e., decreases in the Brier score) may even lead to decreases in value. This reversal in the usual accuracy/value relationship (in which value increases or remains constant as accuracy increases) is possible because a range of economic payoffs, as measured by VF, exists for a given numerical value of accuracy, as measured by BS. (This possibility arises because the Brier score is necessarily an incomplete measure of multidimensional quality.) When this result is considered in conjunction with the fact that quality/value relationships are inherently nonlinear (e.g., see Katz and Murphy, 1987), it is clear that forecast accuracy is *not* an appropriate surrogate for forecast value.

The existence of accuracy/value envelopes also has important implications for the information that the meteorological community should provide to actual and potential users of the forecasts. Clearly, users need more than overall measures of accuracy (such as the Brier score) to judge both the scientific quality of the forecasts and their operational utility. The inadequacy of such measures from the perspective of users of the forecasts has also been noted by Murphy and Winkler (1987).

Acknowledgments. We would like to thank Richard W. Katz and two anonymous reviewers for helpful comments on an earlier version of this paper. This work was supported in part by the National Science Foundation (Division of Atmospheric Sciences) under Grant ATM-8507495.

APPENDIX A

Expected Expenses for Imperfect, Climatological, and Perfect Forecasts

In this appendix we present expressions for the expected expenses associated with the use of imperfect, climatological, and perfect forecasts, under the assumption that the decision maker chooses the action ($a = 1$ or $a = 0$) that minimizes expected expense. If EF denotes the expected expense (per unit loss) associated with imperfect forecasts, then $\mathrm{EF} = \min(C/L, p_{11})$ when $f = 1$ and $\mathrm{EF} = \min(C/L, p_{10})$ when $f = 0$. The overall expected expense is the weighted average of these (conditional) expected expenses, where the weights are the probabilities of the respective forecasts. Thus,

$$\mathrm{EF} = \pi_1 \min(C/L, p_{11}) + \pi_0 \min(C/L, p_{10}). \quad \text{(A1)}$$

Recall that climatological forecasts represent the special case of imperfect forecasts in which $p_{11} = p_{10} = p_1$. If EC denotes the expected expense associated with climatological forecasts, then

$$EC = \min(C/L, p_1) \tag{A2}$$

(recall that $\pi_1 + \pi_0 = 1$). Analogously, perfect forecasts represent the special case of imperfect forecasts in which $p_{11} = 1$, $p_{10} = 0$, $\pi_1 = p_1$, and $\pi_0 = p_0$. Thus, if EP denotes the expected expense associated with perfect forecasts, then

$$EP = p_1(C/L). \tag{A3}$$

Since climatological forecasts and perfect forecasts represent lower and upper bounds, respectively, on the quality of imperfect forecasts, it follows that EP $\leqslant$ EF $\leqslant$ EC. Thus, EC and EP represent upper and lower bounds, respectively, on EF.

APPENDIX B

Relationships among Conditional, Predictive, and Climatological Probabilities

Consideration of the definitions of the conditional, predictive, and climatological probabilities and the basic laws of probability indicates that certain relationships exist among these three types of probabilities. Specifically,

$$p_1 = \pi_1 p_{11} + \pi_0 p_{10} \tag{B1}$$

$$p_0 = \pi_1 p_{01} + \pi_0 p_{00}. \tag{B2}$$

[Since addition of (B1) and (B2) yields an identity, (B2) is redundant with respect to (B1), and vice versa.] In other words, the climatological probabilities (p_1 or p_0) are the weighted averages of the conditional probabilities (p_{11} and p_{10} or p_{01} and p_{00}), where the weights are the predictive probabilities (π_1 and π_0). When (B1) and (B2) are considered in conjunction with the assumption that $p_{11} \geqslant p_{01}$ ($p_{10} \leqslant p_{00}$) (see section 2), it follows that $0 \leqslant p_{10} \leqslant p_1 \leqslant p_{11} \leqslant 1$ ($0 \leqslant p_{01} \leqslant p_0 \leqslant p_{00} \leqslant 1$).

APPENDIX C

Measure of Forecast Accuracy

In the context of this paper, the expected half Brier score (BS) can be written as follows:

$$BS = \pi_1[p_{11}(p_{11} - 1)^2 + (1 - p_{11})p_{11}^2] + \pi_0[p_{10}(p_{10} - 1)^2 + (1 - p_{10})p_{10}^2]. \tag{C1}$$

It has been assumed in (C1) that p_{11} represents both the conditional probability of adverse weather given a forecast of adverse weather and the relative frequency with which the event $w = 1$ occurs when the forecast is $f = 1$ (analogous statements hold for p_{01}, p_{10}, and p_{00}). This assumption is equivalent to assuming that the forecasts are primitive (i.e., two-value) probabilistic forecasts and that they are completely reliable. (A probability forecast p is completely reliable when the event of concern occurs on a fraction p of the occasions on which the forecast p is assigned to this event.) As defined in (C1), BS ranges from zero for perfect forecasts ($p_{11} = p_{00} = 1$ and $p_{01} = p_{10} = 0$) to $p_1(1 - p_1)$ for climatological forecasts ($p_{11} = p_{10} = p_1$ and $p_{01} = p_{00} = p_0$). Thus, smaller values of BS are indicative of more accurate forecasts.

We have carefully distinguished in this paper between the quality of the imperfect forecasts, as characterized by basic parameters of the forecasting system, and the accuracy of these forecasts, as measured by the Brier score. To clarify this distinction, we note here that BS in (C1) can be rewritten as follows:

$$BS = p_1(1 - p_1) - [\pi_1/(1 - \pi_1)](p_{11} - p_1)^2, \tag{C2}$$

since $p_{10} = (p_1 - \pi_1 p_{11})/\pi_0$ [see (B1)] and $\pi_1 = 1 - \pi_0$. In (C2), we have expressed BS in terms of two basic determinants of forecast quality, p_{11} and π_1 (recall that the climatological probability, p_1, is assumed to be known).

APPENDIX D

Measures of Forecast Value

If it is assumed that climatological forecasts would always be available to the decision maker, then the expected value of imperfect (perfect) forecasts can be defined as the difference in expected expenses between the situation involving climatological forecasts and the situation involving imperfect (perfect) forecasts. Thus, if VF denotes the value of imperfect forecasts (per unit loss), then VF = EC − EF, or from (A1) and (A2),

$$VF = \min(C/L, p_1) - \pi_1 \min(C/L, p_{11}) - \pi_0 \min(C/L, p_{10}). \tag{D1}$$

Analogously, if VP denotes the value of perfect forecasts, then VP = EC − EP, or from (A2) and (A3),

$$VP = \min(C/L, p_1) - p_1(C/L). \tag{D2}$$

In view of the fact that EP $\leqslant$ EF $\leqslant$ EC (see appendix A), it follows that VF $\leqslant$ VP. That is, the value of imperfect forecasts is nonnegative and is less than or equal to the value of perfect forecasts.

REFERENCES

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Brown, B. G., R. W. Katz and A. H. Murphy, 1986: On the economic value of seasonal precipitation forecasts: The fallowing/planting problem. *Bull. Amer. Meteor. Soc.*, **67**, 833–841.

Chen, Y.-S., M. Ehrendorfer and A. H. Murphy, 1987: On the relationship between the quality and value of forecasts in the generalized cost–loss ratio situation. *Mon. Wea. Rev.*, **115**, 1534–1541.

Katz, R. W., and A. H. Murphy, 1987: Quality/value relationship for imperfect information in the umbrella problem. *The American Statistician*, **41**, (in press).

——, —— and R. L. Winkler, 1982: Assessing the value of frost forecasts to orchardists: A dynamic decision-making approach. *J. Appl. Meteor.*, **21**, 518–531.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost–loss ratio situation. *Mon. Wea. Rev.*, **105**, 803–816.

——, 1986: Comparative evaluation of categorical and probabilistic forecasts: Two alternatives to the traditional approach. *Mon. Wea. Rev.*, **114**, 245–249.

——, and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 379–437.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.

——, R. W. Katz, R. L. Winkler and W.-R. Hsu, 1985: Repetitive decision making and the value of forecasts in the cost–loss ratio situation: A dynamic model. *Mon. Wea. Rev.*, **113**, 801–813.

Thompson, J. C., 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. *J. Appl. Meteor.*, **1**, 13–17.

Winkler, R. L., and A. H. Murphy, 1985: Decision analysis. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 493–524.