Chapter 2

FORECASTING AND DECISION THEORY

CLIVE W.J. GRANGER and MARK J. MACHINA

Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508

Contents

Handbook of Economic Forecasting, Volume 1
Edited by Graham Elliott, Clive W.J. Granger and Allan Timmermann

DOI: 10.1016/S1574-0706(05)01002-5

Abstract

When forecasts of the future value of some variable, or the probability of some event, are used for purposes of ex ante planning or decision making, then the preferences, opportunities and constraints of the decision maker will all enter into the ex post evaluation of a forecast, and the ex post comparison of alternative forecasts. After a presenting a brief review of early work in the area of forecasting and decision theory, this chapter formally examines the manner in which the features of an agent's decision problem combine to generate an appropriate *decision-based loss function* for that agent's use in forecast evaluation. Decision-based loss functions are shown to exhibit certain necessary properties, and the relationship between the functional form of a decision-based loss function and the functional form of the agent's underlying utility function is characterized. In particular, the standard squared-error loss function is shown to imply highly restrictive and not particularly realistic properties on underlying preferences, which are *not* justified by the use of a standard local quadratic approximation. A class of more realistic loss functions ("location-dependent loss functions") is proposed.

Keywords

forecasting, loss functions, decision theory, decision-based loss functions

JEL classification: C440, C530

Preface

This chapter has two sections. Section 1 presents a fairly brief history of the interaction of forecasting and decision theory, and Section 2 presents some more recent results.

1. History of the field

1.1. Introduction

A decision maker (either a private agent or a public policy maker) must inevitably consider the future, and this requires forecasts of certain important variables. There also exist forecasters – such as scientists or statisticians – who may or may not be operating independently of a decision maker. In the classical situation, forecasts are produced by a single forecaster, and there are several potential users, namely the various decision makers. In other situations, each decision maker may have several different forecasts to choose between.

A decision maker will typically have a payoff or utility function $U(x, \alpha)$, which depends upon some uncertain variable or vector x which will be realized and observed at a future time T, as well as some decision variable or vector α which must be chosen out of a set $\mathcal{A}$ at some earlier time $t < T$. The decision maker can base their choice of α upon a current scalar forecast (a "point forecast") x_F of the variable x, and make the choice $\alpha(x_F) \equiv \arg\max_{\alpha \in \mathcal{A}} U(x_F, \alpha)$. Given the realized value x_R, the decision maker's *ex post* utility $U(x_R, \alpha(x_F))$ can be compared with the maximum possible utility they could have attained, namely $U(x_R, \alpha(x_R))$. This shortfall can be averaged over a number of such situations, to obtain the decision maker's average loss in terms of foregone payoff or utility. If one is forecasting in a stochastic environment, perfect forecasting will not be possible and this average long-term loss will be strictly positive. In a deterministic world, it could be zero.

Given some measure of the loss arising from an imperfect forecast, different forecasting methods can be compared, or different combinations selected.

In his 1961 book *Economic Forecasts and Policy*, Henri Theil outlined many versions of the above type of situation, but paid more attention to the control activities of the policy maker. He returned to these topics in his 1971 volume *Applied Economic Forecasting*, particularly in the general discussion of Chapter 1 and the mention of loss functions in Chapter 2. These two books cover a wide variety of topics in both theory and applications, including discussions of certainty equivalence, interval and distributional forecasts, and non-quadratic loss functions. This emphasis on the links between decision makers and forecasters was not emphasized by other writers for at least another quarter of a century, which shows how farsighted Theil could be. An exception is an early contribution by White (1966).

Another major development was Bayesian decision analysis, with important contributions by DeGroot (1970) and Berger (1985), and later by West and Harrison (1989, 1997). Early in their book, on page 14, West and Harrison state "A statistician, economist or management scientist usually looks at a decision as comprising a forecast or belief, and a utility, or reward, function". Denote Y as the outcome of a future random quantity which is "conditional on your decision α expressed through a forward or probability function $P(Y|\alpha)$. A reward function $u(Y,\alpha)$ expresses your gain or loss if Y happens when you take decision α". In such a case, the expected reward is

$$r(\alpha) = \int u(Y,\alpha)\,\mathrm{d}P(Y|\alpha) \tag{1}$$

and the optimal decision is taken to be the one that maximizes this expected reward. The parallel with the "expected utility" literature is clear.

The book continues by discussing a dynamic linear model (denoted DLM) using a state-space formulation. There are clear similarities with the Kalman filtering approach, but the development is quite different. Although West and Harrison continue to develop the "Bayesian maximum reward" approach, according to their index the words "decision" and "utility" are only used on page 14, as mentioned above. Although certainly important in Bayesian circles, it was less influential elsewhere. This also holds for the large body of work known as "statistical decision theory", which is largely Bayesian.

The later years of the Twentieth Century produced a flurry of work, published around the year 2000. Chamberlain (2000) was concerned with the general topic of econometrics and decision theory – in particular, with the question of how econometrics can influence decisions under uncertainty – which leads to considerations of distributional forecasts or "predictive distributions". Naturally, one needs a criterion to evaluate procedures for constructing predictive distributions, and Chamberlain chose to use risk robustness and to minimize regret risk. To construct predictive distributions, Bayes methods were used based on parametric models. One application considered an individual trying to forecast their future earnings using their personal earnings history and data on the earnings trajectories of others.

1.2. The Cambridge papers

Three papers from the Department of Economics at the University of Cambridge moved the discussion forward. The first, by Granger and Pesaran (2000a), first appeared as a working paper in 1996. The second, also by Granger and Pesaran (2000b), appeared as a working paper in 1999. The third, by Pesaran and Skouras (2002), appeared as a working paper in 2000.

Granger and Pesaran (2000a) review the classic case in which there are two states of the world, which we here call "good" and "bad" for convenience. A forecaster provides a probability forecast $\hat{\pi}$ (resp. $1-\hat{\pi}$) that the good (resp. bad) state will occur. A decision maker can decide whether or not to take some action on the basis of this forecast, and a completely general payoff or profit function is allowed. The notation is illustrated in Table 1. The Y_{ij}'s are the utility or profit payoffs under each state and action, net of any costs of the action. A simple example of states is that a road becoming icy and dangerous

Table 1

Action	State	
	Good	Bad
Yes	Y_{11}	Y_{12}
No	Y_{21}	Y_{22}

is the bad state, whereas the road staying clear is the good state. The potential action could be to add sand to the road. If $\hat{\pi}$ is the forecast probability of the good state, then the action should be undertaken if

$$\frac{\hat{\pi}}{1-\hat{\pi}} > \frac{Y_{22}-Y_{12}}{Y_{11}-Y_{21}}. \tag{2}$$

This case of two states with predicted probabilities of $\hat{\pi}$ and $1-\hat{\pi}$ is the simplest possible example of a predictive distribution. An alternative type of forecast, which might be called an "event forecast", consists of the forecaster simply announcing the event that is judged to have the highest probability. Granger and Pesaran (2000a) show that using an event forecast will be suboptimal compared to using a predictive distribution. Although the above example is a very simple case, the advantages of using an economic cost function along with a decision-theoretic approach, rather than some statistical measure such as least squares, are clearly illustrated.

Granger and Pesaran (2000b) continue their consideration of this type of model, but turn to loss functions suggested for the evaluation of the meteorological forecasts. A well-known example is the Kuipers Score (*KS*) defined by

$$KS = H - F \tag{3}$$

where H is the fraction (over time) of bad events that were *correctly* forecast to occur, and F is the fraction of good events that had been incorrectly forecast to have come out bad (sometimes termed the "false alarm rate"). Random forecasts would produce an average *KS* value of zero. Although this score would seem to be both useful and interpretable, it turns out to have some undesirable properties. The first is that it cannot be defined for a one-shot case, since regardless of the prediction and regardless of the realized event, one of the fractions H or F must take the undefined form $0/0$. A generalization of this undesirable property is that the Kuipers Score cannot be guaranteed to be well-defined for *any* prespecified sample size (either time series or cross-sectional), since for any sample size n, the score is similarly undefined whenever all the realized events are good, or all the realized events are bad.

Although the above properties would appear serious from a theoretical point of view, one might argue that any practical application would involve a prediction history where incorrect forecasts of both types had occurred, so that both H and F would be well-defined. But even in that case, another undesirable property of the Kuipers Score can manifest itself, namely that the neither the score itself, nor its ranking of alternative

Table 2

Year	Realized event	A's forecast	B's forecast	A's 5-year score	B's 5-year score	A's 10-year score	B's 5-year score
1	good	good	good	$H^A_{1-5} = 1$	$H^B_{1-5} = 0$		
2	good	bad	good				
3	good	bad	good	$F^A_{1-5} = \frac{3}{4}$	$F^B_{1-5} = \frac{1}{4}$		
4	good	bad	bad				
5	bad	bad	good	$KS^A_{1-5} = \frac{1}{4}$	$KS^B_{1-5} = -\frac{1}{4}$	$H^A_{1-10} = \frac{2}{5}$	$H^B_{1-10} = \frac{3}{5}$
6	bad	bad	bad	$H^A_{5-10} = \frac{1}{4}$	$H^B_{5-10} = \frac{3}{4}$	$F^A_{1-10} = \frac{3}{5}$	$F^B_{1-10} = \frac{2}{5}$
7	bad	good	bad			$KS^A_{1-10} = -\frac{1}{5}$	$KS^B_{1-10} = \frac{1}{5}$
8	bad	good	bad	$F^A_{5-10} = 0$	$F^B_{5-10} = 1$		
9	bad	good	good				
10	good	good	bad	$KS^A_{5-10} = \frac{1}{4}$	$KS^B_{5-10} = -\frac{1}{4}$		

forecasters, will exhibit the natural uniform dominance property with respect to combining or partitioning sample populations. We illustrate this with the following example, where a 10-element sample is partitioned into two 5-element subsamples, and where the history of two forecasters, A and B, are as given in Table 2. For this data, forecaster A is seen to have a higher Kuipers score than forecaster B for the first five-year period, and also for the second five-year period, but A has a lower Kuipers score than B for the whole decade – a property which is clearly undesirable, whether or not our evaluation is based on an underlying utility function. The intuition behind this occurrence is that the two components H and F of the Kuipers score are given equal weight in the formula $KS = H - F$ even though the number of data points they refer to (the number of periods with realized bad events versus the number of periods with realized good events) needn't be equal, and the fraction of bad versus good events in each of two subperiods can be vastly different from the fraction over the combined period. Researchers interested in applying this type of evaluation measure to situations involving the aggregation/disaggregation of time periods, or time periods of different lengths, would be better off with the simpler measure defined by the overall fraction of events (be they good *or* bad) that were correctly forecast.

Granger and Pesaran (2000b) also examine the relationship between other statistical measures of forecast accuracy and tests of stock market timing, and with a detailed application to stock market data. Models for stock market returns have emphasized expected risk-adjusted returns rather than least-squares fits – that is, an economic rather than a statistical measure of quality of the model.

Pesaran and Skouras (2002) is a survey paper, starting with the above types of results and then extending them to predictive distributions, with a particular emphasis on the role of decision-based forecast evaluation. The paper obtains closed-form results for a variety of random specifications and cost or utility functions, such as Gaussian distributions combined with negative exponential utility. Attention is given to a general survey of the use of cost functions with predictive distributions, with mention of the possible use of scoring rules, as well as various measures taken from meteorology. See also Elliott and Lieli (2005).

Although many of the above results are well known in the Bayesian decision theory literature, they were less known in the forecasting area, where the use of the whole distribution rather than just the mean, and an economic cost function linked with a decision maker, were not usually emphasized.

1.3. Forecasting versus statistical hypothesis testing and estimation

Although the discussion of this chapter is in terms of forecasting some yet-to-be-realized random variable, it will be clear to readers of the literature that most of our analysis and results also apply to the statistical problem of testing a hypothesis whose truth value is already determined (though not yet known), or to the statistical problem of estimating some parameter whose numerical value is also already determined (though not yet observed, or not directly observable). The case of hypothesis testing will correspond to the forecasting of binary events as illustrated in the above table, and that of numerical parameter estimation will correspond to that of predicting a real-valued variable, as examined in Section 2 below.

2. Forecasting with decision-based loss functions

2.1. Background

In practice, statistical forecasts are typically *produced* by one group of agents ("forecasters") and *consumed* by a different group ("clients"), and the procedures and desires of the two groups typically do not interact. After the fact, alternative forecasts or forecast methods are typically *evaluated* by means of statistical loss functions, which are often chosen primarily on grounds of statistical convenience, with little or no reference to the particular goals or preferences of the client.

But whereas *statistical science* is like any other science in seeking to conduct a "search for truth" that is uninfluenced by the particular interests of the end user, *statistical decisions* are like any other decision in that they should be driven by the goals and preferences of the particular decision maker. Thus, if one forecasting method has a lower bias but higher average squared error than a second one, clients with different goals or preferences may disagree on which of the two techniques is "best" – or at least, which one is best for them. Here we examine the process of forecast evaluation from the point of view of serving clients who have a need or a use for such information in making some upcoming decision. Each such situation will generate its own loss function, which is called a *decision-based loss function*.

Although it serves as a sufficient construct for forecast evaluation, a decision-based loss function is *not* simply a direct representation of the decision maker's underlying preferences. A decision maker's ultimate goal is not to achieve "zero loss", but rather, to achieve maximum utility or payoff (or expected utility or expected payoff). Furthermore, decision-based loss functions are not derived from preferences alone: Any decision problem that involves maximizing utility or payoff (or its expectation) is subject to certain opportunities or constraints, and the nature and extent of these opportunities or constraints will also be reflected in its implied decision-based loss function.

The goal here is to provide a systematic examination of the relationship between decision problems and their associated loss functions. We ask general questions, such as "Can every statistical loss function be derived from some well-specified decision problem?" or "How big is the family of decision problems that generate a given loss function?" We can also ask more specific questions, such as "What does the use of *squared-error loss* reveal or imply about a decision maker's underlying decision problem (i.e. their preferences and/or constraints)?" In addressing such questions, we hope to develop a better understanding of the use of loss functions as tools in forecast evaluation and parameter estimation.

The following analysis is based Pesaran and Skouras (2002) and Machina and Granger (2005). Section 2.2 lays out a framework and derives some of the basic categories and properties of decision-based loss functions. Section 2.3 treats the reverse question of deriving the family of underlying decision problems that generate a given loss function, as well as the restrictions on preferences that are implicitly imposed by the selection of specific functional forms, such as squared-error loss or error-based loss. Given that these restrictions turn out to be stronger than we would typically choose to impose, Section 2.4 describes a more general, "location-dependent" approach to the analysis of general loss functions, which preserves most of the intuition of the standard cases. Section 2.5 examines the above types of questions when we replace point forecasts of an uncertain variable with distribution forecasts. Potentially one can extend the approach to partial distribution forecasts such as moment or quantile forecasts, but these topics are not considered here.

2.2. Framework and basic analysis

2.2.1. Decision problems, forecasts and decision-based loss functions

A decision maker would only have a material interest in forecasts of some uncertain variable x if such information led to "planning benefits" – that is, if their optimal choice in some intermediate decision might depend upon this information. To represent this, we assume the decision maker has an *objective function* (either a utility or a profit function) $U(x, \alpha)$ that depends upon the realized value of x (assumed to lie in some closed interval $\mathcal{X} \subset \mathbf{R}^1$), as well as upon some *choice variable* α to be selected out of some closed interval $\mathcal{A} \subset \mathbf{R}^1$ after the forecast is learned, but before x is realized. We thus define a *decision problem* to consist of the following components:

$$\begin{aligned}&\text{uncertain variable} \quad x \in \mathcal{X},\\ &\text{choice variable and choice set} \quad \alpha \in \mathcal{A},\\ &\text{objective function} \quad U(\cdot,\cdot) : \mathcal{X} \times \mathcal{A} \to \mathbf{R}^1.\end{aligned} \tag{4}$$

Forecasts of x can take several forms. A forecast consisting of a single value $x_F \in \mathcal{X}$ is termed a *point forecast*. For such forecasts, the decision maker's *optimal action function* $\alpha(\cdot)$ is given by

$$\alpha(x_F) \equiv \arg\max_{\alpha \in \mathcal{A}} U(x_F, \alpha) \quad \text{all } x_F \in \mathcal{X}. \tag{5}$$

The objective function $U(\cdot,\cdot)$ can be measured in either utils or dollars. When $U(\cdot,\cdot)$ is posited exogenously (as opposed from being derived from a loss function as in Theorem 1), we assume it is such that (5) has interior solutions $\alpha(x_F)$, and also that it satisfies the following conditions on its second and cross-partial derivatives, which ensure that $\alpha(x_F)$ is unique and is increasing in x_F:

$$U_{\alpha\alpha}(x, \alpha) < 0, \quad U_{x\alpha}(x, \alpha) > 0 \quad \text{all } x \in \mathcal{X}, \text{ all } \alpha \in \mathcal{A}. \tag{6}$$

Forecasts are invariably subject to error. Intuitively, the "loss" arising from a forecast value of x_F, when x turns out to have a realized value of x_R, is simply the loss in utility or profit due to the imperfect prediction, or in other words, the amount by which utility or profit falls short of what it would have been if the decision maker had instead possessed "perfect information" and been able to exactly foresee the realized value x_R. Accordingly, we define the *point-forecast/point-realization loss function* induced by the decision problem (4) by

$$L(x_R, x_F) \equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(x_F)\big) \quad \text{all } x_R, x_F \in \mathcal{X}. \tag{7}$$

Note that in defining the loss arising from the imperfection of forecasts, the realized utility or profit level $U(x_R, \alpha(x_F))$ is compared with what it would have been *if the forecast had instead been equal to the realized value* (that is, compared with $U(x_R, \alpha(x_R))$), and *not* with what utility or profit would have been *if the realization had instead been equal to the forecast* (that is, compared with $U(x_F, \alpha(x_F))$). For example, given that a firm faces a realized output price of x_R, it would have been best if it had had this same value as its forecast, and we measure loss relative to this counterfactual. But given that it received and planned on the basis of a price forecast of x_F, it is *not* best that the realized price also come in at x_F, since any *higher* realized output price would lead to *still higher* profits. Thus, there is no reason why $L(x_R, x_F)$ should necessarily be symmetric (or skew-symmetric) in x_R and x_F. Under our assumptions, the loss function $L(x_R, x_F)$ from (7) satisfies the following properties:

$$\begin{aligned}&L(x_R, x_F) \geqslant 0, \quad L(x_R, x_F)|_{x_R = x_F} = 0,\\ &L(x_R, x_F) \text{ is increasing in } x_F \text{ for all } x_F > x_R,\\ &L(x_R, x_F) \text{ is decreasing in } x_F \text{ for all } x_F < x_R.\end{aligned} \tag{8}$$

As noted, forecasts of x can take several forms. Whereas a point forecast x_F conveys information on the general "location" of x, it conveys no information as to x's potential variability. On the other hand, forecasters who seek to formally communicate their own extent of uncertainty, or alternatively, who seek to communicate their knowledge of the stochastic mechanism that generates x, would report a *distribution forecast* $F_F(\cdot)$ consisting of a cumulative distribution function over the interval $\mathcal{X}$. A decision maker receiving a distribution forecast, and who seeks to maximize expected utility or expected profits, would have an optimal action function $\alpha(\cdot)$ defined by

$$\alpha(F_F) \equiv \arg\max_{\alpha \in \mathcal{A}} \int U(x, \alpha)\, \mathrm{d}F_F(x) \quad \text{all } F_F(\cdot) \text{ over } \mathcal{X} \tag{9}$$

and a *distribution-forecast/point-realization loss* function defined by

$$L(x_R, F_F) \equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(F_F)\big) \quad \text{all } x \in \mathcal{X},\ \text{all } F_F(\cdot) \text{ over } \mathcal{X}. \tag{10}$$

Under our previous assumptions on $U(\cdot,\cdot)$, each distribution forecast $F_F(\cdot)$ has a unique *point-forecast equivalent* $x_F(F_F)$ that satisfies $\alpha(x_F(F_F)) = \alpha(F_F)$ [e.g., Pratt, Raiffa and Schaifer (1995, 24.4.2)]. Since the point-forecast equivalent $x_F(F_F)$ generates the same optimal action as the distribution forecast $F_F(\cdot)$, it will lead to the same loss, so that we have $L(x_R, x_F(F_F)) \equiv L(x_R, F_F)$ for all $x_R \in \mathcal{X}$ and all distributions $F_F(\cdot)$ over $\mathcal{X}$.

Under our assumptions, the loss function $L(x_R, F_F)$ from (10) satisfies the following properties, where "increasing or decreasing in $F_F(\cdot)$" is with respect to first order stochastically dominating changes in $F_F(\cdot)$:

$$\begin{aligned} &L(x_R, F_F) \geqslant 0, \quad L(x_R, F_F)|_{x_R = x_F(F_F)} = 0, \\ &L(x_R, F_F) \text{ is increasing in } F_F(\cdot) \text{ for all } F_F(\cdot) \text{ such that } x_F(F_F) > x_R, \\ &L(x_R, F_F) \text{ is decreasing in } F_F(\cdot) \text{ for all } F_F(\cdot) \text{ such that } x_F(F_F) < x_R. \end{aligned} \tag{11}$$

It should be noted that throughout, these loss functions are quite general in form, and are not being constrained to any specific class.

2.2.2. Derivatives of decision-based loss functions

For point forecasts, the optimal action function $\alpha(\cdot)$ from (5) satisfies the first-order conditions

$$U_\alpha\big(x, \alpha(x)\big) \underset{x}{\equiv} 0. \tag{12}$$

Differentiating this identity with respect to x yields

$$\alpha'(x) \equiv -\frac{U_{x\alpha}(x, \alpha(x))}{U_{\alpha\alpha}(x, \alpha(x))} \tag{13}$$

and hence

$$\begin{aligned}\alpha''(x) \equiv &-\frac{U_{xx\alpha}(x,\alpha(x))\cdot U_{\alpha\alpha}(x,\alpha(x)) - U_{x\alpha}(x,\alpha(x))\cdot U_{x\alpha\alpha}(x,\alpha(x))}{U_{\alpha\alpha}(x,\alpha(x))^2}\\ &-\frac{U_{x\alpha\alpha}(x,\alpha(x))\cdot U_{\alpha\alpha}(x,\alpha(x)) - U_{x\alpha}(x,\alpha(x))\cdot U_{\alpha\alpha\alpha}(x,\alpha(x))}{U_{\alpha\alpha}(x,\alpha(x))^2}\cdot\alpha'(x)\\ \equiv &-\frac{U_{xx\alpha}(x,\alpha(x))}{U_{\alpha\alpha}(x,\alpha(x))} + 2\cdot\frac{U_{x\alpha}(x,\alpha(x))\cdot U_{x\alpha\alpha}(x,\alpha(x))}{U_{\alpha\alpha}(x,\alpha(x))^2}\\ &-\frac{U_{x\alpha}(x,\alpha(x))^2\cdot U_{\alpha\alpha\alpha}(x,\alpha(x))}{U_{\alpha\alpha}(x,\alpha(x))^3}. \end{aligned}\tag{14}$$

By (7) and (12), the derivative of $L(x_R, x_F)$ with respect to small departures from a perfect forecast is

$$\left.\frac{\partial L(x_R,x_F)}{\partial x_F}\right|_{x_F=x_R} \equiv -U_\alpha\big(x_R,\alpha(x_F)\big)\big|_{x_F=x_R}\cdot\alpha'(x_F)\big|_{x_F=x_R} \equiv 0. \tag{15}$$

Calculating $L(x_R, x_F)$'s derivatives at general values of x_R and x_F yields

$$\begin{aligned}\frac{\partial L(x_R,x_F)}{\partial x_R} &\equiv U_x\big(x_R,\alpha(x_R)\big) + U_\alpha\big(x_R,\alpha(x_R)\big)\cdot\alpha'(x_R) - U_x\big(x_R,\alpha(x_F)\big),\\ \frac{\partial L(x_R,x_F)}{\partial x_F} &\equiv -U_\alpha\big(x_R,\alpha(x_F)\big)\cdot\alpha'(x_F),\\ \frac{\partial^2 L(x_R,x_F)}{\partial x_R^2} &\equiv U_{xx}\big(x_R,\alpha(x_R)\big) + U_{x\alpha}\big(x_R,\alpha(x_R)\big)\cdot\alpha'(x_R)\\ &\quad + U_{x\alpha}\big(x_R,\alpha(x_R)\big)\cdot\alpha'(x_R) + U_{\alpha\alpha}\big(x_R,\alpha(x_R)\big)\cdot\alpha'(x_R)^2\\ &\quad + U_\alpha\big(x_R,\alpha(x_R)\big)\cdot\alpha''(x_R) - U_{xx}\big(x_R,\alpha(x_F)\big),\\ \frac{\partial^2 L(x_R,x_F)}{\partial x_R\partial x_F} &\equiv -U_{x\alpha}\big(x_R,\alpha(x_F)\big)\cdot\alpha'(x_F),\\ \frac{\partial^2 L(x_R,x_F)}{\partial x_F^2} &\equiv -U_{\alpha\alpha}\big(x_R,\alpha(x_F)\big)\cdot\alpha'(x_F)^2 - U_\alpha\big(x_R,\alpha(x_F)\big)\cdot\alpha''(x_F). \end{aligned}\tag{16}$$

2.2.3. Inessential transformations of a decision problem

One can potentially learn a lot about decision problems or families of decision problems by asking what changes can be made to them without altering certain features of their solution. This section presents a relevant application of this approach.

A transformation of any decision problem (4) is said to be *inessential* if it does not change its implied loss function, even though it may change other attributes, such as the formula for its optimal action function or the formula for its ex post payoff or utility. For point-forecast loss functions $L(\cdot,\cdot)$, there exist two types of inessential transformations:

Inessential relabelings of the choice variable: Given a decision problem with objective function $U(\cdot,\cdot):\mathcal{X}\times\mathcal{A}\rightarrow\mathbf{R}^1$, any one-to-one mapping $\varphi(\cdot)$ from $\mathcal{A}$ into an arbitrary space $\mathcal{B}$ will generate what we term an *inessential relabeling* $\beta=\varphi(\alpha)$ of the choice variable, with objective function $U^*(\cdot,\cdot):\mathcal{X}\times\mathcal{B}^*\rightarrow\mathbf{R}^1$ and choice set $\mathcal{B}^*\subseteq\mathcal{B}$ defined by

$$U^*(x,\beta)\equiv U\big(x,\varphi^{-1}(\beta)\big),\qquad \mathcal{B}^*=\varphi(\mathcal{A})=\big\{\varphi(\alpha)\mid\alpha\in\mathcal{A}\big\}. \tag{17}$$

The optimal action function $\beta(\cdot):\mathcal{X}\rightarrow\mathcal{B}^*$ for this transformed decision problem is related to that of the original problem by

$$\begin{aligned}\beta(x_F)&\equiv\arg\max_{\beta\in\mathcal{B}^*}U^*(x_F,\beta)\equiv\arg\max_{\beta\in\mathcal{B}^*}U\big(x,\varphi^{-1}(\beta)\big)\\&\equiv\varphi\big(\arg\max_{\alpha\in\mathcal{A}}U(x_F,\alpha)\big)\equiv\varphi\big(\alpha(x_F)\big).\end{aligned} \tag{18}$$

The loss function for the transformed problem is the same as for the original problem, since

$$\begin{aligned}L^*(x_R,x_F)&\equiv U^*\big(x_R,\beta(x_R)\big)-U^*\big(x_R,\beta(x_F)\big)\\&\equiv U\big(x_R,\varphi^{-1}\big(\beta(x_R)\big)\big)-U\big(x_R,\varphi^{-1}\big(\beta(x_F)\big)\big)\\&\equiv U\big(x_R,\alpha(x_R)\big)-U\big(x_R,\alpha(x_F)\big)\equiv L(x_R,x_F).\end{aligned} \tag{19}$$

While any one-to-one mapping $\varphi(\cdot)$ will generate an inessential transformation of the original decision problem, there is a unique "most natural" such transformation, namely the one generated by the mapping $\varphi(\cdot)=\alpha^{-1}(\cdot)$, which relabels each choice α with the forecast value x_F that would have led to that choice – we refer to this labeling as the *forecast-equivalent labeling* of the choice variable. Technically, the map $\alpha^{-1}(\cdot)$ is not defined over the entire space $\mathcal{A}$, but just over the subset $\{\alpha(x)\mid x\in\mathcal{X}\}\subseteq\mathcal{A}$ of actions that are optimal for some x. However, that suffices for the following decision problem to be considered an inessential transformation of the original decision problem:

$$\hat{U}(x,x_F)\underset{x,x_F}{\equiv}U\big(x,\alpha(x_F)\big),\qquad \hat{\mathcal{B}}=\varphi(\mathcal{A})=\big\{\varphi(\alpha)\mid\alpha\in\mathcal{A}\big\}. \tag{20}$$

We refer to (20) as the *canonical form* of the original decision problem, note that its optimal action function is given by $\hat{\alpha}(x_F)\equiv x_F$, and observe that $\hat{U}(x,x_F)$ can be interpreted as the formula for the amount of *ex post* utility (or profit) resulting from a realized value of x when the decision maker had optimally responded to a point forecast of x_F.

Inessential transformations of the objective function: A second type of inessential transformation consists of adding an arbitrary function $\xi(\cdot):\mathcal{X}\rightarrow\mathbf{R}^1$ to the original objective function, to obtain a new function $U^{**}(x,\alpha)\equiv U(x,\alpha)+\xi(x)$. Since $U_\alpha(x_F,\alpha)\equiv U^{**}_\alpha(x_F,\alpha)$, the first order condition (12) is unchanged, so the optimal action functions $\alpha^{**}(\cdot)$ and $\alpha(\cdot)$ for the two problems are identical. But since the ex post utility levels for the two problems are related by $U^{**}(x,\alpha^{**}(x_F))\equiv$

$U(x, \alpha(x_F)) + \xi(x)$, their canonical forms are related by $\hat{U}^{**}(x, x_F) \equiv \hat{U}(x, x_F) + \xi(x)$ and $\hat{\mathcal{B}} = \mathcal{A}$, which would, for example, allow $\hat{U}^{**}(x, x_F)$ to be increasing in x when $\hat{U}(x, x_F)$ was decreasing in x, or vice versa. However, the loss functions for the two problems will be identical, since:

$$\begin{aligned} L^{**}(x_R, x_F) &\equiv U^{**}\big(x_R, \alpha^{**}(x_R)\big) - U^{**}\big(x_R, \alpha^{**}(x_F)\big) \\ &\equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(x_F)\big) \equiv L(x_R, x_F). \end{aligned} \tag{21}$$

Theorem 1 below will imply that these two forms, namely inessential relabelings of the choice variable and inessential additive transformations of the objective function, exhaust the class of loss-function-preserving transformations of a decision problem.

2.3. *Recovery of decision problems from loss functions*

In practice, loss functions are typically *not* derived from an underlying decision problem as in the previous section, but rather, are postulated exogenously. But since we have seen that decision-based loss functions inherit certain necessary properties, it is worth asking precisely when a given loss function (or functional form) can or cannot be viewed as being derived from an underlying decision problem. In cases when they can, it is then worth asking about the restrictions this loss function or functional form implies about the underlying utility or profit function or constraints.

2.3.1. *Recovery from point-forecast loss functions*

Machina and Granger (2005) demonstrate that for an arbitrary point-forecast/point-realization loss function $L(\cdot,\cdot)$ satisfying (8), the class of objective functions that generate $L(\cdot,\cdot)$ has the following specification:

THEOREM 1. *For arbitrary function $L(\cdot,\cdot)$ that satisfies the properties* (8)*, an objective function $U(\cdot,\cdot): \mathcal{X} \times \mathcal{A} \to \mathbf{R}^1$ with strictly monotonic optimal action function $\alpha(\cdot)$ will generate $L(\cdot,\cdot)$ as its loss function if and only if it takes the form*

$$U(x, \alpha) \equiv f(x) - L\big(x, g(\alpha)\big) \tag{22}$$

for some function $f(\cdot): \mathcal{X} \to \mathbf{R}^1$ and monotonic function $g(\cdot): \mathcal{A} \to \mathcal{X}$.

This theorem states that an objective function $U(x, \alpha)$ and choice space $\mathcal{A}$ are consistent with the loss function $L(x_R, x_F)$ if and only if they can be obtained from the function $-L(x_R, x_F)$ by one or both of the two types of inessential transformations described in the previous section. This result serves to highlight the close, though not unique, relationship between decision makers' loss functions and their underlying decision problems.

To derive the canonical form of the objective function (22) for given choice of $f(\cdot)$ and $g(\cdot)$, recall that each loss function $L(x_R, x_F)$ is minimized with respect to x_F when

x_F is set equal to x_R, so that the optimal action function for the objective function (22) takes the form $\alpha(x) \equiv g^{-1}(x)$. This in turn implies that its canonical form $\hat{U}(x, x_F)$ is given by

$$\hat{U}(x, x_F) \equiv U\big(x, \alpha(x_F)\big) \equiv f(x) - L\big(x, g\big(\alpha(x_F)\big)\big) \equiv f(x) - L(x, x_F). \tag{23}$$

2.3.2. Implications of squared-error loss

The most frequently used loss function in statistics is unquestionably the *squared-error form*

$$L_{\text{Sq}}(x_R, x_F) \equiv k \cdot (x_R - x_F)^2, \quad k > 0, \tag{24}$$

which is seen to satisfy the properties (8). Theorem 1 thus implies the following result:

COROLLARY 1. *For arbitrary squared-error function $L_{Sq}(x_R, x_F) \equiv k \cdot (x_R - x_F)^2$ with $k > 0$, an objective function $U(\cdot,\cdot) : \mathcal{X} \times \mathcal{A} \to \mathbf{R}^1$ with strictly monotonic optimal action function $\alpha(\cdot)$ will generate $L_{\text{Sq}}(\cdot,\cdot)$ as its loss function if and only if it takes the form*

$$U(x, \alpha) \equiv f(x) - k \cdot \big(x - g(\alpha)\big)^2 \tag{25}$$

for some function $f(\cdot) : \mathcal{X} \to \mathbf{R}^1$ and monotonic function $g(\cdot) : \mathcal{A} \to \mathcal{X}$.

Since utility or profit functions of the form (25) are not particularly standard, it is worth describing some of their properties. One property, which may or may not be realistic for a decision setting, is that changes in the level of the choice variable α do not affect the *curvature* (i.e. the second and higher order derivatives) of $U(x, \alpha)$ with respect to x, but only lead to uniform changes in the *level* and *slope* with respect to x – that is to say, for any pair of values $\alpha_1, \alpha_2 \in \mathcal{A}$, the difference $U(x, \alpha_1) - U(x, \alpha_2)$ is an affine function of x.[1]

A more direct property of the form (25) is revealed by adopting the forecast-equivalent labeling of the choice variable to obtain its canonical form $\hat{U}(x, x_F)$ from (20), which as we have seen, specifies the level of utility or profit resulting from an actual realized value of x and the action that would have been optimal for a realized value of x_F. Under this labeling, the objective function implied by the squared-error loss function $L_{\text{Sq}}(x_R, x_F)$ is seen (by (23)) to take the form

$$\hat{U}(x, x_F) \equiv f(x) - L_{\text{Sq}}(x, x_F) \equiv f(x) - k \cdot (x - x_F)^2. \tag{26}$$

In terms of our earlier example, this states that when a firm faces a realized output price of x, its shortfall from optimal profits due to *having planned* for an output price of x_F only depends upon the difference between x and x_F (and in particular, upon the

[1] Specifically, (25) implies $U(x, \alpha_1) - U(x, \alpha_2) \equiv -k \cdot [g(\alpha_1)^2 - g(\alpha_2)^2] + 2 \cdot k \cdot [g(\alpha_1) - g(\alpha_2)] \cdot x$.

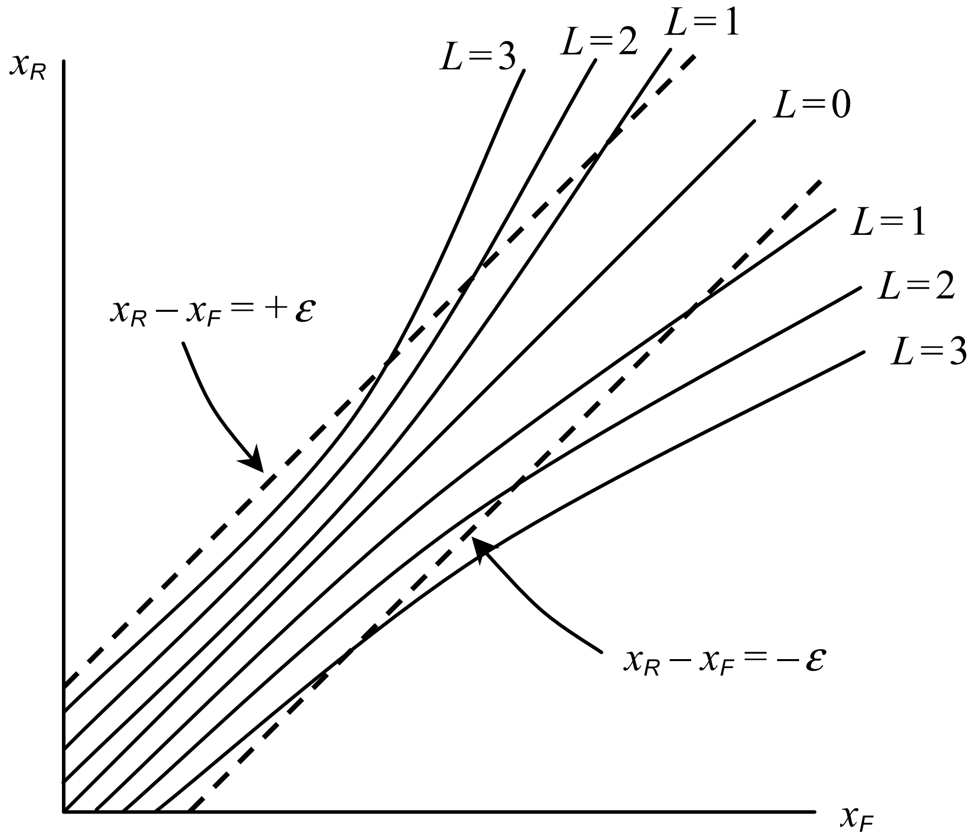

Figure 1. Level curves of a general loss function $L(x_R, x_F)$ and the band $|x_R - x_F| \leqslant \varepsilon$.

square of this difference), and not upon how high or how low the two values might both be. Thus, the profit shortfall from having underpredicted a realized output price of $10 by one dollar is the same as the profit shortfall from having underpredicted a realized output price of $2 by one dollar. This is clearly unrealistic in any decision problem which exhibits "wealth effects" or "location effects" in the uncertain variable, such as a firm which could make money if the realized output price was $7 (so there would be a definite loss in profits from having underpredicted the price by $1), but would want to shut down if the realized output price was only $4 (in which case there would be no profit loss at all from having underpredicted the price by $1).

2.3.3. Are squared-error loss functions appropriate as "local approximations"?

One argument for the squared-error form $L_{\text{Sq}}(x_R, x_F) \equiv k \cdot (x_R - x_F)^2$ is that if the forecast errors $x_R - x_F$ are not too big – that is, if the forecaster is good enough at prediction – then this functional form is the natural second-order approximation to any smooth loss function that exhibits the necessary properties of being zero when $x_R = x_F$ (from (8)) and having zero first-order effect for small departures from a perfect forecast (from (15)).

However, the fact that $x_R - x_F$ may always be close to zero *does not* legitimize the use of the functional form $k \cdot (x_R - x_F)^2$ as a second-order approximation to a general smooth bivariate loss function $L(x_R, x_F)$, even one that satisfies $L(0,0) = 0$ and $\partial L(x_R, x_F)/\partial x_F|_{x_R=x_F} = 0$. Consider Figure 1, which illustrates the level curves of some smooth loss function $L(x_R, x_F)$, along with the region where $|x_R - x_F|$ is less than or equal to some small value ε, which is seen to constitute a constant-width band about the 45° line. This region does *not* constitute a small neighborhood in $\mathbf{R}^2$, even as $\varepsilon \to 0$. In particular, the second order approximation to $L(x_R, x_F)$ when x_R and x_F are both small and approximately equal to each other is *not* the same as the second-order approximation to $L(x_R, x_F)$ when x_R and x_F are both large and approximately equal to each other. Legitimate second-order approximations to $L(x_R, x_F)$ can only be taken in over *small neighborhoods of points* in $\mathbf{R}^2$, and not over *bands* (even narrow bands) about the 45° line. The "quadratic approximation" $L_{\mathrm{Sq}}(x_R, x_F) \equiv k \cdot (x_R - x_F)^2$ over such bands is not justified by Taylor's theorem.

2.3.4. *Implications of error-based loss*

By the year 2000, virtually all stated loss functions were of the form (27) – that is, a single-argument function of the forecast error $x_R - x_F$ which satisfies the properties (8):

$$L_{\mathrm{err}}(x_R, x_F) \equiv H(x_R - x_F), \quad H(\cdot) \geqslant 0, H(0) = 0, H(\cdot) \text{ quasiconcave.} \tag{27}$$

Consider what Theorem 1 implies about this general *error-based form*:

COROLLARY 2. *For arbitrary error-based function $L_{\mathrm{err}}(x_R, x_F) \equiv H(x_R - x_F)$ satisfying* (27)*, an objective function $U(\cdot,\cdot) : \mathcal{X} \times \mathcal{A} \to \mathbf{R}^1$ with strictly monotonic optimal action function $\alpha(\cdot)$ will generate $L_{\mathrm{err}}(\cdot,\cdot)$ as its loss function if and only if it takes the form*

$$U(x, \alpha) \equiv f(x) - H\big(x - g(\alpha)\big) \tag{28}$$

for some function $f(\cdot) : \mathcal{X} \to \mathbf{R}^1$ and monotonic function $g(\cdot) : \mathcal{A} \to \mathcal{X}$.

Formula (28) highlights the fact that the use of an error-based loss function of the form (27) implicitly assumes that the decision maker's underlying problem is again "location-independent", in the sense that the utility loss from having made an ex post nonoptimal choice $\alpha \neq g^{-1}(x_R)$ only depends upon the *difference* between the values x_R and $g(\alpha)$, and not their general levels, so that it is again subject to the remarks following Equation (26). This location-independence is even more starkly illustrated in formula (28)'s canonical form, namely $\hat{U}(x, x_F) \equiv f(x) - H(x - x_F)$.

2.4. *Location-dependent loss functions*

Given a loss function $L(x_R, x_F)$ which is location-dependent and hence does *not* take the form (27), we can nevertheless retain most of our error-based intuition by defining $e = x_R - x_F$ and defining $L(x_R, x_F)$'s associated *location-dependent error-based form* by

$$H(x_R, e) \equiv L(x_R, x_R - e) \tag{29}$$

which implies

$$L(x_R, x_F) \equiv H(x_R, x_R - x_F). \tag{30}$$

In this case Theorem 1 implies that the utility function (22) takes the form

$$U(x, \alpha) \equiv f(x) - H\big(x, x - g(\alpha)\big) \tag{31}$$

for some $f(\cdot)$ and monotonic $g(\cdot)$. This is seen to be a generalization of Corollary 2, where the error-based function $H(x - g(\alpha))$ is replaced by a location-dependent form $H(x, x - g(\alpha))$. Such a function, with canonical form $\hat{U}(x, x_F) \equiv f(x) - H(x, x - x_F)$, would be appropriate when the decision maker's sensitivity to a unit error was different for prediction errors about high values of the variable x than for prediction errors about low values of this variable.

2.5. Distribution-forecast and distribution-realization loss functions

Although the traditional form of forecast used was the point forecast, there has recently been considerable interest in the use of distribution forecasts. As motivation, consider "forecasting" the number that will come up on a biased (i.e. "loaded") die. There is little point to giving a scalar point forecast – rather, since there will be irreducible uncertainty, the forecaster is better off studying the die (e.g., rolling it many times) and reporting the six face probabilities. We refer to such a forecast as a *distribution forecast*. The decision maker bases their optimal action upon the distribution forecast $F_F(\cdot)$ by solving the first order condition

$$\int U_\alpha(x, \alpha)\, \mathrm{d}F_F(x) = 0 \tag{32}$$

to obtain the optimal action function

$$\alpha(F_F) \equiv \arg\max_{\alpha \in \mathcal{A}} \int U(x, \alpha)\, \mathrm{d}F_F(x). \tag{33}$$

For the case of a distribution forecast $F_F(\cdot)$, the reduced-form payoff function takes the form

$$R(x_R, F_F) \equiv U\Big(x_R, \arg\max_{\alpha \in \mathcal{A}} \int U(x, \alpha)\, \mathrm{d}F_F(x)\Big) \equiv U\big(x_R, \alpha(F_F)\big). \tag{34}$$

Recall that the *point-forecast equivalent* is defined as the value $x_F(F_F)$ that satisfies

$$\alpha\big(x_F(F_F)\big) = \alpha(F_F) \tag{35}$$

and in the case of a single realization x_R, the *distribution-forecast/point-realization loss function* is given by

$$L(x_R, F_F) \equiv U\big(x_R, \alpha(x_R)\big) - U\big(x_R, \alpha(F_F)\big). \tag{36}$$

In the case of T successive throws of the same loaded die, there is a sense in which the "best case scenario" is when the forecaster has correctly predicted each of the successive realized values $x_{R1}, \ldots, x_{RT}$. However, when it is taken as given that the successive throws are independent, and when the forecaster is restricted to offering a single distribution forecast $F_F(\cdot)$ which must be provided prior to any of the throws, then the "best case" distribution forecast is the one that turns out to match the empirical distribution $F_R(\cdot)$ of the sequence of realizations, which we can call its "histogram". We thus define the *distribution-forecast/distribution-realization loss function* by

$$L(F_R, F_F) \equiv \int U\big(x, \alpha(F_R)\big)\, \mathrm{d}F_R(x) - \int U\big(x, \alpha(F_F)\big)\, \mathrm{d}F_R(x) \tag{37}$$

and observe that much of the above point-realization based analysis can be extended to such functions.

References

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, second ed. Springer-Verlag, New York.

Chamberlain, G. (2000). "Econometrics and decision theory". Journal of Econometrics 95, 255–283.

DeGroot, M.H. (1970). Optimal Statistical Decisions. McGraw Hill, New York.

Elliott, G., Lieli, R., (2005). "Predicting binary outcomes". Working Paper, Department of Economics, University of California, San Diego.

Granger, C.W.J., Pesaran, M.H. (2000a). "A decision-theoretic approach to forecast evaluation". In: Chan, W.S., Li, W.K., Tong, H. (Eds.), Statistics and Finance: An Interface. Imperial College Press, London.

Granger, C.W.J., Pesaran, M.H. (2000b). "Economic and statistical measures of forecast accuracy". Journal of Forecasting 19, 537–560.

Machina, M.J., Granger, C.W.J., (2005). "Decision-based forecast evaluation". Manuscript, Department of Economics, University of California, San Diego.

Pesaran, M.H., Skouras, S. (2002). "Decision-based methods for forecast evaluation". In: Clements, M.P., Hendry, D.F. (Eds.), A Companion to Economic Forecasting. Blackwell, Oxford.

Pratt, J.W., Raiffa, H., Schlaifer, R. (1995). Introduction to Statistical Decision Theory, second ed. MIT Press, Cambridge, MA.

Theil, H. (1961). Economic Forecasts and Policy, second ed. North-Holland, Amsterdam.

Theil, H. (1971). Applied Economic Forecasting. North-Holland, Amsterdam.

West, M., Harrison, J. (1989). Bayesian Forecasting and Dynamic Models. Springer-Verlag, New York.

West, M., Harrison, J. (1997). Bayesian Forecasting and Dynamic Models, second ed. Springer-Verlag, New York.

White, D.J. (1966). "Forecasts and decision making". Journal of Mathematical Analysis and Applications 14, 163–173.