Problems in Measuring the Quality of Investment Information: The Perils of the Information Coefficient

Author(s): T. Daniel Coggin and John E. Hunter

Source: *Financial Analysts Journal*, May - Jun., 1983, Vol. 39, No. 3 (May - Jun., 1983), pp. 25+27-33

Published by: Taylor & Francis, Ltd.

Stable URL: https://www.jstor.org/stable/4478642

To limit the sampling error associated with the information coefficient (spearman's rank correlatin or rho) the model needs to transact with thousands of stocks (the model needs to be very large) or thousands or samples of stock returns/prices.

by T. Daniel Coggin and John E. Hunter

Problems in Measuring the Quality of Investment Information: The Perils of the Information Coefficient

Many investment firms and services use the "information coefficient"—the correlation between predicted and actual returns—to assess the performance of analysts or valuation models. What many may fail to realize is that any IC based on fewer than several thousand stocks will suffer from significant sampling error.

For example, an IC based on predictions for all the stocks on the New York Stock Exchange is unlikely to be higher than 0.10. But an IC based on 30 of those stocks will vary randomly by plus or minus 0.35! Thus sampling error in the 30-stock IC is likely to create an illusion of massive variation over time, or over analysts or over investment models.

Analysis of IC data on the Wells Fargo and Value Line valuation models reveals that the large differences in the models' ICs over time are due solely to sampling error. Analysis of analysts' ICs reveals, similarly, that the differences between analysts' ICs can be attributed entirely to sampling error. If there are differences in analysts' performances, they must be measured in some manner other than comparing ICs.

SEVERAL recent articles have recommended that the quality of research information in investment analysis be measured by the information coefficient (IC).[1] The IC is the (Pearson) correlation between the return on a common stock predicted by a valuation model or analyst and the actual return. Like any correlation, the IC is subject to sampling error—i.e., chance fluctuations in the value of the statistic due to randomness in the selection of the numbers used in the analysis. The extent of random error depends on the number of stocks correlated. In the current applications of the IC, the number of stocks is usually small, hence the amount of sampling error is large. Sampling error can seriously distort the conclusions drawn from ICs.

This article outlines *meta-analysis,* a new methodology for assessing the degree to which the difference in correlations cumulated across studies or time periods is due to sampling error.[2] Meta-analysis is used to compare the ICs of the Wells Fargo and Value Line valuation methods and the ICs of professional investment analysts. In each case, *all* the observed variation in ICs is due to sampling error.

1. Footnotes appear at end of article.

T. Daniel Coggin is Manager of the Investment Systems Department at Centerre Trust Company of St. Louis. John Hunter is Professor of Industrial Psychology at Michigan State University. *The authors thank Steven Einhorn, James Farrell and Stephen Levkoff for their helpful comments.*

An earlier version of this article was presented at the 31st Annual Meeting of the Midwest Finance Association, Chicago, 1982.

Sampling Error and Meta-Analysis

The ideal evaluation of any forecasting procedure would be to test it against a large population of stocks, such as all stocks traded on the New York Stock Exchange. This is not usually feasible, however, and most empirical research is confined to a sample of stocks from a larger population. It has long been known that the correlation between forecast returns and actual returns for a sample of stocks (denoted "r" in statistics) will differ from the correlation between forecast returns and actual returns for the entire population of stocks (denoted "ρ" in statistics). The difference is called "sampling error."

The sampling error is usually random in sign and specific magnitude. One can, however, compute its probability range from known formulas. The results of many such computations indicate that the size of the sampling error will depend largely on the number of stocks in the sample: The larger the sample of stocks, the smaller the typical sampling error. (The formulas developed in this article use the sample correlation coefficient rather than Fisher's z transformation. The reasons for this are given in the appendix.)

Suppose we wish to use the IC to evaluate investment analysts at Acme Investment Sorcerers, Inc. An examination of Clara Clairvoyant's predictions for every stock on the NYSE might tell us that her IC is $\rho = 0.20$. But an examination of Clara's predictions for only 30 stocks will give us an observed IC that departs from the true value of $\rho = 0.20$ by some unknown random amount.

The value of Clara's IC will have a normal distribution with a mean of $\rho = 0.20$ and a standard deviation of:

$$\sigma_e = \frac{1-\rho^2}{\sqrt{N-1}} = \frac{1-0.20^2}{\sqrt{30-1}} = 0.18 \,, \qquad (1)$$

where N is the number of stocks in the sample. This tells us that, 16 per cent of the time, Clara's observed IC will lie above 0.38 (=0.20 + 0.18), and 5 per cent of the time it will lie above 0.50 (=0.20 + 1.64 (0.18)). Another 16 per cent of the time, her observed IC will lie below 0.02 (=0.20 − 0.18), and 5 per cent of the time it will lie below −0.10 (=0.20 − 1.64 (0.18)). Thus 90 per cent of the time, Clara's observed IC will fall in the interval −0.10 to +0.50, but there is a 10 per cent chance of even greater error.

Suppose we calculate Clara's IC for each quarter of 1982, arriving at values of 0.44, −0.04, 0.28 and 0.12. At first glance, these values might appear to be radically different from one another. We might ask why Clara did so well in the first quarter, but so poorly in the second quarter. Yet meta-analysis would show that the variance in these values is no larger than would be expected, given the extent of the sampling error. In fact, the standard deviation of these four values is exactly the value of the sampling error (0.18) computed from Equation (1).

The same conclusion would apply if we were dealing with the performance of four analysts, each of whom evaluated 30 stocks. Even if the population IC were $\rho = 0.20$ for all analysts (i.e., no actual difference in job performance between sorcerers), the observed ICs would vary at least as far as from −0.10 to +0.50 for samples of 30 stocks. There would appear to be huge differences between analysts, but this observed variation would actually be due solely to sampling error, hence would be of no real value in comparing the analysts' performances.

Meta-analysis suggests that there is no reason to believe there is any real variation in the ICs for the different analysts. This result would be even stronger if the meta-analysis had covered a larger number of analysts. For example, if the variation in the ICs for 30 analysts were exactly equal to the value predicted by sampling error, then there is little likelihood that any other conclusion would be possible.

Estimating Real Variance

There is no way to estimate the sampling error in a sample IC from the information in that sample. The sampling error in a *sequence* of ICs can, however, be estimated using meta-analysis. The procedure is this:

(1) compute the mean and variance of observed ICs,
(2) compute the variance in ICs that is attributable to sampling error,
(3) test the observed variance against the chance value to see if there is any real variance in population values,
(4) subtract the sampling error variance from the observed variance to produce an estimate of the variance in population ICs, and
(5) take the square root of the corrected variance to estimate the standard deviation of population ICs.

If the ICs in a sequence are based on differing numbers of stocks, they should be weighted accordingly. The weighted mean and the weighted variance of a sequence of observed ICs are given by:

$$\overline{IC} = \frac{\Sigma(N_i IC_i)}{\Sigma N_i} \qquad (2)$$

and

$$s^2_{IC} = \frac{\Sigma[N_i(IC_i - \overline{IC})^2]}{\Sigma N_i} , \tag{3}$$

where N_i is the number stocks on which the ith IC is based. The variance due to sampling error is approximately:

$$s^2_e = \frac{K(1-\overline{IC}^2)^2}{\Sigma N_i} , \tag{4}$$

where K is the number of ICs and ΣN_i is the total number of stocks underlying all the coefficients. The *real* variance of population ICs is estimated by subtracting the sampling error variance from the observed variance:

$$s^2_\rho = s^2_{IC} - s^2_e \quad . \tag{5}$$

To test the hypothesis that there is no real variation in ICs, we can use the following statistic:

$$X^2_{K-1} = \frac{K(s^2_{IC})}{s^2_e} , \tag{6}$$

which has a chi-square distribution with K−1 degrees of freedom if the null hypothesis is true. This statistic is used as a formal test of no variation, although it has so much power that it will reject the null hypothesis (of no variation) even if only trivial amounts of variation exist. If the chi-square value is not significant, there is truly no variation across studies; but if it is significant, then the variation may still be negligible in magnitude.

Consider now the IC values used in the above illustrations—0.44, −0.04, 0.28 and 0.12. Suppose that the number of stocks in each IC was 32, 31, 30 and 29. Then the weighted mean and variance of the sequence would be:

$$\overline{IC} = \frac{32(0.44)+31(-0.04)+30(0.28)+29(0.12)}{32+31+30+29}$$

$$= \frac{24.72}{122} = 0.2026 ,$$

$$s^2_{IC} = \frac{32(0.44-0.2026)^2+31(-0.04-0.2026)^2+\ldots}{32+31+30+29}$$

$$= \frac{4.0053}{122} = 0.0328 .$$

The sampling error variance would be:

$$s^2_e = \frac{4(1-0.2026^2)^2}{122} = 0.0301 .$$

Thus our best estimate of the true variance in ICs would be the difference:

$$s^2_\rho = s^2_{IC} - s^2_e = 0.0328-0.0301 = 0.0027 .$$

For these four values, the observed mean IC is 0.20, and the observed standard deviation is 0.18. Meta-analysis suggests that the true mean is 0.20 and the true standard deviation 0.05 (i.e., $\sqrt{0.0027}$). Applying the significance test to see if the observed variation is greater than sampling error alone would suggest, we have:

$$X^2_3 = \frac{K(s^2_{IC})}{s^2_e} = \frac{4(0.0328)}{0.0301} = 4.36 ,$$

which, for three degrees of freedom, is not statistically significant at the 5 per cent level. Thus even the corrected variance of 0.0027 may be attributable to sampling error. The chi-square test serves as an important reminder that even the cumulated statistics suffer from sampling error, the total sample size being only 122.[3]

Variation Between Models: Wells Fargo Versus Value Line

Table I presents the results of a study by Ambachtsheer comparing ICs for two valuation models over six independent six-month periods from 1973 to 1976.[4] As Ambachtsheer notes, the Wells Fargo model is essentially a dividend discount model that incorporates the Security Market Line concept; the Value Line model is based on historical price and earnings relationships and incorporates short-term earnings surprise factors.[5] Ambachtsheer contends that these data demonstrate "considerable period-to-period variability in the ICs for both valuation methodologies."

Table II presents the results of applying Equations (3), (4) and (5) to Ambachtsheer's data. In both cases, the observed variance (s^2_{IC}) is greater than the error variance (s^2_e), which would suggest that some of the variability in ICs is real.

Table I Wells Fargo and Value Line Model ICs

Period	*N*	IC: *Wells Fargo*	IC: *Value Line*
9/1973–3/1974	200	0.12	0.17
3/1974–9/1974	200	0.16	0.04
9/1974–3/1975	200	0.01	−0.09
3/1975–9/1975	200	0.13	0.16
9/1975–3/1976	200	0.08	0.11
3/1976–9/1976	200	0.31	0.01
Weighted Average		0.135	0.067

Table II Sampling Error in Model ICs

	Wells Fargo	*Value Line*
s^2_{IC}	0.0083	0.0083
s^2_e	0.0048	0.0049
s^2_ρ	0.0035	0.0034

However, the magnitude of the real variance (s^2_ρ) may itself be due to chance.

Applying Equation (6) to test the significance of the real variances yields values of 10.37 for Wells Fargo and 10.16 for Value Line (with five degrees of freedom). Neither is significant at the 5 per cent level. Thus the chi-square test shows that the observed variances are not significantly greater than the error variances. This analysis refutes Ambachtsheer's claim of significant period-to-period variability.

Our analysis of the data for each model separately shows that the variation over time is no greater than would be produced by sampling error. If the variation over time is an illusion, then the true IC values for the two models are best estimated by the average values. That is, our best estimates of the six successive values for Wells Fargo are not the observed 0.12, 0.16, etc., but rather 0.135, 0.135, etc.

The Confidence Interval for the IC

The mean IC value for Wells Fargo (0.135) is just over twice as great as the value for Value Line (0.067). Is there any possibility this difference can be attributed solely to sampling error? Each mean is based on a total sample size of only 1,200, so there is room for sampling error.

Sampling error in a single value can only be described as a probability distribution. Thus, associated with a given IC, there is an error band or confidence interval that will contain the population IC with a given probability. If the observed value of the information coefficient is IC, then the width of the 95 per cent confidence band is:

$$w = \frac{1.96(1-IC^2)}{\sqrt{N-1}} \tag{7}$$

where N is the number of stocks in the IC. The endpoints of the confidence interval are $IC-w$ and $IC+w$, and the population IC will fall within those endpoints 95 per cent of the time.

The 95 per cent confidence intervals for Wells Fargo and Value Line can be obtained by using 1,200 as the sample size. They are 0.079 to 0.191 for Wells Fargo and 0.011 to 0.123 for Value Line. Because the confidence intervals do overlap, a possibility exists that the population values are the same for the two models. About four times as much data would be needed to establish the difference between these models if the true difference is no larger than that observed.

The confidence interval is an unbiased way of describing sampling error. The alternative method of handling sampling error is the significance test, which asks if the observed IC is significantly different from 0.0. The appendix shows that the significance test is contained in the confidence interval and represents a biased treatment of sampling error if used alone.

Combining Predictions

Ambachtsheer and Farrell have argued that predictive ability can be improved by combining the predictions made by two investment models.[6] Their basic argument is as follows. Suppose two models are available, both work equally well (i.e., $IC_1 = IC_2$), and the predictions of the two models are uncorrelated. According to multiple regression theory, the optimum prediction would be obtained by averaging the predictions made by the two models. The IC for the average prediction (equal-weighted) would be $\sqrt{2}(IC_{1\text{ or }2})$, meaning that the averaged prediction should do 41 per cent better than the predictions of either model taken separately.

This procedure works if one assumes there is no correlation between the predictions of the two models. In fact, however, most valuation models rely on information derived from the same sources (e.g., the same industry reports and economic statistics). One would therefore expect a substantial correlation between the predicted values of two investment models, hence only modest improvement from averaging the predictions. Certainly, the recent emergence of the dividend discount model as, perhaps, the dominant model of investment value does not weaken this expectation.[7]

Apparent Variation in ICs across Analysts

A number of firms use ICs to evaluate the performance of their investment analysts. At the end of each quarter, for example, they compute the IC for the predictions of each analyst and use these coefficients to rank their analysts. One commercial service even calculates daily ICs for analysts.

Table III Mean Quarterly Analysts' ICs

Analyst	*1979*	*(N)*	*1980*	*(N)*	*1981*	*(N)*
	Mean IC					
1	0.18	(72)	−0.08	(76)	0.08	(101)
2	0.12	(65)	0.02	(92)	0.11	(86)
3	−0.06	(66)	0.23	(73)	−0.04	(99)
4	−0.03	(128)	0.16	(130)	0.17	(118)
5	0.08	(32)	0.18	(44)	0.24	(37)
6	0.12	(79)	−0.06	(78)	0.27	(80)
7*	—		0.23	(88)	0.06	(108)
8*	—		0.17	(108)	0.16	(112)
9*	—		−0.02	(54)	−0.04	(60)
Weighted Average	0.06	(74)	0.10	(83)	0.11	(89)

*Analysts 7, 8 and 9 were not employed for the full year 1979.

Table IV Real Variance in Analysts' ICs

Year	*Real Variance* (s_ρ^2)	*Chi-Square* (X^2)	*Degrees of Freedom*	*Significant* ($\alpha = 0.05$)
1979	−0.0055	3.55	5	NO
1980	0.0013	9.98	8	NO
1981	−0.0019	7.44	8	NO

We analyzed the quarterly ICs of the investment analysts at a large regional trust investment department over a three-year period, 1979 to 1981. There were six analysts in 1979 and nine in 1980 and 1981. Each analyst covered two to three industries, following, on average, 18 companies in 1979, 21 in 1980 and 22 in 1981.

Each analyst predicted total return on the basis of a three-phase dividend discount model, providing the model with near-term forecasts and long-term growth rates for earnings and dividends. Table III presents the mean quarterly IC for each analyst for each year. The quarterly ICs were averaged over a year to reduce sampling error; that is, the sampling error in the one-year average of quarterly ICs based on 20 stocks is approximately the same as an IC based on 80 stocks. Averaging over a year is also consistent with management's annual review of the analysts' performance. The weighted average quarterly IC for all analysts for each year and the 95 per cent confidence interval are as follows: 1979 = 0.06 (−0.03, 0.15), 1980 = 0.10 (0.03, 0.17), 1981 = 0.11 (0.04, 0.18).

Table IV presents the results of meta-analysis of the data in Table III. There is no significant real variation in ICs across analysts for any of the three years. Indeed, the chi-square test was unnecessary for 1979 and 1981, since the error variance in these years was larger than the observed variance.[8] To make distinctions between these analysts on the basis of these ICs is equivalent to throwing dice.

A Replication of the Analysts' IC Study

After hearing of our results, an independent source made available to use two samples of its IC data. Meta-analysis showed that, for both samples, the variance in ICs was due to sampling error; in each case, there was no real difference in analysts' ICs. The first sample, covering 185 analysts, had an observed variance (s_{IC}^2) of 0.00107, but with an error variance (s_e^2) of 0.00228. The second sample, covering 168 analysts, had an observed variance of 0.00062 and an error variance of 0.00075.

Both these samples had for the two time periods a mean IC of 0.0, which differs sharply (by more than sampling error) from our analysts' three-year average of about 0.10. Other research has indicated a mean analysts' IC of about 0.12. This suggests to us that mean analysts' ICs may be period-specific, having to do with stock market cycles.[9]

Our findings make moot a worrisome question concerning the IC—the potential noncomparability of ICs based on different stock groups. Had there been significant variation across the analysts' ICs for a particular period, we would have had to address the question of the comparability of ICs across industries or stock groups—that is, whether differences might have been due to differences in stock groups, rather than differences in analysts' ability. The fact that no such differences were found suggest that there is simply no empirical validity to the use of ICs for the evaluation of analysts' relative performance.

ICs and Portfolio Performance

The theoretical relationship between the magnitude of the IC and portfolio performance has been derived in Treynor and Black and Ferguson.[10] Several empirical studies have discussed the concept of a minimum IC level necessary to indicate that active portfolio management produces a return greater than that of a market index.[11] Unfortunately, these studies too have failed to address properly the problem of the amount of sampling error present in the suggested minimum IC level. Failure to estimate the degree of sampling error can lead to highly suspect and possibly misleading IC guidelines for successful active management. ■

Appendix

Distribution of the Sample Correlation

Some readers may be surprised that meta-analysis does not use the Fisher z transformation of the correlation coefficient. Actually, early work by Schmidt and Hunter did use the Fisher z transformation.[12] But later work discovered that the Fisher z transformation is biased and that the analysis of the simple correlation coefficient is both more understandable and more statistically efficient.

There is widespread misunderstanding about the purpose of the Fisher transformation. Many believe that the transformation is intended to make the distribution of the sample correlation more normal in shape. But except for extreme cases, the distribution of r is already normal. Furthermore, the Fisher z transformation was derived for a very different purpose. The purpose of the Fisher z transformation was to develop a statistic whose standard error was independent of the population value.

The standard error of r is:

$$\sigma_e = (1 - \rho^2) / \sqrt{N-1} ,$$

which is known exactly only if the population correlation ρ is known. Thus if r^2 is used to estimate ρ^2, then there is a small error in the confidence interval corresponding to the sampling error in r^2. On the other hand, there is no error in the standard error used to form the confidence interval for Fisher's z. The transformed value z has a known standard error of $1/\sqrt{N-3}$. The problem is that no one really wants to know the confidence interval for the population value of z. Therefore the confidence interval for z is backtransformed to create a confidence interval for ρ. Alas, this confidence interval for ρ has an upward bias.

For the purpose of creating confidence intervals for single correlations, there is only a decimal dust difference between the use of r and the use of z in most circumstances (i.e., $\rho < 0.80$ or $N > 20$). The two will be numerically identical for the tiny correlations that occur in the present research using the information coefficient.

For meta-analysis, the situation is different. Averaging across studies eliminates the effect of sampling error. The bias of Fisher's z then becomes a systematic overestimate of the population correlation. Schmidt, Gast-Rosenberg and Hunter discovered this in a study on computer programmers.[13] The estimated population correlation using the Fisher z transformation was about 0.90, while the estimated population correlation using sample correlations was about 0.80. Simulation and series approximations showed that it was Fisher's z that was wrong.

It is ironic that the standard error for z of $1 / \sqrt{N-3}$ is actually only an approximation. The conditions under which the approximation is accurate are exactly the same conditions under which the sample correlation is normally distributed (such as $\rho < 0.80$ or $N > 20$). Thus for meta-analytic purposes, it is never appropriate to use the Fisher z transformation, although the error is not large unless the average population correlation is greater than 0.50.

Significance Tests versus Confidence Intervals

Many people believe that the use of significance tests guarantees an error rate of 5 per cent or less. This is just not true. Mathematical statisticians have known (and have tried to teach practitioners) this for years, with the possibility of higher error rates shown in discussions of the power of statistical tests. The 5 per cent error rate is a Type I error rate that is guaranteed only if the null hypothesis ($\rho = 0.0$) is true. If the null hypothesis is false, then the error rate is a Type II error rate that can go up to 95 per cent!

Consider the case of an "average" portfolio (i.e., IC = 0.15, N = 200). The one-tailed significance test for a correlation coefficient at the 5 per cent level is $\sqrt{N-1}$ (r) ≥ 1.64. In our case, this implies that the IC must be greater than 0.12 to be significantly different from 0.0. If the population correlation is 0.15 and the sample size is 200, then the mean of the sample correlations is 0.15, while the standard deviation is $(1-\rho^2)/\sqrt{N-1} = 0.07$. Thus the probability that the observed correlation will be significant is the probability that the sample correlation will be greater than 0.12 when it has a mean of 0.15 and a standard deviation of 0.07:

$$P(IC \geq 0.12) = P\left(\frac{IC-0.15}{0.07} \geq \frac{0.12-0.15}{0.07}\right)$$
$$= P(z \geq -0.43) = 0.67 \ .$$

That is, if all studies were done with a sample size of 200, then a population correlation of 0.15 would imply an error rate of 33 per cent!

What does this illustration tell us? Specifically, it shows that if the population IC is 0.15 and the sample size is 200, there will be positive information in only 67 per cent of the studies, even though the null hypothesis ($\rho = 0.0$) is false in all cases. Generally, and more importantly, it shows that the t-statistic is not the appropriate measure of the amount of information contained in an IC. Fortunately, there are two alternatives to the significance test. At the level of review studies, there is meta-analysis, and at the level of the single study, there is the confidence interval for the IC.

If we consider our example from the point of view of the confidence interval for the population correlation from a single study, we find $0.01 \leq \rho \leq 0.29$, with 95 per cent confidence. Confidence intervals are superior to the significance test for at least three reasons. First, the confidence interval is correctly centered about the observed value rather than the hypothetical value of the null hypothesis. Second, the confidence interval is more appropriate for comparing analysts or models, since overlapping confidence intervals show the extent of uncertainty surrounding the true rank order of the ICs. The significance test instead compares each analyst separately to the hypothetical value $\rho = 0.0$. Furthermore, the significance test treats all values above the minimum t-value the same (i.e., "significant"), despite large potential differences in the magnitude of the ICs.

Consider, for example, a situation in which one analyst is just barely higher than the t-test cutoff value, whereas a second analyst is just barely below the cutoff value. The performance of the analyst above the value is judged significantly better than chance, whereas the performance of the analyst below the cutoff is not significantly better than chance. Thus the first analyst is judged better than the second analyst, even if their observed ICs are virtually identical in value and far from significantly different from each other. The confidence intervals for the two analysts, however, would almost completely overlap, hence show the true uncertainty about the actual rank order.

Third, the confidence interval gives the researcher a correct picture of the degree of uncertainty in small samples. For example, suppose that we want the confidence interval for the IC to define the population correlation to the first digit—i.e., to have a width of ± 0.05. Then it can be shown that the minimum sample size is 1,538. In order for a sample size of 1,000 to be sufficient, the population correlation must be ≥ 0.44. Thus all single correlation studies with less than 1,000 stocks are dealing with a small sample size.

As we have noted, confidence intervals give the correct picture of the degree of uncertainty in correlations computed from small samples. The only way to reduce this uncertainty is to construct large sample single studies or combine results across several small sample studies. With IC analysis, the most feasible solution is to combine results across studies or test periods and perform meta-analysis.

Footnotes

1. See, for example, Keith P. Ambachtsheer, "Where Are All the Customers' Alphas?" *The Journal of Portfolio Management*, Fall 1977; Ambachtsheer and James L. Farrell, Jr., "Can Active Management Add Value?" *Financial Analysts Journal*, November/December 1979; and Ambachtsheer, "Evaluating the Quality of Investment Information," in Sumner N. Levine, ed., *Investment Manager's Handbook* (Homewood, IL: Dow Jones-Irwin, 1980).
2. For an introduction to meta-analysis, see Gene V. Glass, Barry McGaw and Mary Lee Smith, *Meta-Analysis in Social Research* (Beverly Hills: Sage Publications, 1981). For a more advanced treatment, including a discussion of sampling error and error of measurement, see John E. Hunter, Frank L. Schmidt and Gregg B. Jackson, *Meta-Analysis: Cumulating Research Findings Across Studies* (Beverly Hills: Sage Publications, 1982).
3. The formulas given here assume that the IC is the Pearson correlation between predicted and actual values. However, some researchers convert predicted and actual values to ranks before correlating. This means their correlation is the Spearman rank correlation coefficient—Spearman's rho. (For a definition of Spearman's rho, see Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1956) or W.J. Conover, *Practical Nonparametric Statistics*, Second Edition (New York: John Wiley, 1980).) Assuming bivariate normality, Spearman's rho is less than Pearson's r by an amount ranging from 0.0, when both are 0.0, up to about 0.02, when the correlations are in the 0.70 to 0.80 range. Thus Spearman's rho slightly understates the strength of the relationship. The sampling error is also slightly different for $\rho \neq 0.0$ (see Maurice G. Kendall, "Rank and Product-Moment Correlation," *Biometrika*, Vol. 36, 1949 and M.G. Kendall, *Rank Correlation Methods*, Fourth Edition (New York: Hafner Publishing Company, 1962)). Within the range of ICs in the investment literature, the sampling error in Spearman's rho is the same as the sampling error in Pearson's r. Therefore the formulas given in this article work equally well for both. If a study is done on correlations greater than 0.70, there will be a discrepancy between the meta-analysis formulas given here and the corresponding formulas for Spearman's rho.
4. Ambachtsheer, "Evaluating the Quality of Investment Information."
5. The Wells Fargo model is discussed in detail in William F. Fouse, "Risk, Liquidity and Common Stock Prices," *Financial Analysts Journal*, May/June 1976 and Fouse, "Risk and Liquidity Revisited," *Financial Analysts Journal*, January/February 1977. The Value Line model is discussed in Arnold Bernhard, *Value Line Methods of Evaluating Common Stocks* (New York: Arnold Bernhard & Company, 1979).
6. Ambachtsheer and Farrell, "Can Active Management Add Value?"
7. Evidence for our contention can be found in T. Daniel Coggin and John E. Hunter, "Forecasting Annual Earnings Per Share: A Comparison of Three Statistical Models to a Consensus of Analysts," *The Journal of Business Forecasting*, Winter 1982/83. This study found (using IBES analysts' consensus earnings forecast data for 1978 and 1979) that the ideosyncratic component in analysts' forecasts is very small in proportion to the systematic (consensus) error component.
8. Our analysis of the reliability of the quarterly ICs for the analysts (not presented in this article) yields

an average Spearman-Brown reliability coefficient of 0.0 for the three years, thus adding confirmation to the evidence that all the variability in these ICs is due to sampling error. For a discussion of the Spearman-Brown reliability coefficient, see J.P. Guilford and Benjamin Fruchter, *Fundamental Statistics in Psychology and Education*, Fifth Edition (New York: McGraw-Hill, 1973), Chapter 17.

9. For a full description of the underlying model with an empirical test, see J.E. Hunter, T.D. Coggin and M.K. Conway, "Forecast Data and Systematic Risk" (Paper presented at the annual meeting of the Midwest Finance Association, St. Louis, 1983).
10. See Jack L. Treynor and Fischer Black, "How to Use Security Analysis to Improve Portfolio Selection," *The Journal of Business*, January 1973 and Robert Ferguson, "Active Portfolio Management," *Financial Analysts Journal*, May/June 1975.
11. See Stewart D. Hodges and Richard A. Brealey, "Portfolio Selection in a Dynamic and Uncertain World," *Financial Analysts Journal*, March/April 1973 and Ambachtsheer, "Where Are All the Customers' Alphas?"
12. Frank L. Schmidt and John E. Hunter, "Development of a General Solution to the Problems of Validity Generalization," *Journal of Applied Psychology*, Vol. 62, 1977.
13. F.L. Schmidt, I. Gast-Rosenberg and J.E. Hunter, "Validity Generalization for Computer Programmers," *Journal of Applied Psychology*, Vol. 65, 1980.

Books

concluded from page 80

bad. Existence of a bad system implies the existence of a superior one, and one will prove as elusive as the other. Therefore, how-to books on investing are to be read with caution. And we are cautious.

Mr. Bigus, an MIT-trained management engineer, faces the investing problem squarely. He analyzes a long catalogue of systems that investors use that do not work. Then he presents his own solution, low-risk investing—pure Graham and Dodd. His first requirement is a very strong and conservative balance sheet. Next he builds an investment value for the stock based on an average price/earnings ratio and historical earnings. This investment value is the sell price; the buy price is reached by applying a severe haircut to the investment value, based on company size and on the degree of past earnings fluctuation. That's it, one compact chapter.

This reviewer automatically challenged that chapter because of his feeling that reducing the downside risk might reduce the upside potential commensurately; the method looked as if the investor would be out of the stock market much of the time. However, Mr. Bigus answered the important questions in the appendix and elsewhere in the text. He tested his results throughout. It seems that he has reduced fundamental risks through fundamentals, rather than blind diversification. On the record, he seems to have tied systematic risk to individual stock fundamentals rather than to overall market timing. Best of all, he has insulated his decision-making from a host of investment methods that have not worked.

Not that the book is perfect. The organization is somewhat confusing. The tables are not always clearly labeled. The investment method, in our view, is worth a try, but it is possible that its successful execution may prove to be too slow and too demanding for most would-be millionaires.

—*R.I.C.*