Assessing Predictive Performance of an Investment Portfolio

Malek Lashgari
University of Hartford

Performance of an investment portfolio is often viewed in comparison to a benchmark return, within a similar risk class. When a risk adjusted measure is desired, given a normally distributed return, the usual notion of risk is the standard deviation, or beta as the degree of association of the portfolio with respect to the market. Investment plans, however, are goal oriented, depend on the prevailing circumstances and are designed based on a set of expectations. Thus, performance should be assessed accordingly. Within this framework an average information index value, denoting the degree of divergence between investment objectives and the later observed outcomes, appears to be a good criterion for appraising performance of an investment portfolio. The resulting information index value is in line with non-replicable, path dependent nature of investment portfolio designs.

INTRODUCTION

Scholars in finance have searched for a risk adjusted minimum required return for an investment portfolio. This desired rate of return plays an important role in the determination of capital asset prices as well as appraisal of an investment outcome. The literature in finance pays much attention to the standard deviation of returns as a measure of total risk as it is easily known from historical observations. A further refinement is made by considering the risk due to the reaction in price of a financial asset with respect to changes in the market. This is known as beta. The standard deviation measures variations around the average returns from the past observations and is most useful when the distribution is normal. Beta measures the co-movements or the degree of association of one security with the market. There are at times problems associated with these notions of risk as the distribution of the outcomes may not be normal and standard deviation may not exist. This is, for example, the case with Pareto distribution. Numerical values of beta of an asset may change over time and it further depends on the choice of the time interval used in a regression analysis.

An important notion of risk in investment is the risk of shortfall from the desired goal. Furthermore, in cases where investment plans are based on a set of expectations regarding the course of the market, deviations away from expectations may show in poor performance. The notion of risk and adequate compensation for it, as used in this paper in the context of information theory, arises from the comparison of the predicted return with the outcome and is referred to as the average information. The predictive performance power is then shown by time series analysis using autoregressive moving average and statistical filtering designs. If the expectation regarding the likely course of the market constitutes the prior state of knowledge and the result of an action is the posterior one, then the difference between the two is known as the information embedded in the outcome. This amount would be zero for a case in which the outcome is the same as what was expected.

Economic Views of Information

Financial Economists view uncertainty as the lack of information regarding factors of production such as capital, labor and technology. For example, the lack of information regarding changes in prices appears to play a role in developing a futures market. With imperfect, noisy information market equilibrium may be characterized by a price distribution rather than by a single price. Hirshleifer (1973) discusses dissemination of information in a market context where uncertainty is summarized by the dispersion of individuals' subjective probability distributions. Information is used as a negative measure of uncertainty which consists of events tending to change these probability distributions; a dispersed probability distribution is called less "informative" than a concentrated one. Stigler (1961) views price dispersion as a measure of ignorance in the market.

Shackle (1969) analyses decisions which are unique and are unlikely to be replicated. He defines surprise as what we feel when the outcome of an action is not what was expected. The greatest surprise is caused by the occurrence of the impossible. Zero degree of surprise corresponds to perfect possibility. Shackle hypothesizes that the decision maker first focuses on the maximum gain or the maximum loss. In between the two outcomes there is an outcome which would be considered neither good nor bad, but neutral causing no surprise. However, to either side of which there are desirable and undesirable surprising outcomes if they occur (Watkins, 1957).

Statistical Views of Information

The error rate in a message transmission, in the context of communications theory, constitutes the degree of inaccuracy or risk associated with the system. Shannon (1948) provides an easily measureable and operational definition of information as the logarithm of the reciprocal of the probability; H(p) = log $(1/p_i)$ = - log p_i. Where H(p) is the amount of information transmitted of the probability distribution, p_i. Average information is further measured as the expected value of the outcome as shown in equation (1).

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \quad or \quad \mathrm{H}(p) = -K\sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

K refers to the constant amount for the choice of the measurement unit; when logarithm on base two is used (binary) the resulting measure is in "bits," and when the base of e, the "natural number," is used the resulting measure is in "nits." One bit equals .693 nits, one nit equals 1.443 bits.

Average information as measured in equation (1), is informative and describes the degree of uncertainty associated with an outcome. Consider, for example, the following three probability distributions.

I E = (A_1, A_2) p = (1/256, 255/256)
II E = (B_1, B_2) p = (1/2, 1/2)
III E = (C_1, C_2) p = (7/16, 9/16)

Their average information can be shown as follows.

$H_1 = -(1/256 \log 1/256 + 255/256 \log 255/256) = .0369$ bits
$H_2 = -(1/2 \log 1/2 + 1/2 \log 1/2) = 1$ bit
$H_3 = -(7/16 \log 7/16 + 9/16 \log 9/16) = .989$ bits.

All computations are binary or with logarithm in base two. In event I, it is expected that A_2 will occur with high confidence. In event III this guess is much harder for C_1 and C_2 , and in event II it is most difficult to predict the occurrence of either B_1 or B_2. Thereby event II has a larger average information (uncertainty) than event III (Reza 1961, p. 76-80). This can also be regarded as our degree of ignorance in terms of the outcome. Thereby, the maximum ignorance or uncertainty relates to the case of equally likely events.

Selected Applications in Financial Economics

Garner and McGill (1956) show the equivalency of the information theoretic design and the analysis of variance. Philippatos and Gressis (1975) further show that when information theoretic measure is used, the results are comparable to those of mean-variance, and the stochastic dominance criteria for portfolios distributed according to the normal and uniform distributions. Philippatos and Wilson (1972) find that the expected information can be used in investment portfolio building models as a measure of uncertainty since it is not dependent upon any particular distribution. For normal distribution, however, it is measured as $H(x) = \text{Ln}\ (\sqrt{2\pi\sigma}) + \frac{1}{2}$, which is a linear function of standard deviation σ. For uniform distribution in the form $f(x) = 1/(b-a)$, $a \leq x \leq b$, variance is computed as $\sigma^2 = (b-a)^2 / 12$, and information is measured as $H(x) = \text{Ln}\ (b-a)$, which shows their interrelationship.

Nawroki (1984) reviews applications of information theory to the random walk processes which assume an infinite speed of information dissemination in the market. Thereby, the return generating process will be independent over time. A random walk process is described as $I(t) = \bar{I} + V(t)$, where $\bar{I}$ is the mean of a stationary information process, V(t) is the error process distributed according to a stationary normal distribution, independent over time. This is consistent with the concept that information arrival is independent over time and normally distributed. Black (1976) however argues that the assumption that information arrives continuously in small random doses is not realistic. He states that information arrives periodically and in large doses. Therefore, the information system is non-continuous and follows a sporadic jump process.

Dinkel and Kochenberger (1979), and Nicholis and Prigogine (1977) argue that the speed of the price adjustment process is constrained to some finite speed by taxes, transaction costs, information costs and noisy information channels. The greater the amount of un-disseminated information, the greater the divergence from equilibrium. When a disequilibrium in the market becomes too severe, the market can restructure itself to improve the flow of information (e.g. temporary suspension of trading). This concept of restructuring to operate at a more acceptable disequilibrium level is known as bifurcation. Thereby, bifurcation theory describes a non-stationary disequilibrium process in the market system. Grossman and Stiglitz (1976) state that prices never fully adjust because of noisy information system, the cost of acquiring and evaluating information, and the continuing need to adjust to new information shocks to the economy.

MEASURING PERFORMANCE IN THE CONTEXT OD INFORMATION THEORY

The information content of the probability distribution of returns as shown in equation (1) can be compared to the distribution of the later observed returns, as shown in equation (2). In this manner, the original probability is denoted as p_i and its revision as q_i. The amount of change in information can then be obtained by taking the difference between the average information provided by the prior and posterior probability distributions, respectively.

$$H(q:p) = H(p_i) - H(q_i) = -p_i \ \log p_i - (-q_i \ \log q_i) \tag{2}$$

where q_i is the posterior probability and p_i is the prior probability. Using the most recent set of observations, q_i, for averaging, the expected information can be shown as:

$$H:(q:p) = \sum_{i=1}^{n} q_i \ \log(q_i / p_i) \tag{3}$$

where:
H(q:p) denotes average information,
$p_{i,}$ denotes the prior probability or the desired objective,
$q_{i,}$ denotes the posterior probability or the resulting outcome.
H (q:p) is always positive when the prior and posterior probabilities are not pair wise equal and that it vanishes if, and only if, all prior and posterior probabilities are pair wise equal ($p_i = q_i$ for each i). Hence, no information is to be expected from the message when it does not change any of the prior probabilities (Shannon 1948, Theil 1967).

For good performance in a goal oriented investment, the minimum degree of the average information, shown in equation (3), is desired. Advantages of an information theoretic design for measuring performance are as follows.

First, the p_i can be associated with any probability distribution. Multivariate Normal or Gaussian distributions required by mean-variance portfolio theory becomes simply a case of a more general framework.

Second, the probability distribution can cover any span, short (e.g., daily or weekly where much randomness typically occurs), long (e.g., five or ten year trends where economic growth contributes significantly), or sampled (e.g., turning points of the economy, bull, or bear markets).

Third a portfolio can be assessed relative to its own expectations.

Consistency of the Average Information Measure with Stochastic Dominance

Under stochastic dominance for x to be preferred to y, we should have $F(x) - G(y) \leq 0$, for non-intersecting cumulative distributions. F and G stand for cumulative distribution functions for random variables x, and y, respectively. That is, in order for x to be preferred to y, it should have a lower cumulative distribution function. Values of F(x) denote the probability of obtaining x percent or less return.

For the proof of the consistency of the information index value, as shown in equation (3) with the stochastic dominance, note that if F(x) and G(y) denote cumulative distributions and f(x) and g(y) the probability density functions of x and y, respectively, then for x to be preferred to y under stochastic dominance the following relations must hold.

	$F(x) - G(y) < 0$
or	$F(x) < G(y)$
Multiply by (-1)	$-F(x) > -G(y)$
adding a unity	$1 - F(x) > 1 - G(y)$
we have	$f(x) > g(y)$
taking logarithm	$\log f(x) > \log g(y)$
	$f(x) \log f(x) > g(y) \log g(y)$
	$-f(x) \log f(x) < -g(y) \log g(y)$

taking integrals of both sides will result in equation (4) as follows.

$$\int_{-\infty}^{+\infty} -f(x)\log f(x)\,dx < \int_{-\infty}^{+\infty} -g(y)\log g(y)\,dy$$

$$-\int_{-\infty}^{+\infty} f(x)\log f(x)\,dx < -\int_{-\infty}^{+\infty} g(y)\log g(y)\,dy \qquad (4)$$

That is: $H(x) < H(y)$

Note that equation (4), which depicts average information index values for x and y, is in line with equation (3) showing that for random variable x to dominate random variable y, its average information index value should be lower. Thus, for good performance, a minimal value of the average information index is desired.

Data Analysis

An example of portfolio performance using time series analysis for generating expectations is shown in TABLE 1 for the values of the average information resulting from the auto-regressive moving average return generating function (ARIMA), as well as adaptive filtering and Kalman filtering against a random-walk presumption of equal probability of "up," "down," or "no change." The lower numerical values for performance predicted by ARIMA, adaptive filtering and Kalman filtering return generating models, as compared with a case of random character, shows their superiority. That is, later outcomes or returns are dependent and can be predicted using historical information.

TABLE 1
PREDICTIVE PERFORMANCE OF AN INVESTMENT PORTFOLIO

	ARIMA	Adaptive Filtering	Kalman Filtering	Random Character
Information Index Value	0.08438	0.12504	0 .06622	0.20569

ARIMA model is ARIMA (1,1,1). This is based on an autoregressive moving average smoothing design. Adaptive filtering is based on AR (2). In adaptive filtering, the true pattern is distinguished and separated from the noise or random error in an auto-regressive design. Kalman filtering is a method of separating the noise or random error from the true pattern or persistent component.

Total, Joint, and Conditional Information

Correlation analysis as well as conditional probabilities can be performed in the context of information theory. These are pertinent factors playing a role in investment portfolio management.

Considering two random variables x and y, we have:

a. Total information H(x,y) as the total uncertainty about the system.
b. Joint or mutual information I(x;y) as the information of the mutual interactions of x and y.
c. Conditional information $H(x|y)$ which is the relevant measure of uncertainty in the Bayesian analysis.

$$H(x,y) = H(x) + H(y) - I(x;y) \tag{5}$$

For independent events, $H(x,y) = H(x) + H(y)$.

Perez and Tondel apply mutual information as a measure of the degree of correlation between two random variables and indicate that the correlation coefficient may be zero even in cases characterized by a strong stochastic dependence. Thus, as an alternative approach, the mutual information can be used as a measure of the stochastic dependence. The notion of dependence (in this manner) is of an asymmetric character; the dependence of x on y is not necessarily the same as the dependence of y on x.

The measure of information dependence of x on y denoted as z(x,y) is shown as:

$$z(x,y) = \frac{I(x;y)}{H(x)},\ 0 \le (z(x,y)) \le 1 \tag{6}$$

And the measure of information dependence of y on x denoted as z(y,x) is shown as:

$$z(y,x) = \frac{I(x;y)}{H(y)},\ 0 \le (z(y,x)) \le 1 \tag{7}$$

where I(x;y) is the mutual information, H(x) and H(y) are the average information values for x and y, respectively. A value of zero for z(x,y) is indicative of independence of x and y. If the information necessary to determine y proves to be greater than that required to determine x, then x will depend on y to a greater degree than y will depend on x. It is to be noted that z(x,y) plays the role of correlation between x and y, whereas z(y,x) is the correlation between y and x but relations (6) and (7), show that the direction of the movement plays an important role in the determination of the correlation coefficient, implying its path dependent nature. (Perez and Tondel in *Information and Prediction in Science* edited by Dockx et al. 1965, pp. 15-37.)

MANAGERIAL IMPLICATIONS

The investment portfolio performance process needs to be viewed in regards to its stated objectives. This is because deviations from desired goals are a sign of shortfall and poor performance. In contrast, portfolio performance criteria in the context of modern portfolio theory pay no attention to reaching goals and objectives. Instead, results are compared with a benchmark or the overall market. The concept of average information as shown here is applicable to measuring performance of a goal oriented investment portfolio. Conceptually, for good performance, the average information resulting from the prior and posterior probability distributions of returns must be minimal. In this manner, the ex-post results would be in line with the ex-ante expectations. Consequently, an investment portfolio performance can be assessed relative to its own expectations. This approach is of particular importance when investment results are not replicable, depend on a particular market condition and are path dependent.

REFERENCES

Black, S. W. (1976). Rational Response to Shocks in a Dynamic Model of Capital Assets Prices. *American Economic Review*. December, 767-779.

Dinkel, J. and G. Kohenberger (1979). Constrained Entropy Models: Solvability and Sensitivity. *Management Science*. June, 555-64.

Dockx, S. and Bernays (1965). *Information and Prediction in Science.* Academic Press.

Garner, W. R. and W. J. McGill (1956). The Relations Between Information and Variance Analysis. *Psychometrica.* No. 21, September, 219-228.

Grossman, S. and J. Stiglitz. (1976). Information and Competitive Price Systems. *American Economic Review*. 246-53.

Hirshleifer (1973). Where are We in the Theory of Information? *American Economic Review, Papers and Proceedings. May, 31-39.*

Nawrocki, D. (1984). Entropy, Bifurcation and Dynamic Market Disequilibrium. *The Financial Review*. Vol. 19, No. 2. May, 266-284.

Nicolis, G. and I. Prigogine. (1977). *Self Organization in Non-Equilibrium Systems.* New York: John Wiley.

Philippatos, G. and N. Gressis (1975). Conditions of Equivalence Among E-V, SSD, and E-H Portfolio Selection Criteria: The Case for Uniform, Normal, and Lognormal Distribution. *Management Science.* February. Vol. 21, No. 6.

Philippatos, G. and G. Wilson (1972). Entropy as a Measure of Dispersion in the Portfolio Selection Problem. *Decision Science.* 142-152.

Reza, Fazlollah (1961). *An Introduction to Information Theory*. New York: McGraw-Hill.

Shackle, G. (1969). *Decisions, Order and Time in Human Affairs*. Cambridge: The University Press.

Shannon, C. E. and W. Weaver (1948). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Stigler, George J. (1961). The Economics of Information. *The Journal of Political Economy*. Vol. 69, No. 3. June, 213-225.

Theil, Henri (1967). *Economics and Information Theory.* Amsterdam: North-Holland Publishing Co.

Watkins, J. W. N. (1957). Decisions Under Uncertainty. *Uncertainty and Business Decisions.* Edited by Carter, Meredith and Shackle. Liverpool University Press, 107-121.