ETF Insight: Giving backtests the once over

By Scott Longley
13 January 2020

Backtesting
ETF Insight
ETFs
Research Affiliates
smart beta
The ‘zoo of factors’ has been well-documented by various critics, foremost among them the analysts at Research Affiliates who have been railing against the overfitting of backtests for some time now.

Indeed, with the list of factors that have been ‘discovered’ now reaching over 400, the team at the Newport Beach-based asset allocation consultancy suggest that many of the claims for much smart beta performance do not stand up to scrutiny (https://www.etfstream.com/feature/10043_research-affiliates-not-so-happy-days-for-backtests/).

Or as Campbell Harvey, partner and senior advisor at the firm, suggests many of these factors will have “no economic meaning whatsoever” (https://www.etfstream.com/feature/9387_research-affiliates-harvey-backtesting-is-a-huge-problem-in-the-industry/).
“There are various fundamental variables that can be twisted and given the incentives in academia of the need to be published and the need for significant findings, it means that data-mining has been used much more,” he says.

It leads to the issue of overfit – that is, a backtest which performs “in sample” i.e. in the research phase, but then fails to perform out-of-sample or when it has gone live.

But as Harvey points out, backtests are still a vital and necessary element of all factor research. “You have an idea and hopefully this idea is based on a solid economic foundation, indeed, that is the first step,” he says. “You need to have an economic story as to why this factor strategy makes sense. So before you look at the data, you come up with the economic foundation.”

But other more dangerous approaches are possible and this is where the reputation of backtests has been dragged through the mud. One such is the use of machine-learning techniques to sift through the data and “discover” a trading strategy or factor. “That sort of implementation is much more likely to deliver a factor without any economic foundation and is likely to look good purely by luck,” Harvey says.

Another tell-tale sign of an untrustworthy backtest methodology lies in the process of splitting the sample. As Harvey explains, this is where a researcher will estimate a model with the first half of the data and then validate it in the second half. “Now you have to be very careful with that,” he says. The second half of the data will be historic, covering a period where “we all know what happened, what did well and what did not do well”.

“So it is no big deal to specify a model that does well in the first half of the sample with variables that you know will do well in the second.”

**Raking over the coals**

But how can you discriminate between what might be a valid backtest and one that is suspect? In this instance, there is no replacing a thorough due diligence process, one that can ask the right questions to uncover spurious factors and questionable backtesting.
It comes down to ensuring that first, the economic foundations of the strategy are in place and second that the way the statistics are applied stands up to scrutiny.

These are what Harvey sketches out as the rules for backtests. It starts with the economic foundation of the idea. Or as Harvey put it, “you produce a result, it looks good but what is the mechanism to show it should work?” Is this a mechanism that was true before the backtest or was it cooked up after?

“That is very important,” he says. “Often people will data mine, they will find something, then they ex-post rationalise it with some economic framework for why this worked when potentially this is just a lucky discovery.”

Then he suggests questions need to be asked about how many of these factors were tried. “Is there any evidence that this is just a data-mining expedition? Within the research programme, how many things did they actually try? The more you try, the higher the chance that what you have discovered is a false discovery.”


Then there are the questions about validation and whether and how the sample was split. In particular, Harvey says a common mistake is where there is an in-sample period and an out-of-sample period. “You try something in sample, it fails in the other sample, then you try number two strategy and so on until, say, strategy 20 works in both sample. That is false cross-validation.”

Finally, there are questions around the date of the sample period. Harvey says he has seen some backtests where the Global Financial Crisis in 2008 is deliberately left out. “Why do that? The reason often is the strategy does not work in the crisis, but there is no way you can do that. It doesn’t make sense to exclude it.”

Giving diligence its due
As he says, there are many questions that can be asked but he admits that the job of “policing” backtests is very difficult. There are some checks and balances, however. In academia, Harvey says editors are now “very much more sceptical of new factors. In the literature the hurdle has gone up very significantly,” he says. “It is a self-policing mechanism.”

For practitioners, meanwhile, incentives much more aligned. “In their world, it is different,” he adds. “If you data-mine a factor and it looks really good and you set up a fund or ETF based on it, and it fails, then your reputation is damaged. This is particularly true for hedge funds. There is a self-enforcing mechanism that you only want to take strategies that are really robust. Your reputation is on the line.”

However, to ensure robustness, Harvey says the culture of the company needs to have “dis-incentives for overfitting”.

“The worst thing is a high reward for research that does well and that potentially penalises those that do not work. That is an incentive for overfitting.”

But at heart, this is an issue of due diligence on the part of the investor and their advisers. “You need to be sceptical,” he says. “Was this backtest part of many other strategies that were tried? What is the economic foundation of the backtest? Was the economic foundation developed after the fact or was the research founded on the idea originally proposed?”

Moreover, questions need to be asked of the research firm or fund provider. “What is the research quality at the firm? Its reputation? The track record of the firm and the CIO?

“This due diligence is more important today than it was in the past,” Harvey concludes. “There is less low-hanging fruit for researchers and it is much more difficult to discover a new factor. Hence the degree of scepticism should be high and this has been made worse by technological advances.”

ETF Insight is a new series brought to you by ETF Stream. Each week, we shine a light on the key issues from across the European ETF industry, analysing and interpreting the latest trends in the space. For last week’s insight, click here.
Most popular Features

01 BlackRock’s decision to merge mutual fund and ETF sales teams shows direction of travel for ........

02 Australia’s best ETF: Vanguard’s VAS or VanEck’s MVW? (https://www.etfstream.com/feature/8009_australias-best-etf-...)

03 ETF Insight: Inside First Trust, the silent ETF powerhouse (https://www.etfstream.com/feature/8117_etf-insight-is-jim-bowe...)

04 Five top UK ETFs (https://www.etfstream.com/feature/6274_five-top-uk-etfs/)

05 Five top Asian ETFs (https://www.etfstream.com/feature/7736_five-top-asian-etfs/)
06 Can BlackRock match Vanguard’s brave fee cut? (https://www.etfstream.com/feature/7905_can-blackrock-match-...)


09 Four top property ETFs (https://www.etfstream.com/feature/6119_four-top-property-etfs/)

10 Five top Europe ETFs (https://www.etfstream.com/feature/6486_five-top-europe-etfs/)

Never miss an update again
First Name Last Name Email

Company

- Weekly Newsletter - ETF Insights

Submit

By submitting your email address, you agree to receive email updates from ETF Stream in accordance with our Privacy Policy. Unsubscribe at any time

Privacy Policy (/privacy-policy/)
Cookie Policy (/cookie-policy/)
Terms of Service (/terms-of-service/)
Terms and Conditions (/terms-and-conditions/)

News (/news/)
Features (/features/)
Events (/events/)