BARRON'S

Focus on Funds

News and analysis on ETFs, mutual funds and hedge funds.

April 17, 2014, 8:58 A.M. ET

Is a Too-Perfect ETF Backtest Fraud?

By Brendan Conway

It's negligence, or worse, when an investment manager's innovative-looking strategy is the result of too much quantitative trial-and-error.

That's the argument in a notable new study flagged by **Stephen Foley** of the *Financial Times*. "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance" argues that what happens behind the scenes in the development of quantitative strategies is a major problem in investment management.

"Backtest" simply means reviewing historical returns to try to divine how a new strategy might perform in the future. The method has become bread-and-butter in the launch of many new ETFs.

Investors don't know how many hypotheses managers examined before finding the perfect-looking backtest, a process which turns out to matter greatly, write **David H. Bailey, Jonathan M. Borwein, Marcos Lopez de Prado** and **Qiji Jim Zhu**. "The higher the number of configurations tried, the greater is the probability that the backtest is overfit," they write. "Overfit" means the data has been tortured until it yielded something that looks nice.

If an investment process is driven by what looks good historically, there's a greater chance the attractive-looking result is just a fluke.

Getty Images

Not quite like this … But still troubling

Sure enough, a **Vanguard Group** study found a while back that backtested ETFs — which look great in the historical data — on average lagged the market after the real-world launch.

From Foley's discussion:

> [By] failing to apply mathematical rigour to their methods, many purveyors of quantitative investment strategies are, deliberately or negligently, misleading clients.
>
> It is reasonable to want to test a promising investment strategy to see how it would have performed in the past. The trap comes when one keeps tweaking the strategy until it neatly fits the historical data. Intuitively, one might think one has finally hit upon the most successful investment strategy; in fact, one is likely to have hit only upon a statistical fluke, a false positive.

This is the problem of "over-fitting", and even checks against it – such as testing in a second, discrete historical data set – will continue to throw up many false positives, the mathematicians argue.

How do you get from here to fraud? The study's authors liken the process to an investment newsletter editor who plays the odds in a mendacious manner. First he emails a forecast of a market gain to one large audience. He also predicts a market loss to another set of email recipients. Next, after one call succeeds and the other fails, he winnows the list down to recipients who got the winning side. He repeats this shenanigan until he looks like a genius to a small audience.

Such a process, of course, is fraudulent:

> To half of them he predicts that markets will go up, and to the other half that markets will go down. After the month passes, he drops from his list the names to which he sent the incorrect forecast, and sends a new forecast to the [remaining investors]. He repeats the same procedure n times, after which only x names remain. These x investors have witnessed n consecutive infallible forecasts and may be extremely tempted to give this investment manager all of their savings. Of course, this is a fraudulent scheme based on random screening: The investment manager is hiding the fact that for every one of the x successful witnesses, he has tried [2 to the n power] unsuccessful ones[.]