

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: <http://www.nber.org/books/kmen80-1>

Publication Date: 1980

Chapter Title: Prediction Analysis of Economic Models

Chapter Author: David K. Hildebrand, James D. Laing, Howard Rosenthal

Chapter URL: <http://www.nber.org/chapters/c11695>

Chapter pages in book: (p. 91 - 122)

## Prediction Analysis of Economic Models

*DAVID K. HILDEBRAND*

DEPARTMENT OF STATISTICS  
UNIVERSITY OF PENNSYLVANIA  
PHILADELPHIA, PENNSYLVANIA

*JAMES D. LAING*

SCHOOL OF PUBLIC AND URBAN POLICY  
UNIVERSITY OF PENNSYLVANIA  
PHILADELPHIA, PENNSYLVANIA

*and*

*HOWARD ROSENTHAL*

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION  
CARNEGIE-MELLON UNIVERSITY  
PITTSBURGH, PENNSYLVANIA

This paper discusses an approach to evaluating a broader class of predictions than traditionally has been considered in econometric analysis.<sup>1</sup> Although a rich variety of predictions may be derived from economic theory, econometric models are generally restricted to *point* predictions specifying a single value of a dependent variable for each observation. However, the event predicted by economic theory is not always unique. For example, any of a *set* of outcomes typically is consistent with the prediction that the result of

<sup>1</sup> For a general treatment of the methods discussed in this paper when all variables are discrete, see Hildebrand, Laing, & Rosenthal (1977a). An earlier version of this paper was presented at the Conference on Criteria for Evaluation of Econometric Models, Ann Arbor, Michigan, June 9-10, 1977, sponsored by the National Bureau of Economic Research. We gratefully acknowledge support for this work provided Laing by a National Fellowship (1976-77) at the Hoover Institution on War, Revolution, and Peace, Stanford University, and by grant SOC72-05245 from the National Science Foundation. We also thank an anonymous referee and the editors for thoughtful comments.

bilateral bargaining will be Pareto optimal, that the outcome of a cooperative game will belong to the core, or that the vector of strategies chosen by the players of a noncooperative game will be in equilibrium.

Methods are described in this paper for evaluating models which predict that each observation's value on the dependent variable belongs to a specified *set*. Clearly, as the predicted set increases in size from a single point, the prediction becomes increasingly imprecise. Since imprecise predictions are more easily correct, the precision of a set prediction should be considered in addition to its error rate in evaluating the model's performance. This is accomplished by providing measures of both prediction precision and success.

These methods can be extended to deal with models that generate actuarial or probabilistic predictions about each observation's value on the dependent variable. Thus extended, this approach can be used to interpret the coefficient of determination  $r^2$ , thereby placing the evaluation of standard regression models within a more general framework.

The argument in this paper begins with some general concepts about prediction then presents specific methods for statistical analysis of data to evaluate a prediction's performance. In Section 1 some alternative styles of prediction are identified to distinguish: propositions that make an event prediction for each observation versus those that do not, *a priori* versus *ex post* predictions, absolute versus actuarial predictions, and point versus set predictions. In Section 2 methods are developed for evaluating the precision and accuracy of absolute predictions that each observation belongs to some specified set of states of the dependent variable. We apply the methods to a linear prediction with squared error. In Section 3 the methods are elaborated to deal with actuarial predictions and to interpret  $r^2$  within this context. Section 4 covers the evaluation of both the overall performance of multivariate predictions based on several independent variables and the contribution of each predictor in the model. In Section 5 statistical methods are presented for making inferences from sample data. Finally, Section 6 suggests some directions for further development.

## 1. Forms of Prediction

### 1.1. EVENT PREDICTION

This paper is devoted entirely to event predictions. A proposition is an *event prediction* if it specifies, for each observation, one or more values on the dependent variable. The error committed by such a proposition can be

assessed for each observation, one at a time, by comparing the predicted and observed values for the dependent variable. For example, a linear equation for which the values of all coefficients are specified represents an event prediction. However, the proposition that a set of regression coefficients will be positive is not an event prediction for each observation.

This paper discusses an approach to stating and evaluating the performance of event predictions about a dependent variable based on a set of independent variables. Variables may be qualitative or quantitative. The latter may be discrete or continuous. The approach is applied here when all (qualitative or quantitative) variables in the prediction are discrete. The continuous analogies generally follow by replacing probabilities with probability densities and summations with integrals. For bivariate propositions relating discrete variables, we will represent the dependent variable by an exhaustive set of  $R$  mutually exclusive states:  $Y = \{y_1, \dots, y_i, \dots, y_R\}$ . The independent variable has  $C$  states and is represented as  $X = \{x_1, \dots, x_j, \dots, x_C\}$ . When the variables are continuous, we omit subscripts on the variable states: for example,  $Y = \{y | y \in R^1\}$ . When the meaning is clear, we also omit subscripts for discrete variable states. Multivariate propositions can be translated into bivariate equivalents by forming Cartesian products. For example, two independent variables, say  $X$  and  $W$ , can be transformed into a single composite variable,  $V = X \times W$ , with typical state  $v_i = (x_j, w_k)$ . When a proposition pertains to several dependent variables, it can be evaluated with respect to its success in predicting each of the variables, and also for its overall success in predicting the composite dependent variable formed by the Cartesian product set. Initially, we focus on bivariate propositions that predict  $Y$  from  $X$ . Although either of these variables might be a composite, we do not focus explicitly on multivariate prediction until Section 4.

## 1.2. A PRIORI VERSUS EX POST PREDICTION

The usual regression prediction is *ex post*, since the coefficients are estimated from the data being analyzed. On the other hand, a linear model such as "predicted  $y = 4.2 + .6x$ " is *a priori* for the data to be analyzed if all coefficient values are specified by theory or prior estimation from an earlier data set. In general, an event prediction is *a priori* for a data set if it can be applied to predict  $Y$  for each observation in the set before the conditional and unconditional distributions on  $Y$  are known. In *ex post* prediction, information about the observed distributions on  $Y$  must be supplied before the specific event prediction is selected or applied. Frequently, of course, a prediction is selected *ex post* from one data set then applied as an *a priori*

prediction in a replication study. The methods discussed in this paper can be used to evaluate both *a priori* and *ex post* event predictions. Nonetheless, we give primary attention to evaluating event predictions stated *a priori*.

### 1.3. ABSOLUTE VERSUS ACTUARIAL PREDICTION

Many predictions are stated in *absolute* (or deterministic) terms. In absolute prediction, any particular value for the independent variable always yields the same predicted value(s) for the dependent variable. Absolute predictions may be stated in the form, "If this, then always predict that." Actuarial propositions are stated in probabilistic terms, such as, "If this, then that with probability .6." Section 2 defines methods for evaluating absolute predictions. In Section 3 these methods are extended for the evaluation of actuarial predictions, and this extension provides a basis for interpreting the coefficient of determination.

### 1.4. POINT VERSUS SET PREDICTION

Most current methods for evaluating event predictions at least tacitly rest on an assumption that a single value of the dependent variable must be predicted for any given value of the independent variable(s). In contrast, we treat set predictions as admissible. A *set prediction* states that the dependent variable value of each observation having a particular value of the predictor variable(s) belongs to some specified set.

Set predictions play a central role in economics. For one example, the prediction that the outcome will be Pareto optimal typically is a set prediction. In addition, applications of the theory of games in economics naturally lead to set predictions. The solutions based on game theory typically are sets, not points. Perhaps the two most central game-theoretic solution concepts are the Nash equilibrium of a noncooperative game and the core of a cooperative game. The solutions based on these concepts need not be unique. In economic applications of the core, for example, nonunique solutions have been obtained from game-theoretic analyses of peak load pricing (Sorenson, Tschirhart, & Whinston, 1976), setting aircraft landing fees (Littlechild & Thompson, 1977), determining premium rates for automobile insurance (Borch, 1962–1963), dividing profits among companies forming a business merger (Mossin, 1968), allocating investment and operating costs among states cooperating within a region to develop and distribute electricity (Gately, 1974), and apportioning gains achieved by countries through a common market (Segal, 1970). Thus set predictions derive quite naturally from formal theory of economic behavior. Typically, strong as-

sumptions need to be added, thereby restricting the domain of analysis, before the predicted set can be narrowed to a single point.

Introducing set predictions thus broadens the domain but also creates problems for evaluating a model's performance. It no longer suffices to measure only the error rate observed for the model because a set prediction may be so imprecise that it is guaranteed to have low error. In the limit, the tautological "prediction" which always predicts the entire set of  $Y$  values must be error free and thus has no scientific value for predicting behavior.

## 2. Evaluating Absolute Predictions

This section presents methods for taking both prediction precision and error into account in evaluating the success of an absolute set prediction. To distinguish issues in the evaluation of prediction success from those of statistical inference, we assume until Section 5 that the data to be analyzed constitute the entire population or, at least, a sufficiently large random sample that questions of statistical inference may be ignored.

### 2.1. PREDICTION LOGIC

The first task is to present a formal language for stating absolute set predictions in a way that reveals their basic structure. In the bivariate context, any such prediction,  $\mathcal{P}$ , may be written in the form

$$\mathcal{P}: x_1 \rightsquigarrow \mathcal{S}(x_1), \dots, x_j \rightsquigarrow \mathcal{S}(x_j), \dots, \& x_c \rightsquigarrow \mathcal{S}(x_c), \quad (2.1)$$

where each  $\mathcal{S}(x_j)$  is a ("success") set of  $Y$ -states, and the symbol " $\rightsquigarrow$ " may be read "tends to be sufficient for" or "predicts." Thus the prediction  $\mathcal{P}$  includes a set prediction  $\mathcal{S}(x_j)$  about  $Y$  for each state of the independent variable. Equivalently, any such prediction  $\mathcal{P}$  identifies the set of error events

$$\mathcal{E}_{\mathcal{P}} = \{(x_j, y_i) \mid x_j \in X, y_i \notin \mathcal{S}(x_j)\}. \quad (2.2)$$

We require that  $\mathcal{E}_{\mathcal{P}} \neq \emptyset$ .

For example, the approach discussed in this paper has been applied in a number of cross-national tests of the monetarists' impulse theory about the effect of the money supply on inflation rates (Dutton, 1978; Fourcans, 1978; Korteweg, 1978). Korteweg has provided us with the cross classification shown in Table 1 of 24 annual observations for the Netherlands. (While these observations are treated here as population data, statistical tests and confidence intervals can be found in Hildebrand, Laing, & Rosenthal 1977a,

TABLE 1  
INFLATION AND THE MONEY SUPPLY IN THE NETHERLANDS<sup>a,b</sup>

		(A) Major Diagonal Prediction, $\mathcal{D}$			(B) Asymmetric Prediction, $\mathcal{A}$		
		Change in money supply			Change in money supply		
		+	0	-	+	0	-
Change in consumer prices	+	11	2	0	11	2	0
	0	0	5	1	0	5	1
	-	3	1	1	3	1	1

<sup>a</sup> Error cells (shaded) for two propositions.

<sup>b</sup> Actual numbers of cases shown for annual data provided by Pieter Korteweg.

pp. 210–211.) Both the money supply variable (the predictor) and the inflation variable (the criterion) are categorized as increase (+), little or no change (0), and decrease (-).

Korteweg has treated monetarist theory as predicting that the data lie on the main diagonal. That is,

$$\mathcal{D}: + \rightsquigarrow +, 0 \rightsquigarrow 0, \& - \rightsquigarrow -.$$

An alternative, but less precise, theory might specify that the change in the money supply determines only a lower bound for a change in the inflation rate. Thus, for example, when the money supply was unchanged (0), prices would either remain unchanged (0) or increase (+). In “prediction logic” notation, this asymmetric prediction can be written as

$$\mathcal{A}: + \rightsquigarrow +, 0 \rightsquigarrow (0 \text{ or } +), \& - \rightsquigarrow (-, 0, \text{ or } +).$$

This rule involves set prediction; for instance, the 0 change in money supply predicts only (0 or +), that is, no decrease in price.

Korteweg’s data illustrate the importance of considering prediction precision as well as error. Note that proposition  $\mathcal{A}$  for the Korteweg data is more accurate than proposition  $\mathcal{D}$ , in the sense that  $\mathcal{A}$  makes only four errors while  $\mathcal{D}$  makes seven. Yet  $\mathcal{A}$  is also clearly a less precise prediction than  $\mathcal{D}$  in that the error set of  $\mathcal{A}$  is just a proper subset of  $\mathcal{D}$ ’s error set. The remaining tasks of this section are, first, to provide numerical measures of accuracy,

precision, and overall prediction success and, second, to discuss the trade-off between prediction success and precision that arises in the comparative evaluation of alternative set predictions.

## 2.2. EVALUATING PREDICTION SUCCESS

We now consider the analysis of cross-classified data to evaluate predictions such as those posed in Table 1. Borrowing the notation of continuous variables to simplify the notation for discrete variable states, let  $f(x, y)$  denote the joint probability in the cross classification or contingency table that both  $X = x$  and  $Y = y$ . (Recall that we are ignoring sampling considerations until Section 5.) The marginal or unconditional probabilities of  $x$  and  $y$  are  $f(x)$  and  $f(y)$ . We shall represent conditional probabilities by expressions such as  $f(y|x)$ .

To conduct numerical analysis, the investigator must specify the error weights

$$\begin{aligned} \omega(x, y) &> 0 && \text{if } (x, y) \in \mathcal{E}_\varphi, \\ \omega(x, y) &= 0 && \text{if } (x, y) \notin \mathcal{E}_\varphi. \end{aligned} \tag{2.3}$$

Unweighted errors, where  $\omega(x, y) = 1$  for all  $(x, y) \in \mathcal{E}_\varphi$ , are often a natural choice for categorical variables, but weighted errors (such as the square of the difference between actual  $y$  and predicted  $y$ ) frequently are appropriate for quantitative variables.

We now motivate measures of accuracy,  $K_\varphi$ , and precision,  $U_\varphi$ , as (weighted) error rates from two different prediction tasks. In the first task, one simply applies the prediction  $\mathcal{P}$  when the  $X$ -state is given for each observation drawn at random from the population (with replacement). For each observation drawn having  $X = x$ , the prediction is  $\mathcal{S}(x)$ . The expected error rate when  $X$  is known equals

$$K_\varphi = \sum_x \sum_y \omega(x, y) f(x) f(y|x) = \sum_x \sum_y \omega(x, y) f(x, y). \tag{2.4}$$

Note that in this first task, the prediction  $\mathcal{S}(x)$  is used for an expected fraction  $f(x)$  of the predictions.

The second task provides a benchmark by replicating the same predictions but without knowledge of the  $X$ -state. We continue to draw observations at random. However, for each observation drawn (with unknown  $X$ -state), we instead select each prediction  $\mathcal{S}(x)$  with probability  $f(x)$ . Since we predict each  $\mathcal{S}(x)$  with the same probability in the two tasks, the predictions in the two tasks are equally precise. The only difference is whether or not the  $X$  value is known when the prediction about  $Y$  is made. Since in

the second task observations are drawn randomly from the entire distribution, error probabilities are governed by the unconditional rather than conditional  $Y$  distribution. Therefore, the weighted error rate expected when  $X$  is unknown equals

$$U_{\mathcal{P}} = \sum_x \sum_y \omega(x, y) f(x) f(y). \quad (2.5)$$

Statistical independence is sufficient (but not necessary) for  $U = K$ . Under statistical independence, since there is no gain from knowing  $X$ -states, the prediction has no more accuracy than its benchmark. We interpret  $U$  as a measure of precision. With unweighted errors, if  $\sum_y \omega(x, y) f(y)$  is large,  $\mathcal{P}(x)$  is precise in that a success would be rare if  $\mathcal{P}(x)$  were applied to an observation drawn at random from the entire population. The benchmark  $U$  in turn is just a weighted average of the precision of each component prediction  $\mathcal{P}(x)$ .

The Korteweg example can be used to contrast  $K$  and  $U$ . From Table 1(B) we see that for  $\mathcal{A}$  the prediction “+” will be made with probability  $\frac{1}{24}$  in both the  $K$  and  $U$  tasks. For this prediction the error probabilities are  $\frac{3}{14}$  under  $K$  but  $(\frac{6}{24} + \frac{5}{24}) = \frac{11}{14}$  under  $U$ . Similarly, the prediction is “+ or 0” with probability  $\frac{8}{24}$ ; the corresponding  $K$  and  $U$  error probabilities are  $\frac{1}{8}$  and  $\frac{5}{24}$ , respectively. Finally, with probability  $\frac{2}{24}$ , the prediction is the vacuous “+, 0, or -” and no error is possible. Combining the above for prediction  $\mathcal{A}$ , we find

$$\begin{aligned} K_{\mathcal{A}} &= \left(\frac{1}{24}\right)\left(\frac{3}{14}\right) + \left(\frac{8}{24}\right)\left(\frac{1}{8}\right) + 0 = .167, \\ U_{\mathcal{A}} &= \left(\frac{1}{24}\right)\left(\frac{11}{14}\right) + \left(\frac{8}{24}\right)\left(\frac{5}{24}\right) + 0 = .337. \end{aligned}$$

As argued at the outset of this section, the success of a prediction involves not only the actual error rate or accuracy but also the precision. A prediction with actual error rate  $K = .01$  against a benchmark figure of  $U = .80$  is highly successful; a prediction with actual error rate .01 against a benchmark figure of .01 is highly unsuccessful. Our measure of prediction success is thus the proportionate reduction in error

$$\nabla_{\mathcal{P}} = (U_{\mathcal{P}} - K_{\mathcal{P}})/U_{\mathcal{P}} = 1 - (K_{\mathcal{P}}/U_{\mathcal{P}}). \quad (2.6)$$

In general,  $\nabla_{\mathcal{P}}$  is undefined if  $U_{\mathcal{P}} = 0$ . Otherwise,  $-\infty < \nabla_{\mathcal{P}} \leq 1$ . If  $\nabla_{\mathcal{P}} = 1$ , then there is no error when  $\mathcal{P}$  is applied given each observation's  $X$ -state. If  $\nabla = 0$ , then the prediction is in error with the same probability under the two information conditions. Any value of  $\nabla > 0$  represents a proportionate reduction in error. For the data shown in Table 1(A),  $\nabla_{\mathcal{A}} = (.337 - .167)/.337 = +.505$ . Thus the asymmetric prediction  $\mathcal{A}$  achieves a

50.5% reduction of the error expected for the prediction under statistical independence. If  $\nabla < 0$ , then use of  $X$  information actually leads to more error for the prediction than had it been applied randomly without that information, and  $|\nabla|$  indicates the proportionate *increase* in error. Negative values of  $\nabla$  result from an unfortunate choice of a prediction specified *a priori*.

### 2.3. DECOMPOSITION OF PREDICTION SUCCESS AND COMPARATIVE EVALUATION OF ALTERNATIVE PREDICTIONS

One useful property of  $\nabla_{\mathcal{P}}$  is that it can be expressed as a weighted average of  $\nabla$ -measures for components of the prediction  $\mathcal{P}$ . This property can be used in evaluating the micro-level support for the prediction and for the comparative analysis of this prediction with an alternative. Consider, first, the elementary proposition that identifies only  $(x, y)$  as an error event (and recall that, for now, we are dealing only with discrete variables). Then the measure for this elementary proposition may be written

$$\nabla_{xy} = 1 - [f(x, y)/f(x)f(y)].$$

Similarly, given the error weights specified for the component prediction,  $\mathcal{P}(x): x \rightsquigarrow \mathcal{P}(x)$ , we can compute

$$\nabla_{\mathcal{P}(x)} = 1 - [K_{\mathcal{P}(x)}/U_{\mathcal{P}(x)}] = 1 - \sum_y \omega(x, y)f(x, y)/\sum_y \omega(x, y)f(x)f(y).$$

If, for any component,  $U = 0$  so that  $\nabla$  is undefined, set  $U\nabla (= U - K)$  equal to zero. Then it follows that  $\nabla_{\mathcal{P}}$  can be expressed as a weighted average of  $\nabla$ -measures for component predictions (or for any partition of the prediction's error set), where the weight given each component  $\nabla$  in this weighted average is the fraction of the overall precision,  $U_{\mathcal{P}}$ , contributed by the component.

$$\nabla_{\mathcal{P}} = \sum_x \sum_y \frac{\omega(x, y)f(x)f(y)}{U_{\mathcal{P}}} \nabla_{xy} = \sum_x \frac{U_{\mathcal{P}(x)}}{U_{\mathcal{P}}} \nabla_{\mathcal{P}(x)}. \tag{2.7}$$

To illustrate, for the asymmetric prediction  $\mathcal{A}$  and the data of Table 1(B),  $\nabla_{\mathcal{A}(+)} = 1 - .125/.267 = .532$ ,  $\nabla_{\mathcal{A}(0)} = 1 - .042/.069 = .400$ ,  $U_{\mathcal{A}} = U_{\mathcal{A}(+)} + U_{\mathcal{A}(0)} = .337$ ; hence by (2.7),  $\nabla_{\mathcal{A}} = [U_{\mathcal{A}(+)}/U_{\mathcal{A}}]\nabla_{\mathcal{A}(+)} + [U_{\mathcal{A}(0)}/U_{\mathcal{A}}]\nabla_{\mathcal{A}(0)} = (.794)(.532) + (.206)(.400) = .505$ .

The weighted average expression (2.7) also can be used in the comparative evaluation of two alternative predictions,  $\mathcal{P}$  and  $\mathcal{P}'$ , with associated error weights,  $\{\omega(x, y)\}$  and  $\{\omega'(x, y)\}$ . The alternative  $\mathcal{P}'$  assigns *greater* error weight for the set of events  $\mathcal{E}_+ = \{(x, y) | \omega'(x, y) > \omega(x, y)\}$ , and assigns *less*

weight for the set of events  $\mathcal{E}_-$  defined analogously. Let  $\Delta\omega_{xy} = \omega'(x, y) - \omega(x, y)$ . Then define the measures

$$\begin{aligned} \nabla_+ &= 1 - \frac{\sum_{\mathcal{E}_+} |\Delta\omega_{xy}| f(x, y)}{\sum_{\mathcal{E}_+} |\Delta\omega_{xy}| f(x) f(y)}, \\ \nabla_- &= 1 - \frac{\sum_{\mathcal{E}_-} |\Delta\omega_{xy}| f(x, y)}{\sum_{\mathcal{E}_-} |\Delta\omega_{xy}| f(x) f(y)}, \end{aligned}$$

and let  $U_+$  and  $U_-$  denote the respective denominators in these expressions. Note that  $U_{\mathcal{P}'} = U_{\mathcal{P}} + U_+ - U_-$ . Then the success of the alternative proposition  $\mathcal{P}'$  with weights  $\{\omega'(x, y)\}$  can be expressed as a weighted average of the success of the prediction  $\mathcal{P}$  and correction terms that account for differences in the two sets of error weights:

$$\nabla_{\mathcal{P}'} = \frac{U_{\mathcal{P}}}{U_{\mathcal{P}'}} \nabla_{\mathcal{P}} + \frac{U_+}{U_{\mathcal{P}'}} \nabla_+ - \frac{U_-}{U_{\mathcal{P}'}} \nabla_- \quad (2.8)$$

For example, suppose we consider changing from the asymmetric prediction  $\mathcal{A}$  to the alternative main diagonal prediction  $\mathcal{D}$  for the data shown in Table 1. Then, prediction  $\mathcal{D}$  adds the set of cells ( $\mathcal{E}_+$ ) above the main diagonal to the error events. Note that  $\mathcal{E}_{\mathcal{D}} \supset \mathcal{E}_{\mathcal{A}}$ ; therefore,  $\mathcal{E}_-$  is empty. Then by (2.8), the success of proposition  $\mathcal{D}$  may be expressed as the weighted average  $\nabla_{\mathcal{D}} = (U_{\mathcal{A}}/U_{\mathcal{D}})\nabla_{\mathcal{A}} + (U_+/U_{\mathcal{D}})\nabla_+ = (.337/.583).505 + (.247/.583).493 = .500$ . Note that, by adding the cells above the diagonal ( $\mathcal{E}_+$ ) to the error set, proposition  $\mathcal{D}$  makes a substantial increase in precision over that of proposition  $\mathcal{A}$ . Moreover, this increase in precision is purchased with only a small decrement to overall success (via  $\nabla_+$ ), since the value of  $\nabla_{\mathcal{D}}$  does not differ much from  $\nabla_{\mathcal{A}}$ . This seems to be a rather small price to pay for the resulting increase in precision. It is not always so easy to decide about such trade-offs.

#### 2.4. TRADE-OFFS BETWEEN SUCCESS AND PRECISION

More than one criterion should be considered in selecting one prediction over another. Clearly, parsimony and theoretical relevance are important bases for these decisions. In the following discussion we focus on two of the important dimensions for evaluating propositions: prediction success and precision. Other things being equal, there is no difficulty in choosing one proposition over an alternative if the first *dominates* the second, in that it has higher values for both  $U$  and  $\nabla$ . Typically, however, success can be increased by decreasing precision, and trade-offs between these two dimensions of prediction performance must be considered.

Although the primary focus in this paper is on the evaluation of predictions stated *a priori*, the trade-off issue can be illustrated by considering the

*ex post* selection of a prediction. Assume the entire population is given (we will not consider *ex post* selection from samples in this paper). One might seek to maximize  $\nabla$ , subject to a minimum constraint on  $U$ , or vice versa, reversing the roles of maximand and constraint. Suppose that each of the  $2^{RC} - 1$  alternative (and nonempty) error sets for an  $R \times C$  table is eligible for consideration. Then, as suggested by Eq. (2.7) expressing  $\nabla$  as a weighted average of  $\nabla_{xy}$ s, the maximization procedure basically involves

- (1) ranking cells in order of  $\nabla_{xy}$  values, and
- (2) constructing the error set, iteratively, by forming the union of events already included and that remaining cell  $(x, y)$  for which  $\nabla_{xy}$  is the greatest, until the constraint is satisfied.

Alternatively, one can maximize total reduction in error,  $U\nabla = U - K$ , subject to a constraint that, for every cell admitted to the error set,  $\nabla_{xy}$  does not fall below a specified value.<sup>2</sup>

However the investigator chooses to make it, the  $\nabla$  versus  $U$  trade-off appears central whenever set predictions are analyzed. For example, by choosing to maximize  $U\nabla$ , one has decided to accept the trade-off between  $U$  and  $\nabla$  that equally weights  $\ln U$  and  $\ln \nabla$ . Of course, caveats that lead econometricians to consider other criteria in addition to least squares in evaluating linear models can also apply here, even with full population data. If we looked at quantitative measures of inflation and money supply rather than Korteweg's trichotomizations, we would, for a finite population, find a perfectly or nearly perfectly fitting polynomial. Both  $\nabla$  and  $U$  would be high, but the prediction rules might have little theoretical appeal. Consequently, *ex post* selection will be influenced by a variety of considerations as well as  $\nabla$  and  $U$  values.

Although we acknowledge that model selection is a multifaceted problem, we have emphasized  $\nabla$  and  $U$  because they provide a direct means for comparing the prediction success of models of widely different structure. Elsewhere (Hildebrand, Laing, & Rosenthal 1977a,b) we demonstrate equivalences that allow  $\nabla$  and  $U$  to be used for comparative analysis of predictions, whether they concern single observations or paired comparisons of observations, or are expressed as actuarial or absolute predictions. As a result, many of the well-known proportionate-reduction-in-error measures for nominal and ordinal variables are readily interpretable as  $\nabla$ -measures. In this paper we develop similar equivalences for the correlation ratio and  $r^2$ , the two standard proportionate-reduction-in-error measures for interval variables.

<sup>2</sup> For details on the maximization procedures and equivalences among certain results, modifications of these procedures when theoretical or other considerations can be applied to reduce the set of eligible alternatives, and *ex post* selection from samples, see Hildebrand, Laing, & Rosenthal (1977a, pp. 132-145 and Chap. 6).

Traditionally, the correlation ratio and  $r^2$  have been interpreted in terms of pure strategy predictions with associated error weights equal to the squared difference between predicted and actual  $Y$  values. In the remainder of this section, we develop a measure for these pure strategy predictions with squared error, and show that these measures differ from the correlation ratio and  $r^2$ . The  $\nabla$ -equivalents of these measures are based on the mixed strategy approach of Section 3.

### 2.5. THE CORRELATION RATIO AND A PURE STRATEGY PREDICTION BASED ON CONDITIONAL MEANS

The conventional interpretation of the correlation ratio is based on *ex post*, pure strategy predictions. First, given  $X = x$ , one predicts the conditional mean,  $\mu_{Y|x} = E(y|x)$ . Second, when  $X$  is unknown, one always predicts the unconditional mean,  $\mu_Y = E(y)$ . These *ex post* point predictions minimize the total (sum of squared) error under the corresponding information conditions:  $K_{\eta^2} = E_X[\text{Var}(Y|X)]$  and  $U_{\eta^2} = \text{Var}(Y)$ . Then, the proportionate-reduction-in-error expression,  $1 - (K/U)$ , yields the correlation ratio

$$\eta^2 = 1 - \frac{E_X[\text{Var}(Y|X)]}{\text{Var}(Y)}. \quad (2.9)$$

Thus the conventional interpretation is based on the *ex post*, pure strategy prediction  $\mathcal{M} = \{\mathcal{M}(x): x \rightsquigarrow \mu_{Y|x}\}$  when  $X$  is known. We could follow the same approach used earlier in this paper for pure strategy predictions by applying each component prediction,  $\mathcal{M}(x)$ , and the associated set of error weights  $\{(y - \mu_{Y|x})^2\}$  with probability  $f(x)$  under both of the two information conditions. Note that this differs from the development of  $\eta^2$  in the preceding paragraph. In particular, when  $X$  is unknown, the conventional approach is to predict  $\mu_Y$ ; in contrast, we randomly select each prediction  $\mu_{Y|x}$  with probability  $f(x)$  so that this prediction is applied with equal probability in the two information conditions. In the appendix (A.1) we show that the expected (squared) error when  $\mathcal{M}$  is applied under each of the two information conditions is

$$K_{\mathcal{M}} = \sum_x \sum_y (y - \mu_{Y|x})^2 f(x) f(y|x) = E_X[\text{Var}(Y|X)]$$

$$U_{\mathcal{M}} = \sum_x \sum_y (y - \mu_{Y|x})^2 f(x) f(y) = \text{Var}(Y) + E_X(\mu_{Y|x} - \mu_Y)^2.$$

Therefore,

$$\nabla_{\mathcal{M}} = 1 - \frac{E_X[\text{Var}(Y|X)]}{\text{Var}(Y) + E_X(\mu_{Y|x} - \mu_Y)^2}. \quad (2.10)$$

As a consequence of differences in the predictions when  $X$  is unknown,  $\nabla_{\mathcal{M}}$  is not equivalent to  $\eta^2$ : either  $\nabla_{\mathcal{M}} = \eta^2 = 0$  or  $1$ , or  $U_{\mathcal{M}} > U_{\eta^2}$  and  $\nabla_{\mathcal{M}} > \eta^2$ .

2.6. BIVARIATE LINEAR PREDICTION WITH SQUARED ERROR

The prediction analysis framework may be applied to evaluate a linear prediction as a pure strategy with squared error. Consider the bivariate linear prediction  $\mathcal{L} = \{x \rightsquigarrow \gamma + \delta x\}$  with associated error weights  $\{\omega(x, y)\} = \{(y - \gamma - \delta x)^2\}$ . We write  $\gamma + \delta x$  rather than the familiar  $\alpha + \beta x$  to emphasize that the linear prediction may be selected either *a priori* or *ex post*; the parameters  $\gamma$  and  $\delta$  are not required to be functions of the joint distribution. Following the same approach as before, the prediction  $\gamma + \delta x$  is applied with probability  $f(x)$  under each of the two information conditions. In the  $X$ -known condition, given  $X = x$ , the prediction is  $\gamma + \delta x$ , and so the value  $y$  and associated error  $(y - \gamma - \delta x)^2$  occur with probability  $f(y|x)$ . Therefore, the total squared error expected when  $X$  is known equals

$$K_{\mathcal{L}} = \sum_x \sum_y (y - \gamma - \delta x)^2 f(y|x) f(x).$$

When  $X$  is unknown, the prediction  $\gamma + \delta x$  is randomly chosen with probability  $f(x)$ , and the value  $y$  and associated error  $(y - \gamma - \delta x)^2$  occur with probability  $f(y)$ . Consequently,

$$U_{\mathcal{L}} = \sum_x \sum_y (y - \gamma - \delta x)^2 f(y) f(x).$$

In the appendix (A.2) we show that

$$\nabla_{\mathcal{L}} = 1 - \frac{K_{\mathcal{L}}}{U_{\mathcal{L}}} = 1 - \frac{E_x[\text{Var}(Y|X)] + E_x(\gamma + \delta x - \mu_{Y|x})^2}{\text{Var}(Y) + E_x(\gamma + \delta x - \mu_Y)^2}. \quad (2.11)$$

Thus  $K_{\mathcal{L}}$  is the average conditional variance (i.e., the average squared error when predicting that  $Y$  will be at its conditional mean,  $\mu_{Y|x}$ ) plus the average squared deviation of the selected prediction from the conditional mean. The first term in  $K_{\mathcal{L}}$  as given by (2.11) is the minimum squared error possible for any point prediction (linear or otherwise) in this distribution. The second term in  $K_{\mathcal{L}}$ , of course, is zero if the linear prediction always chooses the conditional mean. Analogously, the  $X$ -unknown error given by (2.11) equals the squared error for the prediction that  $Y = \mu_Y$ , plus the average squared deviation of the prediction  $\mathcal{L}$  from  $\mu_Y$ . Thus the value of  $U_{\mathcal{L}}$  is large if the  $Y$  variance is large or if the prediction  $\gamma + \delta x$  tends to deviate substantially from the mean of  $Y$ .

The resulting value of  $\nabla_{\mathcal{L}}$  is determined not only by the extent to which the actual relation is linear, but also by what coefficients were chosen for linear prediction  $\gamma + \delta x$ . In the appendix (A.2) we show that

$$\nabla_{\mathcal{L}} \gtrless 0 \quad \text{iff} \quad \text{correlation}(\gamma + \delta x, \mu_{Y|x}) \gtrless 0. \quad (2.12)$$

Thus the sign of  $\nabla_{\mathcal{L}}$  depends on the sign of the correlation between the linear prediction  $\gamma + \delta x$  with the optimal point prediction  $\mu_{Y|x}$ .

The least squares regression model  $\mathcal{L}^* = \{x \rightsquigarrow \alpha + \beta x\}$  minimizes  $K_{\mathcal{L}}$ . If the distribution is, indeed, linear, so that  $\alpha + \beta x = \mu_{Y|x}$  for all  $x \in X$ , then the second term in the numerator of (2.11) is zero and

$$\nabla_{\mathcal{L}^*} = 1 - \frac{E_x[E(y - \alpha - \beta x)^2 | x]}{\text{Var}(Y) + E_x(\alpha + \beta x - \mu_Y)^2},$$

whereas the correlation ratio reduces to

$$r^2 = 1 - \frac{E_x[E(y - \alpha - \beta x)^2 | x]}{\text{Var}(Y)}.$$

Note that these two measures differ only in their denominators. The denominator of  $r^2$  is a constant for a given distribution, so that the best fitting model maximizes  $r^2$  by minimizing  $K_{\mathcal{L}}$ . On the other hand, the denominator of (2.11) depends on what linear model is chosen.

In general,  $\nabla_{\mathcal{L}^*} = r^2 = 0$  if the regression line has zero slope; otherwise,  $\nabla_{\mathcal{L}^*} > r^2$ . Thus  $r^2$  cannot be interpreted as a  $\nabla$ -measure for the least-squares, pure strategy (absolute), linear prediction  $\mathcal{L}^*$ .

In this section we have developed a model for generating measures for pure strategy event predictions. One advantage of this model is that it permits direct comparisons of a wide variety of alternative predictions about the same data. In the conventional interpretation, the correlation ratio and  $r^2$  are based on (*ex post*) pure strategy predictions. The  $\nabla$ -measures developed above can be used in comparative evaluations of these *ex post* predictions with other predictions within the same basic framework. However, the  $\nabla$ -measure for the corresponding pure strategy prediction does not equal the correlation ratio or  $r^2$ , and thus cannot be used to interpret these measures.

This gap is bridged in the next section. We extend the method to actuarial predictions and show that the correlation ratio and  $r^2$  can each be interpreted as a  $\nabla$ -measure for an *ex post* actuarial proposition. Also, we identify an equivalence that permits comparisons within the same framework of actuarial propositions with absolute predictions. Consequently, the  $\nabla$ -measure provides a basis for comparing an ordinary least squares model with a large variety of alternatives, including other actuarial propositions and models based on set predictions.

### 3. Evaluating Actuarial Predictions: An Interpretation of the Correlation Ratio and $r^2$

Prediction analysis methods can be extended to actuarial propositions—that is, propositions that assign probabilities to various states of the dependent variable. In this section we discuss alternative ways of applying actuarial

propositions, show how to evaluate event predictions based on a "mixed strategy" application of actuarial propositions, identify an equivalence between mixed strategy predictions and absolute event predictions that facilitates comparative analysis, and, finally, interpret the coefficient of determination within this framework.

### 3.1. ACTUARIAL PROPOSITIONS: MIXED STRATEGIES OVER POINT PREDICTIONS

In their simplest form bivariate actuarial propositions assign a conditional probability, given  $X = x$ , to each state of the dependent variable  $Y$ . For example, under the conditions described by  $X = x$ , the probability of rain is  $\frac{2}{3}$ , and the probability of no rain is  $\frac{1}{3}$ . Such a simple actuarial proposition may be expressed as a statement of the form  $\mathcal{Q}: \{\mathcal{Q}(x) | x \in X\}$ , where

$$\mathcal{Q}(x): \left\{ q(y|x) | y \in Y, q(y|x) \geq 0, \sum_y q(y|x) = 1 \right\} \quad (3.1)$$

We discuss two ways in which an actuarial proposition can be applied. First, the proposition may be interpreted as a prediction about proportions. To evaluate this prediction, one compares the goodness of fit between  $\{q(y|x)\}$  and the observed fractions  $\{f(y|x)\}$  if the true population distributions are known (or calculates the likelihood of  $\{f(y|x)\}$  appearing in the sample if the true probabilities were  $\{q(y|x)\}$ , as asserted). Thus interpreted,  $\mathcal{Q}$  is a prediction about proportions in aggregates of observations, rather than an event prediction about  $Y$  for each observation in these aggregates. Consequently, there need be no relation between goodness of fit and the effectiveness with which  $\mathcal{Q}$  may be used to predict each observation's  $Y$ -state. For example, suppose  $\mathcal{Q}$  correctly asserts that the conditional distributions over  $Y$  are uniform for every value of  $X$ . Then, even if it perfectly fits the observed distributions,  $\mathcal{Q}$  offers little assistance in predicting  $Y$  for each observation. In this paper we focus on event prediction, and thus will not deal with goodness of fit procedures.

Second, borrowing a term from game theory, a simple actuarial proposition may be interpreted as a *mixed strategy* over alternative point predictions. [From this viewpoint, an absolute event prediction represents a *pure strategy*: if  $X = x$ , then always predict  $\mathcal{S}(x)$ .] That is, the set of statements "If  $X = x$ , then  $Y = y$  with probability  $q(y|x)$ " can be replaced with statements of the form "If  $X = x$ , then, with probability  $q(y|x)$ , predict  $Y = y$ ." By this interpretation, when  $X = x$ , the actuarial proposition implies a lottery  $\mathcal{Q}(x)$  over the alternative point predictions about  $Y$ . We next extend the prediction analysis methods to deal with a more general class of mixed strategy predictions.

### 3.2. MIXED STRATEGIES OVER SET PREDICTIONS WITH WEIGHTED ERRORS

In its most general form, a mixed strategy prediction represents a probability mixture of set predictions and associated error weights. Each component  $\mathcal{Q}(x)$  of this mixed strategy prediction consists of a probability mixture over  $M$  alternative set predictions of the form "If  $X = x$ , then, with probability  $q_x^{(m)}$ , predict  $\mathcal{S}^{(m)}(x)$ ," where  $q_x^{(m)} \geq 0$ ,  $\sum_m q_x^{(m)} = 1$ ,  $\mathcal{S}^{(m)}(x) \subseteq Y$  for every  $x \in X$  and  $m = 1, \dots, M$ . Associated with each prediction  $\mathcal{S}^{(m)}(x)$  is a set of error weights  $\{\omega^{(m)}(x, y)\}$ . In applying this mixed strategy prediction,  $\mathcal{Q}(x)$  is selected for an expected fraction  $f(x)$  of the observations. Given that  $\mathcal{Q}(x)$  is selected, the event  $Y = y$  is assigned error weight  $\omega^{(m)}(x, y)$  with probability  $q_x^{(m)}$ . Then, the expected weighted error rate for the mixed strategy  $\mathcal{Q}$  with associated error weights is, when  $X$  is known,

$$K_{\mathcal{Q}} = \sum_x \sum_y \sum_m q_x^{(m)} \omega^{(m)}(x, y) f(x, y), \quad (3.2)$$

and, when  $X$  is unknown,

$$U_{\mathcal{Q}} = \sum_x \sum_y \sum_m q_x^{(m)} \omega^{(m)}(x, y) f(x) f(y). \quad (3.3)$$

Finally, as before, define the measure

$$\nabla_{\mathcal{Q}} = [U_{\mathcal{Q}} - K_{\mathcal{Q}}]/U_{\mathcal{Q}}. \quad (3.4)$$

By comparing these results to the corresponding expressions of Section 2, it is clear that the pure strategy proposition  $\mathcal{P}$  with associated error weights  $\{\omega(x, y)\}$  is  $\nabla$ -equivalent to the mixed strategy  $\mathcal{Q}$  with associated error weights if, for every  $(x, y) \in X \times Y$ ,

$$\omega(x, y) = \sum_m q_x^{(m)} \omega^{(m)}(x, y).$$

This equivalence allows us to compare mixed strategy predictions with pure strategy predictions using the same framework.<sup>3</sup> The comparisons based on  $\nabla$  are unaffected by proportional rescaling of error weights. On the other hand, values for the measure of precision,  $U$ , cannot be compared without adopting a numeraire for error weights. A similar need for a numeraire arises in comparing standard regression models where  $r^2$  values are independent of the units of measurement but variances are not.

<sup>3</sup> This equivalence has a simple form if  $\mathcal{Q}$  is a mixed strategy over point predictions with unweighted errors, the type discussed in Section 3.1: in this case, the pure strategy equivalent has error weights  $\{\omega(x, y)\} = \{1 - q(y|x)\}$ . The restrictions on the  $q(y|x)$  given by (3.1) imply that  $0 \leq \omega(x, y) \leq 1$  and  $\sum_y \omega(x, y) = R - 1$ , where  $R$  is the number of discrete  $Y$  values or states.

The equivalence noted above indicates that the mixed strategy model can be used to interpret or explain the error weights associated with pure strategy prediction. The mixed strategy approach also enables us to assess probabilistic models on an observation by observation basis. One basic probabilistic model is indeed the bivariate linear model with an additive random disturbance. Applying the mixed strategy approach, we find, perhaps surprisingly, that the correlation ratio and  $r^2$  have a  $\nabla$  interpretation.

### 3.3. A MIXED STRATEGY INTERPRETATION OF THE CORRELATION RATIO AND $r^2$

This interpretation of the correlation ratio and  $r^2$  is based on  $\nabla$  for the *ex post* best fitting mixed strategy prediction  $\mathcal{Q}^* = \{\mathcal{Q}^*(x) | x \in X\}$ ,

$$\mathcal{Q}^*(x) = \{f(y|x) | y \in Y\}. \tag{3.5}$$

Given  $X = x$ , the component  $\mathcal{Q}^*(x)$  predicts each  $Y$ -state with its conditional probability in the distribution,  $f(y|x)$ . The error weight may be written as the squared difference between actual ( $a$ ) and predicted ( $p$ )  $Y$  values,  $(a - p)^2$ . Given  $X = x$ , then with probability  $f(p|x)$ , the prediction is  $Y = p$ ; thus the event  $Y = a$  is assigned error weight  $(a - p)^2$  with probability  $f(p|x)$  and occurs with probability  $f(a|x)$ . Altogether, each of the component predictions  $\mathcal{Q}^*(x)$  is applied with probability  $f(x)$ . We show in the appendix (A.3) that the expected error in applying the mixed strategy prediction  $\mathcal{Q}^*$  when  $X$  is known equals twice the average conditional variance:

$$K_{\mathcal{Q}^*} = \sum_x \sum_{a \in Y} \sum_{p \in Y} (a - p)^2 f(p|x) f(a|x) f(x) = 2E_x[\text{Var}(Y|X)]. \tag{3.6}$$

When  $X$  is known, the component  $\mathcal{Q}^*(x)$  is again selected with probability  $f(x)$ . Since observations are being drawn at random from the entire distribution, the probability that  $Y = a$  is the unconditional probability. Otherwise, the same argument as above applies. As shown in the appendix (A.3), the expected error for the mixed strategy prediction  $\mathcal{Q}^*$  when  $X$  is unknown equals twice the variance:

$$U_{\mathcal{Q}^*} = \sum_x \sum_{a \in Y} \sum_{p \in Y} (a - p)^2 f(p|x) f(a) f(x) = 2 \text{Var}(Y). \tag{3.7}$$

Then define the measure

$$\nabla_{\mathcal{Q}^*} = 1 - \frac{2E_x[\text{Var}(Y|X)]}{2 \text{Var}(Y)} = \eta^2. \tag{3.8}$$

Therefore, the correlation ratio and, under the standard linearity assumption of regression analysis,  $r^2$  may be interpreted as a  $\nabla$ -measure when the best-fitting actuarial prediction is applied as a mixed strategy with squared error.<sup>4</sup>

When all conditional variances are zero, the  $\nabla$  predictions become identical to the conventional  $E(y|x)$ . Also, the  $K$  error becomes zero, but the  $U$  error for  $\nabla$  remains double the variance. This doubling reflects the descriptive or replication role of  $\nabla$ 's benchmark where predictions are matched with the  $K$  predictions instead of selecting a single predicted value such as  $\mu_Y$  for all observations.

The descriptive aspect of the mixed strategy  $K$  predictions with nonzero conditional variances is that they are unique for every distinct conditional probability distribution, whereas the conventional  $E(y|x)$  reflects only the conditional means.

The equivalences we have established between the  $\nabla$ -measure and the more conventional correlation measures might seem at first to be little more than formal curiosities. The existence of this equivalence, however, increases our confidence in using the prediction analysis approach to assess predictions that cannot be evaluated with the conventional linear-model framework.

#### 4. Multivariate Prediction Analysis

The previous discussion has been exclusively bivariate. Obviously, to be of general use, the formulation should have a multivariate extension. At present, we consider only the case where a *vector* of independent variables  $V$  is used to predict a unique dependent variable  $Y$ . How "simultaneous equation" prediction should be treated within a prediction analysis framework is an open question.

##### 4.1. MULTIPLE $\nabla$

The elementary part of the extension is trivial. If the description of an observation by the vector of independent variables is  $v \in V$ , then predict that its  $Y$  value belongs to the set  $\mathcal{S}(v)$ . The derivation of multiple  $\nabla$  then

<sup>4</sup> The same approach can be used to interpret Goodman & Kruskal's (1954)  $\tau_{ab}$  and  $\tau_{bc}$  measures for nominal data and Kendall's  $\tau_{bc}^2$  measure for ordinal data (see Hildebrand, Laing, & Rosenthal 1977a, pp. 55–56, 123, 175; 1977b, pp. 52–57). Our rule  $K$  for the correlation ratio is equivalent to Goodman and Kruskal's prediction given  $X$  for the tau measures. An analogous prediction, this time relating pairs of observations on the two variables, can also be used to develop Kendall's  $\tau_{bc}^2$  for ordinal variables. Both of these measures are also special cases of  $\nabla$ . Thus the mixed strategy  $\nabla$  approach enables us to interpret these measures for nominal and ordinal variables and the correlation ratio within a common framework.

follows directly from that of bivariate  $\nabla$ , as given by (2.4–2.6), leading to

$$\nabla_{\mathcal{P}} = \frac{U - K}{U} = 1 - \frac{\sum_v \sum_y \omega(v, y) f(v, y)}{\sum_v \sum_y \omega(v, y) f(v) f(y)} \quad (4.1)$$

On the surface there is really no change.

#### 4.2. ASSESSING PARTIAL CONTRIBUTIONS TO ERROR REDUCTION

The subtleties arise in actually specifying predictions and in “partialing out” the effects of various independent variables. The prediction rule possibilities are completely general. Each component prediction  $\mathcal{S}(v)$  may be anything—a single point, an interval, two disjoint intervals, several widely spaced points, the whole space, even the empty set. This level of generality can certainly be awkward. When the statistical method does not impose arbitrary restrictions, such as linearity and additivity, theory may be hard pressed to specify an *a priori* prediction. On the other hand, linearity, or even polynomial regression, can be unnecessarily constraining from a more general perspective.

To emphasize that the issue we are raising is fundamentally one of prediction analysis and not one of estimation, we now turn to an example. For both the linear equation  $y = b_0 + b_1x + b_2w + u$ , and for  $y = c_0 + c_1 \max(x, w) + u$ , estimating the coefficients *ex post* is an easy task for econometricians (given standard assumptions about the errors  $u$ ). Similarly, in the case of the first equation, econometricians have long been able to evaluate the marginal or partial contribution of the variable, say  $W$ , by comparing (in essence) the simple regression of  $Y$  on  $X$  with multiple regression of  $Y$  on  $X$  and  $W$ . But for the second equation we know of no way of using the conventional methods to evaluate the contribution of each independent variable to overall error reduction.

This section concerns the evaluation of partial contributions of the independent variables to the performance of a multivariate prediction. We discuss only predictions incorporating two independent variables; the extension to more than two predictors requires simply a sequential application of the basic ideas (see Hildebrand, Laing, & Rosenthal, 1975, p. 168; 1977a, pp. 275–276, 281).

Consider first the following hypothetical example. Suppose that an oligopoly theorist identifies a “price leader” and a “price ratifier” in certain industries. The claim is that a price change proposed by the leader must be matched at least in part by the ratifier before it is adopted throughout the industry. Suppose that the price change in any month by the leader  $L$  or the ratifier  $R$  is described by one of four categories: ++ (increase by at least 2%), + (increase, but less than 2%), 0 (no change), and – (decrease).

TABLE 2  
 HYPOTHETICAL DATA ON PRICE CHANGES IN OLIGOPOLY<sup>a,b</sup>

PRICE CHANGE BY-- LEADER, L: RATIFIER, R:	++			+			0			-			FOLLOWER'S PRICE CHANGE, F:
	++	+	0	++	+	0	++	+	0	++	+	0	
++	8	7	0	2	3	0	0	0	0	0	0	0	20
+	1	3	5	0	7	9	2	2	1	0	0	0	32
0	2	0	10	2	1	2	4	6	1	2	12	4	48
-	0	1	2	0	0	0	2	0	0	0	0	1	20
	11	11	17	2	2	10	14	8	8	3	12	5	120

<sup>a</sup> Data presented as number of observations.

<sup>b</sup> Vertical shading identifies error events for prediction  $\mathcal{P}_{FLA}$  in text, horizontal shading identifies error events for prediction  $\mathcal{P}_{FL}$ .

The dependent variable  $F$  is the average price change of other members (“followers”) of the industry described with the same four categories. Suppose the population data to be as in Table 2.

The prediction may be written formally as a set of component predictions in which  $L$  states appear before the ampersand,  $R$  states after it, and  $F$  states after the prediction symbol  $\rightsquigarrow$ :

$$\begin{aligned} \mathcal{P}_{FLR}: \quad & ++ \& ( ++ \text{ or } + ) \rightsquigarrow ++, & ++ \& ( 0 \text{ or } - ) \rightsquigarrow 0, \\ & + \& ( ++ \text{ or } + ) \rightsquigarrow +, & + \& ( 0 \text{ or } - ) \rightsquigarrow 0, \\ & 0 \& ++ \rightsquigarrow ( + \text{ or } 0 ), & 0 \& ( +, 0, \text{ or } - ) \rightsquigarrow 0, \\ & - \& ++ \rightsquigarrow ( + \text{ or } 0 ), & - \& ( + \text{ or } 0 ) \rightsquigarrow 0, \\ & - \& - \rightsquigarrow -. \end{aligned}$$

The prediction  $\mathcal{P}_{FLR}$  identifies the set of error cells indicated by the vertical shading in Table 2. For these hypothetical data, the trivariate proposition  $\mathcal{P}_{FLR}$  achieves  $K = .317$ ,  $U = .676$ , so that  $V = .532$ .

A natural question concerns the value of the ratifier firm’s action in predicting the action of followers once the leader’s action has been specified. To answer this question, we need a prediction rule which involves only the leader’s action. A natural choice is a simple follow-the-leader model,

$$\mathcal{P}_{FL}: \quad ++ \rightsquigarrow ++, + \rightsquigarrow +, 0 \rightsquigarrow 0, \& - \rightsquigarrow -.$$

This rule identifies as error events those cells in Table 2 indicated by horizontal shading. For the prediction  $\mathcal{P}_{FL}$  with these data,  $K = .433$ ,  $U = .755$ , so  $V = .426$ . Comparing  $V$  values, the “ratifier” variable seems to have modest predictive value since it, together with the “leader” action, yields a proportionate reduction in error of 53.2% as opposed to the 42.6% value for the “leader” alone. However, the precision  $U$  of the trivariate prediction  $\mathcal{P}_{FLR}$  is lower than that of the simple follow-the-leader prediction  $\mathcal{P}_{FL}$ . Given our previous comments about the trade-off of precision and success, it is possible that the higher  $V$  value is largely an artifact of the lower precision. We need a method for assessing the value of the “ratifier” variable to the trivariate proposition which does not merely reflect this trade-off.

To “partial out” the value of the information about the leader’s price change in order to assess the extra value of the ratifier price information to the trivariate proposition  $\mathcal{P}_{FLR}$ , we want to look at *conditional* probabilities, given the leader’s move. This is, of course, directly analogous to the definition of a partial correlation in the multivariate normal situation.

### 4.3. SUBPOPULATION PARTIALS

Rather than evaluate the entire two-predictor model, as in the calculation of multiple  $V$ , it is natural to focus here on the subpopulations defined by the states of the “leader” variable. In each subpopulation, we consider

TABLE 3  
LEADER SUBPOPULATION PARTIALS  
FOR THE  $\mathcal{P}_{FLR}$  PREDICTION

	<i>Leader's price change</i>			
	++	+	0	-
<i>K</i>	.341	.350	.087	.500
<i>U</i>	.645	.570	.157	.500
$\nabla$	.471	.386	.446	.000

how successfully the model  $\mathcal{P}_{FLR}$  predicts for the various price changes by the "ratifier" variable. (This subpopulation perspective also has its analog in partial correlation analysis.) For example, in the subpopulation where "leader" is ++, we use the expression for bivariate  $\nabla$  given by Eqs. (2.4)–(2.6) to define the (subpopulation partial)  $\nabla$  for the  $\mathcal{P}_{FLR}$  predictions based on the various states of the "ratifier" variable. Given  $L = ++$ , the  $\mathcal{P}_{FLR}$  predictions based on the  $R$  predictor are (++) or (+)  $\rightsquigarrow$  ++, & (0 or -)  $\rightsquigarrow$  0. More generally, rewriting the error measure for the trivariate prediction as  $\omega(v, y) = \omega(w, x, y)$ , the subpopulation partial for the prediction  $\mathcal{P}_{YXW}$ , given  $X = x$ , is

$$\begin{aligned} \nabla_{\mathcal{P}_{YXW|x}} &= 1 - \frac{K_{\mathcal{P}_{YXW|x}}}{U_{\mathcal{P}_{YXW|x}}} \\ &= 1 - \frac{\sum_w \sum_y \omega(w, x, y) f(w, y|x)}{\sum_w \sum_y \omega(w, x, y) f(w|x) f(y|x)}. \end{aligned} \quad (4.2)$$

Table 3 shows the subpopulation results for the  $\mathcal{P}_{FLR}$  example.

When the leader does not decrease his price (++ , + , or 0), the ratifier variable adds a modest amount of predictive value for  $\mathcal{P}_{FLR}$ , but, for the last category (-), there is no reduction in error as a result of using the  $\mathcal{P}_{FLR}$  prediction for the various ratifier variable states. (One can easily construct examples where there is a mixture of positive and negative subpopulation partials.) The ratifier information has different value to  $\mathcal{P}_{FLR}$  across the leader categories; how can we summarize its overall value?

#### 4.4. OVERALL PARTIAL $\nabla$

To develop the overall partial  $\nabla$  for  $\mathcal{P}_{YXW}$ , controlling for  $X$ , we use the same basic strategy as before by comparing error rates when the  $\mathcal{P}_{YXW}$  predictions are made under two information conditions. Since we want to control for  $X$ , the state of  $X$  is always known under both information conditions. When  $W$  also is known, the error for  $\mathcal{P}_{YXW}$  is  $K =$

$\sum_w \sum_x \sum_y \omega(w, x, y) f(w, y|x) f(x)$ . When  $\mathcal{P}_{YXW}$  is applied to randomly drawn observations for which  $X$  but not  $W$  is known, the expected error is  $U = \sum_w \sum_x \sum_y \omega(w, x, y) f(w|x) f(y|x) f(x)$ . Then the proportionate reduction in error for  $\mathcal{P}_{YXW}$  attributable to  $W$  (controlling for  $X$ ) is given by the overall partial

$$\begin{aligned} \nabla_{\mathcal{P}_{YXW|X}} &= 1 - \frac{\sum_x f(x) K_{\mathcal{P}_{YXW|X}}}{\sum_x f(x) U_{\mathcal{P}_{YXW|X}}} \\ &= 1 - \frac{\sum_w \sum_x \sum_y \omega(w, x, y) f(w, y|x) f(x)}{\sum_w \sum_x \sum_y \omega(w, x, y) f(w|x) f(y|x) f(x)}. \end{aligned} \tag{4.3}$$

It follows that the overall partial is a weighted average of the subpopulation partials

$$\nabla_{\mathcal{P}_{YXW|X}} = \sum_x \left[ \frac{f(x) U_{\mathcal{P}_{YXW|X}}}{\sum_{x'} f(x') U_{\mathcal{P}_{YXW|X'}}} \right] \nabla_{\mathcal{P}_{YXW|X'}} \tag{4.4}$$

where a prime indicates a separate summation.

In the example, the value of this overall partial equals .376. Thus, overall, the ratifier information makes a nontrivial contribution to the performance of  $\mathcal{P}_{FLR}$  in predicting the followers' moves. This indicates that the larger  $\nabla$  value for the leader-ratifier-follower model is not merely an artifact of differential precision rates, but rather reflects the actual predictive value of the ratifier variable to the trivariate proposition.

#### 4.5. PREDICTION SHIFTS

General results for moving from bivariate to trivariate analysis are not simple. Partial  $\nabla$  can be 0 or negative, even when trivariate  $\nabla$  exceeds bivariate  $\nabla$ . The reason for this is that while one is adding information in the form of another variable, one is also shifting predictions from, say,  $\mathcal{P}_{FL}$  to  $\mathcal{P}_{FLR}$ . The trivariate prediction will have a different precision, and will also use the original independent variable differently. For a complete accounting of changes in going from a bivariate to a trivariate model, one needs to assess "U-shift"—the ratio of trivariate precision to bivariate precision—and "K-shift"—the ratio of error expected when only the original independent variable information is used in the trivariate prediction to the error expected when it is used in the bivariate prediction.

The numerator and denominator in the U-shift and the denominator in the K-shift are by now familiar quantities. The K-shift numerator is in fact equal to the partial  $U$  error. In our two predictor example, this is

$$\sum_l f(l) U_{\mathcal{P}_{FLR|l}}$$

The shift values for the ratifier problem are

$$U\text{-shift} = \frac{U_{\mathcal{P}_{FLR}}}{U_{\mathcal{P}_{FL}}} = \frac{.676}{.755} = .895,$$

$$K\text{-shift} = \frac{\sum_l f(l)U_{\mathcal{P}_{FLR|l}}}{K_{\mathcal{P}_{FL}}} = \frac{.507}{.433} = 1.171.$$

The full accounting equation is

$$(1\text{-trivariate } \nabla) \\ = (1\text{-bivariate } \nabla)(1\text{-overall partial } \nabla)(K\text{-shift}/U\text{-shift}),$$

or

$$(1\text{-}\nabla_{\mathcal{P}_{FLR}}) = (1\text{-}\nabla_{\mathcal{P}_{FL}})(1\text{-}\nabla_{\mathcal{P}_{FLR|L}})(K\text{-shift}/U\text{-shift}). \quad (4.5)$$

In the example, the ratio of shifts equals 1.308. Values of the shift ratio that differ substantially from 1.0 can lead to some strange sounding results. It is not enough merely to “partial out” a variable; the effect of shifting predictions must also be assessed.

#### 4.6. MULTIPLE AND PARTIAL CORRELATION

The problem of accounting for prediction shifts can be understood relative to the baseline provided by standard correlation analysis. Our development of mixed strategy predictions that led to a bivariate  $\nabla$  equal to the correlation ratio and  $r^2$  generalizes readily. For example, if there are two predictors, the multiple prediction is “With probability  $f(y|x, w)$  predict  $Y = y$ .” The corresponding  $\nabla$  measure is

$$\eta^2 = 1 - (2E_{x,w}[\text{Var}(Y|X, W)]/2 \text{Var}(Y)), \quad (4.6)$$

and the partial  $\nabla$ , controlling, say, for  $X$ , is

$$\eta_{Y|W|X}^2 = 1 - (2E_{x,w}[\text{Var}(Y|X, W)]/2E_x[\text{Var}(Y|X)]). \quad (4.7)$$

If linearity is satisfied conditional on  $X$  and on  $W$  separately as well as on  $X$  and  $W$  jointly, as is true when  $W$ ,  $X$ , and  $Y$  have a multivariate normal distribution, then the multiple and partial correlation ratios are equal to multiple and partial correlations.

The accounting equations for correlation models are much simpler than (4.5). It is well known that the accounting equation for a trivariate linear model is

$$(1 - R^2) = (1 - r_{YX}^2)(1 - r_{Y|W|X}^2).$$

There are no shifts. There is no  $U$ -shift because both the trivariate and bivariate models lead to  $U$  errors that are both equal to (twice, in the mixed strategy interpretation) the variance of  $Y$ . This is easy to see in the standard approach where, to provide a benchmark  $U$ , one always predicts the unconditional mean of the dependent variable, no matter how many predictors are used. With squared error, taking expectations over the mixed strategies leads to the same result. There is no  $K$ -shift because, in the standard approach, the conditional mean of  $Y$  given  $X$ , say, is both the  $K$  prediction in the bivariate model and the  $U$  prediction (with knowledge of  $X$  but not  $W$ ) for the partial.

## 5. Estimation of $V$ from Sample Data

In the previous sections we have assumed explicitly that the population joint distribution was known. In practice, of course, one must estimate  $V$  based on limited amounts of data. In this section, we turn to issues of estimation and hypothesis testing. We emphasize that this section pertains to estimating  $V$  values for predictions that are fully specified *a priori* and not to estimating parameters of prediction rules as, for example, in ordinary least squares regression.<sup>5</sup> For simplicity, we assume that the data are bivariate and constitute a simple random sample with neither independent nor dependent variable controlled by the researcher. The general principles of this section apply equally well to more complex sampling schemes and to the multiple, partial, and shift statistics of the multivariate methods.

### 5.1. THE DISCRETE APPROACH

Once again, we present the argument in terms of cross-tabulated data. Quantitative variables such as those common in economics can be treated as discrete. For example, one might identify the categories \$.00, \$.01, . . . , \$9,999,999.99. Thus the principles for cross classifications also apply to quantitative variables which are treated as discrete. In Section 5.2 we indicate that these principles extend naturally to continuous variables.

The natural (and maximum likelihood) estimator of a cell probability is  $\hat{f}(x_j, y_i) = n_{ij}/n$ , where  $n_{ij}$  is the number of sample points having  $X = x_j$  and  $Y = y_i$ . Similarly, the unconditional (marginal) probabilities may be estimated by the observed fractions  $\hat{f}(y_i) = n_{i.}/n$ , and  $\hat{f}(x_j) = n_{.j}/n$ . The most

<sup>5</sup> For methods of statistical inference about predictions developed *ex post* from the joint probability distribution of discrete variables and the relation of these procedures to chi square, see Hildebrand, Laing, & Rosenthal (1977a, pp. 221-230, 243-248, 289-292).

obvious estimator is obtained by replacing true probabilities by these estimates in the definition of  $\hat{V}$ :

$$\hat{V} = 1 - \left( \sum_x \sum_y \omega(x, y) \hat{f}(x, y) / \sum_x \sum_y \omega(x, y) \hat{f}(x) \hat{f}(y) \right). \quad (5.1)$$

We have shown (Hildebrand, Laing, & Rosenthal 1977a, pp. 232–236) that this estimator is consistent, asymptotically unbiased, and asymptotically normal; we also show that the approximate variance of  $\hat{V}$  is

$$\text{Var}(\hat{V}) \simeq (n - 1)^{-1} \left\{ \sum_x \sum_y [a(x, y)]^2 f(x, y) - \left[ \sum_x \sum_y a(x, y) f(x, y) \right]^2 \right\},$$

where

$$\begin{aligned} a(x, y) &= U^{-1} \{ \omega(x, y) - (1 - \nabla) [\pi(x) + \pi(y)] \}, \\ \pi(x) &= \sum_y \omega(x, y) f(y), \quad \pi(y) = \sum_x \omega(x, y) f(x). \end{aligned} \quad (5.2)$$

Since probabilities in Eq. (5.2) can be replaced by sample estimates without affecting asymptotic normality, the usual normal-distribution methods for hypothesis testing and confidence intervals can be used, at least for large samples.

As in any asymptotic theory, the obvious question is How large must a sample be? In other words, how adequate is the normal approximation for realistic sample sizes? We have done extensive Monte Carlo studies and, for special cases, comparisons with exact distributions (Hildebrand, Laing, & Rosenthal 1977a, pp. 211–216). The conclusions are compatible with general principles of such statistical approximations, so we feel reasonably confident of their general applicability. A good index seems to be  $n \times \min[K, 1 - K]$ . When this index exceeds 5, the approximation is reasonably good, particularly when a continuity correction is made. The point about this index is that it depends only on the *aggregate* quantity  $K$  and the full sample size  $n$ . There is *no* requirement that the expected number in each cell be at least one or five or whatever, as there is in a chi-square test. Therefore, for *a priori* theory, it is possible to use normal approximations even when the data are sparse. Thus our suggestion that quantitative economic data might be analyzed in terms of discrete variables (with possibly enormous numbers of values) is not so absurd as it might seem.

## 5.2. THE CONTINUOUS APPROACH

The direct estimation approach, treating the underlying variables as continuous, has not yet been worked out fully. We can sketch the natural estimation procedure for the continuous case. In this section a subscript ( $t$ )

indexes observations (whereas previously subscripts indexed variable states). To do this, note that

$$\begin{aligned} \nabla &= 1 - (K/U), \\ K &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(x, y)f(x, y) dx dy = E[\omega(X, Y)], \\ U &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(x, y)f(x)f(y) dx dy = E[\omega(X, Y^*)], \end{aligned} \tag{5.3}$$

where  $f(x, y)$  is the joint density of  $X$  and  $Y$ , while the marginal densities are

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

The interpretation of  $U$  requires random variables which are statistically independent with respective densities  $f(x)$  and  $f(y)$ . For these random variables, a pair is denoted  $(X, Y^*)$  to avoid confusion with the  $(X, Y)$  pair which has the joint distribution  $f(x, y)$ . Now assume that we have  $n$  independently sampled bivariate observations,  $(x_1, y_1), \dots, (x_t, y_t), \dots, (x_n, y_n)$ . The natural estimator of  $K$  is

$$\hat{K} = \frac{1}{n} \sum_{t=1}^n \omega(x_t, y_t), \tag{5.4}$$

which is merely the sample mean estimating the population mean. The estimator of  $U$  is not quite so obvious; the natural estimator, by direct analogy to the categorized data estimator, is

$$\hat{U} = \frac{1}{n^2} \sum_{t=1}^n \sum_{t'=1}^n \omega(x_t, y_{t'}). \tag{5.5}$$

We then estimate  $\nabla$  by

$$\hat{\nabla} = 1 - (\hat{K}/\hat{U}). \tag{5.6}$$

By the law of large numbers and the central limit theorem, it follows that  $\hat{K}$  is consistent and asymptotically normal. We conjecture (in part because the result holds for the categorical case) that the same holds true for  $\hat{U}$  and hence for the estimator of  $\nabla$ .<sup>6</sup> Further, the fact that this estimator is identical

<sup>6</sup> Note that one can easily show that  $[n(n-1)]^{-1} \sum_{t=1}^n \sum_{t' \neq t} \omega(x_t, y_{t'})$  is an unbiased estimator of  $U$ . The estimator of  $K$  also is unbiased. However, since use of two unbiased estimators in the ratio does not imply an unbiased estimator for  $\nabla$ , there is no compelling reason to use this alternative estimator of  $U$ .

to that for a “fine-grained” categorization of the data suggests that the same normal approximation which is appropriate for cross classifications also applies to continuous variables. However, we have not proved such a result.

## 6. Summary and Directions for Research

This paper outlines an approach for evaluating the performance of models which make set predictions about a dependent variable and establishes some links between this approach and econometric methods. Set predictions vary in precision in the sense that the prediction becomes increasingly less precise as the size of the predicted set increases from a single point. The approach provides measures for assessing the performance of a model relative to its precision and facilitates the comparative evaluation within the same framework of alternative predictions about the same criterion, even though these predictions might differ in precision.

The approach has been developed for analyzing cross classifications of qualitative variables or discrete quantitative variables. An initial extension of these methods to continuous variables was described, but further extensions are clearly in order.

While this paper has emphasized the evaluation of predictions stated *a priori*, the framework also has been applied to select predictions *ex post facto* (Hildebrand, Laing, & Rosenthal 1977a, pp. 132–145, 221–230, 289–291). The related statistical theory developed for evaluating predictions selected *ex post* provides hypothesis tests that are closely related to chi-square tests of association. Unfortunately, the tests are extremely conservative. A more refined approach is needed to generate more powerful tests.

Although there are other important questions related to statistical inference that deserve attention, the critical questions that are raised by the framework discussed in this paper pertain not to statistical inference, but rather to the evaluation of predictions even when the data constitute the entire population of interest. The prediction analysis methods have been developed extensively for bivariate analysis. Some essential foundations for multivariate analysis—including the multiple and partial  $\nabla$  measures and associated statistical theory—also have been developed. However, many important issues in multivariate analysis have not been touched. For example, our work has emphasized a “single-equation” model with one dependent variable. Although some very limited results for recursive systems have appeared (Hildebrand, Laing, & Rosenthal 1975), the important questions concerning simultaneous systems of set predictions have not yet been addressed. Progress on these topics will be of use to economists if economic theory continues to generate set predictions, whether about qualitative or quantitative variates.

**Appendix**

We first develop  $\nabla$ -measures for the pure strategy, *ex post* prediction based on conditional means and for a bivariate linear prediction, then derive  $\eta^2$  as the  $\nabla$ -measure for the best-fitting actuarial proposition applied as a mixed strategy prediction. Throughout this appendix, error is measured as the square of the difference between the actual and predicted value of  $Y$ .

**A.1. EVALUATING THE ABSOLUTE PREDICTION BASED ON CONDITIONAL MEANS**

This prediction may be written  $\mathcal{M}: \{x \rightsquigarrow \mu_{Y|x} | x \in X\}$ . The total error for this prediction when  $X$  is known equals

$$\begin{aligned} K_{\mathcal{M}} &= \sum_x \sum_y (y - \mu_{Y|x})^2 f(y|x) f(x) = \sum_x \text{Var}(Y|x) f(x) \\ &= E_X[\text{Var}(Y|X)]. \end{aligned}$$

When  $X$  is unknown for any randomly drawn observation, then with probability  $f(x)$  predict  $Y = \mu_{Y|x}$ . The expected error under this condition equals

$$U_{\mathcal{M}} = \sum_x \sum_y (y - \mu_{Y|x})^2 f(x) f(y).$$

Expanding the square around  $\mu_Y$ ,

$$\begin{aligned} U_{\mathcal{M}} &= \sum_x \sum_y (y - \mu_Y + \mu_Y - \mu_{Y|x})^2 f(x) f(y) \\ &= \sum_x \sum_y (y - \mu_Y)^2 f(x) f(y) + 2 \sum_x \sum_y (y - \mu_Y)(\mu_Y - \mu_{Y|x}) f(x) f(y) \\ &\quad + \sum_x \sum_y (\mu_Y - \mu_{Y|x})^2 f(x) f(y) \\ &= \sum_y (y - \mu_Y)^2 f(y) \sum_x f(x) + 2 \sum_x (\mu_Y - \mu_{Y|x}) f(x) \sum_y (y - \mu_Y) f(y) \\ &\quad + \sum_x (\mu_Y - \mu_{Y|x})^2 f(x) \sum_y f(y) \\ &= \text{Var}(Y) + E_X(\mu_{Y|x} - \mu_Y)^2. \end{aligned}$$

Therefore,

$$\nabla_{\mathcal{M}} = 1 - \frac{E_X[\text{Var} Y | X]}{\text{Var}(Y) + E_X(\mu_{Y|x} - \mu_Y)^2}.$$

as given by (2.10).

## A.2. THE V-MEASURE FOR A BIVARIATE LINEAR PREDICTION

Given the linear prediction  $\mathcal{L} = \{x \rightsquigarrow \gamma + \delta x\}$  and squared error, the total error expected when  $X$  is known, as given in the text, is

$$K_{\mathcal{L}} = \sum_x \sum_y (y - \gamma - \delta x)^2 f(y|x) f(x).$$

Expanding the square around  $\mu_{Y|x}$ ,

$$\begin{aligned} K_{\mathcal{L}} &= \sum_x \sum_y (y - \mu_{Y|x})^2 f(y|x) f(x) + 2 \sum_x \sum_y (y - \mu_{Y|x})(\mu_{Y|x} - \gamma - \delta x) f(y|x) f(x) \\ &\quad + \sum_x \sum_y (\mu_{Y|x} - \gamma - \delta x)^2 f(y|x) f(x) \\ &= \sum_x \text{Var}(Y|x) f(x) + \sum_x (\mu_{Y|x} - \gamma - \delta x)^2 f(x) \\ &= E_X[\text{Var}(Y|X)] + E_X(\gamma + \delta x - \mu_{Y|x})^2. \end{aligned}$$

Similarly, expanding the square around  $\mu_Y$  yields

$$\begin{aligned} U_{\mathcal{L}} &= \sum_x \sum_y (y - \gamma - \delta x)^2 f(y) f(x) \\ &= \text{Var}(Y) + E_X(\gamma + \delta x - \mu_Y)^2. \end{aligned}$$

Therefore,

$$\nabla_{\mathcal{L}} = 1 - \frac{E_X[\text{Var}(Y|X)] + E_X(\gamma + \delta x - \mu_{Y|x})^2}{\text{Var}(Y) + E_X(\gamma + \delta x - \mu_Y)^2}.$$

To show that the sign of  $\nabla_{\mathcal{L}}$  is determined by the correlation between  $\gamma + \delta x$  and  $\mu_{Y|x}$ , note that  $\nabla_{\mathcal{L}} \{>, =, <\} 0$  if and only if  $U_{\mathcal{L}} \{>, =, <\} K_{\mathcal{L}}$ . Using the standard identity

$$\text{Var}(Y) = E_X[\text{Var}(Y|X)] + E_X(\mu_{Y|x} - \mu_Y)^2,$$

it follows that  $\nabla_{\mathcal{L}} \{>, =, <\} 0$  if and only if

$$\begin{aligned} E_X(\mu_{Y|x} - \mu_Y)^2 \{>, =, <\} \sum_x [(\gamma + \delta x - \mu_{Y|x})^2 - (\gamma + \delta x - \mu_Y)^2] f(x) \\ = \sum_x (\mu_{Y|x}^2 - \mu_Y^2) f(x) - 2 \sum_x (\gamma + \delta x)(\mu_{Y|x} - \mu_Y) f(x), \end{aligned}$$

where the last equality follows by expanding the square in the bracketed term. But by another expand-the-square argument,

$$\sum_x (\mu_{Y|x}^2 - \mu_Y^2) f(x) = E_X(\mu_{Y|x} - \mu_Y)^2.$$

Consequently,

$$\nabla_{\mathcal{L}} \begin{matrix} \geq \\ \leq \end{matrix} 0 \quad \text{iff} \quad \sum_x (\gamma + \delta x)(\mu_{Y|x} - \mu_Y) f(x) \begin{matrix} \geq \\ \leq \end{matrix} 0.$$

The sum in this expression can be replaced by the covariance of the prediction with the conditional mean:

$$\begin{aligned} \text{Cov}(\gamma + \delta x, \mu_{Y|x}) &= \sum_x [\gamma + \delta x - E(\gamma + \delta x)](\mu_{Y|x} - \mu_Y)f(x) \\ &= \sum_x (\gamma + \delta x)(\mu_{Y|x} - \mu_Y)f(x). \end{aligned}$$

Since the sign of the correlation is identical to the sign of the covariance, it follows that

$$\nabla_{\mathcal{Q}} \geq 0 \quad \text{iff} \quad \text{correlation}(\gamma + \delta x, \mu_{Y|x}) \geq 0.$$

### A.3. THE CORRELATION RATIO AS A $\nabla$ -MEASURE

The best fitting actuarial proposition may be written  $\mathcal{Q}^* : \{\mathcal{Q}^*(x) | x \in X\}$ , where  $\mathcal{Q}^*(x) = \{f(y|x) | y \in Y\}$ . Applying  $\mathcal{Q}^*$  as a mixed strategy, the component  $\mathcal{Q}^*(x)$  is selected with probability  $f(x)$  under either information condition. Given that  $\mathcal{Q}^*(x)$  is selected, we predict, with probability  $f(p|x)$ , that  $Y = p$ ; therefore, the event  $Y = a$  is assigned error weight  $(a - p)^2$  with probability  $f(p|x)$ . These comments apply whether or not the  $X$  value is known when the prediction is made for an observation.

In the first information condition, each observation's  $X$  value is known when the prediction is made. Since  $Y = a$  occurs with conditional probability  $f(a|x)$  in the  $X = x$  subpopulation, the total error expected for the mixed strategy when  $X$  is known equals

$$K_{\mathcal{Q}^*} = \sum_x \sum_{a \in Y} \sum_{p \in Y} (a - p)^2 f(p|x) f(a|x) f(x).$$

Expanding the square around  $\mu_{Y|x}$  yields

$$K_{\mathcal{Q}^*} = 2E_X[\text{Var}(Y | X)].$$

If  $X$  is unknown, then for any observation randomly drawn from the entire distribution, the component prediction  $\mathcal{Q}^*(x)$  is selected with probability  $f(x)$ , and  $Y = a$  occurs with its unconditional probability,  $f(a)$ . Replacing  $f(a|x)$  with  $f(a)$  in the initial expression given for  $K_{\mathcal{Q}^*}$  above, we may express the total error expected for the mixed strategy when  $X$  is unknown as

$$\begin{aligned} U_{\mathcal{Q}^*} &= \sum_x \sum_{a \in Y} \sum_{p \in Y} (a - p)^2 f(p|x) f(a) f(x) \\ &= \sum_a \sum_p (a - p)^2 f(p) f(a). \end{aligned}$$

Expanding the square around  $\mu_Y$  yields

$$U_{\mathcal{Q}^*} = 2 \text{Var}(Y).$$

Therefore,

$$\nabla_{\varrho^*} = 1 - (2E_X[\text{Var}(Y|X)]/2 \text{Var}(Y)).$$

Consequently,  $\nabla_{\varrho^*} = \eta^2$ , the correlation ratio, as given by (2.9).

#### REFERENCES

- Borch, K. Application of game theory to some problems in automobile insurance. *ASTIN Bulletin*, 1962–1963, 2, 208–221.
- Dutton, D. S. The economics of inflation and output fluctuations in the United States, 1952–74. In K. Brunner & A. H. Meltzer (Eds.), *The problem of inflation* (Carnegie–Rochester conference series on public policy, Vol. 8). Amsterdam: North-Holland Publ., 1978. Pp. 203–232.
- Fourcans, A. Inflation and output growth: The French experience, 1960–75. In K. Brunner & A. H. Meltzer (Eds.), *The problem of inflation* (Carnegie–Rochester conference series on public policy, Vol. 8). Amsterdam: North-Holland Publ., 1978. Pp. 81–140.
- Gately, D. Sharing the gains from regional cooperation: A game-theoretic application to planning investment in electric power. *International Economic Review*, 1974, 15, 195–208.
- Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications. *Journal of the American Statistical Association*, 1954, 49, 732–764.
- Hildebrand, D. K., Laing, J. D., & Rosenthal, H. A prediction logic approach to causal models of qualitative variates. In D. R. Heise (Ed.), *Sociological methodology*. San Francisco: Jossey-Bass, 1975. Pp. 146–175.
- Hildebrand, D. K., Laing, J. D., & Rosenthal, H. *Prediction analysis of cross classifications*. New York: Wiley, 1977. (a)
- Hildebrand, D. K., Laing, J. D., & Rosenthal, H. *Analysis of ordinal data* (Sage University Paper series on Quantitative Applications in the Social Sciences, Series No. 07–008). Beverly Hills and London: Sage, 1977. (b)
- Korteweg, P. The economics of inflation and output fluctuations in the Netherlands 1954–75: A test of some implications of the dominant impulse-cum-rational expectations hypothesis. In K. Brunner and A. H. Meltzer (Eds.), *The problem of inflation* (Carnegie–Rochester conference series on public policy, Vol. 8). Amsterdam: North-Holland Publ., 1978. Pp. 17–68.
- Littlechild, S. C., & Thompson, G. F. Aircraft landing fees: A game theory approach. *Bell Journal of Economics*, 1977, 8, 186–204.
- Mossin, J. Merger agreements: Some game-theoretic considerations. *Journal of Business*, 1968, 41, 460–471.
- Segal, D. On making custom unions fair: An East African example, *Yale Economic Essays*, 1970, 10, 115–160.
- Sorenson, J. R., Tschirhart, J. T., & Whinston, A. B. A game-theoretic approach to peak-load pricing. *Bell Journal of Economics*, 1976, 7, 497–520.