Factor investing

How many factors are there? Or how to navigate the ‘factor zoo’

- The literature has become a 'factor zoo' with hundreds of alleged factors
- We find evidence of dozens of distinct factors driving future stock returns
- We classify these into a handful of strategic composite factors

In the 1980s and 1990s, factors such as Size, Value and Momentum were shown to generate returns that could not be explained by the capital asset pricing model (CAPM). Since then, hundreds of other ‘factors’ have emerged in the literature, all of which allegedly offer a unique new source of return. This has given rise to a heated debate regarding the actual number of distinct factors and the number of findings that can be attributed to data mining.

Data mining (or data dredging, or p-hacking, or factor fishing) means that if brute force computing power is used to test hundreds, thousands or even millions of possible factors, many false positive results can be expected. In other words, simply due to chance many strategies, even completely random ones, will appear to be statistically significant. To illustrate, Arnott, Harvey, and Markowitz (2018) give the example of testing all combinations of the first three letters of a firm’s ticker symbol. They find that a strategy which goes long stocks with S as the third letter and short stocks with U as the third letter shows a spectacular back-tested performance and passes all common tests of statistical significance with flying colors. This is a clear example of how an economically nonsensical strategy can still appear to be profitable from a statistical point of view. Data mining is arguably the biggest pitfall of empirical research in finance.

Several recent studies take a comprehensive look at the ‘factor zoo’ (a term introduced by Cochrane, 2011). McLean and Pontiff (2016) report results for the US that are consistent with over-fitting or factor decay, as the performance of 97 factors is on average 26% lower out-of-sample and 58% lower post-publication. However, Jacobs and Müller (2020) find that these results do not carry over

Article
For professional investors
March 2020

David Blitz, Head of quantitative research
Matthias Hanauer, Senior researcher

to international equity markets, as none of the other 39 investigated countries shows a noticeable decline in post-publication performance. Harvey, Liu, and Zhu (2016) propose to counter p-hacking by raising thresholds for statistical significance in empirical research, specifically by requiring t-statistics to be at least 3, instead of the conventional threshold of 2. Applying this higher threshold and using a methodology which prevents illiquid micro-caps from dominating results, Hou, Xue and Zhang (2019) find that most previously documented factors become insignificant.

Academics aim to bring order to the factor zoo by looking for the smallest number of factors with the highest explanatory power. The best-known model is the Fama and French (1993) model, which includes the market, size, and value factors. Fama and French (2015) extend this 3-factor model to a 5-factor model by adding the profitability and investment factors. Interestingly, the widely accepted momentum factor of Jegadeesh and Titman (1993) and Carhart (1997) is still left out of this 5-factor model. However, most researchers prefer to also include momentum, so that the 3-factor model becomes a 4-factor model and the 5-factor model a 6-factor model. The Fama-French model competes with, amongst others, the Hou, Xue, and Zhang (2015) 4-factor model (also known as the q-factor model), the Stambaugh and Yuan (2017) 4-factor model, the Barillas and Shanken (2018) 6-factor model, and the Daniel, Hirshleifer, and Sun (2019) 3-factor model.

Although the factors used in these models are different, there does seem to be a consensus among leading academics that the entire factor zoo can be tamed with parsimonious models consisting of just three to six factors. This would imply that a handful of factors capture all information in the factor zoo. Other academics, however, find more independent dimensions in stock returns. Green, Hand, and Zhang (2017) find that 12 out of 94 characteristics are reliably independent determinants of return among non-microcap stocks, and that 11 of the 12 independent characteristics lie outside prominent benchmark models. Taking an international perspective can also increase the number of required factors, as Jacobs and Müller (2018) find that the dominant variables for a global sample differ from the ones in models fitted for the US market.

If three to six factors is indeed all one needs, then a factor-based investment approach becomes fairly straightforward. In the investment management industry this view is advocated by Edhec's SciBeta division. Goltz and Luyten (2019) argue that investors should only consider the widely accepted Fama-French factors, and that everything else is basically data mining and not evidence-based. They illustrate this with some appealing examples showing that other index providers use all kinds of unconventional factor definitions or come up with a different choice of factors and methodologies every time they launch a new index.

Based on our own research, we conclude that out of the hundreds of factors proposed in the literature many are indeed redundant, lack robustness, or cannot even be replicated to begin with. We do not find, however, that the entire factor zoo can be reduced to just a handful of factors. Although the small set of factors used in academic asset pricing models can serve as a very good starting point, that is not the end of the story. In our research, we find evidence of dozens of factors because of the following reasons:

- Some factors are wrongly rejected by the academic community.
- Multiple factors are often needed to capture one broader phenomenon.
- Factors based on non-standard data sources or with limited history should not be ignored.
- We expect a next generation of factors based on big data, machine learning and artificial intelligence.

Although we identify dozens of factors that are distinct drivers of future stock returns, we typically group these factors according to their data source and economic rationale, resulting in a handful of composite factors. For instance, the new Fama-French profitability and investment factors are independent phenomena, with a correlation that is close to zero, but they are commonly regarded as different dimensions of a more general quality factor. It is important to realize that any definition of composite factors is somewhat arbitrary, as the grouping can be more or less granular. An analogy can be drawn here with sector and industry classifications, which can be very broad (e.g. financials versus non-financials), or extremely detailed (there are over 150 GICS level 4 industries), or anything in between. For instance, the question whether analyst revisions and price momentum are separate factors, or different dimensions of one overarching sentiment or trend factor, is a similar kind of question as whether real estate investment trusts (REITs) should be seen as part of the Financials sector or as a separate one: there is no right or wrong here. In the remainder of this article, we describe our factor identification and classification process.

Factors which are wrongly rejected or ignored

Some of the key factors are omitted in academic asset pricing models, but wrongly so in our view. The most striking example is the low-risk factor, which challenges the fundamental notion of efficient markets that higher risk should be rewarded with higher returns. Despite abundant evidence that the empirical relation between risk and return is flat or even inverted, many academics dismiss the low-risk anomaly or believe that it is explained by other factors. In

Blitz, van Vliet, and Baltussen (2019) we review all this evidence and conclude that low-risk is a strong and distinct factor, that should not be ignored.

A second example is the classic short-term reversal factor. This factor is typically dismissed in academic research, because the profits are believed to disappear after accounting for market micro-structure issues (such as bid/ask effects) and transaction costs. In De Groot, Huij, and Zhou (2012), however, we show that the short-term reversal factor can be turned into a profitable strategy after costs when using a smart implementation approach. Moreover, we find that it can add value as a trade-timing indicator, so when the decision to trade is already made and trading costs are a sunk cost, but the best moment of trading still needs to be determined.

A third example is our residual momentum factor. Blitz, Huij, and Martens (2011) show that the residual momentum factor, which cleans the standard momentum factor for time-varying systematic risks, generates a much higher risk-adjusted performance. Blitz, Hanauer, and Vidojevic (2017) take a further look and find that residual (or idiosyncratic) momentum and traditional momentum are distinct phenomena which coexist next to each other. In other words, residual momentum is not explained by standard asset pricing models which include the standard momentum factor. This is empirically illustrated in Table 1 by regressing a Fama-French style residual momentum factor on the classic Fama-French factors, including standard momentum. The results show that standard momentum can only explain about half of the raw alpha of residual momentum. The unexplained alpha is still sizable and highly significant.

A fourth example is the accruals factor, which is another dimension of the quality factor. Although Fama and French (2008) still concluded that accruals is a separate factor, it is not included in the Fama and French (2015) 5-factor model, nor in competing academic models such as the Hou, Xue, and Zhang (2015) model. In Kyosev, Hanauer, Huij, and Lansdorp (2020), we find that the accruals factor exists next to other quality factors such as profitability and investment, i.e. that there are at least three distinct dimensions to quality.

Multiple factors needed to capture one broader phenomenon

Another reason we get to more than a handful of factors is because we find that, oftentimes, multiple factors are needed to capture one broader phenomenon. A nice example is the value factor. Fama and French (1992, 1993) choose to measure value with the ratio of book value to market value. This choice appears to be motivated by convenience rather than a strong theoretical preference for that particular metric. Fama and French (1996) empirically justify their choice by showing that the performance of alternative value metrics, such as earnings-to-price or cashflow-to-price, is entirely subsumed by their book-to-market factor. Since then, however, we have seen a rather weak performance for book-to-market and a considerably stronger performance for these alternative metrics.

To illustrate this point, we measure how much of the performance of book-to-market (BtM) is already captured by earnings-to-price (EtP) and vice versa. The results for these 'spanning regressions' are shown in Table 2 (see p.4). Our analysis confirms that BtM is the superior value measure for the original sample period investigated in Fama and French (1996), as BtM fully subsumes EtP, while the opposite is not true. However, if Fama and French would have conducted their analysis over the post-1994 period (the out-of-sample period for their 1996 study), they would probably have come to a different conclusion. Since 1994, EtP fully subsumes BtM, while BtM does not explain EtP anymore.

The analysis above clearly illustrates that the best-performing measure changes from one sample period to another. The economic rationale of book value is also increasingly being questioned, given that book values seem to become less relevant and meaningful. Various blue-chip firms even have negative book values nowadays. At the same time, we cannot rule out a strong comeback

Table 1 | Spanning regression for residual momentum

		CAPM alpha	4-factor alpha	Market (Rm-Rf)	Size (SMB)	Value (HML)	Momentum (WML)
07/1929-12/2018							
	ResMom	8.10% (10.07)	4.23% (6.83)	0.03 (2.52)	0.04 (2.21)	0.12 (7.52)	0.35 (28.97)

Source: The analysis is based on monthly total returns in U.S. dollars for the period July 1929 to December 2018. The residual momentum factor is self-constructed; all the other data comes from the Kenneth French online data library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. T-statistics are reported between brackets.

Table 2 | Spanning regressions for value defined based on BtM or on EtP

	CAPM Alpha	3-factor alpha	Market (Rm-Rf)	Size (SMB)	Value (EtP)	Value (BtM)
07/1963-12/1993						
Value (BtM)	6.67% (4.36)	1.48% (1.97)	-0.03 (-1.78)	0.05 (2.48)	0.89 (34.46)	
Value (EtP)	5.61% (3.76)	0.01% (0.02)	-0.03 (-1.84)	-0.04 (-1.78)		0.86 (34.46)
01/1994-06/2019						
Value (BtM)	2.22% (1.07)	-1.18% (-1.08)	0.03 (1.25)	0.09 (3.01)	1.03 (28.34)	
Value (EtP)	3.29% (1.88)	1.75% (1.95)	-0.05 (-2.83)	-0.11 (-4.34)		0.71 (28.34)

Source: The analysis is based on monthly total returns in U.S. dollars for the period July 1963 to June 2019. All the data comes from the Kenneth French online data library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Market is the value-weight return on the market portfolio minus the one-month Treasury bill rate. The size factor (SMB, small minus big) and the two value factors (BtM, EtP) are constructed based on the 2x3 value-weighted sorts. The first panel shows the results for the original sample period of Fama and French (1996) while the second panel shows the results for the period afterwards. T-statistics are reported between brackets.

scenario for book-to-market. Therefore, instead of trying to pinpoint which one value factor is the best among all, we prefer to have a diversified mix of value factors which all show strong long-term results, and which provide effective diversification in the shorter run.

The same argument can be made for other factors. For instance, if we consider the low-risk factor, then popular metrics for risk are volatility and beta, measurement periods vary from less than one year to over five years, and commonly used data frequencies are daily, weekly, and monthly. Instead of selecting one specific low-risk factor, with the risk of data mining, we again prefer to use a diversified mix of low-risk factors measured in various ways.

Factors based on non-standard data sources or with limited history

Leading academics tend to show little interest in factors which require non-standard data sources, or which are only available with a limited data history. An example is our factor which looks at earnings estimate revisions by analysts. This factor has been a key pillar of our models for more than twenty years and has proven to be one of the most powerful alpha generators over this period. However, the academic community treats it with a lot of skepticism because of data reliability concerns. We find that these concerns can be addressed by keeping proper track of security identifier changes, e.g. in case of mergers or acquisitions. Other inconveniences with the analyst revisions factor are that it has a shorter history than standard factors, with data starting only in the 1980s, and that it does not have coverage for all stocks, in particular very small stocks.

Signals from the credit market are another example of a non-standard source with limited history that gets little academic attention. However, we find that using the spread on credit default swaps as a forward-looking low-risk indicator for stocks and the momentum of the corporate bonds of a firm as a signal that augments stock momentum both add value. Another potentially fruitful source of alpha is short interest data, e.g. short interest levels and fees. However, this involves again a non-standard source with limited history.

Another example in this regard is our recently introduced index arbitrage factor. Smart beta factor indices tend to have high turnover at just a few rebalancing days of the year. We find that this results in significant price distortions on the specific stocks involved, in particular in emerging markets, where liquidity is relatively low. We also observe strong reversal effects in subsequent months, which we have been able to exploit with an innovative new alpha factor. The data history for this factor is even shorter than for the previous examples, because large-scale replication of smart beta factor indices is a relatively recent phenomenon. Nevertheless, the empirical results are strong and the economic rationale is compelling.

In sum, factors outside the comfort zone of the academic community are a particularly fruitful hunting ground for alpha. Non-standard factors can be considered as the icing on the factor cake.

The next generation of factors

The last category of factors worth mentioning is the next generation factors, such as those based on big data, machine learning and/or artificial intelligence. An example is our recently introduced news sentiment factor, which is based on the analysis of millions of news articles, that are converted into a quantitative sentiment signal using linguistic algorithms. Although news sentiment is correlated with other (price- and analyst-based) sentiment indicators, we find that it contains unique information above and beyond these more conventional factors.

Big data and ML/AI clearly take 'non-standard' factors to the next level, but also raise concerns about data mining, as these techniques have much more degrees of freedom for fitting the data. We expect abundant research activity in this area in the coming years, and we expect more new alpha factors to emerge from this.

Summary

In this note, we have explained our stance on the debate regarding the factor zoo. We share the concerns expressed in several recent academic papers that many of the hundreds of factors that have been proposed over the past decades either are redundant, lack robustness, or cannot even be replicated. We do not find, however, that the entire factor zoo can be reduced to just a handful of factors. Although the small set of factors used in academic asset pricing models can serve as a very good starting point, that is not the end of the story. In our research, we find evidence of dozens of factors. These include factors that are wrongly dismissed or rejected, multiple factors to capture one broader phenomenon, factors based on non-standard data sources or with limited history, and 'next generation' factors, based on big data, machine learning or artificial intelligence.

That said, for practical implementation purposes, it is common to categorize factors into a small number of strategic composite factors. The **low-risk** factor contains metrics such as volatility and beta, measured using different lookback periods and different data frequencies, but also distress risk indicators, such as distance-to-default and credit spreads. The **value** factor consists of all variables which measure price relative to fundamentals, such as book value, earnings and cash-flows. These ratios can be adjusted for e.g. distress risk or environmental footprint.

The **quality** factor is essentially a mixed bag of company fundamentals, such as profitability, earnings quality and investment patterns. We agree with the academic perspective that these actually appear to be separate, distinct factors, but follow the industry convention to combine them under a common 'quality' header.

For the **momentum** factor it is basically the other way around. Academics and smart beta index providers tend to see momentum as a single factor (price momentum), but in our research we find that it is better understood as an entire family of different sentiment-related factors, most notably **price momentum** and **analyst revisions**. Finally, there is another broad set of **short-term factors**. These are typically ignored by academics and index providers altogether, but we find them to be highly effective for trade timing purposes. This theme includes various reversal phenomena, signals based on short interest and signals derived from trading volume patterns.

In sum, starting from an academic zoo consisting of hundreds of alleged factors, we narrow it down to several dozens that really work, which we organize into a small number of composite factors.

References

Arnott, Rob, Campbell R. Harvey, and Harry Markowitz. 2019. “A Backtesting Protocol in the Era of Machine Learning” Journal of Financial Data Science, vol. 1, no. 3 (Winter): 64-74.

Barillas, Francisco, and Jay Shanken. 2018. “Comparing Asset Pricing Models.” Journal of Finance, vol. 73, no. 2 (April): 715–754.

Blitz, David, Joop Huij, and Martin Martens. 2011. “Residual Momentum.” Journal of Empirical Finance, vol. 18, no. 3 (June): 506–521.

Blitz, David, Matthias X. Hanauer, and Milan Vidojevic. 2017. “The Idiosyncratic Momentum Anomaly”. SSRN working paper, no. 2947044.

Blitz, David, Pim van Vliet, and Guido Baltussen. 2020. “The Volatility Effect Revisited.” Journal of Portfolio Management vol. 46, no. 2 (January): 45-63.

Carhart, Mark M. 1997. “On Persistence in Mutual Fund Performance.” Journal of Finance, vol. 52, no. 1 (March): 57-82.

Cochrane, John H. 2011. “Presidential Address: Discount Rates.” Journal of Finance, vol. 66, no. 4 (August): 1047-1108.

Daniel, Kent D., David Hirshleifer, and Lin Sun. 2019. “Short- and Long-Horizon Behavioral Factors.” Review of Financial Studies (forthcoming).

De Groot, Wilma, Joop Huij, and Weili Zhou. 2012. “Another look at trading costs and short-term reversal profits.” Journal of Banking and Finance, vol. 36: 371-382.

Fama, Eugene F., and Kenneth R. French. 1992. “The Cross-Section of Expected Stock Returns.” Journal of Finance, vol. 47, no. 2 (April): 427–465.

Fama, Eugene F., and Kenneth R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” Journal of Financial Economics, vol. 33, no. 1 (February): 3–56.

Fama, Eugene F., and Kenneth R. French. 1996. “Multifactor Explanations of Asset Pricing Anomalies.” Journal of Finance, vol. 51, no. 1 (March): 55–84.

Fama, Eugene F., and Kenneth R. French. 2008. “Dissecting Anomalies.” Journal of Finance, vol. 63, no. 4 (August): 1653–1678.

Fama, Eugene F., and Kenneth R. French. 2015. “A Five-Factor Asset Pricing Model.” Journal of Financial Economics, vol. 116, no. 1 (April): 1–22.

Goltz, Felix, and Ben Luyten. 2019. “The Risks of Deviating from Academically-Validated Factors.” Scientific Beta White Paper.

Green, Jeremiah, John R.M. Hand, and X. Frank Zhang. 2017. “The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns.” Review of Financial Studies, vol. 30, no. 12 (December): 4389–4436.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. “...and the Cross-Section of Expected Returns.” Review of Financial Studies, vol. 29, no. 1: 5-68.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. “Digesting anomalies: An investment approach.” The Review of Financial Studies, vol 28, no. 3 (March): 650-705.

Hou, Kewei, Chen Xue, and Lu Zhang. 2019. “Replicating Anomalies.” Review of Financial Studies (forthcoming).

Kyosev, Georgi, Matthias X. Hanauer, Joop Huij, and Simon Lansdorp. 2020. “Does Earnings Growth Drive the Quality Premium?.” Journal of Banking and Finance (forthcoming).

Jacobs, Heiko, and Sebastian Müller. 2018. " ...And Nothing Else Matters? On the Dimensionality and Predictability of International Stock Returns." SSRN working paper, no. 2845306.

Jacobs, Heiko, and Sebastian Müller. 2020. "Anomalies across the globe: Once public, no longer existent?." Journal of Financial Economics, vol. 135, no. 1 (January): 213–230.

McLean, R. David, and Jeffrey Pontiff. 2016. “Does Academic Research Destroy Stock Return Predictability?” Journal of Finance, vol. 71, no. 1 (February): 5-32.

Stambaugh, Robert F., and Yu Yuan. 2017. “Mispricing Factors.” Review of Financial Studies, vol. 30, no. 4 (April): 1270–1315.

Important Information

Additional Information for US investors

Additional Information for investors with residence or seat in Canada