Arnott, Harvey: machine learning dangerous when data thin

Experts warn ML should be used “for its correct purpose”

Rob Arnott and Campbell Harvey – two of the best known experts in quant investing – have warned investors against using machine learning to derive investment strategies from too-thin data.

According to Arnott, using sparse data to train “powerful” machine learning algorithms is akin to driving a Ferrari on an off-road dirt track.

“If you visit the data often enough and in enough depth [using machine learning], you'll find all sorts of things that look marvellous,” Arnott says. “It doesn’t mean they are marvellous. It means you've ignored the things that don't work and you're data mining.”

Arnott, founder of smart beta pioneer Research Affiliates, has commented in the past about data mining in statistically-grounded factor investing, claiming popular quant strategies look better in backtests than they should. His comments have been controversial at times, drawing a stinging response from AQR Capital Management founder Cliff Asness, who accused Arnott of treating AQR as a “whipping boy” in comments that appeared in a Bloomberg article in 2017.
Recently, he has collaborated with Harvey, a professor of finance at Duke University, and Nobel-winning economist Harry Markowitz, on a backtesting protocol for machine learning.

Arnott and Harvey were speaking together in an interview with Risk.net.

Machine learning has been successfully applied in physical sciences, in some cases generating new discoveries directly from data. It is now being widely used – or explored – by buy- and sell-side firms in financial markets as a way to build investment and trading strategies. But Arnott and Harvey note that scientists in fields such as particle physics have been able to draw on information sets consisting of quadrillions of data points. In contrast, monthly data for most financial markets may sum to fewer than 700 data points.

“One commonality in these applications is a massive amount of data. In financial economics we are not blessed with the amount of data these other sciences have,” Harvey says.

**Don’t torture the data**

The two experts say the use of more stringent statistical tests in research is critical to account for the likelihood of finding patterns by fluke. These tests should consider the probability of so-called false positives, bearing in mind the number of strategies tested – even sometimes accounting for strategies the researchers did not test.

At the same time, researchers must strike a balance: setting tougher tests to avoid data mining but making sure not to throw out ideas that work. In some instances – such as where studies are grounded in strong economic reasoning – researchers can justifiably lower the threshold for approving a strategy, Arnott and Harvey say.

“The simpler the model, the less you’ve tortured the data, the lower the threshold for significance that you should feel comfortable using,” Arnott says.

“The threshold that you apply depends upon the economic foundation of the idea,” Harvey adds. “If you apply the same high threshold for everything, you're going to miss some good strategies.”

The two experts also warn poor research culture will encourage the misuse of machine learning. The incentives in investment research “are all aligned towards maximising backtest efficacy, not live asset efficacy”, Arnott says.

Academics and practitioners are often rewarded for coming up with strategies that work rather than carrying out thorough research regardless of the results.
The “most toxic” thing managers can do is to reward researchers whose strategies worked over those whose research has come to nothing, Harvey says. “In scientific discovery there are more things that don’t work [than do]. You need to reward the quality of the actual work.”

In a 2014 study Harvey and his collaborators found 158 false positives in only 296 published papers explaining returns in equities. Harvey has since led calls for financial researchers to rethink how they establish the statistical significance of their findings.

**Alternative potential**

Despite their warnings, Arnott and Harvey are optimistic about some aspects of machine learning.

The use of machine learning to generate signals from alternative data, for example by gauging sentiment from social media posts is “potentially very promising”, Harvey says, though he adds a note of caution for investors about the “millions of ways potentially to do this” and the high probability that “something’s going to work by chance”.

For high-frequency traders and those using machine learning to help manage trading, there is “just enough data” to make machine learning techniques applicable, Arnott says. “There’s not data like you find in the Cern experiments that identified the Higgs boson. But it is millions of data samples. And it is enough to be potentially pretty interesting.”

Machine learning as a tool must be applied thoughtfully, he concludes: “You wouldn’t use a hammer to adjust the fit of a windshield on a car. You’d use it to drive in a nail. Use the tool for its correct purpose.”

Copyright Infopro Digital Limited 2017. All rights reserved.

You may share using our article tools. Printing this article is for the sole use of the Authorised User (named subscriber), as outlined in our terms and conditions - [https://www.infopro-insight.com/terms-conditions/insight-subscriptions/](https://www.infopro-insight.com/terms-conditions/insight-subscriptions/)