Estimating Expected Return

Fischer Black

The key issue in investments is *estimating* expected return. It is neither explaining return nor, as Fama and French suggest, explaining average return.[1] These topics combine estimating expected return with explaining variance, which is a completely different matter.

For example, the Capital Asset Pricing Model (CAPM) is a model of expected return. The "market model," which says that the market residuals of all stocks are independent, is a model of variance. The CAPM and the market model are almost wholly unrelated. One helps in estimating expected return, while the other helps in explaining variance.

Similarly, the Sharpe-Lintner-Black (SLB) model that Fama and French cite is a model of expected return. It says that expected return is a linear function of beta with a positive slope, and that only beta matters in explaining expected return.

In contrast, Arbitrage Pricing Theory (APT) is a model of variance. It says that the number of independent factors influencing return is limited, but it is silent on the pricing of these factors, so it is silent on expected return.

Explaining variance is easy. We can use daily (or more frequent) data to estimate covariances. Our estimates are accurate enough that we can see the covariances change through time. Explaining return or average return is easy too, because that's just a way of explaining variance.

Estimating expected return is hard. Daily data hardly help at all. Only longer time periods help. We need decades of data for accurate estimates of average expected return. We need such a long period to estimate the average that we have little hope of seeing changes in expected return.

THEORY AND DATA

Fama and French do not seem to believe much in theory when they estimate expected return. They (and many others) rely heavily on data. They look at average returns on certain factors as evidence of expected returns for those factors.

Similarly, people who use the APT framework in estimating expected return usually use past average return as an estimate. When we have no theory for how a factor should be priced, past average return can give the best estimate, but it's normally a highly inaccurate estimate.

People who rely on data sometimes assume that factor expected returns (or expected returns in excess of the riskless interest rate) are constant, but they often condition expected returns on various observables. In that case, they are using conditional average return to estimate conditional expected return. Using a complex multivariate analysis does not change the essence of the approach. Estimates of conditional expected return are about as inaccurate as estimates of ordinary expected return.

The people who use data often think of the factors as rationally priced, because the factors represent risks that people care about, as in Merton's intertemporal asset pricing model.[2] But they rarely tell us *how* the factors should be priced; they usually don't even predict the *signs* of the factor expected excess returns.

Sometimes (but less commonly in the academic world), people speak of the factors as "irrationally priced" or "mispriced." Factor mispricing can be consistent, as when "institutions" matter a lot to asset prices, or it can shift rapidly, as when "noise" matters a lot.

One version of the "noise" theory predicts that ratios of price to accounting estimates of value can help predict expected return. That's a possible reason for the good performance of Fama and French's price-to-book factor. But Fama and French hardly mention this explanation. They don't want to hear about theory, especially theory suggesting that certain factors and securities are mispriced.

"Investor psychology," which I feel is important, can give various kinds of irrational pricing. Mispricing may be consistent, as when a desire for yield causes people to bid up the prices of high-dividend stocks, or inconsistent, as when fads (like the belief that stocks of small firms are underpriced) sweep the market.

Reprinted from Financial Analysts Journal *(September/October 1993):36–38.*

If we are willing to use theory (and data other than past returns), we can estimate expected return without even looking at past returns. We can explain why people may care about long or short exposures to particular factors, and we can explain how investor psychology or institutional behavior may affect factor pricing.

DATA MINING

We have two ways to estimate a factor's expected return—*theory* and *data*. Both have problems.

Theory has problems because, for example, we may not know what portfolio to use to represent a factor. We don't even know what to use for the first factor.

In the U.S., the problem of not knowing what portfolio to use is not too severe (except, as I note later, it tends to flatten the line relating expected return and beta). In my view, all the candidates for the U.S. market portfolio are highly correlated.

The world market portfolio and the U.S. domestic market portfolio are highly correlated. Broader and narrower market indexes are highly correlated. Equally weighted and value-weighted portfolios are highly correlated. Even human capital and real estate are highly correlated (in my view) with portfolios of traded assets.

The problems with data, though, are severe, both because the high ratio of risk to expected return means we need decades of data to estimate the pricing of most factors, and because of *data mining*. We often don't have the decades of data we need. Even when we do, we can only estimate the average pricing of a factor over the whole period; if factor prices are changing, today's price may be far from that average.

In principle, conditioning factor prices on various observables can help. But unless we have strong theory to guide us, we need many more decades to estimate conditional factor prices than to estimate unconditional ones. Each new observable we add tends to weaken our estimates.

It's ironic. We seem to have so much data, with monthly and daily returns on thousands of securities. That does help in analyzing variance; for expected returns, however, it gives us false security and makes us think we're better off than we are. Normally, the best we can do with data alone is to create a portfolio out of traded securities that captures each factor and use the average return on that portfolio as an estimate of the expected return on the factor.

When people use Fama and MacBeth's methods instead of constructing a single portfolio that represents a factor, I think it's out of hope that so many cross-sectional regressions, with so many data points each, will somehow add precision to the estimates.[3] It's similar when people use elaborate multivariate tests. Normally, the loss of degrees of freedom offsets any potential gains, and we end up with worse estimates than the ones that simple portfolio tests give us.

And as if these problems weren't bad enough, we have the data mining problem. (Data mining is also known as "data snooping," "data dredging" or just "hindsight.")

The simplest kind of data mining is when a researcher gives a table filled with t-statistics and labels the big ones "significant" at the 5% level, when only 5% of the ones he gives fall in that range. Another kind is when a researcher fails to report everything he tried in analyzing his data. I'm thankful we have that kind. Without it, the "overpublication problem" would be even worse than it is now.

Yet another kind is when a researcher chooses what to do and how to do it in the light of what others have done using similar data. In a less formal version of this, a researcher designs his studies with the knowledge of past patterns of security returns. He knows these patterns because he reads newspapers and magazines, or because he invests.

All these forms of data mining are made worse by the huge number of miners, both academic and nonacademic. "There's gold in them thar hills," since people who find good ways to estimate expected returns can make a lot of money.

Moreover, all these kinds of data mining are worse when we try to estimate conditional means than when we estimate unconditional ones. We are mining when we choose the observables to condition on. So multivariate tests are especially suspect. (Even time is a suspicious variable.)

The result is that conventional tests of statistical significance are almost completely invalid (and multivariate tests are more invalid than simple ones). We don't know what area was mined, or what mining tools were used.

As I have done more theoretical work than empirical work, you may have anticipated my conclusion: I find theory to be far more powerful than data when we're trying to estimate expected return. When I read an empirical paper, I usually seek out the theory section and ignore the tables.

This means that most so-called "anomalies" don't seem anomalous to me at all. They seem like nuggets from a gold mine, found by one of the thousands of miners all over the world.

The "small-firm effect" or "size effect" may be the best example. There's actually some theory suggesting that such firms may be consistently underpriced, because not many analysts or investors follow them. But most academic researchers downplay this theory, because it implies mispricing rather than a rationally priced factor.

Still, it's a curious fact that just after the small-firm effect was announced, it seems to have vanished. What this sounds like is that people searched over thousands of rules until they found one that worked in the past. Then they reported it, as if past performance were indicative of future performance. As we might expect, in real-life, out-of-sample data, the rule didn't work any more.

WHAT THEORY SAYS

Now let's leave the messy world of data and return to the almost-clear waters of theory.

The strongest prediction of theory is that there is at least one priced factor. It's something like the market portfolio, though we don't know whether to use the world market portfolio or a domestic portfolio, or exactly what assets to include with what weights.

Even Fama and French agree that the market's expected excess return is positive. It's positive enough to make people willing to bear the risk of the first factor.

Can theory tell us what to use for the second factor? One theory can. What Fama and French call the "SLB model" identifies the "beta factor" as a second factor that should be priced. We can define the beta factor as the minimum-variance, zero-beta portfolio of risky assets, where beta is defined using whatever market portfolio we use to represent the first factor.

Curiously, Fama and French don't mention this factor at all. They talk only about a small-firm factor and a price-to-book factor. I think of these as less important in explaining variance than the beta factor and refer to them as the Fama-French "third" and "fourth" factors.

When Fama and French say that the line relating expected return and beta is flat, they are just saying that the expected excess return on the second factor is large. If we believe it's as large as they say, we won't fool around with their third and fourth factors, for which they give no theory. We'll go for the gold in the second factor!

A rational investor who believes the line is flat today should switch out of bonds (if he has any) and move toward low-beta stocks. If he doesn't have bonds, he should borrow (if he can) and do the same thing.

Moreover, a corporation that believes the line is flat can increase its stock price by emphasizing low-beta corporate assets and using lots of leverage. (This can also have tax advantages.)

If beta had been dead, the Fama-French results would have revived it!

But is the line flat today? Is it flatter than the CAPM suggests? Or is the flat line just another offering from the worldwide data mine?

I first wrote on this subject (with Jensen and Scholes) in 1972.[4] We found that the line relating average return (up to that time) and beta was flatter than the CAPM would predict, and we discussed some possible explanations for this result. But we were aware of certain patterns in stock market returns when we started our work, so we may have been mining the data ourselves, at least in part.

The years since 1972 have been "out of sample" in a sense. But I would not be writing this if the average excess return on the second factor had been zero over this 20-year period. Again, I may be mining the data by highlighting the return on the second factor over the last 20 years. So I think we must rely on theory.

THREE THEORIES

Mismeasuring the market portfolio (like using the domestic market when we should be using the world market) tends naturally to give stocks with low measured betas high alphas.

Imagine, for example, an extreme case where all stocks in our universe have true betas of 1.0 and have positive but varying amounts of a "noise factor" that's independent of the true market. (This also works if they all have negative amounts of the independent factor.) Imagine that we use the portfolio of all stocks in this universe for our measured market portfolio.

In this case, the true line is flat. A stock's alpha is proportional to the difference between 1.0 and its beta. Moreover, it pays for an investor to emphasize low-beta stocks from this universe, because that gives him less of the unpriced noise factor. And it pays for a corporation to use high leverage and emphasize low-beta assets, if it is restricted to assets in the same universe.

Another reason for a flatter line is restricted borrowing. Margin requirements, borrowing rates that are higher than lending rates, and limited deductibility of interest costs all tend to make the line flatter. Those who can't borrow at good rates bid up the prices of high-beta stocks instead.

Yet another reason for a flatter line, I believe, is investor psychology, in particular "reluctance to borrow" even when the rules allow it and the rates are good. Many people seem to dislike the idea of borrowing or the trading needed to adjust borrowing amounts to the values of their securities portfolios.

That makes three theoretical reasons for a flatter line in the future—"mismeasurement," "restrictions" and "reluctance." Plus we have some data, for anyone who thinks the data haven't been mined too much.

What do *you* think? That the line will be flat in the future? That it will be as steep as the CAPM says it should be? That it will be flatter, but not completely flat?[5]

FOOTNOTES

1. E.F. Fama and K.R. French, "The Cross-Section of Expected Stock Returns," *Journal of Financial Studies* 47 (1992):427–65.
2. R.C. Merton, "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41 (1973):867–87.
3. E.F. Fama and J.D. MacBeth, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy* 81 (1973):607–36.
4. F. Black, M. Jensen, and M. Scholes, "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets*, M.C. Jensen, ed. (New York: Praeger, 1972):79–121.
5. This article is based on a talk prepared for the September 1992 Berkeley Program in Finance, *Are Betas Irrelevant? Evidence and Implications for Asset Management*. I am grateful for comments by participants in that program.