Information Analysis

A two-step approach to information ratios, information coefficients, and the value of investment information.

Richard C. Grinold and Ronald N. Kahn

(*Reprinted with permission from The Journal of Portfolio Management*)

RICHARD C. GRINOLD is Executive Vice President, and **RONALD N. KAHN** is Director of Research, at BARRA in Berkeley (CA 94704).

The authors wish to thank Keith Ambachtsheer, Stewart Hodges, and Bob Jones for valuable suggestions.

Information is the vital input into any active management strategy. Information separates active management from passive management. Information, properly applied, allows active managers to outperform their informationless bench marks.

Information analysis is the science of evaluating information content, and refining information to build portfolios. Information analysis works both for managers who use a non-quantitative process and for those who use a quantitative investment process. The only requirement is that there is a process.

Information is a fuzzy concept. Information analysis begins by transforming information into something concrete: investment portfolios. Then it analyzes the performance of those portfolios to determine the value of the information.

Information analysis can work with something as simple as an analyst's buy and sell recommendations. Or it can work with alpha forecasts for a broad universe of stocks. Information analysis is not concerned with the intuition or process used to generate stock recommendations, only with the recommendations themselves.

Information analysis can be precise. It can determine whether information is valuable on the upside, the downside, or both. It can determine whether information is valuable over short horizons or long horizons. It can determine whether information is adding value to your investment process.

The science of information analysis began in the 1970s with work by Treynor and Black [1973], Brealey and Hodges [1973], Ambachtsheer [1974], Rosenberg [1976], and Ambachtsheer and Farrell [1979]. These authors all investigate the role of active management in investing: its ability to add value and measures for determining this.

Treynor and Black, and Hodges and Brealey, were the first to examine the role of security analysis and active management within the context of the capital asset pricing model. They investigate the requirements for active management to outperform the market, and identify the importance of correlations between return forecasts and outcomes among these requirements.

Ambachtsheer, alone and with Farrell, provides further insights into the active management process, specifically, turning information into investments. He coined the term "information coefficient," or IC, to describe this correlation between forecasts of residual returns (alphas) and subsequent realizations. Rosenberg investigates the active management process and measures of its performance, as part of his analysis of the optimal amount of active management for institutional investors.

This article relies on the pioneering work of these authors, but focuses explicitly on the task of information analysis itself. It presents a unified treatment of information analysis, with both theoretical discussions and concrete examples. Given that information analysis is a comprehensive subject, we cover the general approach, recommending as well a specific approach to best analyze investment information.

First we describe how and where information appears in the active management process, then describe the two-step process of information analysis. Next we discuss the first step: turning information into portfolios. The second step is analyzing the performance of those portfolios. Here we focus on one particularly convenient statistic — the information ratio — which can summarize how the information can generate investment value-added. We also explain the information coefficient, a close relative of the information ratio. Finally, we step back from the details of information analysis to discuss some precautions. Information analysis is a tool, and, as with a hammer, one must distinguish between thumb and nail.

INFORMATION AND ACTIVE MANAGEMENT

Where and how does information arise in active management? Active managers, as opposed to passive managers, apply information to achieve superior returns relative to a benchmark. Passive managers simply try to replicate the performance of the benchmark. They have no information.

Active managers use information to predict the future exceptional return on a group of stocks. The emphasis is on predicting alpha, or residual return: beta-adjusted return relative to a benchmark.[1] We want to know what stocks will do better than average, and what stocks will do worse, on a risk-adjusted basis.

So, when we talk about information in the context of active management, we are really talking about alpha predictors. For any set of data pertaining to stocks, we can ask: Do these data help predict alphas? We will even call this data a *predictor*.

In general, any predictor is made up of signal plus noise. The signal is linked with future stock returns. The noise masks the signal and makes the task of information analysis both difficult and exciting. Random numbers contain no signal, only noise. Information analysis is an effort to find the signal-to-noise ratio.

A predictor will cover a number of time periods and a number of stocks in each time period. The information at the beginning of period t is a data item for each stock. The data item can be as simple as +1 for all stocks on a recommended buy list and -1 for all stocks on a sell list. On the other hand, the data can be a precise alpha, 2.15% for one stock, -3.72% for another, and so on.

Other predictors might be scores. Crude scores can be a grouping of stocks into categories, a more refined version of the buy and sell idea. Other scores might be a ranking of the stocks along some dimension. Notice it is possible to start with alphas and produce a ranking. It is possible to start with a ranking and produce other scores such as four for the stocks in the highest quartile, down to one for the stocks in the lowest quartile.

The predictors can be publicly available information such as consensus earnings forecasts, or they can be derived data, such as a change in consensus earnings forecasts. Predictors are limited only by availability and imagination.

In examples that we follow throughout the article, we use book-to-price data in the United Kingdom to generate return predictors according to various standard schemes. For instance, we can generate a buy list and a sell list by ranking all U.K. stocks according to book-to-price ratio, and placing the top half on the buy list and the bottom half on the sell list. The intent of this and other examples is not to suggest novel new strategies, but simply to illustrate information analysis techniques.

Underlying the book-to-price examples is the hypothesis that book-to-price ratios contain information concerning future stock returns, and, in particular, that high book-to-price stocks will outperform low book-to-price stocks. Is this hypothesis true? How much information is contained in book-to-price ratios? We will apply information analysis and find out.

INFORMATION ANALYSIS

Information analysis is a two-step process:

Step 1: Turn predictions into portfolios, and
Step 2: Evaluate the performance of those portfolios.

Step 1 transforms the information into a concrete object: a portfolio. Step 2 then analyzes the performance of the portfolio.

Information analysis is flexible. There are a great many ways to turn predictions into portfolios and a great many ways to evaluate performance. We will explore many of these alternatives below.

Step 1: Information into Portfolios

Let's start with Step 1: turning predictions into portfolios. As we have predictions for each time period, we generate portfolios for each time period.[2] Now there are a great many ways to generate portfolios from predictions, and the procedure selected can depend on the type of prediction. Here are six possibilities. For each case, we provide the general idea, and then discuss how to apply this to data concerning book-to price ratios in the U.K. Later we analyze the performance of these portfolios.

PROCEDURE 1

With buy and sell recommendations we can equal- (or value-) weight the buy group and the sell group.

Using book-to-price ratios, we can generate the buy and sell lists, as described above, by first ranking stocks by book-to-price, and putting the top half on the buy list and the bottom half on the sell list.

PROCEDURE 2

With scores we can build a portfolio for each score by equal- (or value-) weighting within each score category.

We can generate scores from book-to-price ratios by ranking stocks by book-to-price ratio, as before, and then, for example, give the top fifth of the list (by number or capitalization) a score of five, the next fifth a score of four, ..., down to the bottom fifth with a score of one. This is simply dividing stocks into quintiles, by book-to-price.

PROCEDURE 3

With straight alphas we can split the stocks into two groups, one group with higher than average alphas, and one group with lower than average alphas. Then we can weight the stocks in each group by how far their alpha exceeds (or lies below) the average. This is an elaboration of Procedure 1.

One way to generate alphas from book-to-price ratios is to assume that they are linearly related to the book-to-price ratios. So we can weight each asset in our buy and sell list by how far its book-to-price ratio lies above or below the average.

PROCEDURE 4

With straight alphas we can rank the assets according to alpha, and then group the assets into quintiles (or deciles or quartiles or halves) and then equal- (or value-) weight within each group. This is an elaboration of Procedure 2.

For alphas linearly related to book-to-price ratios, this is a straightforward extension of Procedure 3.

PROCEDURE 5

With any numerical score we can build a factor portfolio that bets on the prediction and does not make a market bet. The factor portfolio consists of a long portfolio and a short portfolio. The long and short portfolios have equal value and equal beta, but the long portfolio will have a unit bet on the prediction, relative to the short portfolio. Given these constraints, the long portfolio will track the short portfolio as closely as possible.

For book-to-price data, we can build long and short portfolios with equal value and beta, with the long portfolio exhibiting a book-to-price ratio one standard deviation above that of the short portfolio, and designed so that the long portfolio will track the short portfolio as closely as possible.

PROCEDURE 6

With any numerical score we can build a factor portfolio, consisting of a long and a short portfolio, designed so that the long and short portfolios are matched on a set of prespecified control variables. For example, we could make sure the long and short portfolios match on industry, sector, or small-capitalization stock exposures. This is a more elaborate form of Procedure 5 where we controlled only for beta (as a measure of exposure to market risk).

Using the book-to-price data, this is an extension of Procedure 5.

Examples

The general idea should be clear. We are trying to establish some sort of relative performance. In each case we will produce two or more portfolios. In the first, third, fifth, and sixth procedures we will have a long and a short portfolio. The long bets on the information, the short bets against it. In Procedure 2, we have a portfolio for each score, in Procedure 4, a portfolio for each quintile.

Procedures 5 and 6 are more elaborate and “quantitative” than the first four procedures. They do require more sophisticated technology, but the basic inputs — the information being analyzed — needn't be based on a quantitative strategy. Numerical scores derived through any method will work.

While Procedures 5 and 6 are more elaborate, they are also more precise in isolating the information contained in the data. These procedures build portfolios based solely on new information in the data, controlling for other important factors in the market. Because they set up a controlled experiment, we recommend Procedure 5 or Procedure 6 as the best approach for analyzing the information contained in any numerical scores.

To be explicit about Step 1, let's apply some of these procedures in two separate examples based on book-to-price ratios in the U.K. from March 1983 through December 1990.

EXAMPLE A

For Example A, we build portfolios according to Procedure 2. Every month we rank assets in the *Financial Times* All Shares Index by book-to-price ratio, and then divide them into quintiles defined so that each quintile has equal capitalization. We turn these quintiles into portfolios by capitalization-weighting the assets in each quintile.

EXAMPLE B

For Example B, we build portfolios according to Procedure 5. Every month we build two portfolios, a long portfolio and a short portfolio. The two portfolios will have equal value and beta. The long portfolio will exhibit a book-to-price ratio one standard deviation above that of the short portfolio. And given these constraints, the long portfolio will track the short portfolio as closely as possible. What can these examples tell us about the investment information contained in book-to-price ratios?

Step 2: Performance Evaluation

We have turned the data into two or more portfolios. Now we must evaluate the performance of those portfolios. The standard issues of performance analysis come into play here. What is the appropriate benchmark against which to judge performance? How do we analyze performance, and separate skill and value-added from luck?

The comparison benchmark is usually a marketwide index or prespecified normal. In the United States, the S&P 500 is the generic benchmark. For our example strategy in the United Kingdom, we chose the *Financial Times* All Shares Index as a benchmark.

Once we decide on a benchmark, we can calculate the returns to the portfolios and benchmark. The simplest form of performance analysis is just to calculate the cumulative returns to the portfolios and the benchmark, and plot them. Some summary statistics, like the means and standard deviations of the returns, can augment this analysis.

Figure 1 illustrates this basic analysis for Example A. It shows the cumulative active return to each of the five quintile portfolios.[3] While the two highest quintiles show the highest returns over this period, the middle quintile exhibits the lowest.

Figure 1 suggests that book-to-price ratios provide investment information, but the results are not entirely clear. But remember that we constructed the quintile portfolios based only on book-to-price ratios, without controlling for other factors. The quintiles may include incidental bets that muddy the analysis.

FIGURE 1

CUMULATIVE ACTIVE RETURNS BY BOOK-TO-PRICE QUINTILES CAPITALIZATION-WEIGHTED PORTFOLIOS COMPARED TO *FINANCIAL TIMES* ALL SHARES INDEX (FTA)

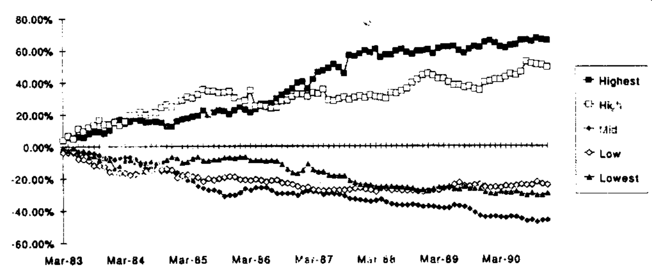

FIGURE 2

CUMULATIVE RETURNS LONG, SHORT, AND NET BOOK-TO-PRICE FACTOR PORTFOLIOS

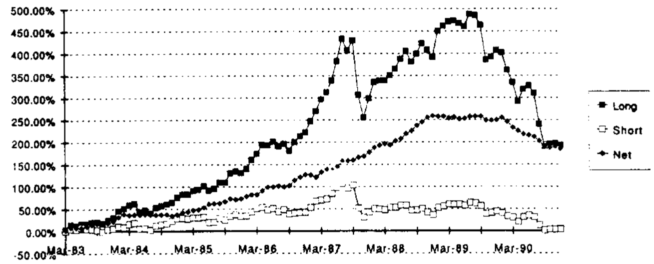

Figure 2 shows the cumulative returns for the long, short, and net (long minus short) portfolios for Example B. The cumulative net return is higher than any of the quintile returns from Example A over this test period. This cumulative net return is the cumulative long return minus the cumulative short return. In this example, the cumulative short return over the entire period lies close to zero, so the cumulative net return is very close to the cumulative long return.

Figure 2 also shows that the information embodied in book-to-price ratios, upside information in particular, disappears starting in 1989. Comparing Figures 1 and 2, it is clear that different approaches for constructing portfolios from the same basic information lead to different observed performance and different estimates of information content.

T-STATISTICS, INFORMATION RATIOS, AND INFORMATION COEFFICIENTS

So far, we have discussed only the simplest form of performance analysis: looking at the returns. More sophisticated analyses go beyond this to investigate statistical significance, value-added, and skill. These more sophisticated analyses rely on three important statistics describing investment performance: t-statistics, information ratios, and information coefficients. All are related, as you will see. The information ratio in particular, however, most directly captures the investment value added offered by the information, and so is the most important statistic for investment information analysis.

More sophisticated performance analysis begins with regression analysis on the portfolio returns. The idea is to regress the excess portfolio returns against the excess benchmark returns, to separate the portfolio return into a benchmark-related and non-benchmark related component:

$$(r_t - rf_t) = \alpha + \beta \cdot (r_{benchmark,t} - rf_t) + \varepsilon_t \qquad (1)$$

This regression will estimate the portfolio's alpha and beta. It will evaluate, via the t-statistic, whether the alpha is significantly different from zero.

The t-statistic for the portfolio's alpha is:

$$t - \text{stat} = \frac{\alpha}{SE(\alpha)}, \qquad (2)$$

simply the ratio of the estimated alpha to the standard error of the estimate. This statistic measures whether the alpha differs significantly from zero. If the t-statistic exceeds 2, and alphas are normally distributed, then the probability that simple luck generated these returns is less than 5%.

Applying regression analysis to Example A, we find:

Quintile	Alpha	t-Alpha	Beta	t-Beta
Highest	0.68%	3.43	0.93	26.02
High	0.47%	2.57	1.00	29.91
Mid	-0.48%	-2.99	1.05	36.02
Low	-0.27%	-1.70	0.98	34.12
Lowest	-0.39%	-2.20	1.06	32.96

This analysis corroborates the visual results in Figure 1. Only the two highest quintiles exhibit positive alphas, and the middle quintile underperforms the fourth quintile. As for statistical significance, using the 95% confidence level as a cutoff, all but one of the quintile alphas are significant. The analysis of beta shows that the quintiles differ significantly in their exposure to the market.

Applying regression analysis to Example B, we find:

Portfolio	Alpha	t-Alpha	Beta	t-Beta
Long	0.80%	2.75	1.03	19.50
Short	-0.41%	-2.64	1.00	35.24
Net	1.21%	4.90	0.03	0.82

This analysis is consistent with Figure 2. All the alphas are significant, with the alpha for the net portfolio exhibiting a t-statistic far above 2. Note that, even though the long and short portfolios are designed to have equal betas, the observed betas from the time series analysis do not exactly match. The long and short portfolios were constructed monthly to match exactly on forecast betas. Still, the net portfolio exhibits a beta of only 0.03, which, according to the t-statistic, is not significantly different from zero.

So far, this analysis has focused only on t-statistics. What about information ratios? Let's assume that the active manager attempts to maximize risk-adjusted annual alpha:

$$VA = \alpha_{annual} - \lambda \cdot [\omega_{annual}]^2. \quad (3)$$

This value-added is the alpha minus a risk tolerance (lambda) times the annualized residual risk. Now, the manager is able to achieve a certain amount of alpha for any given level of risk. Depending on the investor's risk aversion, the manager will choose an appropriate level of aggressiveness to maximize this value-added.

Detailed analysis shows that the maximum value-added rises in proportion to the square of the manager's information ratio, IR, the ratio of annual alpha to annual residual risk:

$$VA_{max} = \frac{1}{4\lambda} \cdot \left(\frac{\alpha_{annual}}{\omega_{annual}} \right)^2, \quad (4)$$

with $IR = \frac{\alpha_{annual}}{\omega_{annual}}$. (5)

The information ratio is essentially an investment signal-to-noise ratio. An information ratio of 0.75 means that we can expect 3% alpha per year if we take a 4% per year residual risk. As the analysis demonstrates, value-added rises with the manager's information ratio, regardless of the level of risk aversion. Because of this direct connection to investment value added, the information ratio is the best overall statistic to use for information analysis.

We can calculate the information ratios arising in examples A and B.

Portfolio		Information Ratio
Example A:	Highest Quintile	1.21
	High Quintile	0.91
	Mid Quintile	-1.06
	Low Quintile	-0.60
	Lowest Quintile	-0.78
Example B:		
	Long Portfolio	0.97
	Short Portfolio	-0.93
	Net Portfolio	1.73

The net portfolio implemented in Example B can achieve the highest value-added from the book-to price information.

By their definitions the t-statistic and the information ratio certainly look closely related. They are. If we observe returns over a period of T years, the information ratio is approximately the t-statistic divided by the square root of the number of years of observation:

$$IR \approx \frac{t - stat}{\sqrt{T}} \tag{6}$$

This relationship becomes more exact as the number of observations increases.

Overall, the t-statistic measures the statistical significance of the return, but the information ratio also captures the risk-reward trade-off of the strategy and the manager's value-added. (For a more detailed discussion of the information ratio and its relationship to skill and value-added, see Grinold [1989].)

An information ratio of 0.5 observed over five years may be statistically more significant than an information ratio of 0.5 observed over one year, but their value-added will be equal.[4] The distinction between t-statistic and information ratio arises because we define value-added based on risk over a particular horizon, in this case one year.

We have now discussed the t-statistic and the information ratio. What about the information coefficient, the grandfather of all information analysis statistics? Remember that the information coefficient measures the correlation between forecast and realized alphas. In the context of information analysis, the information coefficient is the correlation between our data and realized alphas.

If the data item is all noise and no signal, the information coefficient is zero. If the data item is all signal and no noise, the information coefficient is one. If there is a perverse relationship between the data item and the subsequent alpha, the information coefficient can be negative. The information coefficient has the nice property that it must lie between +1 and -1.

By itself, the information coefficient does not tell us the potential value-added of the information. It testifies mainly to the skill of the forecaster. It does not tell us how well we can diversify risk in portfolios that bet on the information. This becomes clear when we examine the fundamental law (Grinold [1989]) that relates information coefficients and information ratios:

$$\mathrm{IR} \approx \mathrm{IC} \cdot \sqrt{\mathrm{B}}. \qquad (7)$$

The information ratio and information coefficient are related by a parameter B, which measures the number of independent bets per year that the information will allow. For example, stock-specific information may allow independent quarterly bets on 500 different stocks, with B = 2000. Tactical asset allocation information may allow independent semi-annual bets on ten different asset classes, for a B = 20. In this case, if both sets of information exhibit the same information coefficient, the stock-specific information would have an information ratio ten times larger than the tactical asset allocation information, and would be capable of generating significantly higher value-added.

These examples illustrate the importance of the parameter B, the number of independent bets per year contained in the information. More bets per year leads to more diversification. As a practical matter, the parameter B is more difficult to measure than either the information ratio or the information coefficient.

ADVANCED TOPICS IN PERFORMANCE ANALYSIS

Performance analysis is a broad field; it is not the goal of this article on information analysis to provide a definitive review. Still, some advanced topics that pertain to analyzing information are worth a brief mention.

The first issue concerns portfolio turnover. Our two-step process has turned information into portfolios and then analyzed the performance of those portfolios. Because we have the portfolios, we can also investigate their turnover. In fact, given transactions costs, turnover will directly affect performance.

Other issues concern more detailed levels of analysis, which our approach allows. For instance, when we build long and short portfolios to bet for and against the information, we can also observe whether our information better predicts upside or downside alphas. In Example B, we saw that the long and short portfolios had almost equal information ratios (0.97 versus -0.93). Evidently, book-to-price ratios in the U.K. embody equal amounts of upside and downside information.

Beyond upside and downside information, we can also investigate whether the data contain information pertaining to up-markets or down-markets. How well do the portfolios perform in periods when the market is rising, and in periods when the market is falling?

Finally, some advanced topics here involve the connection between performance analysis and the step of turning information into portfolios. We can investigate the implicit horizon of information by varying the portfolio holding period — i.e., constructing daily portfolios, weekly portfolios, and so on — and analyzing performance in each case. The maximum information ratio should be achieved when the portfolio holding period matches the information horizon. We can also investigate the importance of controlling for different other variables such as industry or size. We can construct portfolios with different controls, and analyze the performance in each case.

THE PITFALLS OF INFORMATION ANALYSIS

Information analysis is a powerful tool. If we can use information analysis to evaluate the investment value of a set of raw data, we can also use it to refine that data. Done correctly, the process separates wheat from chaff. Done incorrectly, it is nothing more than data mining.

Data mining can fool information analysis into believing that information exists when it does not. Data mining can lead managers to bet on information that doesn't exist. Data mining is the bane of information analysis.

Fortunately, analysts can observe four guidelines to help keep information analysis from turning into data mining: intuition, restraint, sensibility, and out-of-sample testing. (For a detailed analysis, see Kahn [1990].)

First, the data analyzed must bear some intuitive relation to exceptional stock returns. Second, restraint should govern the information analysis process. Do not keep refining the data ad infinitum in a blind search for a high information ratio. Third, make sure that the results of information analysis are sensible. Beware if the analysis shows an information content much higher than expected. Fourth, when using information analysis to refine your investment process, augment the analysis with out-of-sample testing. Refine the process on one set of data, and test it on another set of data.

SUMMARY

This article has discussed information analysis. Information analysis proceeds in two steps. First, transform information into investment portfolios. Second, analyze the performance of those portfolios. To quantify that performance — and the value of the information being analyzed — the information ratio most succinctly measures the potential investment value-added contained in the information.

ENDNOTES

[1] To be explicit. we define alpha (residual return) as:

$$\alpha_n = (r_n - rf) - \beta_n \cdot (r_{benchmark} - rf).$$

The alpha for stock n is the excess return to stock n (stock n's return above the risk-free return) minus beta times the benchmark 's excess return.

[2] The choice of time period may affect information analysis. As a general comment, the investment time period should match the information time period. Portfolios based on daily information — information that changes daily and influences daily returns — should be regenerated each day. Portfolios based on quarterly information — information that changes quarterly and influences quarterly returns — should be regenerated each quarter.

[3] The active return is the portfolio return minus the bench mark return. By definition, the active retu rn to the benchmark is zero.

[4] In fact, the standard error of the information ratio is inversely related to the number of years over which we observe the returns:

$$SE[IR] = \frac{1}{\sqrt{T}} .$$

REFERENCES

Ambachtsheer. Keith P. "Profit Potential in an 'Almost Efficient' Market." *Journal of Portfolio Management*, Fall 1974, pp. 84-87.

——. "Where are the Customers' Alphas?'' *Journal of Portfolio Management*, Fall 1977, pp. 52-56.

Ambachtsheer, Keith P., and James L. Farrell, Jr. "Can Active Management Add Value?" *Financial Analysts Journal*, November/December 1979, pp. 39-47.

Grinold, Richard C. "The Fundamental Law of Active Management.'' *Journal of Portfolio Management*, Spring 1989, pp. 30-37.

Hodges, S.D., and R.A. Brealey. "Portfolio Selection in a Dynamic and Uncertain World." *Financial Analysts Journal,* March/April 1973, pp. 50-65.

Kahn, Ronald N. “What Practitioners Need to Know About Back testing.” *Financial Analysts Journal*, July/August 1990, pp. 17-20.

Rosenberg, Barr. “Security Appraisal and Unsystematic Risk in Institutional Investment.” *Proceedings of the Seminar on the Analysis of Security Prices*. Chicago: University of Chicago Press, November 1976, pp. 171-237.

Treynor, Jack, and Fischer Black. “How to Use Security Analysis to Improve Portfolio Selection.” *Journal of Business*, January 1973, pp. 68-86.