

FT Alphaville AQR Capital Management LLC

Are bigger AI models better stock pickers? Maybe, but probably not

Complexity ain't all that, wonks say

$$b = E[x^2]E[x]^{-1}(E[x] + y)^{-2}.$$

Thus,

$$\begin{aligned} & \int x dH(x) \left(z\tilde{m} + \int \frac{c\tilde{m}x}{(1 + \tilde{m}x)^2} \right) - \int \frac{x dH(x)}{1 + \tilde{m}x} \\ &= E[x] \left(z(ac^{-1} + bc^{-2}) + \int c(ac^{-1} + bc^{-2})x(1 - 2\tilde{m}x - (\tilde{m}x)^2 + 4(\tilde{m}x)^2) \right) \\ & - \int x dH(x)(1 - \tilde{m}x + (\tilde{m}x)^2) + O(c^{-3}) \\ &= E[x] \left(cy(ac^{-1} + bc^{-2}) + \int (a + bc^{-1})(x - 2(ac^{-1} + bc^{-2})x^2 + 3(ac^{-1})^2x^3) \right) \quad (\text{IA180}) \\ & - \int dH(x)(x - (ac^{-1} + bc^{-2})x^2 + (ac^{-1})^2x^3) + O(c^{-3}) \\ &= E[x]byc^{-1} + bE[x]^2c^{-1} - 2E[x]a^2c^{-1}E[x^2] + ac^{-1}E[x^2] + O(c^{-2}) \\ &= E[x]E[x^2]E[x]^{-1}(E[x] + y)^{-2}yc^{-1} + E[x^2]E[x]^{-1}(E[x] + y)^{-2}E[x]^2c^{-1} \\ & - 2E[x](E[x] + y)^{-2}c^{-1}E[x^2] + (E[x] + y)^{-1}c^{-1}E[x^2] + O(c^{-2}) \\ &= \frac{2E[x^2]y}{(E[x] + y)^2} > 0. \end{aligned}$$

Thus complete the proof, it suffices to show that the function $-1/\tilde{m}$ is concave in c .

Page 52 of the proof section of [The Virtue of Complexity in Return Prediction](#)

Bryce Elder

Published JUL 3 2025

In December 2021, Bryan Kelly, head of machine learning at quant house AQR Capital Management, put his name to an academic paper that caused quite a stir.

[The Virtue of Complexity in Return Prediction](#) — co-authored by Kelly with Semyon Malamud and Kangying Zhou — found that complex machine-learning models were better than simple ones at predicting stock prices and building portfolios.

The finding was a big deal because it contradicted one of machine learning's guiding principles, the [bias-variance trade-off](#), which says the predictive power of models weakens as they grow beyond some optimum level. Given too many parameters to play with, a bot will tend to overfit its output to random noise in the training data.

But Kelly and his co-authors concluded that, surprisingly, more variables always improve returns. Available computing power is the only limit. Here's a video of Kelly explaining to Wharton School in 2023 that the same principles that apply to the multi-billion-parameter models powering ChatGPT and Claude AI also apply to accuracy in financial forecasting.

A lot of academics hated this paper. It relies on theoretical analysis “so narrow that it is [virtually useless](#) to financial economists,” says Stanford Business School's Jonathan Berk. Performance depends on sanitised data that wouldn't be available in the real world, according to some [Oxford University researchers](#). Daniel Buncic, of Stockholm Business School, says the larger models tested by Kelly et al [only outperform](#) because the simple models are “artificially” made to look bad in two unorthodox ways.

This week, Stefan Nagel of University of Chicago has joined the pile-on. His paper — [Seemingly Virtuous Complexity in Return Prediction](#) — argues that the “stunning” result shown by Kelly et al is . . .

. . . effectively a weighted average of past returns, with weights highest on periods whose predictor vectors are most similar to the current one.

Nagel challenges the paper's central conclusion that a very complex bot can make good predictions based on just a year of stock performance data.

The finding was rooted in an AI concept known as [double descent](#), which says deep learning algorithms make fewer mistakes when they have more variable parameters than training data points. Having a model with a huge number of parameters means it can fit perfectly around the training data.

According to Kelly et al, this all-enveloping-blob approach to pattern matching is able to pick out the predictive signals in very noisy data, such as a single year of US equities trading.

Rubbish, says Nagel:

In short training windows, similarity simply means recency, so the forecast reduces to a weighted average of recent returns — essentially a momentum strategy.

Crucially, the algorithm isn't advising a momentum strategy because it has sensed it'll be profitable. It just has recency bias.

The bot “simply averages the most recent few returns in the training window, which correspond to the predictor vectors most similar to the current one,” Nagel says. It “does not learn from the training data whether momentum or reversal dynamics are present; it mechanically imposes a momentum-like structure regardless of the underlying return process.”

Outperformance shown by the 2021 study “thus reflects the coincidental historical success of volatility-timed momentum, not predictive information extracted from training data,” he concludes.

We're skipping a *lot* of detail. Any reader wanting to know about the mechanics of kernel scaling by Random Fourier Features would be better served by an author who knows what they're talking about. Our main interest is in AQR, the \$136bn-under-management quant, which wears its academic roots with pride.

Kelly acts as [AQR's frontman for better investment through machine learning](#): His “Virtue of Complexity” paper is [on the AQR website](#), alongside some more [circumspect commentary](#) from his boss Cliff Asness on the value of machine generated signals.

The savaging of Kelly et al — including by a professor at the University of Chicago, both his and Asness's alma mater — isn't a great look. But since straightforward momentum strategies have historically been among the things that [AQR does best](#), maybe this demystification of academic AI hype is no bad thing for investors.

Update (16:00 BST):

An AQR spokesperson tells us by email:

The papers referenced do not in any way undermine either the theoretical or empirical claims in KMZ (Kelly, Malamud and Zhou). None attempt a rigorous dispute of the theory in KMZ, which is its main focus. These papers rely on empirical arguments that nitpick narrow special cases of the empirical results in KMZ and ignore the mountain of evidence in theirs and others' financial machine learning research.

The empirical dominance of large models has been shown in every area of ML by heavyweight ML academics' research, which has been conducted throughout the natural and applied sciences. Language and image modeling are most well-known applications that exemplify the success of large models. Do we really think that finance, economics, or other social sciences are special? The work of Kelly and team shows otherwise.

[Copyright](#) The Financial Times Limited 2025. All rights reserved.

Follow the topics in this article

Artificial intelligence

AQR Capital Management LLC

FT Alphaville

Bryce Elder

