

Lu Zhang

'Most reported anomalies fail to hold up'

Lu Zhang is Professor of Finance and The John W. Galbreath Chair at the Fisher College of Business of The Ohio State University. For several years, together with fellow researchers Kewei Hou and Chen Xue, he has been digging deeper into the robustness of dozens of market anomalies reported in the academic literature. In our Great Minds series, a set of interviews with renowned academics and investment experts, we asked him about this work that involved thorough fact-checking reported equity market anomalies. More generally, we also asked him about factor investing and how investors should go about it.

Great Minds

Your recent research has focused on the replication of numerous academically-reported anomalies in equity markets. Could you explain how this idea came about and what led you to undertake this endeavor?

"It took a long while. Kewei, Chen and I first documented some of the evidence when we were working on our q-factor paper back in 2014.¹ At the time, we coded up about 80 anomaly variables, but only 35 were significant. In particular, 12 out of 13 liquidity variables failed to holdup. The editor of our article, Professor Geert Bekaert, deserves a huge amount of credit for guiding our q-factor paper and letting it see the light of day. While editing our work, Geert told us that he found our evidence that so many well-known anomalies are insignificant very important, and wanted us to highlight it more. We did. But since the objective of that article was to establish a new workhorse factor model, we did not make the evidence the centerpiece of the article."

"Back in 2015, Eugene Fama and Kenneth French responded to our q-factor paper by incorporating two factors that resemble our investment and return on equity factors in the q-factor model into their three-factor model to form a five-factor model.² And the Factors War was on. We quickly fired back with the working paper

'A comparison of new factor models', which compares our q-factor model with their five-factor model on both conceptual and empirical grounds.³ Our key evidence is that the q-factors subsume their CMA and RMW factors, but their factors cannot subsume ours in factor spanning tests."

"Alas, that paper met with considerable resistance in the editorial process. Knowing very well what it takes to debate with Fama and French on their home turf, we set out to clear a higher hurdle with respect to incremental contribution, by replicating virtually all of the published literature about anomalies. Our initial thought was to compile the largest set of testing portfolios to test factor models, and to hold up our work against the competitive pressure from Fama and French."

"The tremendous amount of respect we have for Fama and French is borne out in the massive effort we put into 'Replicating anomalies'. It is probably worthwhile pointing out that we did not set out to beat down the literature on anomalies. We were focusing on the right-hand, not the left-hand side of factor regressions. After three years of coding, it finally dawned on us that most anomalies fail to hold up, 64% to be precise. The evidence is undeniable."

"We were aware of Professor Campbell Harvey's work with Yan Liu and Caroline Zhu, as well as Cam's 'Presidential address'.⁴ Looking at our evidence, we realized that Cam was right. We started to dig deeper into his work and the meta-science literature that he cited in his research. After that, the big picture became very clear to us.

¹ K. Hou, C. Xue, and L. Zhang 2015, 'Digesting anomalies: An investment approach,' *Review of Financial Studies* 28, 650-705.

² E. F. Fama, and K. R. French, 2015, 'A five-factor asset pricing model,' *Journal of Financial Economics* 116, 1-22.

³ K. Hou, C. Xue, and L. Zhang, 2014, 'A comparison of new factor models,' NBER Working Paper No. 20682, November 2014.

⁴ C. R. Harvey, Y. Liu, and H. Zhu, 2016, '... and the cross-section of expected returns,' *Review of Financial Studies* 29, 5-68. C. R. Harvey, 2017, 'Presidential address: The scientific outlook in financial economics,' *Journal of Finance* 72, 1399-1440.

Lu Zhang

is Professor of Finance and The John W. Galbreath Chair at the Fisher College of Business of The Ohio State University. He is also a Research Associate at the National Bureau of Economic Research and an Associate Editor of *Journal of Financial Economics* and *Journal of Financial and Quantitative Analysis*. He is Founding President of Macro Finance Society, an international academic society devoted to advancing and disseminating high-quality research at the intersection of financial economics and macroeconomics. Before joining Ohio State in 2010, he taught at Stephen M. Ross School of Business at University of Michigan and William E. Simon Graduate School of Business Administration at University of Rochester. Lu Zhang's research focuses on asset pricing, in connection with macroeconomics, corporate finance, labor economics, and capital markets research in accounting. One of his major contributions is 'The investment CAPM', a unified, economics-based framework for understanding asset pricing anomalies. Lu Zhang has published extensively at leading academic journals. One chapter of his doctoral thesis 'The value premium' won the Smith-Breeden Award for Best Paper for 2005 from American Finance Association (AFA) and *Journal of Finance*.

'Only 36% of the anomalies in our large universe withstood the replication tests. The survival rate is largely in line with those reported in other scientific disciplines such as psychology and oncology.'

Great Minds

Publication bias is real, and it affects everyone, ourselves included. We should all, as a profession, at least be aware of that danger."

The resulting paper⁵ considers almost 450 anomalies. Doing such an extensive groundwork must have required a lot of effort. Can you briefly tell us how you went about it? What were the main parameters you changed compared to the original studies and why?

"Professor Chen Xue at the University of Cincinnati is the real hero behind our 'Replicating anomalies'. I went through the published anomalies literature, and wrote a first draft of our data appendix. I knew a lot of the classic anomalies, but needed a refresher course on those documented in the past ten years, so it was not time-consuming for me. It was Chen who painstakingly coded up all 447 anomalies, one-by-one, making sure that we followed the variable definitions in the original studies, and when our replication results differed from those originally reported, making sure we understood why. Professor Kewei Hou went through Chen's SAS programs to ensure that our empirical execution was of the highest possible quality."

"In our replication, we emphasized a reliable set of empirical procedures that use NYSE breakpoints and value-weighted portfolio returns. This set of procedures is more reliable because it better captures the economic importance of an anomaly. For comparison, in our June 2017 draft, we also reported results from NYSE-Amex-NASDAQ breakpoints and equal-weighted returns, a procedure that gives microcaps excessive weights. We are currently compiling results from a variety of additional procedures, including cross-sectional regressions."

And what are the main conclusions you would highlight?

⁵ K. Hou, C. Xue, and L. Zhang, 'Replicating anomalies,' NBER Working Paper No. 23394, May 2017.

"The main conclusion is that most anomalies fail to replicate. To be precise, only 36% of the anomalies in our large universe withstood the replication tests. The survival rate is largely in line with those reported in other scientific disciplines such as psychology and oncology."

Does that mean markets are not as inefficient as some suggest?

"The short answer is yes. No answer is long enough for the efficient markets/behavioral finance debate. But in terms of the simplistic view that equates cross-sectional predictability with 'free lunches', our evidence does indicate that there are certainly fewer free lunches around in the marketplace."

Still, some factors appear to be significant and persistent enough. Would you consider the 'anomaly glass' half full or half empty?

"Half full. Our replication did confirm the validity of many factors that investors have been loading on for a long time, such as value and momentum."



You argue that p-hacking is widespread in academic research. How would you suggest changing this?

"We are dropping the p-hacking interpretation of our evidence. We are doing a major revision of 'Replicating anomalies'. The next draft will just say: 'most anomalies fail to replicate.' When writing up the two first drafts, we were just using 'p-hacking' as a new name for 'data-mining', which has been around in the academic literature for a long time. Regardless of whether 'p-hacking' is mentioned in our article, we feel that data-mining is widespread. We don't have a good answer to this problem."

"Oftentimes, the line between data-mining and striving for good empirical performance is blurry. But it is healthy to at least be aware of the problem. Nowadays, for every working paper we circulate, we go through multiple rounds of internal replication, to control the quality of execution. We also routinely highlight evidence that goes against the tested hypotheses in our work."

You mention the fact that many academic studies overweight microcaps and the fact that due to the high trading costs associated with these stocks, anomalies in microcaps are too difficult to exploit in practice. What exactly do you mean by microcaps? Is this phenomenon constant over time?

"We first read about microcaps in Fama and French's 2008 article in Journal of Finance.⁶ Over the past 15 years, we have learned a lot from studying Fama and French's articles, especially those from the 1990s. Microcaps are tiny stocks of which the market equity is below the 20th percentile of NYSE stocks. These stocks are not just small, they're tiny. Fama and French showed that microcaps account for only 3% of the total market cap, but 60% of the names. In 'Replicating anomalies', we updated their evidence. As of December 2014, microcaps represented only 1.4% of the total market cap."

⁶ E. F. Fama, and K. R. French, 2008, 'Dissecting anomalies,' Journal of Finance 63, 1653-1678.
⁷ K. Hou, H. Mo, C. Xue, and L. Zhang, 2018, 'q's', working paper, The Ohio State University.
⁸ Robeco, 'The research culture is crucial for the success of an asset manager', Quant Quarterly magazine, October 2017.

"An example would be that the Fama-French five-factor model outperforms the q-factor model in explaining the value-minus-growth anomalies (but not by much).⁷ Finally, we try to take economic theory seriously, and use it to guide our empirical work. Otherwise, one is practicing applied statistics, not empirical economics."

In a recent interview with Robeco⁸, former president of the AFA, Campbell Harvey, advocated a concept called 'registered reports', where researchers would first pitch ideas to editors and these ideas would be peer-reviewed. If reviews were positive, editors would commit to publish the paper, no matter what results were found. What do you think?

"The concept of 'registered reports' is new to me. I have not thought much about it. Cam has thought long and hard about related issues. He also has a wealth of editorial experience. I would defer to Cam in these matters. I do find the idea of 'registered reports' interesting and worthwhile to experiment with, starting out on a limited scale. I also think replication should be more routine in finance."

Your findings deal a particularly severe blow to the academic literature around a purported liquidity factor. Could you explain bit more about them?

“Sure. We certainly were not aiming at the liquidity literature. As mentioned, back in 2014, when working on the q-factor paper, we came across the evidence that 12 out of 13 liquidity variables are not significant. We took notice, but did not make a big deal out of it. We thought the set of 13 is too small.”

“In ‘Replicating anomalies’, we have looked at a total of 102 variables broadly related to liquidity and trading frictions. We find that 95 of them, or 93%, fail to yield significant high-minus-low decile returns on average. The list of 95 includes many influential variables, such as short-term reversal, share turnover, absolute return-to-volume, idiosyncratic volatility, the number of zero trading days, and bid-ask spread.”

Does this mean there is no such thing as a liquidity premium? To what extent should investors care about liquidity?

“No, our findings do not mean that there is no such thing as the liquidity premium. However, they do say that in the value-weighted universe that accounts for 97% of total market cap, liquidity is just not that important. Before writing ‘Replicating anomalies’, I thought liquidity was as important as value and momentum, and probably more important than investment and profitability in the cross section, given the amount of the published literature. That was the impression I got from reading the published liquidity articles.”

“But I was not even close to being right. Value and momentum can be found in value weights, but not liquidity. Investment and profitability also can be found in value weights, and are closely related to value and momentum. To what extent should investors care about liquidity? If one rebalances portfolios monthly, weekly, or even daily, trading costs become important. It is critical to develop a trading system that minimizes transaction costs to harvest factor premiums. If it’s done less frequently, such as quarterly or annually, liquidity becomes less important, as shown in our work.”

So much for academic research. But would you say most research carried out by practitioners (mainly product providers) is just as biased?

“Yes. There are many products in the marketplace, including some very popular ones, which are different from value and momentum. Even the ‘quality’ products come with a variety of different definitions, many of which failed to replicate in our study.”

Does all this mean that mean investors should disregard factor investing altogether and simply go for passive strategies?

“Not at all. First, the line between active and passive strategies has blurred substantially in the past decade. In the old days, ‘passive’ literally meant holding the market portfolio, and ‘active’ meant everything else. Nowadays, ‘passive’ refers to predetermined algorithm-based strategies, and ‘active’ means there is more



‘Our work does not discredit factor investing at all. On the contrary, we document reliable cross-sectional predictability in a universe in which frictions seem to play a negligible role.’

frictions seem to play a negligible role. When you take 36% of 447, you still get 161 significant anomalies even in value-weighted returns. We show that our latest factor models still leave as many as 46 anomalies unexplained. In short, the future of factor investing is bright! The challenge is to figure out which factors are the most relevant to forecast returns, and that’s the essence of the new ‘active’.”

But then the obvious question is: which factors do you consider the most relevant for long-term investors? Why?

“Value and momentum, which have already been adopted by many long-term investors. I am likely biased, but I would also say investment and return on equity underlying our q-factor model. Empirically, we have shown that the investment factor largely subsumes the value factor, and the return on equity factor largely subsumes the momentum factor. Theoretically, investment and profitability have a solid economic foundation, based on the net present value rule in corporate finance. I have been developing an asset pricing theory on this rule, which I dub ‘The investment CAPM’.”⁹

“We can debate what the best empirical measures of investment and profitability are. But the basic economic principles are not controversial at all. They have been taught in business schools for many decades. What remains controversial, however, is whether one can deduce asset prices from the net present value rule, while ignoring investor behavior. But the equations underlying the net present value rule and the investment CAPM are the same. Causality runs both ways from investment to the expected return, and back, meaning no causality. The same applies to risk and the expected return in the consumption CAPM, meaning no causality from risk to the expected return.”

human involvement, I think. One may argue that factor investing built on the cross-sectional predictability in finance literature is passive in nature, according to the new definition.”

“Regardless of the passive-active dichotomy, our work does not discredit factor investing at all. On the contrary, we document reliable cross-sectional predictability in a universe in which

⁹ L. Zhang, 2017, ‘The investment CAPM’, *European Financial Management* 23, 545-603.