

FACTOR / STYLE INVESTING

It's Not Data Mining — Not Even Close

June 2, 2015

Data mining — “discovering” historical patterns that are driven by random, not real, relationships and assuming they'll repeat — is a huge concern in many fields. My focus is, of course, on the field of investing, where those concerns are particularly present. That is true in academic and quantitative studies when great statistical power is brought to the effort, but it's also a concern in the non-quant world (how many would want to imitate Warren Buffett if he had not been so successful and do we give too much weight to that ex post result?). Some critics of the basic findings in quantitative finance — here I refer to the success of the small-cap, value and momentum factors — focus on this problem of data mining. They vary from the [sober, helpful and important](#), to the [less so](#).

One early critic of these results, based on fears of data mining, was [Fischer Black](#).¹ I disagreed with him at the time (in fact, you can find me listed in the thank-yous in his paper²), but his worry about these specific factors was inherently more reasonable in 1990, when many of the results were “in sample.” This will be a very short post as all I'm going to do is look at the out-of-sample results since Fischer's worry (also roughly the out-of-sample results I've experienced since my dissertation studying value and momentum — it's fun to have been around long enough to have a personal out-of-sample period!).³

Our most potent weapon in addressing data mining is the out-of-sample test.^{4,5} If a researcher discovered an empirical result only because she tortured the data until it confessed, one would not expect it to work outside the torture zone. Since the initial papers of Fama and French (1992, 1993), the results for value, momentum and size⁶ have been tested out-of-sample in other places besides U.S. equities, where they were initially uncovered. [Back then](#) and [more recently](#) we found strong empirical evidence for these concepts — particularly value and momentum — in other contexts, geographies and asset classes, providing strong support for the basic factors' efficacy. Subsequent research (for example [here](#) and [here](#) extended some of the basics further back in time, another out-of-sample test if you hadn't looked yet. But, there is probably no substitute for simply looking at how the actual first factors for U.S. equities, constructed very simply and in a highly similar fashion to how they were back then, have performed out-of-sample since their initial publication.

I look at just three factors, SMB (Fama-French's construct measuring the return spread of small versus big stocks), HML⁷ (Fama-French's construct measuring the return spread of low versus high price-to-book stocks, or as others might put it, the spread between cheap and expensive stocks), and UMD (Fama-French's version of the momentum factor measuring the return spread of past winner versus loser stocks), over what I label the “in-sample” periods (both July 1963 to December 1991 and January 1927 to December 1991) and the “out-of-sample” period (January 1992 to March 2015).⁸

Each of these factors, and the market itself, has had crashes, long droughts and bear markets. All have come under fire after these events and gone on to recover. I don't worry about that here as it's to be expected and is consistent with the historical in-sample findings. I focus only on the mean returns to the factors in-sample and out-of-sample.

I'm aiming for a “one-table” post. It shows the mean gross returns of each factor for the two different “in-sample” periods, the “out-of-sample” period since they ended, and a t-statistic testing whether the mean over the “out-of-sample” period is reliably different from that over the longer “in-sample” period of 1927-1991 (I'm rounding to the year, as the monthly starting dates were quoted earlier):

Average Gross Return Spreads Over Different Periods

	SMB	HML	UMD
1927-1991	2.8%	5.1%	8.9%
1963-1991	3.2%	4.7%	10.1%
1992-2015	2.6%	3.6%	6.1%
T-stat on Difference	-0.08	-0.48	-0.71

There is some exceptionally minor support for the cynics. The means are all lower out-of-sample, with momentum dropping the most (but still inducing the best stand-alone gross spread in each sample period). However, the cynics are supping on a thin gruel. All of the spreads are economically meaningful, and none of the out-of-sample versus in-sample differences, given significant out-of-sample data, are statistically significant.^{9,10} If at the end of 1991 you invested in these factors and achieved the above results you would be ecstatic without

reservation. I think Fischer would've been as well! Though I admit he was not the most predictable fellow...

Of course, one is still allowed to be cynical about these factors going forward. You might have a very high estimate of transactions costs (for a good discussion of trading costs, see Frazzini, Israel and Moskowitz's [paper](#)), or think the "world has changed" since these factors are now well-known. These are legitimate concerns for these or any investment strategy, though we would argue they are perhaps reasons to assume less going forward but hardly reasons to assume little to nothing. Furthermore, they are completely different concerns than data mining.

Data mining was a reasonable — if still (imho back then and now) wrong — worry back when Fischer Black wrung his hands over it in the early 1990s. While it is a reasonable worry for the overall field now, it is no longer a reasonable worry for the original research that found these factors. If you're still hawking this story, that the original results of Fama and French, Jegadeesh and Titman, Lakonishok, Vishny and Shleifer — and even yours truly and others — were the result of data mining, you have been completely defeated on the field of financial battle, and you must stop.

[1] Fischer actually thought that the size effect was data mining, the value effect probably data mining but perhaps the result of investor irrationality (he was more negative on the possibility it was a rational risk premium) and didn't address the momentum effect, but one can presume his opinion!

[2] As Fama and French's all-but-dissertation student at the time, I went into a semi-panic when I saw Fischer thanked me on his harsh rebuke of their work. Allaying my fears, they were mildly amused by my worries.

[3] [Moskowitz and Israel](#) cover highly related ground more thoroughly, but less short and "bloggy" than I do here.

[4] Vying in importance is insisting on an economic rationale for why something works — including if possible testable implications that go beyond just historical success.

[5] If one repeatedly iterates looking for in-sample results that also survive out-of-sample, then tests that appear out-of-sample can actually become sort of in-sample. Judgment must be used as to what is, or is not, going on in each specific case. I'm confident the three factors I study here survive this judgment.

[6] We have never been as positive on the size effect as the value and momentum ones. I include it here as not to cherry pick among the "big three" being studied in the early 1990s. For our latest thoughts on the size anomaly see [here](#).

[7] As many of you are aware we favor a slightly different version of HML that uses more up to date prices. [We find](#) it's much more powerful in combination with momentum. We do not use our version here as we didn't publicly document this until far later (we were in fact trading this way since 1995 when our group formed at Goldman Sachs) so it's not as truly out-of-sample, at least not publicly verifiably so.

[8] The monthly factor returns all come from Ken French's website and I refer you there for the full definitions.

[9] While still quite strong they have dropped more in Sharpe than in average return, an issue I expect to address another day in a post tentatively titled (and unwritten as of now) "How Can a Strategy Everyone Knows About Still Work?"

[10] [Harvey et al.](#) tell us that one thing we might do to combat data mining is the simple act of raising our threshold for new factors from a t-statistic of 2.0 to 3.0. T-statistics are a function of both realized Sharpe ratio and time, and none of SMB, HML or UMD are a 3.0 t-statistic out-of-sample (1992–2015 in our tests). But, that threshold was only meant for new factors not for out-of sample tests of old factors. Still, amazingly, a portfolio of the three factors (1/3 invested in each factor) realizes an out-of-sample t-statistic of 2.8. An out-of-sample test of the three known factors together almost reaches Harvey et al.'s threshold for new in-sample factors. I think that's pretty great.

This document is not intended to, and does not relate specifically to any investment strategy or product that AQR offers. It is being provided merely to provide a framework to assist in the implementation of an investor's own analysis and an investor's own view on the topic discussed herein.

This document has been provided to you solely for information purposes and does not constitute an offer or solicitation of an offer or any advice or recommendation to purchase any securities or other financial instruments and may not be construed as such. The factual information set forth herein has been obtained or derived from sources believed by the author and AQR Capital Management, LLC ("AQR") to be reliable but it is not necessarily all-inclusive and is not guaranteed as to its accuracy and is not to be regarded as a

representation or warranty, express or implied, as to the information's accuracy or completeness, nor should the attached information serve as the basis of any investment decision. This document is not to be reproduced or redistributed to any other person. The information set forth herein has been provided to you as secondary information and should not be the primary source for any investment or allocation decision. Past performance is not a guarantee of future performance. Diversification does not eliminate the risk of experiencing investment losses.

This material is not research and should not be treated as research. This paper does not represent valuation judgments with respect to any financial instrument, issuer, security or sector that may be described or referenced herein and does not represent a formal or official view of AQR. The views expressed reflect the current views as of the date hereof and neither the author nor AQR undertakes to advise you of any changes in the views expressed herein.

The information contained herein is only as current as of the date indicated, and may be superseded by subsequent market events or for other reasons. Charts and graphs provided herein are for illustrative purposes only. The information in this presentation has been developed internally and/or obtained from sources believed to be reliable; however, neither AQR nor the author guarantees the accuracy, adequacy or completeness of such information. Nothing contained herein constitutes investment, legal, tax or other advice nor is it to be relied on in making an investment or other decision. There can be no assurance that an investment strategy will be successful. Historic market trends are not reliable indicators of actual future market behavior or future performance of any particular investment which may differ materially, and should not be relied upon as such. Diversification does not eliminate the risk of experiencing investment losses.

The information in this paper may contain projections or other forward-looking statements regarding future events, targets, forecasts or expectations regarding the strategies described herein, and is only current as of the date indicated. There is no assurance that such events or targets will be achieved, and may be significantly different from that shown here. The information in this document, including statements concerning financial market trends, is based on current market conditions, which will fluctuate and may be superseded by subsequent market events or for other reasons.