ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 3

Entropy - Reducing Properties of Predictive Financial Models

N. J. Macleod and J. D. Thomison

August 27, 1992

Abstract

The fundamental objects of communication theory (sources and receivers, channels, the rate of transmission of information) are shown to have natural analogues within economic systems and models. Kelly's interpretation of the information rate [Bell System Technical Journal, 1956] is generalized: his "gambler with a private wire" is replaced by an investor with a predictor model, operating in an arbitrage-free economic system. The potential usefulness of the theory is established by the demonstration of an important predictor model producing strongly positive information rates over extended time periods.

Introduction

Kelly's "A New Interpretation of Information Rate"

Most of the present paper is concerned with properties of a concept known as the "rate of transmission of information" or, more simply, as the "information rate." The importance of the concept derives from Shannon's Second Theorem [6] which asserts that, with suitable encoding, binary digits can be transmitted at this rate over a noisy communication channel, with arbitrarily small probability of error. The result is central in the theory of communication as well as in engineering applications; it is often referred to as the "Fundamental Theorem for Noisy Channels."

Soon after the publication of Shannon's theorem, workers in a wide range of academic and applied disciplines began making informal use of the information rate concept; that is, in ways unrelated to coding schemes. Indeed, the information rate is a simple and appealing measure of the association between two random variables. In 1956, J.L. Kelly [1] described a situation where coding is irrelevant yet the information rate has theoretical significance. Kelly summarizes his result as follows (our italics):

> If the input symbols to a communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities (i.e.,"fair" odds), a gambler can use the knowledge given him by the received symbols to cause his money to grow exponentially. *The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel.*

Although literally thousands of papers have been written on information theory and its applications since 1956, it appears that the characterization of the information rate as the maximum exponential rate of growth of a gambler's capital remains the only theoretically justified application of the concept outside communication theory. In his fascinating 1991 survey [5] of the many ramifications of self-similarity, Manfred Schroeder (a longtime colleague of Kelly's at Bell Labs) writes (again, italics are ours): "[Kelly's result] was the first instance, *and is still the only one, as far as, I know*, in which a benefit can be reaped [from information theory] without the elaborate coding that is necessary to realize the error-free transmission promised by Shannon's theory."

A New - and Improved - Interpretation of Information Rate

Kelly's result is unsatisfactory in several respects. Although the problem he constructs and solves, to give significance to the information rate, is simple and intuitively compelling, it is also artificial: few (honest) gamblers are so fortunate as to possess a "private wire" continually providing inside information on infinitely repeated trials of the same experiment. Second, the rather narrow gambling context itself limits the relevance of the result. But most importantly, Kelly's hypothetical conditions are *unnecessarily* restrictive: the central conclusion can be reached in a much more general and interesting setting - that of an arbitrage-free economic system. This is the basic result of the present paper.

Organization of this paper

The remainder of the paper is organized as follows.

Section 1, "Entropy, information, and financial models", gives the definitions, terminology, and results we will need. (There are many excellent texts and treatises for the reader who is interested in pursuing the subject; please refer to the bibliography for some suggestions). We then show that the fundamental objects of the theory (sources and receivers, channels, the information rate) have natural analogues in the context of financial markets and predictive models.

Section 2, "Information and market dynamics" contains the main result of the paper. Here we establish the equivalence of arbitrage-free economic systems and a generalized betting model in such a way that the interpretation of the information rate, as the maximum rate of increase of an investor's capital, is preserved. As a corollary to this result we show that the *difference* in the information rate between models represents the achievable incremental return associated with the superior predictor model.

The theoretical material of Section 2 is of little practical significance unless we can point to at least one important economic system with a strongly positive information rate. This is the main result in Section 3, "Application to financial models." The section also contains, partly to illustrate the calculations and partly to show the versatility of information-theoretic techniques, an example in which the information rate is used as a means of comparing several predictors of the future volatility of the S&P 500 index.

1. Entropy, information, and financial models

The entropy, H(X), of a random variable X which takes the values $x_1,\dots,x_m$ with respective probabilities $p_1,\dots,p_m$, is defined as

$$H(X) = -\sum_{s} p_s \log p_s \qquad (1.1)$$

where $s = 1,\dots,m$

The base to which logarithms are taken determines the units in which entropy is expressed; if base 2 is used, the unit is the binary digit, or bit. H(X) is a measure of the average uncertainty associated with the outcome of a single trial of an experiment with probabilities $p_1,\dots,p_m$. Knowledge of the outcome eliminates the uncertainty and thereby conveys an equal amount of information. Thus H(X) may be taken to represent the average amount of information given by knowledge of the outcome of a single such trial.

We can extend the definition of entropy to joint and conditional probability distributions. Let X and Y be discrete random variables with joint probability function f(x,y) and marginal probability functions

$$h(x) = \sum_{y} f(x,y) \qquad (1.2)$$

$$g(y) = \sum_{x} f(x,y) \qquad (1.3)$$

Let h(x/y) be the conditional probability of X when the value of Y is known and let g(y/x) be the conditional probability of Y given X. The entropies of X,Y, the pair (X,Y), and the conditional variables (X/Y) and (Y/X) are given respectively by

$$H(X) = -\sum_{x} h(x) \log h(x) \qquad (1.4)$$

$$H(Y) = -\sum_{y} g(y) \log g(y) \qquad (1.5)$$

$$H(X,Y) = -\sum_{x}\sum_{y} f(x,y) \log f(x,y) \qquad (1.6)$$

$$H(X/Y) = -\sum_{y}\sum_{x} g(y)\, h(x/y) \log h(x/y) \qquad (1.7)$$

$$H(Y/X) = -\sum_{x}\sum_{y} h(x)\, g(y/x) \log g(y/x) \qquad (1.8)$$

and the following relationships hold

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y) \qquad (1.9)$$

The entropy H(Y/X) measures the average uncertainty associated with the value of Y when the value of X is known, and is thus an inverse measure of the degree of dependence of Y upon X. If Y is independent of X, then H(Y/X) = H(Y) and H(X,Y) = H(X) + H(Y); if Y is strictly determined by X, then H(Y/X) = 0 and H(X,Y) = H(X). In general, $0 \leq H(Y/X) \leq H(Y)$, and similarly for H(X/Y).

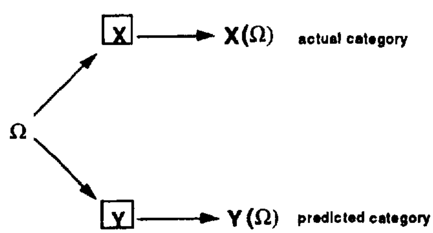
Many aspects of financial modelling may be viewed usefully in terms of entropy or information. For example, the question of whether the value of adding a new factor to a single or multi-factor interest rate model justifies the increase in the complexity of the model may be approached by calculating (or estimating) the contribution of the added factor to the overall entropy of the model. In what follows, however, we are concerned primarily with predictive financial models, a category in which we include any models which purport to predict or forecast some aspect of market behavior.

In the present context, we are not interested in the nuts and bolts of such models; the approach we take will be seen to be very general, applying equally well to rough estimates of future market volatility and to the output of sophisticated quantitative models used in active portfolio management. The central idea is to treat a predictive model as a *communication channel* between the market and the model-user. Input to the channel consists of the actual values of some specified market variable while the corresponding predicted values represent the output. In this formulation, failures of prediction are treated as errors of communication.

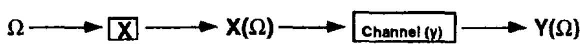
More formally, suppose that a particular aspect of market behavior can be classified into a number of distinct categories and that we have a model which forecasts the category. As an example, an active management model might translate its input data into categories given by "buy", "do nothing" or "sell." In retrospect, the appropriate course of action can also be expressed as one of these categories, and a comparison of the actual and predicted categories can be made.

Let Ω represent the state of the financial markets at the time a forecast is made. Ω is taken to include the totality of market data, investor intentions, government policy, etc., much of which cannot be known. We can describe the action of the market as an unknown random function X which maps Ω to the categories we have defined (say A, B and C in the example above). Similarly we can treat the forecast as a function Y which maps Ω to the same categories. (Obviously any practical model will use only a small part of Ω).

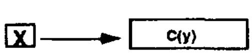
The situation may be represented schematically as follows.



Alternatively, treating the model Y as a communication channel, we have the following representation.



Thus the predictive model Y is treated as the system



This representation leads naturally to the idea of measuring the effectiveness of the model Y in terms of the efficiency with which the channel C(y) transmits information from the market to the model-user. Specifically, we express the effectiveness of Y as the rate R(y) of transmission of information over the channel C(y).

For any communication channel, with input and output given respectively by the values of random variables X and Y, the *rate R of transmission of information* (also referred to as the mutual information and denoted I(X,Y)) is given by

$$R = H(X) - H(X/Y) \qquad (1.10)$$

which in view of (1.9) can be written as

$$R = H(X) + H(Y) - H(X,Y) \qquad (1.11)$$

(1.10) states that the rate of transmission of information over a channel is equal to the entropy of the source X reduced by the conditional entropy of X when the output value of Y is known. For a noiseless channel, knowledge of the value of Y tells us the value of X with certainty so that $H(X/Y) = 0$, and $R = H(X)$. When X and Y are independent (i.e., knowledge of the value of Y conveys no information about the value of X), $H(X/Y) = H(X)$ and $R = 0$.

In section 3 we measure R(y) for a particular active portfolio management model over different time periods and discuss the results. The next section is devoted to placing R(y) in an economic context.

2. Information and market dynamics

In this section we will attempt to provide an intuitive feel for the idea of the information rate of a financial model by relating it to more familiar economic concepts. We consider the effect of proprietary information on an arbitrage-free economy, first in qualitative terms, and then by applying Kelly's result to show how, under specific conditions, the information rate is related to the long-term rate of return achievable in such an economy.

A betting game

We begin by considering a hypothetical game in which a finite number of participants bet upon an experiment with m mutually exclusive outcomes according to the following rules:

(i) Each participant has a fixed, finite initial capital. Initial capital may vary by participant.

(ii) Each participant must bet all of his capital, but need not bet it all on a single outcome.

(iii) The odds offered on the various outcomes are the same for all participants.

(iv) Whatever the outcome of the experiment, all capital is paid out in winnings.

The game is a wealth redistribution mechanism which differs from standard betting systems in that, in view of rules (iii) and (iv), the odds can only be calculated after all bets have been placed. This contrasts with the more familiar parimutuel arrangement where odds are posted beforehand and different bettors bet at different odds. In addition, we allow an iterative procedure whereby after initial bets have been placed and the associated odds have been calculated, the participants have an opportunity to adjust their positions, leading to new odds, and so on. Under mild regularity assumptions such a system will settle into a stable equilibrium in which all participants are satisfied with their bets at odds which conform to rules (iii) and (iv).

Let $s = 1, \ldots, m$ denote the outcome of the experiment, and let the aggregate capital bet on outcome s be w(s). In order to satisfy rules (iii) and (iv), the odds, α_s, associated with outcome s are given by

$$w(s)\ \alpha_s = w \qquad (2.1)$$

for each value of s, where

$$w = \sum_s w(s) \qquad (2.2)$$

That is,

$$\alpha_s = \frac{w}{w(s)} \qquad s = 1, \ldots, m \qquad (2.3)$$

and we note that

$$\sum_s \frac{1}{\alpha_s} = 1 \qquad (2.4)$$

It is important to observe that rule (ii), which requires each participant to bet his entire capital, is not as restrictive as it may appear: at given odds a participant may hold back any part of his capital by offsetting bets. For example, if the experiment has two outcomes, A and B, at odds of 3 and 3/2 respectively, a participant with $400 capital who wishes to place $250 on A and to hold back the remaining $150 may do so by betting $300 on A and $100 on B.

An alternative interpretation of the betting game

A bettor who has placed b(s) on outcome s at odds α_s ($s=1, \ldots, m$) may be regarded as the holder of a portfolio in an economic system governed by the variable s. Before the experiment has been performed, the value of the portfolio is equal to the bettor's initial capital, $\Sigma\, b(s)$. When the outcome s of the experiment is known, the value of the portfolio is equal to $b(s)\ \alpha_s$.

Denoting the state of the economic system before the experiment is performed by ϕ, the state after the outcome has been determined by s, and the value of the portfolio in state x by P(x), we have

$$P(\phi) = \sum_s b(s) \qquad (2.5)$$

and

$$P(s) = b(s)\,\alpha_s \qquad (2.6)$$

so that

$$P(\phi) = \sum \frac{P(s)}{\alpha_s} \qquad (2.7)$$

Writing

$$\theta_s = \frac{1}{\alpha_s} \qquad s = 1,\ldots, m \qquad (2.8)$$

we see that for any such portfolio, the initial value is related to the potential subsequent values by

$$P(\phi) = \sum_s P(s)\,\theta_s \qquad (2.9)$$

where the values $\{\theta_s\}$ are the same for all portfolios, and

$$\sum_s \theta_s = 1 \qquad (2.10)$$

from (2.4).

It follows that each experiment in the betting game is equivalent to a single transition in an arbitrage-free economic system governed by the state-variable s, where the equilibrium parameters $\{\theta_s\}$ of the economic system are the reciprocals of the odds $\{\alpha_s\}$ of the betting game.

While real economic systems do not satisfy rules (iii) and (iv) of the betting game in a strict sense, the process by which the ultimate odds are arrived at provides an instructive model for the response of prices to trading pressure in financial markets. The final odds (which reflect an equilibrium allocation of bets) balance the participants' individual assessments of the probability distribution of the outcome of the experiment and their tolerances for risk. Similarly, in an arbitrage-free economic system, the parameters θ_s reflect investors' collective risk tolerance and probability assessment. In a betting game, depending upon the nature of the experiment, there may be a consensus among the bettors with respect to the probability distribution of the outcome, in which case the final odds are readily separated into probability and aggregate utility components. In economic systems, however, a consensus regarding the relative likelihood of potential future states is rare, and the values θ_s are more appropriately seen simply as the values which balance the subjective probability assessments and individual utilities of the various market participants.

The discussion in the remainder of this section is phrased for the most part in terms of the betting game, with the presumption that economic systems, though more complex, behave in an analogous manner. From time to time, however, we will shift the perspective from that of a betting game to that of an arbitrage-free economic system.

The effect of proprietary information

Suppose that the betting game has reached an equilibrium, with aggregate capital w(s) placed on outcome s. A bettor whose bets are given by b(s) gains access to a model or some other source of information which predicts the outcome of the experiment, and adjusts hi‹ bets as a result. At the current odds, the game will remain in equilibrium only if the other bettors adjust their bets to restore the initial aggregate allocation of bets. If they make no changes, the odds will simply adjust to reflect the revised allocation. In general, though, the changed odds which follow the actions of the informed bettor will set off a series of adjustments to bets and odds which will terminate eventually at a new equilibrium. We can best illustrate some of the possibilities with an example.

Example

The experiment has two possible outcomes, A and B, and the initial aggregate equilibrium bets are

$$w(A) = 100$$

$$w(B) = 200$$

so that the initial odds are

$$\alpha_A = 3$$

$$\alpha_B = {}^3/_2$$

Included in the aggregate bets is the bet of an individual who has capital of 10, and has effectively bet 6 on A and held back 4 by betting $7^1/_3$ on A and $2^2/_3$ on B.

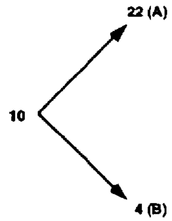
That is,

$$b(A) = 7^1/_3$$

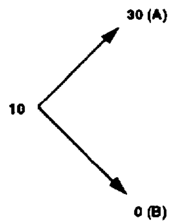
$$b(B) = 2^2/_3$$

Suppose that the individual's source of information predicts outcome A and that he has sufficient confidence in its accuracy to add the 4 he has held back to his bet on A.

If the other bettors change their bets to restore the initial allocation (perhaps an unlikely possibility), the initial odds will remain unchanged. In that case, the individual's prospects have shifted, in portfolio terms, from



to



where the equilibrium parameters are given by

$\theta_A = {}^1/_3$

$\theta_B = {}^2/_3$

On the other hand, if the individual's changed bet prompts no response from the other bettors, the odds themselves must change. In that case, aggregate bets following the individual's shifted bet are

$w'(A) = 102^2/_3$

$w'(B) = 197^1/_3$

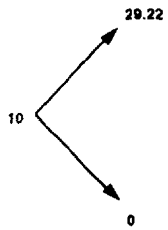
and the new odds are therefore

$$\alpha'_A = \frac{300}{102^2/_3} = 2.922$$

and

$$\alpha'_B = \frac{300}{197^1/_3} = 1.520$$

In this equilibrium, the individual's portfolio may be represented as



and the new equilibrium parameters are

$$\theta'_A = 0.342$$

$$\theta'_B = 0.658$$

If the other bettors have no access to the information source and are not influenced by the individual's actions to shift their own bets towards A, they will have a slight incentive to shift their bets towards B, since the odds on A have shortened and those on B have lengthened. We might expect, therefore, that the final equilibrium will lie somewhere between the two cases illustrated above.

It is clear that the greater the proportion of total capital controlled by bettors with access to, and confidence in, the information source, the greater the influence of the information on the final odds, and the smaller the advantage which can be gained from access to it. In the example above, where access to the information is restricted, the informed bettor is able to shift bets without perturbing the system unduly, thereby obtaining odds close to the initial odds.

Interpreting the information rate of a financial model: Kelly's result

Kelly considers the case of a bettor who is allowed to bet repeatedly at fixed odds on the outcome of an experiment with a fixed and known probability distribution. The bettor is assumed to have access to a source which predicts the outcome, but which does not affect the odds. By allocating his capital in proportion to the conditional probability of each outcome given the predicted outcome, the bettor can add an increment to his maximum long-term rate of return equal to the information rate of the source. In a discrete, arbitrage-free economic system, where the information source is regarded as a predictive model of some kind, Kelly's result may be translated to the following terms.

Let s = 1, ..., m denote the possible future states at the end of a transition. Let p(s,r) be the joint probability of an actual future state s and a model prediction of state r, and let q(r) be the marginal probability of a prediction of state r. The betting strategy which maximizes the long-term rate of return is equivalent to an investment strategy whereby at each transition, given a prediction of state r, the portfolio is constructed so that its value in state s, P(s,r), is given by

$$P(s,r) = P(\phi) \frac{p(s,r)}{\theta_s \, q(r)} \qquad (2.11)$$

Note that portfolios which satisfy (2.11) satisfy

$$P(\phi) = \sum_s P(s,r)\, \theta_s \qquad (2.12)$$

In real economic systems, of course, the "odds" and probabilities are not stationary over time, (although many economic models assume that they are), and as pointed out earlier, the true probability distribution of future states may not be known. Nor, for that matter, is the investment strategy described by (2.11) necessarily optimal in any sense other than that of maximizing long-term return with respect to the variable s. Nevertheless, Kelly's result provides, under static "ideal" conditions, a concrete interpretation of the information rate of a predictive financial model.

In terms of the dynamic betting game model discussed above, Kelly's conditions correspond to a situation in which the capital controlled by bettors with access to the information source is small enough in relation to total capital to have a negligible effect on the equilibrium odds.

As a practical matter, a large part of the information conveyed by any model, however proprietary, is available to other investors through their own models (which can include intuition and judgement as well as more objectively developed components). As indicated above, such information will already be incorporated into market prices. Thus it is the *difference in information rate between models* which is significant in terms of achievable incremental return.

3. Application to financial models

The examples chosen in this section demonstrate the versatility of information-theoretic constructs. Essentially similar calculations are applied to distinct problems, each of which regularly confronts the financial or actuarial practitioner:

- developing efficient estimators for the parameters of a given financial model;
- measuring the absolute performance of a given model;
- selecting the most efficient model from a field of candidates.

Before turning to the examples proper, a few comments are in order.

Although the validity of conclusions drawn on the basis of historical data depends upon future realizations of various economic time series resembling past experience, we make absolutely no assumption with respect to the underlying distributions.

In our framework the analyst is attempting to unravel a message arriving through an unreliable channel -the financial model. Whether he or she makes successful use, or even sense,of the message, is outside the scope of information theory. But it is certainly important to determine whether the analyst has a fighting chance: unless there is some minimal amount of information in the model output it is highly unlikely that this output will be of any genuine use in decision making.

In engineering applications, the various information measures are meaningful in *absolute* terms. For example, they will indicate the amount of redundancy needed to assure, at a high probability level, reliable transmission of a source message. Kelly's result, as we have noted, appears to be the only instance outside the communications engineering context where the information rate has absolute significance.

By contrast, in the examples to follow, the objective is not to produce absolute measures so much as to allow comparison of alternative estimators and models. It is important to note, however, that the information measures we calculate are consistently scaled. That is, if the indices calculated for models A and B are, respectively, 1.05 and .35, then it can be concluded that model A is 3 times as efficient as model B. The validity of this assertion follows from the origin of these measures as (absolute) rates of transmission of information.

Predicting future S&P 500 volatility from historical data

This is a frequently-encountered problem. For example, the Black-Scholes option-pricing model assumes that changes in the price of the underlying stock are lognormally distributed; determination of the standard deviation σ of the associated normal distribution is critical to successful application of the model.

Perhaps the most straightforward approach to predicting the volatility which will prevail in the near future is to use a value based upon the closing averages of a recent experience period. In the sequel we will refer to this predictor as the "simple extrapolation" model. It will provide a benchmark against which to compare other forecasting models.

With 92 months of historical data, we can apply the framework of section 1 to evaluate the simple extrapolation predictor. The calculations are summarized in Exhibit 1a; see also Appendix 1, which gives details of the exploratory and preliminary data analysis. Of course, the performance of any predictor can also be evaluated by more familiar statistical techniques. What we offer here is a fundamentally different perspective.

The mutual information I(X.in, X.out), where X.in is the actual volatility observed in month n, and X.out is the predicted volatility, is about .2553 for the simple extrapolation model. Note that "mutual information" is synonymous here with "information rate." The alternative terminology captures the usefulness of the common concept as a measure of association. Since the source entropy, H(X.in) = 2.3216 (which is approximately equal to $-5 \times {}^1/_5 \log ({}^1/_5)$ with logarithms taken to base 2), we see that the model transmits about 11% of the total information generated at the source.

Before turning to alternative models, it is interesting to see how much information is lost by introducing a one month reporting lag. The calculations are summarized in Exhibit 1b. Here I(X.in, X.out) = .2061. The reporting delay has the effect of reducing the information content of the model by about 20%.

We consider two alternatives to the simple extrapolation model, both of which are based on mean reversion, a property which is suggested by inspection of the time series of monthly σ values. The first model (Exhibit 1c) assumes symmetric mean reversion; the second (Exhibit 1d) assumes that months of high volatility tend to be followed by months of lower volatility but that months of low volatility are not under a mean reverting influence. The asymmetric model appears slightly more effective than the symmetric model; neither is superior to the simple extrapolation model, however. Nor is either significantly inferior, which is interesting in view of the fact that each of these alternatives is based on a reduced alphabet.

Predicting changes in the S&P 500 index

Our objective here is to evaluate the performance of a model which forecasts daily fluctuations in the S&P 500 index. The model is actually one component of a well known tactical asset allocation model which recommends transfers between an equity fund designed to replicate the S&P 500 index, a corporate bond fund, and Treasury bills.

Daily changes in the S&P 500 index forecast by the model are sorted into three categories and compared with actual changes in the index. The method is very similar to that used in the previous example; see Appendix II for details.

From scanning the raw data, it appeared that the model was quite adept at anticipating significant increases in the S&P 500 index. The mutual information calculation given in Exhibit II provides objective support for this hypothesis. During the period January 1, 1989 through June 30, 1989, the model transmitted 23% of the information generated at the source [I(X.in,X.out) = .3651; H(X.in) = 1.5841].

This observation is not inconsistent with the assumption that consecutive index changes are independent. It does suggest, however, that important elements of the process which determines the changes between days n and n+1 are executed on or before day n, and that analysts had measurable success discerning and interpreting some of these early indicators.

The mutual information I(X.in, X.out) fell markedly between the first and second six-month observation periods, and then appears to have stabilized at a rate equal to about 13% of the entropy in the source.

In contrast to the previous example, where we had alternative models to compare, here we are limited to comparing the performance of the same model over different time periods. We encourage the reader to apply the calculations in Exhibit II to his or her own price change forecasting model.

We close the discussion with a couple of suggestions for future consideration and possible research.

It appears to us that there is more meaning in individual information-theoretic measures than we are currently able to exploit. In fact, it seems that these indices should have more significance in an absolute sense than do competing measures of the strength of association of two random variables derived using more conventional statistical techniques.

Thus, when we see a financial model transmitting a significant portion (23% in the example above) of the information generated by a complex economic system, we feel that we are observing a significant, perhaps even noteworthy, event. Unfortunately, there are no benchmarks against which to set the 23% figure.

Contrast this with situations in other fields. An epidemiologist, for example, is familiar with the strength of association between various epidemiological variables (e.g., reduction in longevity vs. number of cigarettes smoked daily) and can thus, when confronted with the results of a new experiment, say "that's a surprising finding" or turn the page. Economists, sociogists, and educators all have similar benchmark experiments for their discipline.

What is doubly frustrating about this situation is that information-theoretic indices provide a fundamental measure of the association between random variables. These measures are invariant under any transformation of the underlying variables. To see the significance of this observation, consider the following: if a researcher is unable to transform the variables entering a linear regression model successfully, conventional analysis may conclude that there is little association between the variables, when there is actually a direct functional connection. Barring distortions introduced by discretizing continuous variables, an information-theoretic framework will yield the same measure of association for all (one-to-one) transformations of the original data. We thus close the paper by asking the reader to take its main result with the proverbial grain of salt! Provided a few key information-theoretic elements are present, as described herein, the practitioner should not permit the lack of more formal theoretical justification of information rate to discourage its use.

Appendix 1. Forecasting S&P 500 volatility

Initial processing of S&P data

Our data consisted of a compilation of S&P daily closing averages from September 1986 through August 1991. We began by grouping the data into 92 consecutive 20 business-day blocks. The sample standard deviation of each block was then calculated.

Discretization

Information theory is easiest to apply when both the input and output signals belong to finite alphabets of symbols. Inspection of the sample standard deviations suggested that a 5 letter alphabet captured the essential uncertainty of the time series. To maximize the entropy in the discretized sample, we assigned the letters a,b,c,d, and e to continuous data based on quintile breaks. Thus, a sample standard deviation with magnitude in the bottom 20% of the 92 observations became an "a", etc.

Mutual information

The analysis described below was performed for each of the predictors we examined; it will suffice to consider one of these - the "asymmetric mean reversion" predictor.

A preliminary examination of the graphical representation of the time series of sample standard deviations suggested an ambient noise level disturbed periodically by positive "pulses." This observation in turn suggests that we might have success predicting future volatility by replacing the estimate based on the most recent month by an estimate which reflects this "pulse" phenomenon. Thus a current month observation in category e is taken to indicate that the next month's volatility will fall in category d. A current observation in category d leads to a prediction of category c, while current observations of a, b and c each lead to a prediction of an unchanged volatility category for the upcoming month.

We have 91 observations of the joint distribution of (X.in, X.out), where X.in is the actual volatility observed in month n and X.out is the volatility predicted by the asymmetric mean reversion model. Calculation of mutual information is now straightforward: refer to Exhibit 1d.

Appendix 2. Predicting changes in the S&P 500

The raw data

For each trading day between January 1, 1989 and June 30, 1990, we had the following data.

(i) The expected daily return (predicted by the model) on $1 invested in the S&P 500.

(ii) The number of units and the unit values in a portfolio managed to replicate the S&P 500.

(iii) A record of all trades in and out of the portfolio.

By combining (ii) and (iii) we calculated actual daily returns.

Data reduction and discretization

Our concern is with the ability of the model to forecast changes in the S&P 500 index. Following an approach similar to that used with the volatility analysis, we grouped the day to day changes into categories (here three rather than five) with equal frequency representation. Thus a model output of "a" is a predictor that the S&P index will fall materially, a "b" represents insignificant change, and a "c" represents a material increase.

Mutual information

The calculations are summarized in Exhibit II. Note that symbols d and e are not used (we simply borrowed the spreadsheet created for the volatility analysis). The analysis is given separately for the three six-month time periods beginning January 1, 1989. As noted in the text, we observed a substantial deterioration in model performance after the first six months.

References

1. Kelly, J. L., Jr.	*A New Interpretation of Information Rate*	Bell System Technical Journal 35 (1956): 917-926
2. Khinchin, A.I.	*Mathematical Foundations of Information Theory*	Dover, 1957
3. Pierce, J.R.	*An Introduction to Information Theory* (2nd edition)	Dover, 1980
4. Rozanov, Y.A.	*Probability Theory: A Concise Course*	Dover, 1969
5. Schroeder, M.R.	*Fractals,Chaos, Power Laws: Minutes from an Infinite Paradise*	W.H.Freeman, 1991
6. Shannon, C.E. and W. Weaver	*The Mathematical Theory of Communication*	University of Illinois Press, 1949
7. Sveshnikov, A.A.	*Problems in Probability Theory, Mathematical Statistics, and Theory of Random Functions*	Dover, 1968
8. Thorp, E.O.	*Portfolio Choice and the Kelly Criterion*	Proceedings of the American Statistical Association (1971): 599-619

Exhibit Ia. S&P Volatility -- Simple Extrapolation

Observations

		X.out					
		a	b	c	d	e	total
X.in	a	7	4	4	2	1	18
	b	3	6	5	4	0	18
	c	2	3	6	6	2	19
	d	4	3	2	2	7	18
	e	2	2	2	3	9	18
	total	18	18	19	17	19	91

Joint and marginal probabilities

		X.out					
		a	b	c	d	e	p(x.in)
X.in	a	0.0769	0.0440	0.0440	0.0220	0.0110	0.1978
	b	0.0330	0.0659	0.0549	0.0440	0.0000	0.1978
	c	0.0220	0.0330	0.0659	0.0659	0.0220	0.2088
	d	0.0440	0.0330	0.0220	0.0220	0.0769	0.1978
	e	0.0220	0.0220	0.0220	0.0330	0.0989	0.1978
	p(x.out)	0.1978	0.1978	0.2088	0.1868	0.2088	1.0000

Total Entropy: H(X.in, X.out)

		X.out					
		a	b	c	d	e	
X.in	a	0.2846	0.1981	0.1981	0.1211	0.0715	0.8735
	b	0.1623	0.2586	0.2300	0.1981	0.0000	0.8491
	c	0.1211	0.1623	0.2586	0.2586	0.1211	0.9217
	d	0.1981	0.1623	0.1211	0.1211	0.2846	0.8872
	e	0.1211	0.1211	0.1211	0.1623	0.3301	0.8556
							4.3870

Input Entropy: H(X.in)

x.in	p(x.in)	plnp
a	0.1978	0.4624
b	0.1978	0.4624
c	0.2088	0.4718
d	0.1978	0.4624
e	0.1978	0.4624
		2.3216

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.1978	0.4624
b	0.1978	0.4624
c	0.2088	0.4718
d	0.1868	0.4521
e	0.2088	0.4718
		2.3207

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = *0.2553*

Exhibit Ib. S&P 500 Volatility --One month reporting lag

__Observations__

		X.out					
		a	*b*	*c*	*d*	*e*	total
	a	2	5	4	4	3	18
	b	5	5	5	2	1	18
X.in	*c*	6	1	5	4	3	19
	d	4	1	3	4	6	18
	e	0	6	2	3	6	17
	total	17	18	19	17	19	90

__Joint and marginal probabilities__

		X.out					
		a	*b*	*c*	*d*	*e*	p(x.in)
	a	0.0222	0.0556	0.0444	0.0444	0.0333	0.2000
	b	0.0556	0.0556	0.0556	0.0222	0.0111	0.2000
X.in	*c*	0.0667	0.0111	0.0556	0.0444	0.0333	0.2111
	d	0.0444	0.0111	0.0333	0.0444	0.0667	0.2000
	e	0.0000	0.0667	0.0222	0.0333	0.0667	0.1889
	p(x.out)	0.1889	0.2000	0.2111	0.1889	0.2111	1.0000

__Total Entropy: H(X.in, X.out)__

		X.out					
		a	*b*	*c*	*d*	*e*	
	a	0.1220	0.2317	0.1996	0.1996	0.1636	0.9165
	b	0.2317	0.2317	0.2317	0.1220	0.0721	0.8892
X.in	*c*	0.2605	0.0721	0.2317	0.1996	0.1636	0.9275
	d	0.1996	0.0721	0.1636	0.1996	0.2605	0.8954
	e	0.0000	0.2605	0.1220	0.1636	0.2605	0.8065
							4.4351

__Input Entropy: H(X.in)__

x.in	p(x.in)	plnp
a	0.2000	0.4644
b	0.2000	0.4644
c	0.2111	0.4737
d	0.2000	0.4644
e	0.1889	0.4542
		2.3210

__Output Entropy: H(X.out)__

x.out	p(x.out)	plnp
a	0.1889	0.4542
b	0.2000	0.4644
c	0.2111	0.4737
d	0.1889	0.4542
e	0.2111	0.4737
		2.3201

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = *0.2061*

Exhibit Ic. S&P 500 Volatility --Mean Regression Model

Observations

		X.out					
		a	b	c	d	e	total
X.in	a	0	7	10	1	0	18
	b	0	3	15	0	0	18
	c	0	2	15	2	0	19
	d	0	4	7	7	0	18
	e	0	2	7	9	0	18
	total	0	18	54	19	0	91

Joint and marginal probabilities

		X.out					
		a	b	c	d	e	p(x.in)
X.in	a	0.0000	0.0769	0.1099	0.0110	0.0000	0.1978
	b	0.0000	0.0330	0.1648	0.0000	0.0000	0.1978
	c	0.0000	0.0220	0.1648	0.0220	0.0000	0.2088
	d	0.0000	0.0440	0.0769	0.0769	0.0000	0.1978
	e	0.0000	0.0220	0.0769	0.0989	0.0000	0.1978
	p(x.out)	0.0000	0.1978	0.5934	0.2088	0.0000	1.0000

Total Entropy: H(X.in, X.out)

		X.out					
		a	b	c	d	e	
X.in	a	0.0000	0.2846	0.3501	0.0715	0.0000	0.7063
	b	0.0000	0.1623	0.4287	0.0000	0.0000	0.5910
	c	0.0000	0.1211	0.4287	0.1211	0.0000	0.6708
	d	0.0000	0.1981	0.2846	0.2846	0.0000	0.7674
	e	0.0000	0.1211	0.2846	0.3301	0.0000	0.7358
							3.4714

Input Entropy: H(X.in)

x.in	p(x.in)	plnp
a	0.1978	0.4624
b	0.1978	0.4624
c	0.2088	0.4718
d	0.1978	0.4624
e	0.1978	0.4624
		2.3216

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.0000	0.0000
b	0.1978	0.4624
c	0.5934	0.4468
d	0.2088	0.4718
e	0.0000	0.0000
		1.3811

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = *0.2313*

Exhibit Id. S&P 500 Volatility --Modified Mean Regression Model

Observations

		X.out					
		a	*b*	*c*	*d*	*e*	total
	a	7	4	6	1	0	18
	b	3	6	9	0	0	18
X.in	*c*	2	3	12	2	0	19
	d	4	3	4	7	0	18
	e	2	2	5	9	0	18
	total	18	18	36	19	0	91

Joint and marginal probabilities

		X.out					
		a	*b*	*c*	*d*	*e*	p(x.in)
	a	0.0769	0.0440	0.0659	0.0110	0.0000	0.1978
	b	0.0330	0.0659	0.0989	0.0000	0.0000	0.1978
X.in	*c*	0.0220	0.0330	0.1319	0.0220	0.0000	0.2088
	d	0.0440	0.0330	0.0440	0.0769	0.0000	0.1978
	e	0.0220	0.0220	0.0549	0.0989	0.0000	0.1978
	p(x.out)	0.1978	0.1978	0.3956	0.2088	0.0000	1.0000

Total Entropy: H(X.in, X.out)

		X.out					
		a	*b*	*c*	*d*	*e*	
	a	0.2846	0.1981	0.2586	0.0715	0.0000	0.8130
	b	0.1623	0.2586	0.3301	0.0000	0.0000	0.7511
X.in	*c*	0.1211	0.1623	0.3854	0.1211	0.0000	0.7898
	d	0.1981	0.1623	0.1981	0.2846	0.0000	0.8432
	e	0.1211	0.1211	0.2300	0.3301	0.0000	0.8022
							3.9993

Input Entropy: H(X.in)

x.in	p(x.in)	plnp
a	0.1978	0.4624
b	0.1978	0.4624
c	0.2088	0.4718
d	0.1978	0.4624
e	0.1978	0.4624
		2.3216

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.1978	0.4624
b	0.1978	0.4624
c	0.3956	0.5293
d	0.2088	0.4718
e	0.0000	0.0000
		1.9260

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = *0.2483*

IIe. S&P 500 changes – first six months 1989

Observations

		X.out					
		a	*b*	*c*	*d*	*e*	total
	a	28	6	6	0	0	40
	b	10	26	7	0	0	43
X.in	*c*	2	11	27	0	0	40
	d	0	0	0	0	0	0
	e	0	0	0	0	0	0
	total	40	43	40	0	0	123

Joint and marginal probabilities

		X.out					
		a	*b*	*c*	*d*	*e*	p(x.in)
	a	0.2276	0.0488	0.0488	0.0000	0.0000	0.3252
	b	0.0813	0.2114	0.0569	0.0000	0.0000	0.3496
X.in	*c*	0.0163	0.0894	0.2195	0.0000	0.0000	0.3252
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p(x.out)	0.3252	0.3496	0.3252	0.0000	0.0000	1.0000

Total Entropy: H(X.in, X.out)

		X.out					
		a	*b*	*c*	*d*	*e*	
	a	0.4861	0.2126	0.2126	0.0000	0.0000	0.9112
	b	0.2944	0.4739	0.2353	0.0000	0.0000	1.0036
X.in	*c*	0.0966	0.3115	0.4802	0.0000	0.0000	0.8883
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
							2.8031

Input Entropy: H(X.in)

x.in	p(x.in)	plnp
a	0.3252	0.5270
b	0.3496	0.5301
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5841

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.3252	0.5270
b	0.3496	0.5301
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5841

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = 0.3651

IIb. S&P Changes -- second six months 1989

Observations

		X.out					
		a	*b*	*c*	*d*	*e*	total
	a	23	14	3	0	0	40
	b	13	19	11	0	0	43
X.in	*c*	4	10	26	0	0	40
	d	0	0	0	0	0	0
	e	0	0	0	0	0	0
	total	40	43	40	0	0	123

Joint and marginal probabilities

		X.out					
		a	*b*	*c*	*d*	*e*	p(x.in)
	a	0.1870	0.1138	0.0244	0.0000	0.0000	0.3252
	b	0.1057	0.1545	0.0894	0.0000	0.0000	0.3496
X.in	*c*	0.0325	0.0813	0.2114	0.0000	0.0000	0.3252
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p(x.out)	0.3252	0.3496	0.3252	0.0000	0.0000	1.0000

Total Entropy: H(X.in, X.out)

		X.out					
		a	*b*	*c*	*d*	*e*	
	a	0.4523	0.3568	0.1307	0.0000	0.0000	0.9398
	b	0.3427	0.4162	0.3115	0.0000	0.0000	1.0704
X.in	*c*	0.1607	0.2944	0.4739	0.0000	0.0000	0.9290
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
							2.9393

Input Entropy: H(X.in)

x.in	p(x.in)	plnp
a	0.3252	0.5270
b	0.3496	0.5301
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5841

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.3252	0.5270
b	0.3496	0.5301
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5841

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = *0.229*

IIc. S&P 500 Changes --first six months 1990

Observations

		X.out					
		a	*b*	*c*	*d*	*e*	total
	a	21	15	4	0	0	40
	b	16	17	10	0	0	43
X.in	*c*	4	10	26	0	0	40
	d	0	0	0	0	0	0
	e	0	0	0	0	0	0
	total	41	42	40	0	0	123

Joint and marginal probabilities

		X.out					
		a	*b*	*c*	*d*	*e*	p(x.in)
	a	0.1707	0.1220	0.0325	0.0000	0.0000	0.3252
	b	0.1301	0.1382	0.0813	0.0000	0.0000	0.3496
X.in	*c*	0.0325	0.0813	0.2114	0.0000	0.0000	0.3252
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p(x.out)	0.3333	0.3415	0.3252	0.0000	0.0000	1.0000

Total Entropy: H(X.In, X.out)

		X.out					
		a	*b*	*c*	*d*	*e*	
	a	0.4354	0.3702	0.1607	0.0000	0.0000	0.9663
	b	0.3828	0.3946	0.2944	0.0000	0.0000	1.0717
X.In	*c*	0.1607	0.2944	0.4739	0.0000	0.0000	0.9290
	d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	e	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
							2.9671

Input Entropy: H(X.In)

x.In	p(x.In)	plnp
a	0.3252	0.5270
b	0.3496	0.5301
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5841

Output Entropy: H(X.out)

x.out	p(x.out)	plnp
a	0.3333	0.5283
b	0.3415	0.5293
c	0.3252	0.5270
d	0.0000	0.0000
e	0.0000	0.0000
		1.5847

Mutual information I(X.in;X.out) = H(X.in) + H(X.out) - H(X.in,X.out) = 0.2017