

7 New Anomalies

This set of notes covers “Dissecting anomalies,” Novy-Marx “Gross Profitability Premium,” Fama and French “Five factor model” and Frazzini et al. “Betting against beta.”

7.1 Big picture: Three rounds of anomalies and factor models.

We have been through three waves. In the beginning there was chaos. Then came the CAPM. Many variables showed the way to higher expected returns. But expected returns always were explained by betas,

$$E(R^{ei}) = \beta_i E(R^{em})$$

Then came the capm anomalies, most prominently value. Here was an $E(R^{ei})$ definitely *not* matched by β_i – beta went the wrong way and there were big alphas. There were many other anomalies too, like sales growth and long term reversal.

Along came Fama and French, and once again put the zoo of anomalies in order. First, there is strong R^2 in the 3 factor time series regressions, so *hml* and *smb* accounted for the 25 portfolio returns and average returns.

$$R_t^{ei} = \alpha_i + b_i r_{mr} f_t + h_i hml_t + s_i smb_t + \varepsilon_t^i \leftarrow \text{high } R^2$$

More importantly, lots of other variables gave spreads in $E(R^{ei})$, but three factor betas accounted for them,

$$E(R^{ei}) = \alpha_i + b_i E(r_{mr} f) + h_i E(hml) + s_i E(smb) \leftarrow \text{low } \alpha$$

FF hoped momentum would go away, but it didn’t, so the standard factor model became the four-factor model,

$$R_t^{ei} = \alpha_i + b_i r_{mr} f_t + h_i hml_t + s_i smb_t + u_i umd_t + \varepsilon_t^i$$
$$E(R^{ei}) = \alpha_i + b_i E(r_{mr} f) + h_i E(hml) + s_i E(smb) + u_i E(umd)$$

One could argue whether the factors themselves were “rational” – why don’t more people buy value stocks? Momentum stocks? But *given* these anomalies most of the other anomalies seemed to fall in line. Four factors may seem like a lot, but they organize 20 or more anomalies.

And once again the world fell apart. A whole range of new anomalies cropped up, $E(R^{ei})$ strategies that were not explained even by four-factor betas. The world fell apart under our feet again.

7.2 Identities, and earnings forecasts to forecast returns.

Many of these anomalies make sense in the context of our identity,

$$pd_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

Cross sectionally, we don’t use price to dividend, because dividends can be zero. But market/book = price \times shares / book has much the same flavor – it’s all about price over something. Fama and

French and Novy Marx use the related identity

$$\frac{M_t}{B_t} = E_t \sum_{j=1}^{\infty} \frac{Y_{t+j} - dB_{t+j}}{(1+r)^j B_t}$$

where Y = earnings and B = book value. It works the same way – a higher market value comes from higher expected earnings or lower expected returns. (See FF p. 3, this equation makes “three statements about expected stock returns...” and moves each of Y, dB or r in turn.

Remember the lovely interpretation: pd forecasts returns because it *reveals to us* market expectations. If expected returns go up (some signal the market sees and we do not) then p goes down. We see that, and, on average, the higher subsequent returns.

The trouble is, people in the market also have information about cashflows, especially across companies. News of bad Δd will also lower pd . So actual pd is a mixture of the two.

So, *if we had variables that helped to forecast dividends, earnings, etc, those variables would help us to clean up pd , to isolate movements in pd that are just due to expected returns.* That’s a good intuition why

$$R_{t+1} = a + b pd_t + cz_t + \varepsilon_{t+1}$$

will work when z_t is an earning forecast variable. Adding z will make pd a better forecaster.

A second way to see the point. If you find a variable z that helps to forecast cashflows, you are open to the “good stock - good company” fallacy. The market already priced in those great cashflows. But if you hold price constant, and find companies that offer good cashflows without a higher price, you avoid this fallacy. Again, z should work in a multiple regression with z and pd .

Third, the identity seems to pose a puzzle. If we see pd , how can we raise expected returns? Answer, if we also raise expected dividend growth. Thus, holding pd constant, variables can (and must) raise expected returns if and only if they raise expected dividend growth.

In fact, for an extra variable z_t to help forecast returns, it *must* either also forecast dividend growth, or it must help to forecast longer term returns. In “discount rates” I found cay helped to forecast near term returns by changing the forecast of long term returns. But the potential to find variables that forecast returns and cashflows is there.

So all this motivates a search, *let’s look for variables that might forecast cashflows, and see if they help us to forecast returns as well.*

This logic helps to make sense of many of the forecasting variables. Novy Marx gross profits is explicitly there as a measure of expected future profitability. He thinks gross profits are better than the net profits, which FF found did not work, because they do a better job of forecasting future profits. If you’re investing a lot today, today’s profits may stink, but that’s a good sign of future profits. Accruals, asset growth, and the range of “dissecting anomalies” variables all make sense this way.

Investment and net stock issues follow an even clearer logic. When do firms invest a lot, and issue stock to fund it? When they think future profits will be higher. Actually there is a deeper theory of investment: Companies should invest more when expected profits are high, or when the cost of capital is low. In fact, the standard “q” model of investment says investment should be exactly proportional to the market to book ratio (which economists call Q)

$$\frac{I}{K} = f\left(\frac{M}{B}\right)$$

In this simple view, I/K shouldn't help M/B at all. But B and K are hard to measure, so I/K might give additional information.

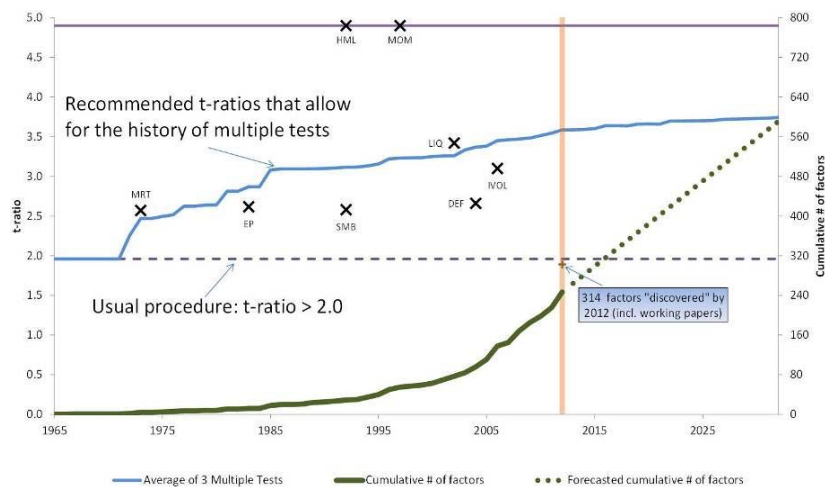
In sum, if the cost of capital or expected return is low, companies should invest more and issue equity. (You might get behavioral and say managers issue equity when they think the market is overvalued and vice versa, but you don't have to – the rational story works just as well.) If expected profits are high, companies should invest more and issue equity, and holding price constant, this can help us to see the manager's evaluation of expected returns and cost of capital.

At this stage we're not going to get too deep in these stories about how pd_t or M_t/B_t are formed. I think it is a vital research topic, and "Discount rates" goes on about how we should do "asset pricing," and make pd_t or M_t/B_t the variable we're trying to explain. But not today. Today, take these stories as motivations for a search, that using earnings forecast type variables might help us to find strategies that yield high average returns.

7.3 Lots of variables, two questions

This search produced lots and lots of return forecasting variables. Interestingly many of these came from the accounting literature, because accountants know how to read the earnings-forecast parts of the compustat.

Here are two cool plots that chart the literature and count up the number of variables that have been found, one at a time, to help to forecast returns. There are 300 such variables and climbing!

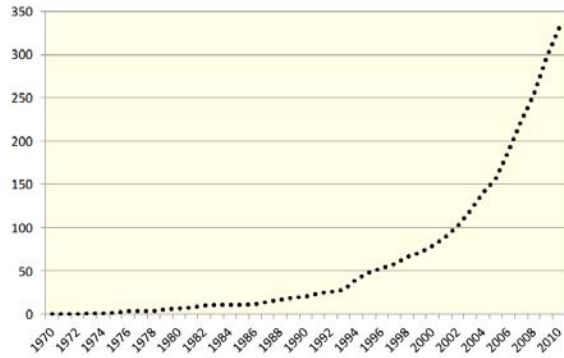


Source: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2249314

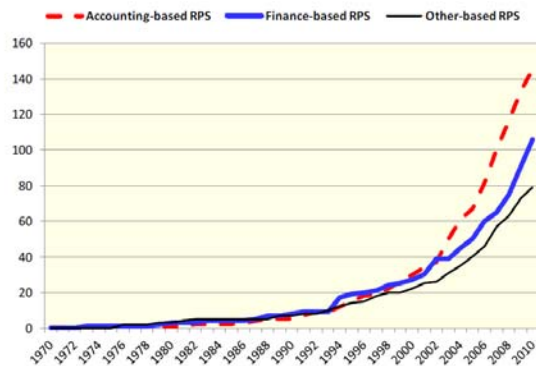
FIGURE 1

Cumulative number of firm-specific return predictive signals (RPS) publicly reported by accounting, finance and other business academics, 1970-2010.

Panel A: Cumulative number of total RPS



Panel B: Cumulative RPS by type (accounting-based, finance-based, other-based)



An accounting-based RPS is one where the signal is in the firm's financial statements (e.g., accruals; cash flows; assets). A finance-based RPS is one where the signal was directly or indirectly dependent on the firm's stock price (e.g., return momentum; implied skewness in returns derived from option prices). The exceptions to these rules are that P/E is classified as an accounting signal, book-to-market is classified as a finance signal, and any accounting signal interacted with a finance signal is classified as a finance signal. Other-based signals are those that are neither accounting nor finance, such as labor mobility, stock repurchases, or a stock's ticker symbol.

Source: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2262374 Green, Hand and Zhang

The point of the papers: Surely some of these magic 2.5 t statistics in the 1-10 portfolio are a bit overstated, no? But that leaves us the question, just which of these variables matters? Also, this literature tries them one at a time. Which subsume the other's information?

So, let's sum up the questions that Fama and French, Novy Marx, etc. are asking

1. What additional variables reliably forecast returns, i.e. produce a spread $E(R^{ei})$ across stocks, that is not accounted for by known variables (size, book/market, past return), or explained by known factor betas?
2. Which of those variables produce spreads in large stocks, not just in the micro caps that are hard to trade?

3. Multiple regressions? Which new sorting variables forecast returns in the presence of the other new forecasting variables? How many new variables do we really need, and how many are just different versions of known ones?

This is really where “dissecting anomalies” stops. There are no betas in “dissecting anomalies!” We need to go on and ask the betas question

1. Do new $E(R^{ei})$ sorts correspond to existing and known betas? (Usually no, but it’s worth checking.)
2. Do 1-10 $E(R^{ei})$ portfolios using a new variable correspond to betas on a new factor constructed from a long-short position in the new variable? (Almost always yes, but it’s worth checking.)
3. Do we really need a new factor for every new variable? Is there please, please, a way to get back to the CAPM or Fama French world that the number of factors is smaller than the number of sorting variables – that of the 300 (!) sorting variables, and the maybe 50 that survive multiple regression analysis, can have expected returns that correspond to a smaller number of betas?

Where we stand: The 2000’s were the era of collecting anomaly variables – new $E(R^{ei})$ sorts. “Dissecting anomalies” shows some of the effort to see which of them really are important given the others, and which are not just dusty corners of microcaps. Finally, with Novy-Marx paper and the FF Five factor model, we are seeing an effort to see if we can find a smaller number of new factors than the number of new expected return variables.

In particular, pay a lot of attention to Novy Marx Table 10 Here he examines how well his model with an earnings factor can account for a range of other sorting variables. Yahoo! OK, many of the sorting variables are other versions of earnings sorts, and many of the others are pretty marginal. Still, this is the first effort to put order back in the zoo that I have seen.

Similarly, Fama and French add two factors to correspond to two new sorting variables, investment and profits. Ugh. But Table 6 implements the test you worked on in a problem set, to see if they can drop any factors. They find that they really don’t need value in the presence of their profitability and investment factors. Hooray!

As you can see, this is all very preliminary. We’re struggling with the forecast variable and factor zoo. It’s the night before Fama and French created size and book to market factors all over again.

7.4 Issues

Our readings reveal lots and lots of hard issues and questions.

7.4.1 Why do I want fewer factors anyway?

This is a good question that came up in class. What’s wrong with a world with 27 dimensions of expected returns and 27 factors?

The basic answer is, nothing really. Still, the world would be a lot prettier if we could reduce 27 dimensions of mean returns to fewer factor betas!

How many factors we want will depend on how we want to use the model. There are factor returns, like industry or *smb*, which help to describe the *variance* and covariance of returns. And there are “pricing factors” which are necessary to understand *expected* returns

Our theory investigation will show that there is always a one factor model. In Deep Theory, There should be an underlying single factor, consumption or macroeconomic risk,

$$E(R_{t+1}^{ei}) = \gamma \times cov(R_{t+1}^{ei}, \Delta c_{t+1}) = \beta_{i\Delta c} \lambda_{\Delta c}$$

The many *hml* *smb* etc. factors show up because they are correlated with consumption growth. If you want to understand why *hml* gets its average return, or debate economic rationality, or think about how you should invest in the *factor* portfolios *hml*, *smb*, etc., you need this kind of model.

Equivalently, we can always write any factor model in terms of a mean-variance frontier. There is a return, a portfolio of *rmrf*, *hml*, *smb*, *umd*, etc. etc. whose return is mean-variance efficient. Then we can express

$$E(R^{ei}) = \beta_{i,mv} E(R^{emv})$$

So we want to know, what is this mean-variance efficient portfolio? That’s interesting for investment purposes as well.

For practical application, however, you more often want to know whether a new anomaly or factor is just a way of gaining exposure to something you understand. For this purpose a larger number of factors, with less economic rationale, is ok. Still, it would be nice if we didn’t have 50 different kinds of risk exposure to consider.

7.4.2 Mean factors and variance factors.

Recall our problem set, in which we found that *smb* was not really necessary to explain average returns – it didn’t really help to change alpha much in the model of *average* returns

$$E(R^{ei}) = \alpha_i + b_i E(rmrf) + h_i E(hml) + s_i E(smb)$$

but it did help a lot to explain return variance – it helped to raise R^2 in

$$R_t^{ei} = \alpha_i + b_i rmrf_t + h_i hml_t + s_i smb_t + \varepsilon_i^t.$$

I gave an example in which industry portfolios might behave the same way for the CAPM. More directly, in the problem set you constructed an ex-post mean-variance efficient portfolio which gave the same α – and disastrously lower R^2 . The consumption model above, even if it worked, would likely produce very small R^2 in

$$R_t^{ei} = a_i + \beta_i \Delta c_t + \varepsilon_{it}$$

We concluded that you might want to keep more factors around for “variance purposes” than for “mean purposes.”

“Variance purposes” can include a lot of “mean purposes.” Suppose you want to benchmark managers, and they have varying s_i loadings. Although in the true long run you might know $E(smb) = 0$, in a 3 year sample *smb* might have had a good return, so the managers with higher s_i loadings did better. If you leave *smb* out of your factor model, you will attribute the great performance of the high s_i manager to alpha! Sure, it might have a big standard error too, but your point estimates will lead you astray.

In the end, we are searching for similar behavior in the N factor universe. Fama and French end their paper by finding that HML can be dropped for mean-return purposes. They ran the test that you explored in the problem set,

$$hml_t = \alpha_h + b_h r_{mr} f_t + s_h smb_t + \dots + \varepsilon_{ht}$$

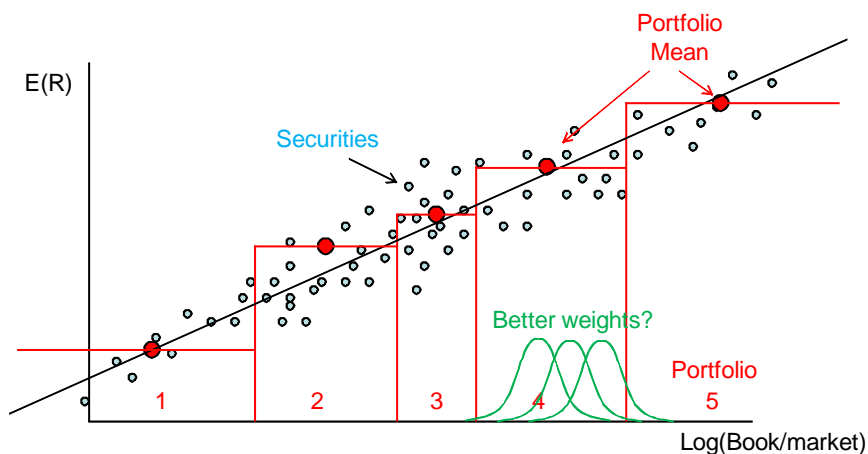
and found $\alpha_h = 0$. But for “variance purposes” they suggest keeping hml anyway.

Finally, HML seems to be a redundant factor in the sense that its high average return is fully captured by its exposures to RM - RF, SMB, and especially RMW and CMA. Thus, in applications where the sole interest is abnormal returns (regression intercepts), our tests suggest that a four-factor model that drops HML performs as well as (no better and no worse than) the five-factor model. But if one is also interested in measuring portfolio tilts toward value, profitability, and investment, the five-factor model is the choice.

7.4.3 Regressions and sorts

“Multifactor explanations” sorts stocks by time t information. In “dissecting anomalies” and Novy-Marx, we meet Fama MacBeth regressions. A lot of “dissecting anomalies” is devoted to thinking hard about the differences between the two techniques.

This graph from “discount rates” should make one point clear:



Sorted portfolios and cross-sectional regressions.

Sorting stocks into portfolios based on book/market ratio at time t, and watching average returns in the next year, provides the same kind of information as running a forecasting regression

$$R_{t+1}^{ei} = a + b \log(B/M_{it}) + \varepsilon_{t+1}$$

and hence

$$E_t(R_{t+1}^{ei}) = a + b \log(B/M_{it})$$

This should be very satisfying. Two totally different techniques in fact are providing the same information – our two big questions are 1) *how* do expected returns vary and 2) *why* do expected

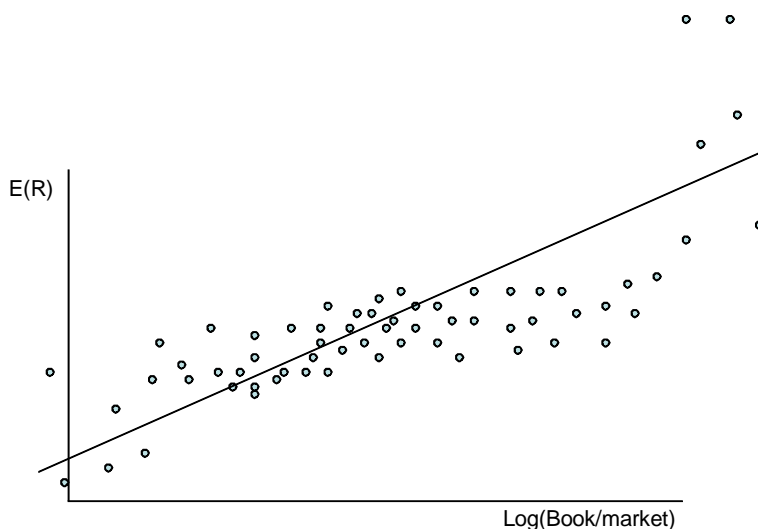
returns vary. These are both “how” techniques. Again, in a problem set I asked you to compare forecasting and cross sectional regressions

$$R_{t+1}^{ei} = a + b \log(B/M_{it}) + c \log(ME_{it}) + \varepsilon_{t+1}$$

to the 5x5 table of average returns. You were supposed to conclude that both techniques give the same information.

Great, but why choose one technique over another? Here are some of the issues explicit or implicit in the papers.

Outliers and small stocks



A warning on OLS equally-weighted cross-sectional regressions

As this graph suggests, linear regressions can be influenced by outliers. Also as FF made clear, half the stocks account for 3% of market value. And small stocks are likely to be outliers. Linear regressions can get you in trouble. (At both classes, students suggested value-weighting the points GLS style. Nobody has tried that. Yet.)

Already, Fama and French warn about equal-weighted 1-10 portfolios, which focus attention on small stocks. Regression coefficients can be worse.

Functional form.

Should we run the regression on BM, or on log BM? Or on the BM percentile? “Dissecting anomalies” Table 2 is worth reading as it wrestles with functional form. It shows the actual value of each forecasting variable in the portfolio bins. And you see how the 1 and 10 bins have extreme values of the sorting variable. Well, duh, in retrospect, if the sorting variable is normally distributed, then the 1/10 and 1/20 portfolios are going to have extreme values of the sorting variable.

In fact, we seem to see frequently that expected returns do not march uniformly across 1-10 portfolios but rather have an s shape. Is this a problem? No, actually, it’s what you expect if $E(R)$ is a linear function of x , and x is normally distributed!

But there is no real reason to expect a linear relation for many of these variables. Means of sorted portfolios let you be agnostic about functional form.

Sexiness

Don't knock marketing. I suspect Fama and French sort into portfolios because it's much more persuasive. If their Table 1A had this instead

$$E(R^{ei}) = a - 0.2ME_i + 0.4BM_i + \varepsilon_i$$

with the requisite t statistics, do you think people would have been excited as seeing the factor of three variation in average returns of Table 1A. Make your work come alive!

Multiple regressions

On the other side of the coin, once we go past 3, 4, 5 and more sorting variables, and we want to know "does sorting variable z help to describe /predict returns *given* sorting variables a, b, c, d, ... x,y" you can see that multiple sorts are just not going to do. OK, we can sort on size and book/market. But not 5 variables!

You can see Fama and French struggling with this in their 5 factor paper. While you were going to sleep looking at table after table of multiple sorts, they were struggling with just how to sort things 5 ways.

Here, regressions shine. A multiple cross sectional regression lets you see immediately "is variable z important given all the others" by the multiple regression coefficient.

But multiple regressions aren't trivial either. It's easy to put 5 variables in a multiple regression, but we see in the data that we may need cross terms as well. Again, you saw that already in a problem set with size and B/M. The size effect was stronger in value stocks than growth stocks. To capture that, you needed

$$\begin{aligned} E(R^{ei}) &= a + (b + dBM_i)ME_i + cBM_i + \varepsilon^i \\ &= a + bME_i + cBM_i + d(BM_i \times ME_i) + \varepsilon^i \end{aligned}$$

If we add arbitrary cross terms we're going to go nuts.

Beta portfolios

Another reason for doing sorts rather than cross-sectional regressions: Once you find

$$\begin{aligned} R_{t+1}^{ei} &= a + bx_{it} + \varepsilon_{it+1} \\ E_t(R_{t+1}^{ei}) &= a + bx_i \end{aligned}$$

how do you check whether that expected return corresponds to a beta? How do we check the beta exposure of a regression coefficient? Actually, I know how to do it, because I'm writing that paper. (Hint: \hat{b} is in fact the average return of a factor portfolio, so we can in fact check if \hat{b} lines up with a beta.) But it's not a common technique. Yet.

I think Fama and French 5 factor went back to portfolio sorts, because they needed to check if portfolio based means line up with betas.

Fama MacBeth and correlated errors.

If you do cross sectional regressions, beware,

$$R_{t+1}^{ei} = a + bx_{it} + \varepsilon_{it+1}$$

The errors in a cross-sectional regression are correlated across securities i. It's not a bad approximation that they are uncorrelated over time $cov(\varepsilon_{it+1}, \varepsilon_{j,t+1}) \neq 0$; $cov(\varepsilon_{it+1}, \varepsilon_{it+2}) = 0$. The OLS

formula for t statistics assumes errors are uncorrelated. You can't use it. Either use Fama MacBeth standard errors, or use a formula for standard errors that corrects for cross sectional correlation $(X'X)^{-1}X'\Omega X(X'X)^{-1}$

Asset Pricing at the cross roads

Why don't I have an answer to "what are, really the important sorting variables?" and "how many factors do we really need to account for the known sorting variables?" The answer is as much methodological as practical. The technique of sorting into 1-10 portfolios and running time-series regressions on the factors – itself a bloody brilliant invention by Fama and French – just is not up to the task of handling 5 or more variables. We need technical innovation as well as time to go look at the data.