Beta and Return

"Announcements of the 'death' of beta seem premature."

Fischer Black

FISCHER BLACK is a partner at Goldman, Sachs & Co. in New York (NY 10004).

Eugene Fama says (according to Eric Berg of *The New York Times*, February 18, 1992) "beta as the sole variable explaining returns on stocks is dead." He also says (according to Michael Peltz of *Institutional Investor*, June 1992) that the relation between average return and beta is *completely* flat.

In these interviews, I think that Fama is misstating the results in Fama and French [1992]. Indeed, I think Fama and French, in the text of that article, misinterpret their own data (and the findings of others).

Black, Jensen, and Scholes [BJS, 1972] and Miller and Scholes [1972] find that in the period from 1931 through 1965 low-beta stocks in the United States did better than the capital asset pricing model (CAPM) predicts, while high-beta stocks did worse. Several authors find that this pattern continued in subsequent years, at least through 1989. Fama and French extend it through 1990.

All these authors find that the estimated slope of the line relating average return and risk is lower than the slope of the line that the CAPM says relates expected return and risk. If we choose our starting and ending points carefully, we can find a period of more than two decades where the line is essentially flat.

How can we interpret this? Why is the line so flat? Why have low-beta stocks done so well relative to their expected returns under the CAPM?

Black [1972] shows that borrowing restrictions (like margin requirements) might cause low-beta stocks to do relatively well. Indeed, Fama and French refer often to the Sharpe-Lintner-Black (SLB) model that includes these borrowing restrictions. This model predicts only that the slope of the line relating expected return and beta is positive.

Fama and French claim to find evidence against this model. They say (for example, on p. 459) that their results "seem to contradict" the evidence that the slope of the line relating expected return and beta is positive.

This is a misstatement, in my view. Even in the period they choose to highlight, they cannot rule out the hypothesis that the slope of the line is positive. Their results for beta and average return are perfectly consistent with the SLB model.

Moreover, if the line is really flat, that implies dramatic investment opportunities for those who use beta. A person who normally holds both stocks and bonds or stocks and cash can shift to a portfolio of similar total risk but higher expected return by emphasizing low-beta stocks.

Beta is a valuable investment tool if the line is as steep as the CAPM predicts. It is even more valuable if the line is flat. No matter how steep the line is, beta is alive and well.

DATA MINING

When a researcher tries many ways to do a study, including various combinations of explanatory factors, various periods, and various models, we often say he is "data mining." If he reports only the more successful runs, we have a hard time interpreting any statistical analysis he does. We worry that he selected, from the many models tried, only the ones that seem to support his conclusions. With enough data mining, all the results that seem significant could be just accidental. (Lo and MacKinlay [1990] refer to this as "data snooping." Less formally, we call it "hindsight.")

Data mining is not limited to single research studies. In a single study, a researcher can reduce its effects by reporting all the runs he does, though he still may be tempted to emphasize the results he likes. Data mining is most severe when many people are studying related problems.

Even when each person chooses his problem independently of the others, only a small fraction of research efforts result in published papers. By its nature, research involves many false starts and blind alleys. The results that lead to published papers are likely to be the most unusual or striking ones. But this means that any statistical tests of significance will be gravely biased.

The problem is worse when people build on one another's work. Each decides on a model closely related to the models that others use, learns from the others' blind alleys, and may even work with mostly the same data. Thus in the real world of research, conventional tests of significance seem almost worthless.

In particular, most of the so-called anomalies that have plagued the literature on investments seem likely to be the result of data mining. We have literally thousands of researchers looking for profit opportunities in securities. They are all looking at roughly the same data. Once in a while, just by chance, a strategy will seem to have worked consistently in the past. The researcher who finds it writes it up, and we have a new anomaly. But it generally vanishes as soon as it's discovered.

Merton [1987, pp. 103-108] has an excellent discussion of these problems. He says (p. 108) "although common to all areas of economic hypothesis testing, these methodological problems appear to be especially acute in the testing of market rationality."

The "size effect" may be in this category. Banz [1981] finds that firms with little stock outstanding (at market value) had, up to that time, done well relative to other stocks with similar betas. Since his study was published, though, small firms have had mediocre and inconsistent performance.

Fama and French [1992] continue studying the small-firm effect, and report similar results on a largely overlapping data sample. In the period since the Banz study (1981-1990), they find no size effect at all, whether or not they control for beta. Yet they claim in their paper that size is one of the variables that "captures" the cross-sectional variation in average stock returns.

Fama and French also give no reasons for a relation between size and expected return. They might argue that small firms are consistently underpriced because they are "neglected" in a world of large institutional investors. But they do not give us that reason or any other reason. Lack of theory is a tipoff: watch out for data mining!

Fama and French also find that the ratio of book value to the market value of the firm's equity helps capture the cross-sectional variation in average stock returns. They favor the idea that this ratio captures some sort of rationally priced risk, rather than market overreaction to the relative prospects of firms. But they say nothing about what this risk might be, or why it is priced, or in what direction.

They mention the possibility that this result is due to "chance," which is another way to describe data mining, but they don't consider that plausible, because the result appears in both halves of their period, and because the ratio predicts a firm's accounting performance.

I consider both those arguments weak. Given that an "effect" appears in a full period, we expect to find it in both halves of the period. We are not surprised when we do.

We know that when markets are somewhat efficient, stock prices react before accounting numbers to events affecting a firm's performance. Thus we are not surprised when firms with high ratios of book-to-market equity show poor subsequent accounting performance. I don't think this is evidence of a priced risk factor at all.

Thus I think it is quite possible that even the book-to-market effect results from data mining, and will vanish in the future. But I also think it may result in part from irrational pricing. The ratio of book-to-market equity may pick up a divergence between value and price across any of a number of dimensions. Thus the past success of this ratio may be due more to market inefficiencies than "priced factors" of the kind that Fama and French favor.

If the subsequent convergence of price and value is gradual, people seeking profit opportunities may not fully eliminate the effect. To capture the gains, they have to spend money on active management, and they must bear the risks of a less-than-fully diversified portfolio.

BETA THEORY

I think most of the Fama and French results are attributable to data mining, especially when they reexamine "effects" that people have discussed for years. Even they note that the ratio of book-to-market equity has long been cited as a measure of the return prospects of stocks.

I especially attribute their results to data mining when they attribute them to unexplained "priced factors," or give no reasons at all for the effects they find.

Strangely, the factor that seems most likely to be priced they don't discuss at all: the beta factor. We can construct the beta factor by creating a diversified portfolio that is long in low-beta stocks and short in smaller amounts of high-beta stocks, so that its beta is roughly zero. The returns to all such portfolios tend to be highly correlated, so we don't have to worry about the details of the "right" way to create the beta factor.

The empirical evidence that the beta factor had extra returns is stronger than the corresponding evidence for the small-stock factor or the book-to-market equity factor. The first evidence was published in 1972, and the factor has performed better since publication than it did prior to publication.

Moreover, we have some theory for the beta factor. Black [1972] showed that borrowing restrictions might cause low-beta stocks to have higher expected returns than the CAPM predicts (or the beta factor to have a higher expected return than interest at the short-term rate). Borrowing restrictions could include margin rules, bankruptcy laws that limit lender access to a borrower's future income, and tax rules that limit deductions for interest expense.

These restrictions have probably tightened in the United States in recent decades. Margin rules have remained in effect, bankruptcy laws seem to have shifted against lenders, and deductions for interest expense have been tightened. Many countries outside the United States seem to have similar restrictions. If they help explain the past return on the beta factor, they will continue to influence its future return.

Moreover, many investors who can borrow, and who can deduct the interest they pay, are nonetheless reluctant to borrow. Those who want lots of market risk will bid up the prices of high-beta stocks. This makes low-beta stocks attractive and high-beta stocks unattractive to investors who have low-risk portfolios or who are willing to borrow.

We can see some evidence for this in the market's reaction to a firm that changes its leverage. An exchange offer of debt for equity generally causes the firm's stock price to increase, while an offer of equity for debt causes it to decrease. This may be because of the tax advantages of debt; or because more debt transfers value from existing bondholders to stockholders; or because buying equity signals manager optimism.

I believe, though, that an important reason is reluctance to borrow: in effect, a firm that adds leverage is providing indirect borrowing for investors who are unwilling to borrow directly. These investors bid up its stock price.

BJS [1972] discuss another possible reason for beta factor pricing: mismeasurement of the market portfolio. If we use a market portfolio that differs randomly from the true market portfolio, stocks that seem to have low betas will on average have higher betas when we use the correct market portfolio to estimate them. Our betas are estimated with error (even in the final portfolio), and we select stocks that seem to have low betas. Such stocks will usually have positive alphas using the incorrect market portfolio. The portfolio method does not eliminate this bias.

Perhaps the most interesting way in which the market portfolio may be mismeasured involves our neglect of foreign stocks. World capital markets are becoming more integrated all the time. In a fully integrated capital market, what counts is a stock's beta with the world market portfolio, not its beta with the issuer country market portfolio. This may cause low-beta stocks to seem consistently underpriced. If investors can buy foreign stocks without penalty, they should do so; if they cannot, stocks with low betas on their domestic market may partly substitute for foreign stocks. If this is the reason the line is flat, they may also want to emphasize stocks that have high betas with the world market portfolio.

Can't we do some tests on stock returns to sort out which of these theoretical factors is most important? I doubt that we have enough data to do that.

We have lots of securities, but returns are highly correlated across securities, so these observations are far from independent. We have lots of days, but to estimate factor pricing what counts is the number of years for which we have data, not the number of distinct observations. If the factor prices are changing, even many years is not enough. By the time we have a reasonable estimate of how a factor was priced on average, it will be priced in a different way.

Moreover, if we try to use stock returns to distinguish among these explanations, we run a heavy risk of data mining. Tests designed to distinguish may accidentally favor one explanation over another in a given period. I don't know how to begin designing tests that escape the data mining trap.

VARYING THE ANALYSIS

While the BJS study covers lots of ground, I am especially fond of the "portfolio method" we used. Nothing I have seen since 1972 leads me to believe that we can gain much by varying this method of analysis.

The portfolio method is simple and intuitive. We try to simulate a portfolio strategy that an investor can actually use. The strategy can use any data for constructing the portfolio each year that are available to investors at the start of that year. Thus we can incorporate into our selection method any "cross-sectional" effects that we think are important.

However, the more complex our portfolio selection method is, the more we risk bringing in a data mining bias. I must confess that when we were doing the original BJS study, we tried things that do not appear in the published article. Moreover, we were reacting to prior work suggesting a relatively flat slope for the line relating average return to beta. Thus our article had elements of data mining too.

To minimize the data mining problem, BJS used a very simple portfolio strategy. We chose securities using historical estimates of beta, and we used many securities to diversify out the factors not related to beta.

But this method does have flaws. For example, beta is highly correlated with both total risk and residual risk across stocks. So what we call the "beta factor" might better be called the "total risk factor" or the "residual risk factor." I can't think of any reliable way to distinguish among these.

When doing the BJS study, we considered estimating the entire covariance matrix for our population of stocks, and using that to improve the efficiency of our test. We realized that this would require us to deal with uncertainty in our estimated covariances. We decided that the potential for improved efficiency was small, while the potential for error in our econometric methods was large. So we did not pursue that route.

Others have used different methods to update our study. My view is that in the presence of data mining and estimate error and changing risk premiums, none of these methods adds enough accuracy to warrant its complexity. I view most of these methods as our method expressed in different language.

For example, Fama and MacBeth [1973] start with cross-sectional regressions of return on beta, and