RISK AND RETURN

. . .

John H. Cochrane and Tobias J. Moskowitz

While early efficient market work could start with the working hypothesis that expected returns are constant over time, the need for risk adjustment and a "model of market equilibrium" is immediately apparent in the cross-section. There are stocks whose average returns are greater than those of other stocks. But are the high-average-return assets really riskier in the ways described by asset pricing models?

This empirical work is not easy. It took lots of thought and creativity between writing down the theory and evaluating it in the data. The four papers in this section are not only a capsule of how understanding of the facts developed. They more deeply show how Gene, alone, with James MacBeth first, and with Ken French later, shaped how we *do* empirical work in finance.

FAMA AND MACBETH

The CAPM, the subject of Fama and MacBeth's famous paper, states that average returns should be proportional to betas,

$$E(R^{ei}) = \beta_i E(R^{em}) + \alpha_j \tag{1}$$

where the betas are defined from the time-series regression

$$R_t^{ei} = \alpha_i + \beta_i R_t^{em} + \varepsilon_t^i,\ t = 1, 2, \dots T. \tag{2}$$

Here, R_t^{ei} is the excess return of any asset or portfolio i, and R_t^{em} is the excess return of the market portfolio.

You run regression (2) first, over time for each i to measure the betas. Then the CAPM relationship (1) says average returns, across i, should be proportional to the betas, with "alphas" as the error term (i.e., the average returns not explained by the model). Though unconventional, we write the alphas as the last term in (1) to emphasize that they are errors to the relationship.

Now, an empiricist faces many choices. First, one could apply this model to data by just running the time-series regression (2). The mean market premium, the betas, and the alphas are all then estimated, and one can see if the alphas are small.

Fama and MacBeth didn't do that. They estimated the cross-section (1) in a second stage. Why? For many good reasons. First, they wanted, we think, to get past formal estimation and generic testing to see how the model behaves in all sorts of intuitive ways. Sure, all models are models, as Gene frequently reminds us, and all models are false. But even if a glass is statistically 5% empty, we want to really understand the 95%.

The paper tries a nonlinear term in beta in (1), idiosyncratic variances, an intercept, and so forth. These are natural explorations of ways that the model might plausibly be wrong. To explore them, we need to run the cross-sectional regression.

Second, betas are poorly estimated, and betas may vary over time. Fama and MacBeth use portfolios to estimate betas, which reduces estimation error. Portfolios, however, also reduce information, in this case cross-sectional dispersion in betas. (It's like measuring your income by measuring the average income on your block.) Fama and MacBeth used portfolios that maintained dispersion in true betas while also reducing estimation error. They sorted stocks into portfolios based on the stocks' historical betas, but then used the subsequent, post-ranking portfolio beta in the regression equation (2). As is typical with Gene's papers, there is a deep understanding of a complicated problem, solved in a simple yet clever way.

The cross-sectional regression framework easily allows one to use rolling regressions (2) so betas can vary over time.

Forty-five years of econometrics later, we know how to estimate models with time-varying betas, and we know other ways to handle the errors-in-variables problem too. But the transparency and simplicity of the Fama-MacBeth approach still reigns.

The use of portfolios itself is a foundational choice in empirical work. In this paper, you see tables with 1 to 10 marching across the top. That isn't in the theory, which just talks about generic assets. And it isn't in formal econometrics either. (Formal econometrics might call it nonparametric estimation with an inefficient kernel.) Yet this is how we all do empirical finance. It is useful and intuitive.

Most famously, Fama and MacBeth dealt with the problem that the errors are correlated across assets. If Ford has an unusually good stock return this month, it's likely GM has one too. Therefore, the usual formulas for standard errors and tests, which assume observations are uncorrelated across companies as well as over time, are wrong.

This was 1970, before the modern formulas for panel data regressions were invented. Fama and MacBeth found a brilliant way around it, by running a cross-sectional regression at each time period and using the in-sample time-variation of the cross-sectional regression coefficients to compute standard errors. In doing so, they allow for arbitrary cross-correlation of the errors. That this procedure remains in widespread use, despite the existence of econometric formulas that can deal with the problem—sometimes successfully, and sometimes not—is a testament to how brilliant the technique was.

More deeply, Fama and MacBeth's approach to cross-correlation was not to adopt GLS or other statistically "efficient" procedures, which every econometrics textbook of the day and up to just a few years ago would advocate, but instead to run robust, reliable OLS regressions and compute corrected standard errors. That practice has since spread far and wide.

Finally, the Fama and MacBeth procedure has a clever portfolio interpretation. The coefficient in the regression of returns on betas represents the return to a portfolio that has zero weight, unit exposure to beta, and is minimum variance among all such portfolios that satisfy the first two constraints. This description is in essence the market portfolio. Hence, an average of this portfolio's returns is an estimate of the market risk premium. Since returns are close to uncorrelated over time, this interpretation justifies the standard error of that mean return as the standard error of the Fama-MacBeth regression coefficient.

This beautiful insight would allow future researchers to look at other characteristics and other betas in the same way: the coefficients associated with other characteristics (e.g., size or book-to-market ratios) or betas on the right-hand side of the regression are minimum-variance returns of zero cost portfolios with unit exposure to the characteristics and zero exposure to all other variables on the right-hand side.

As with the practice of forming 1–10 test portfolios of assets, this insight allowing researchers to easily translate regressions into portfolios and vice versa would spawn a host of empirical facts and models used in academia and practice.

So, while theorists may think of empirical work as easy and a task for lesser minds, here you see Fama and MacBeth dealing with hard issues of how you do empirical work and interpreting empirical results within the confines of theory. The Fama-MacBeth procedure set a pattern that lasted a generation. We still have 10 portfolios marching across the page, we still compute Fama-MacBeth regressions, and we still use those insights to build efficient portfolios.

Fama and MacBeth's influence was so strong, it extends to the omissions. They refer to linearity of the cross-sectional relation, the statistical significance of the market premium, and the absence of other explanatory variables as "tests" of the model. The actual "test" of the model is whether the alphas are jointly zero. They didn't do that test because it hadn't been invented yet. Curiously, though the Gibbons-Ross-Shanken (GRS) test for joint significance of the alphas was developed for time-series regressions and has come into use, the equally easy (in retrospect) cross-sectional version of the GRS test has never, as far as we know, been used. (Just construct a Fama-MacBeth covariance matrix of the errors $\text{cov}(\hat{\alpha}, \hat{\alpha}') = \text{cov}(\varepsilon, \varepsilon')/T$. Then $\hat{\alpha}' \text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1} \hat{\alpha}'$ has a χ^2 distribution. Or use GMM, which gives corrections for estimated betas and autocorrelated residuals. Or bootstrap it.) Well, we follow Fama and MacBeth. And the lack of methodological innovation is understandable. When finding new results, one wants to make sure they come from the data, not the method. So method that was innovative in its day, and is transparent, familiar, robust, and good enough now, survives.

THE CROSS-SECTION OF RETURNS

We include three papers on the cross-section of returns, with its primary workhorse the value premium. They represent a remarkable intellectual journey. Gene was both the Newton and Einstein of finance, presiding over the foundation of the field and development of the CAPM, and then presiding over the second revolution, the incorporation and amalgamation of a plethora of anomalies and the emergence of multifactor models, in this case with Ken French.

The CAPM reigned supreme for about a decade after Fama and MacBeth's article was published. Time and again, someone would come up with a clever technique that seemed to make money, and time and again, when examined carefully, either the average returns came from mismeasurement, overfitting, or survivor/selection bias, or the average returns corresponded to a higher beta, so the profits were just as easily made by investing in the market index.

Yet, starting with the size effect in the late 1970s, more and more cross-sectional anomalies cropped up. These are methods for finding securities which have high average returns, but those returns do not correspond to higher betas. Or, less often, they are techniques for finding securities with low betas that do not have low average returns.

"The Cross-Section of Expected Stock Returns" was a bombshell, for it announced Fama and French's certification, after combing through the entrails of the data, that indeed the CAPM fails. On reexamination—where they painstakingly and cleverly try to maximize the information they glean from the data while simultaneously minimizing the noise/error in the data using what became the standard for estimating betas—they find no association between beta and average returns.

Second, Fama and French dug deeply through the trove of expected-return signals and found that size and book-to-market ratio alone captured the information about expected returns from a plethora of signals.

The exercise is in many ways a multiple regression question. Expected returns across assets depend on a vector of forecasting characteristics. Many of those characteristics are significant return forecasters taken one at a time. But in a multiple regression sense, size and book-to-market ratio encompass the information in the other forecasting characteristics.

This paper nicely connects two ways to understand expected returns as a function of characteristics. One can look at the mean returns of portfolios sorted on the characteristic, or one can run cross-sectional forecasting regressions,

$$R^{ei}_{t+1} = a + b' C^{i}_{t} + \varepsilon^{i}_{t+1}$$

where C denotes a vector of characteristics such as size, book-to-market ratio, or beta. Average returns on a portfolio sorted on the basis of C are no more or less than nonparametric estimates of a nonlinear version of this regression—but one that is very simple and intuitive. Portfolio sorts also assuage the worry that regressions of this sort are driven by outliers—a few extreme values of C that happen to have extreme returns.

The way of thinking about asset returns in these papers carries a deeply influential innovation for how we do empirical work. In looking at the theory, you often think of "asset i" as referring to, well, an asset: a stock or bond. In these papers, Fama and French exploit the idea that average returns and betas attach in a stable and strong way to a set of firm *characteristics*, but not to the firm itself. Expected returns and, later, betas, are stable functions of size, book-to-market ratio, and other characteristics. But an individual firm's expected returns and betas vary over time as the firm's characteristics vary, so these statistics are not a stable function of firm name.

That expected returns, betas, and other statistics are stable functions of characteristics, not firm name, is an auxiliary observation about the data. Nothing about this stability is present in the theory. But this auxiliary assumption seems true of the world and makes asset pricing much more interesting and productive. Since characteristics wander over time, there just isn't that much variation in expected returns or betas across firm names. We see that variation as a function of the characteristics, which have a real-world interpretation as portfolio strategies. This procedure also unites "signals," "managed portfolios," and "assets" as just instances of the same thing.

COMMON RISK FACTORS

"Common Risk Factors in the Returns on Stocks and Bonds" took what is, with ex post hindsight, the next and obvious step. To say expected returns are a function of two characteristics, size and book-to-market ratio, is a fine *description* of average returns, but it cannot stand as an *explanation* of average returns, at least not an explanation of any vaguely "rational" sort. For example, if average returns really are a function of the size of the company, we only have to buy a portfolio of small companies, paying high average returns, and finance the purchase by issuing stock of what is now a big company, paying low average returns. Then we retire rich off the difference. "Explanations" must be betas, which are invariant to portfolio formation.

Fama and French then found two "factors" in the covariance matrix of returns, related to size and book-to-market ratio, and found that the expected returns on size and book-to-market sorted portfolios line up beautifully with betas on these two factors, plus a beta of one on the market portfolio. That insight seems obvious, but it really wasn't. There was no guarantee that the covariance structure of the assets would be captured so easily by these two factors. Each size and book-to-market portfolio could have had its own variance devoid of any common structure. But they didn't. The portfolios' returns were tied together by two common sources of variation that were identified by grouping securities based on the two characteristics: size and book-to-market ratio.

With the advantage of hindsight, you can do the same thing by finding the first three principal components of the covariance matrix of the 25 portfolios. You will find those first three principal components explain the vast bulk of co-movement of the 25 portfolios, as reflected in Fama and French's 90–95% R^2 values. The first three principal components are also clearly a "market" portfolio, one that loads on big minus small portfolios, and another that loads on high minus low book-to-market portfolios, as Fama and French's factors do.

Then, you will find that expected returns are almost completely explained by betas on the three principal components. Since the betas or loadings on the "market" factor are all one, though, variation in market betas across portfolios does nothing to explain the cross-sectional variation in average returns.

Viewed this way, the Fama-French model is an arbitrage pricing theory (APT) model, and it certainly is that at least. What Fama and French did *not* do is interesting in that context. Arbitrage pricing models have been around a long time, and they usually met limited success. The typical approach was to factor analyze the covariance matrix of individual stock returns, and then to see if large factors are important drivers of mean returns. But loadings (betas) on factors so derived never did much to explain the cross-section of average returns. Fama and French *first* formed portfolios on the basis of characteristics known to describe average returns, and *then* found the factors that dominate the covariance of returns. Doing so, they made an APT work nicely. At last.

Fama and French also did not try to build some fundamental asset pricing, starting with consumption or state variables for investment opportunities. In retrospect, we see a natural hierarchy for empirical work: First, find how average returns vary with characteristics such as size and book-to-market ratio. Second, see if there is a factor model based on the same characteristics which explains the average returns. Third, see if more "fundamental" factors such as consumption growth or macroeconomic state variables can explain the risk premiums of the empirically derived factors such as Fama and French's HML and SMB.

Seen this way, Fama and French's three-factor model is a remarkable data summary device. The 25 portfolios capture the spread in average returns across thousands of stocks, using book-to-market and size signals. Loadings on the three factors then explain the average returns of the 25 portfolios. As a result, more fundamental approaches need only explain the three factor risk premiums.

As Fama and French emphasize, the three factors should be proxies for something deeper, such as consumption, marginal utility, "state variables of concern to investors," and so forth. Figuring out what those deeper factors are remains a challenge. That challenge has occupied the attention of academic researchers for the better part of three decades now. But by summarizing all the information in stock markets down to three factors, that challenge is enormously easier for theorists.

Interestingly, most authors seem to have missed this point. Most authors test more fundamental models by pricing the 25 portfolios, or they try to see if macro factors drive out, rather than explain, the three Fama and French factors. (And if authors don't do it, referees demand it!) Blindly copying Fama and French's method for this different purpose misses their underlying point. Fama and French did it so you don't have to!

Factor structure itself is a vital point in evaluating book-to-market ratio, size, and other anomalies. It is easy to build stories about why a class of securities should have prices that are too high or too low, and consequently average returns that are subsequently too low or too high. But why in the world should the underpriced securities all move up or down together the next year? Why should they share strong exposures to some risk?

Well, arbitrage. If not, you could earn a fortune holding a diversified portfolio of value stocks. But that means somebody is thinking about the means and variances of diversified portfolios and has driven prices pretty close to the "rational" point where one must hold undiversifiable risk to earn positive returns.

It seems easy in retrospect. It was not. One of us (Cochrane) was there, and I can attest to the fact. I was thinking about cross-sectional versions of dividend-yield forecasts; I was thinking about beta models to explain it. Nothing like what Fama and French did occurred to me. When I first saw the three-factor model, I asked questions that, if anyone remembered, would go down in the history of stupid seminar questions. Isn't it a tautology to "explain" 25 book-to-market portfolios with three book-to-market factors? No. Their genius came in seeing at the time that this was the most natural and simple thing to do. And once again, Fama and French set the stage. Now everyone sorts portfolios and creates factor portfolios to "explain" expected returns.

MULTIFACTOR EXPLANATIONS

"Multifactor Explanations of Asset Pricing Anomalies" is not as famous as the other three papers, but it should be, in our opinion. It is another example of a paper, later in a series, which explains the basic concepts more clearly than earlier papers, without the forest of robustness tests that early papers must have. It is a good paper to recommend that students read first.

Table 1A–B of "Multifactor Explanations" succinctly distills the three-factor model. Panel A shows a strong pattern of average returns across size and book-to-market dimensions, a description in want of an explanation. Panel B shows how the variation across portfolios in betas (b, h, s) on the three factors lines up with the variation in expected returns.

The point of the table is that variation across portfolios in the b, h, s corresponds to the variation across portfolios in average returns shown in Panel A.

Thus, you should read this as a table of *data* for an implicit cross-sectional regression, of 25 average returns (Panel A) on slopes (Panel B). The intercepts (a) of the time-series regression are the errors in this cross-sectional relationship.

The regression in Table 1 has a secondary direct interpretation. The regression and its R^2 tell you how much movements in the factors account for movements in the portfolio returns. The regression and high R^2 tell you that the three-factor model is a good model of return *variance*. They tell you that most of the actual, ex post, returns of the 25 portfolios can be attributed to the actual, ex post, return of the three factors. The size of the alphas, the pattern of betas, and the implied cross-sectional regression tell you the more important fact that this is a good model of *means*.

Read carefully. When Fama and French say this is a good model of "returns and average returns," they repeat "return" for a reason. A good model of "returns" is a good factor model—high R^2 for the first few principal components. It would be a good model no matter how large the intercepts. A good model of "average returns" is one in which mean returns vary a lot across portfolios, but betas vary in the same way as the means, and the intercepts are small. The R^2 is irrelevant to this point. The Fama-French model has both small alphas and large R^2, which makes it both a good model of "returns" and of "average returns."

Table 1 also summarizes a sea change in empirical procedures that occurred in asset pricing, as well as macroeconomics, in the prior 20 years, with Fama alone and with French playing a leading role. In the late 1970s or early 1980s, people wanted to "test" models. Where is the "test" of the Fama-French model? There is one, and only one, such test: whether the 25 alphas are jointly equal to zero. This is the Gibbons-Ross-Shanken test with normal iid returns in time-series regressions, or the GMM overidentifying restrictions test using pricing errors as moment conditions more generally. As reported by Fama and French, that test blows the model out of the water. Fama and French statistically reject the hypothesis that all the alphas are zero at astronomical levels of significance.

How can it be that this, the most successful asset pricing model of a quarter-century at least, is overwhelmingly statistically rejected? What is the rest of Table 1 even doing there if the model is rejected?

Well, formal rejection is no longer that interesting. All models are rejected if you have enough data. The hypothesis that this model is literally true is just not interesting. As Fama and French point out, the residuals are so small that economically small alphas are statistically different from zero. So the model is not 100% true. But it's 95% true!

So the paper proceeds by showing you the 95% that is true: how average returns vary a lot across portfolios; how betas nicely and smoothly vary in the same way; how alphas are by and large an order of magnitude smaller than the average returns; and so forth. That's how we evaluate models now. The focus on "testing" and "not rejecting" led to a lot of models with much larger alphas, but standard errors larger still, so we couldn't statistically reject that the alphas were zero. Or it led us to "rejecting" good models that explained a lot of data, but could be shown not to be 100% perfect.

"Multifactor Explanations" goes on to explain just how *useful* the three-factor model is. Practical usefulness, rather than great theoretical advance, accounts for the astonishing impact of the three-factor model. The point of the CAPM really never was to settle barroom bets about "rationality" or "irrationality." The point of the CAPM was practical—it gave a procedure for quickly and reliably risk-adjusting new findings. If you have a new clever idea for making money, you want to know, is this really something new, or just a way of getting exposure to a known risk? If the higher average return of a new idea corresponds to a higher beta, with no extra alpha, you find the new idea is no better than just investing more in the market index.

The big payoff of the three-factor model is the same sort of practical utility. You find some new procedure for isolating good returns, some new sort or forecasting variable. Is this just a way of buying value stocks, or, more deeply, buying stocks that *behave like* value stocks, and thus don't give any better performance in a portfolio that already includes value stocks? That's a vital question for practice. It lives quite apart from a deep battle over whether value itself represents macroeconomic risk premiums, "distress," or some collective irrationality. Whatever value is, when I look at something new, I want to take out the known value premium. That's what the CAPM was *useful* for, and that's what the three-factor model is useful for.

In addition, the best way to answer the "tautology" charge is to take the three-factor model out for a spin. If you're still worried about explaining value with value, well, let's explain *other* anomalies with value. For both reasons, the heart of the paper is, as the title suggests, showing how the multifactor model addresses other "multifactor anomalies."

To our mind, the sales growth tables are a shining example. Buying stocks of companies with five years of awful sales turns out to give a lot better return than buying stocks of companies whose sales are growing quickly. Apparently, the great sales growth is "priced in" to the stock. Well, this pattern might just be beta—companies with poor sales are going to go down the tubes in the next downturn, no? Well, no, at least as measured by market beta. But HML betas *do* fully explain the sales growth anomaly. The sales losing firms may not *be* value firms, but they *act like* value firms, and they give you no better performance in a portfolio that already includes value.

This paper is also great for showing the practical limits of the model. Momentum is a bust for the three-factor model—momentum average returns go the *opposite way* from value betas. Momentum portfolios can also be "explained" by a momentum factor, but Fama and French shied away from this specification. They didn't want to certify that every anomaly gets a factor. They now provide a momentum factor, UMD, on Ken French's webpage, and they use it for performance evaluation. But they are still reluctant to add it to their view of risk-based factors.

The abysmally low returns of small growth stocks are also a failure of the model. They account in large part for the statistical rejection, and the fact that characteristics are still a better description (but not explanation) of average returns than Fama-French factor betas. To our minds, they are an interesting anomaly awaiting dissection, potentially related to the firm birth-and-death process alluded to in Dennis Carlton's essay, or the fact that much information trading takes place in these mostly new and dynamic companies.

With these momentous papers, Fama and French put the anomaly zoo of the mid-1990s back in the bottle. Their solution was evolutionary, not revolutionary: yes, the CAPM fails. But one look at its assumptions and you expect it to fail. Multiple factors, long anticipated by theory, finally came to life in their empirical hands. By using three factors, just as you would use the CAPM, you can account for the known anomalies except momentum, and you can perform workaday risk adjustment, portfolio evaluation, and anomaly digestion. Compared to calls to throw out all asset pricing and start from scratch with psychology in place of economics, it is a remarkably conservative solution.

Anomalies have broken out again, however. Momentum did not go away. Now there are literally hundreds of claimed additional variables that describe expected returns, in ways that neither size and value characteristics nor size and value betas can account for. The second Fama-French step, finding additional factors, is slowly emerging. Many of the new return-forecasting variables seem to correspond to factors. For example, 10 momentum-sorted portfolio returns are neatly "explained" by a single winner-minus-loser factor. But adding hundreds of new factors is not a satisfactory approach. It's time to do once again what Fama and French did here, to put some order into the emerging chaos. But as happened last time, current off-the-shelf techniques, including Fama and French's, cannot handle the current empirical situation. We have tens or hundreds of right-hand variables, not two or three. It will take the kind of profound, simplifying insight and profound, simplifying innovation in technique that the Fama and French papers showed to put order back in the empirical asset-pricing universe once again.