

The History of the Cross-Section of Stock Returns

March 13, 2019

Juhani Linnainmaa cautions investors to be wary of alpha decay in newly discovered factors. Using fresh, pre-discovery data, his out-of-sample analysis shows about half of the many factors tested were not robust.

This Conversation is based on

The History of the Cross-Section of Stock Returns (November 2016)

Juhani T. Linnainmaa
Michael R. Roberts

Read in SSRN

Juhani Linnainmaa, PhD Partner & Advisor,
Research Affiliates, and Professor, University of Southern
California

Brent Leadbetter, CFA Senior Vice President,
Client Strategies, Research Affiliates



 Brent Juhani, one of the main takeaways from your paper¹ is that a lot of the return anomalies, or factors, that we hear about so often might be spurious. What prompted you to examine factors?

 Juhani *Our motivation was the belief that some of the many factors identified have been data mined. So Michael Roberts, my coauthor, and I used some very new data to look at this question from a different angle to see what story the data tell.*

 Brent It seems like one of the reasons you wanted to use new data was to get out-of-sample results as a comparison to the in-sample test. Can you define in sample versus out of sample?

 Juhani *Any study, of course, is going to be limited by historical data. A classic study, like the Fama–French 1992 paper, is a very well-cited paper that examines the factors size and value. Fama–French studied monthly stock returns from 1963 to 1989. All the results in the paper are based on that one splice of data. That would be known as the in-sample period.*

The question then is how well does the factor perform out of sample, when we step outside the period 1963 to 1989? We can test it in two ways. We can look at the data that accumulated after 1989, a post-discovery sample, or we can go backward in time before 1963. The idea is to not look at the same data that the original researchers looked at, but to look at the data outside that period, which is the out-of-sample data.

 Brent The purpose is to confirm whether the finding is robust, whether it's actually significant, is that correct?

 Juhani *Yes, that's correct. The data-mining concern is that we can never be sure if a finding is real. The finding will always be a probabilistic statement. We set a null hypothesis that we then test for. When we look for factors, the null hypothesis would be "There is no effect." Then, we look at the data and we make a probabilistic statement of how likely it will be to see these data if there were no effect. So, when we say the t-value is greater than 2 or the p-value is less than 5%, it means it would be unlikely to see an effect like this in the data if the effect wasn't real.*

So, often, the threshold for significance has been 5%, which means that when we find an effect, there's a 5% probability the effect isn't real, that it just happened by luck—and that wouldn't be a big mistake to make. The problem with data mining is that when many researchers try many different things to find a return anomaly, the 5% probability of being lucky adds up. If you try 100 different factors, 5 are going to look great by luck.

If you try 10,000 different factors, 500 are going to look real, even though they just happened by luck. It's not that one researcher tries many different things, but that all the researchers collectively have a huge incentive to find a factor that works. They are independently doing these massive amounts of data mining, so it could be that many "factors" that have been "found" are just not real.

¹See "[The History of the Cross Section of Stock Returns](#)" available at SSRN.

 Brent Is the argument then that the test we've been accustomed to seeing in the industry historically—a t-stat of 2, p-value of less than .05—is valid for a single test, but may be less valid with increases in computing power? That's why the old tests might no longer be sufficient?

 Juhani Yes, that's exactly right. So, if we want to stick to the 5% threshold as being meaningful, we would effectively have to include every single test that has ever been tried. The problem is that we only see the successful tests. The result is that we are then not adjusting the p-values appropriately.

Let me go back to your earlier question about why we test out of sample.

The idea is that if the null hypothesis is "There's no effect at all," then in sample the probability should have been zero, but it was actually 5% because of luck. When we go out of sample, if the "factor" produces no effect, the probability should also be zero. Out of sample, there can be no data mining. Hence, we expect, on average, that the alpha out of sample should be exactly zero. That's what we're testing. That's what we try to do when we go out of sample.

 Brent One of the terms you use frequently in the paper is data snooping. I'm going out on a limb and guessing that you didn't find that the results were significant out of sample, given that snooping doesn't exactly have a positive connotation.

 Juhani Yes, that's correct. The main point of the paper is that even with massive amounts of alpha in sample, whether we look forward or backward in time outside the original study, we find more than a 50% decline in the performance of the factor. Average returns, Sharpe ratios, information ratios, everything declines out of sample. And the fact that it happens both after and before the sample gives us confidence this is probably happening because of data mining and not because, for example, the factor is discovered and then traded on more and more, causing the anomaly to go away. Data mining seems like the simplest explanation.

 Brent Was the impact largely homogenous across factors or did you find some factors that did a better job of surviving out of sample than others?

 Juhani Two things here are important. For any one factor, it's very difficult to say the factor isn't real. The issue is the amount of noise in returns. If we look at any one factor, its out-of-sample alpha may be insignificant but, at the same time, we cannot say that this out-of-sample alpha is significantly different from the in-sample alpha. There is just too much noise in returns. We have to look at multiple anomalies at the same time to draw strong inferences of an anomaly being real or not. That's why, in the paper, we look at a large number of anomalies. Then we have more power to say that in sample the alpha is high, but out of sample, both before and after the sample used in the original study, it is much lower.

 Juhani *It does look like some anomalies are far less powerful out of sample than in sample, even when we account for the fact there's a lot of noise in returns. The investment factor would be one of those anomalies. The idea is that firms that invest conservatively have much higher returns than firms that invest aggressively. That finding only seems to be true in the in-sample period. There's no effect whatsoever when we look at the factor out of sample from 1926 to 1963, and after the original sample ended, I believe, in 2003. That's one example of a factor that doesn't seem quite as good out of sample.*

 Brent The out-of-sample results are not matching up to the in-sample results, that seems clear. What do you think the cause of that disparity is?

 Juhani *That's a great question. Researchers have been thinking that data mining has been going on for some 20 years now. They have discovered that, in many cases, the performance out of sample seems to be worse than in sample. But what they have always done has been to look at the data after the original study—then they can confirm two different effects. The first thing that could be going on is data mining, so that the effect wasn't real in the first place; that would explain why there's a decay in alpha out of sample. The other possibility could be that maybe there was a mispricing in the first place, but now, out of sample, arbitrageurs are trading on it, and that's what makes the alpha go away.*

What we did in our study was to get completely new data that predated the original sample. Most studies use data starting in July 1963 because Standard & Poor's set up the Compustat database in 1962; most researchers never go beyond that wall. But we collected data going back much farther. We have this fresh, really long sample. The benefit of going out of sample in this direction is that now the finding is going to be clearly about data mining. One could always argue that after discovery you're going to have more arbitrageurs. But if we look before discovery, of course, when investors couldn't possibly have known about these discoveries, it's more cleanly attestable to data mining. The fact that we find the same effect post-discovery and pre-discovery gives us the confidence that the decay in alpha is probably due to data mining and not about arbitrageurs destroying the predictability in returns.

 Brent No one was trying to arbitrage momentum away in the 1930s.

 Juhani *Yes, I think that's safe to say.*

 Brent That paper had not yet been written.

 Juhani Yes.

 Brent One of the topics you hit on in your [Advisor Symposium presentation](#) is that for the more recently discovered factors, the latest returns are still in sample. Did you see a similar dynamic when you were working on your paper?

 Juhani *Yes, because no matter what number we see out of sample, we're not going to have any power to say if it's different from the in-sample finding—that would be the limitation. The benefit for us is that we have so many different factors in our analysis that when we put all of them in the pool, we can draw historic inferences.*

 Brent One of the big takeaways for me is that I should be skeptical when I see a new factor introduced. I should look at excess returns with a skeptical eye. What do you think the key takeaways from the paper are?

 Juhani *I would say there are two takeaways. One is that there's going to be significant decay in the alphas of newly discovered factors. So for a new factor, shave something like 50% off of the return estimates. That's how profitable it's going to be out of sample. Of course, the ideal thing to do would be to actually identify the factors that are real and those that are not, but that's exactly the whole point, that we really cannot do it.*

What we can say is that some of the factors are real and others are not real. When we look at all of them in a pool, we can say that about half of them are fake, but we don't know which ones should be trusted and which shouldn't. Your conclusion is exactly right: when you invest in a newly discovered anomaly, you want to be quite cautious, because it might turn out to be one of the ones that doesn't hold up in the future.

 Brent Great. Thanks, Juhani.

 Juhani *Thank you.*

CONTACT US

www.researchaffiliates.com

Americas

phone: +1.949.325.8700
email: info@researchaffiliates.com

Australasia

phone: +6.129.160.2290
email: australia@researchaffiliates.com

Europe

phone: +44.0.203.9299880
email: uk@researchaffiliates.com

Press

phone: +1.212.207.9450
email: hewesteam@hewescomm.com

Disclosure

The material contained in this document is for general information purposes only. It is not intended as an offer or a solicitation for the purchase and/or sale of any security, derivative, commodity, or financial instrument, nor is it advice or a recommendation to enter into any transaction. Research results relate only to a hypothetical model of past performance (i.e., a simulation) and not to actual results or historical data of any asset management product. Hypothetical investor accounts depicted are not representative of actual client accounts. No allowance has been made for trading costs or management fees, which would reduce investment performance. Actual results may differ. Simulated data may have under-or-over compensated for the impact, if any, of certain market factors. Simulated returns may not reflect the impact that material economic and market factors might have had on the advisor's decision-making if the advisor were actually managing clients' money. Simulated data is subject to the fact that it is designed with the benefit of hindsight. Simulated returns carry the risk that the performance depicted is not due to successful predictive modeling. Simulated returns cannot predict how an investment strategy will perform in the future. Simulated returns should not be considered indicative of the skill of the advisor. Investors may experience loss. Index returns represent back-tested performance based on rules used in the creation of the index, are not a guarantee of future performance, and are not indicative of any specific investment. Indexes are not managed investment products and cannot be invested in directly. This material is based on information that is considered to be reliable, but Research Affiliates™ and its related entities (collectively "Research Affiliates") make this information available on an "as is" basis without a duty to update, make warranties, express or implied, regarding the accuracy of the information contained herein. Research Affiliates is not responsible for any errors or omissions or for results obtained from the use of this information. Nothing contained in this material is intended to constitute legal, tax, securities, financial or investment advice, nor an opinion regarding the appropriateness of any investment. The information contained in this material should not be acted upon without obtaining advice from a licensed professional. Research Affiliates, LLC, is an investment adviser registered under the Investment Advisors Act of 1940 with the U.S. Securities and Exchange Commission (SEC). Our registration as an investment adviser does not imply a certain level of skill or training.

Investors should be aware of the risks associated with data sources and quantitative processes used to create the content contained herein or the investment management process. Errors may exist in data acquired from third party vendors, the construction or coding of indices or model portfolios, and the construction of the spreadsheets, results or information provided. Research Affiliates takes reasonable steps to eliminate or mitigate errors, and to identify data and process errors so as to minimize the potential impact of such errors, however Research Affiliates cannot guarantee that such errors will not occur. Use of this material is conditioned upon, and evidence of, the user's full release of Research Affiliates from any liability or responsibility for any damages that may result from any errors herein.

The trademarks Fundamental Index™, RAFI™, Research Affiliates Equity™, RAE™, and the Research Affiliates™ trademark and corporate name and all related logos are the exclusive intellectual property of Research Affiliates, LLC and in some cases are registered trademarks in the U.S. and other countries. Various features of the Fundamental Index™ methodology, including an accounting data-based non-capitalization data processing system and method for creating and weighting an index of securities, are protected by various patents, and patent-pending intellectual property of Research Affiliates, LLC. (See all applicable US Patents, Patent Publications, Patent Pending intellectual property and protected trademarks located at <http://www.researchaffiliates.com/Pages/legal.aspx>, which are fully incorporated herein.) Any use of these trademarks, logos, patented or patent pending methodologies without the prior written permission of Research Affiliates, LLC, is expressly prohibited. Research Affiliates, LLC, reserves the right to take any and all necessary action to preserve all of its rights, title, and interest in and to these marks, patents or pending patents.

The views and opinions expressed are those of the author and not necessarily those of Research Affiliates, LLC. The opinions are subject to change without notice.

©2019 Research Affiliates, LLC. All rights reserved.