1-18-2019

Be Wary of Black-Box Trading Algorithms

Gary N. Smith
Pomona College

Be Wary of Black-Box Trading Algorithms

Gary Smith
Fletcher Jones Professor
Department of Economics
Pomona College
425 N. College Avenue
Claremont CA 91711
gsmith@pomona.edu

Abstract

Black-box algorithms now account for nearly a third of all U. S. stock trades. It is a mistake to think that these algorithms possess superhuman intelligence. In reality, computers do not have the common sense and wisdom that humans have accumulated by living. Trading algorithms are particularly dangerous because they are so efficient at discovering statistical patterns—but so utterly useless in judging whether the discovered patterns are meaningful.

running title: Back-Box Trading Algorithms

keywords: algorithmic trading, black box trading, quants, artificial intelligence

word count: 5,418

Be Wary of Black-Box Trading Algorithms

A computer algorithm is a specific sequence of steps for performing a task, such as finding a square root or spell-checking a word. There are many stock market trading algorithms, including programs that try to reduce the costs of executing trades or try to make a profit by arbitraging price discrepancies across different exchanges.

My focus is on trading algorithms that try to discover profitable statistical patterns, including timing trades (for example, stock prices usually go up after a surge in calm words on Twitter) and convergence trades (for example the term structures of German and French interest rates are related). These kinds of trading algorithms are typically black box in that, once the code is written, humans do not interfere with the algorithm or know why specific trades are made.

A 2017 hedge fund prospectus boasted that their "fully automated portfolio [is] run via computer algorithms…. All trading is conducted through complex computerized systems, eliminating any subjectivity of the manager" (RK Capital 2017). This was evidently thought to be a feature, not a flaw, because computers are smarter than humans. Many investors apparently agree. Black-box algorithms now account for nearly a third of all U. S. stock trades (Zuckerman and Hope 2017).

Computer "intelligence" is, in fact, very different from human intelligence. Trading algorithms do not understand the world in any meaningful sense, and are consequently risky because they are so efficient at discovering statistical patterns—but so utterly useless in judging whether the discovered patterns are consequential or coincidental.

Introduction

The spread of the internet in the 1990s sparked the creation of thousands of internet-based companies, popularly known as dot-coms. Some dot-coms had good business plans and became successful companies. Most did not. In too many cases, the idea was simply to start a company with a *dot-com* in its name, sell it, and walk away rich. Cooper, Orlin, and Rau (2001) found that, on average, companies that did nothing more than add *.com*, *.net*, or *internet* to their names nearly doubled the price of their stock.

The same thing is happening now with artificial intelligence (AI). In 2017 the Association of National Advertisers (2017) chose "AI" as the Marketing Word of the Year. AI has become fashionable in investing, too, with black-box trading algorithms promising more than can be delivered. A dot-com name does not guarantee success, nor does an AI label.

Data Mining

In 2008, Chris Anderson, editor-in-chief of *Wired*, wrote an article with the provocative title, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." Anderson argued that,

> *With enough data, the numbers speak for themselves…. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.*

This declaration seemed intentionally controversial at the time, but it was prescient, as many have abandoned the scientific method and come to believe that correlation supersedes causation.

The scientific method begins with a plausible theory and then collects appropriate data to test this hypothesis. The scientific method was the foundation for the triumph of science over superstition. Today, however, it has become fashionable to turn the scientific method on its head by scrutinizing available data "to reveal hidden patterns and secret correlations" (Sagiroglu and

Sinanc 2013). When a pattern is found, either make up a theory after the fact or assert that theories are unnecessary (Fayyad, Piatetsky-Shapiro, and Smyth, 1996; Cios, Pedrycz, Witold, and Kurgan 2007; Begoli and Horsey, 2012). Some go so far as to argue that using expert knowledge of the phenomena being modeled is not only unnecessary, but limiting (Piatetsky-Shapiro 1991).

After Pepperdine University invested 10% of its portfolio in quant funds in 2016, the director of investments argued that, "Finding a company with good prospects makes sense, since we look for undervalued things in our daily lives, but quant strategies have nothing to do with our lives." (Zuckerman and Hope 2017) In truth, the absence of the wisdom and common sense acquired by being alive is an argument against algorithmic trading.

The now commonplace idea that analyses begin with data rather than expert opinion goes by a variety of names, including data mining, knowledge discovery, knowledge extraction, and information harvesting. The data are mined to discover theories, extract knowledge, and harvest information. Data mining is the cornerstone of black-box trading algorithms.

There was a time when data mining was considered a misdeed, akin to plagiarism. As Nobel Laureate Ronald Coase (1988) lamented decades ago: "If you torture the data long enough, it will confess." His caustic comment is ignored today by people who don't understand that those who ransack data looking for statistical patterns will surely find some—so, their discoveries demonstrate nothing more than that data were ransacked.

In the opening lines to a forward for a book on using data mining for knowledge discovery, a computer scientist (Kecman 2007) wrote, without evident irony,

> *"If you torture the data long enough, [it] will confess," said 1991 Nobel-winning*

> *economist Ronald Coase. The statement is still true. However, achieving this lofty goal is not easy. First, "long enough" may, in practice, be "too long" in many applications and thus unacceptable. Second, to get "confession" from large data sets one needs to use state-of-the-art "torturing" tools. Third, Nature is very stubborn — not yielding easily or unwilling to reveal its secrets at all.*

Coase did not intend his comment to be a lofty goal worth seeking, but as a succinct criticism of the practice of pillaging data in search of statistical significance (Tullock 2001).

The perils of data mining are summarized by the Texas Sharpshooter Fallacy. In one variant, an avowed marksman demonstrates his prowess by painting thousands of bullseyes on the side of a barn. After he fires his gun, he finds the bullseye he hit and paints over all the other bullseyes. Since he will surely hit one bullseye, this proves nothing at all.

In investing, this corresponds to testing a large number of theories and selectively reporting a small fraction of the results. For example, back before it went bankrupt, the L. F. Rothschild investment bank reported that during the preceding six Dragon years in the Chinese zodiac calendar, the U. S. stock market had gone up four times and down twice (Allan 1976). No doubt, the misguided analyst behind this nonsense looked at each of the 12 zodiac signs (rat, ox, tiger, and so on). One sign is bound to have the highest coincidental correlation with up-years in the stock market, and this is the sign that was reported.

In another example, Bolen, Mao, and Zeng (2011) reported that a data-mining analysis of nearly 10 million Twitter tweets during the period February to December 2008 found that an upswing in "calm" words was often followed an increase in the Dow Jones average up to six days later.

These Texas Sharpshooters looked at seven different predictors: an assessment of positive versus negative moods and six mood states (calm, alert, sure, vital, kind, and happy) with, no doubt, considerable flexibility in assigning mood states to various tweets. Is *nice* a calm, kind, or happy word? Is *yes!* an alert, sure, or vital word? The researchers also considered several different days into the future for correlating with the Dow. Finally, why did they use data from February to December 2008? What happened to January? Why did a 2011 paper use 2008 data? Did the discovered patterns only exist during that peculiar period, with words, moods, and days that were selected after the data had been tortured? Even the lead author admitted that he had no explanation.

The second variation of the Texas Sharpshooter Fallacy is when the inept marksman fires his gun at a blank wall, and then draws a bullseye around the bullet hole. Since there is always a bullet hole to draw a bullseye around, this, too, proves nothing at all.

In investing, this corresponds to rummaging around in stock market data with no clear purpose in mind, and discovering a pattern. This was probably the origin of the Super Bowl Stock Market Predictor (Koppett 1978), which claims that the stock market goes up in years when the team that wins the Super Bowl is in the National Football Conference (NFC) or is in the American Football Conference (AFC), but was once in National Football League (NFL).

The stock market has nothing to do with the outcome of a football game. The accuracy of the Super Bowl Indicator is an amusing coincidence bolstered by the fact that the stock market usually goes up and the NFC usually wins the Super Bowl. The correlation is made more impressive by the gimmick of counting the Pittsburgh Steelers, an AFC team, as an NFC team because Pittsburgh won the Super Bowl several times when the stock market went up.

The irony is that Leonard Koppett, the man who created the Super Bowl Indicator intended it to be an amusing demonstration of the fact that correlation is not causation:

> *What does all this mean? Absolutely nothing on any rational level—and that's exactly the point. Just because two sets of numbers coincide in some way, don't leap to the conclusion that one set "causes" the other.* (Koppett 1978)

He was astonished when people took the Super Bowl Indicator seriously: "It's a joke! I meant the whole thing as a satire on the fallibility of human statistical reasoning. It's too stupid to believe." (Zweig, 2011) Among the credulous were two finance professors who published an article in the *Journal of Finance* arguing that, "although the theoretical relationship connecting the Super Bowl and subsequent stock market movements is not obvious," the statistical relationship was highly statistically significant and would have very profitable if followed by investors (Krueger and Kennedy 1990). Spoken like true data miners.

I was told recently that some otherwise sophisticated investors still believe in the Super Bowl Indicator. They are bullish on stocks in 2018 because an NFC team, the Philadelphia Eagles, won the Super Bowl.

People used to have to work hard to torture data in search of patterns. Now it is far too easy. Computer trading algorithms can search for as many patterns in a second as humans can in weeks, months, or even years. This is not useful progress.

Real Intelligence

Computers have perfect memories and can input, process, and output enormous amounts of information at unfathomable speeds. These features allow computers to do truly superhuman feats: to work tirelessly on assembly lines, solve complicated systems of mathematical equations, find detailed directions to bakeries in unfamiliar towns.

Computers can tell us the day of the week Abraham Lincoln was born, the capital of Bulgaria, and the last time Arsenal won the Premier League. Computers are also relentlessly consistent. Asked to calculate the square root of 76,073,284, a computer will give the correct answer (8,722) essentially immediately, every time it is asked. Ask any human who is not a math freak the same question, and the answer will be slow and unreliable. It is tempting to think that computers are smarter than humans because they do some very difficult tasks better than humans.

Some of the allure of algorithmic trading stems from the success of computer programs competing against humans in checkers, chess, Go, and other games. These computer programs perform narrowly defined tasks that have clear goals (in chess, checkmate the opponent) stunningly, but they don't mimic human thinking, which involves a creative recognition of the underlying principles that lead to victory. Instead, game-playing algorithms are built to exploit a computer's strengths—that computers can make calculations quickly, have an infallible memory, and obey rules flawlessly.

Despite their freakish, superhuman skill at board games, computer programs do not possess anything resembling human wisdom and common sense. These programs do not have the general intelligence needed to deal with unfamiliar circumstances, ill-defined situations, vague rules, and ambiguous, even contradictory, goals. Deciding whether to accept a job offer, who to marry, or which stock to buy is very different from recognizing that moving a bishop three spaces will checkmate an opponent—which is why it is perilous to trust computer programs we don't understand to make decisions for us, no matter how fast they calculate square roots or how well they do at board games.

The Winograd Schema Challenge

The Achilles' heel of black-box trading algorithms is that they do not know, in any consequential sense, what words mean; so, they cannot assess whether the patterns they find are real or spurious. Computer algorithms data mine spectacularly well, but have no real understanding of the results of their data mining.

One way to recognize the inadequacies of computer algorithms is to consider the challenges identified by Stanford computer science professor Terry Winograd (1972) that have come to be known as Winograd schemas. Here is an example from a collection compiled by Davis (2017), a computer science professor at New York University:

I can't cut that tree down with that axe; it is too [thick/small].

If the bracketed word is *thick*, then *it* refers to the tree; if the bracketed word is *small*, then *it* refers to the axe. These kind of sentences, with more than one noun and alternate words that identify which noun is being referenced by a pronoun, are understood immediately by humans but are very difficult for computers because computers do not have the real-world experience to place words in context.

When we see a tree, we know it is a tree. We might compare it to other trees and think about the similarities and differences between fruit trees and maple trees. We would not be surprised to see a squirrel run up a pine tree or a bird fly out of a dogwood tree. We might remember planting a tree and watching it grow year by year. We might remember cutting down a tree or watching a tree being cut down.

A computer does none of this. For a computer, there is no significant difference between *tree*, *tiger*, and *eg74w*, other than the fact that they use different symbols. A computer can spellcheck the word *tree*, count the number of times the word *tree* is used in a story, and retrieve facts about trees, but computers do not understand what trees are in any relevant sense, and do not respond to the word *tree* or a picture of a tree the way humans do.

From their life experiences, humans know that it is hard to cut down a tree if the tree is thick or the axe is small. Computers struggle because they have no life experiences to recall. They do not really know what a *tree* is, or an *axe*, or what *cutting down* means.

There is a Winograd Schema Challenge with a $25,000 prize for a computer program that is 90 percent accurate in interpreting Winograd schemas (Levesque, Davis, and Morgenstern 2012). In the 2016 competition, the expected value of the score for guessing was 44 percent correct (some schemas had more than two possible answers). The highest computer score was 58 percent correct, the lowest 32 percent, a variation that may have been due more to luck than to differences in the competing programs' abilities.

If computers do not know what words mean, they cannot possibly evaluate the plausibility of discovered statistical patterns.

Deep Neural Networks

Many computer programs now use deep neural networks (DNNs) that are inspired by the neurons in human brains. However, DNNs do not mimic human brains because we have barely scratched the surface in trying to figure out how human brains work. DNNs are more complicated and sound sexier than earlier algorithms, but they are still just computer programs that identify and manipulate patterns.

DNNs have improved language translation, visual recognition, and other tasks, but they are still limited by the reality that, unlike human brains, computers do not truly understand words, images, life. For example, a language translation program that identifies key words and phrases in a sentence, finds matching words and phrases in another language, and puts the matches in a grammatically correct order is *not* reading or writing and it is not trying to convey meaning. That is why the results are sometimes perfect and, other times, astonishingly bad (Hofstadter 2018).

Similarly, visual-recognition algorithms are very granular, analyzing pixels instead of concepts, and the results are very brittle. Putting graffiti on a photograph of a stop sign or even changing a few pixels in a picture of a stop sign—alterations that would not be noticed by humans—can cause state-of-the-art DNNs to fail miserably (Evtimov, Eykholt, Fernandes, Kohno, et al. 2017; Su, Vargas, and Kouichi 2017). Mapping pixels is not the same as knowing what a stop sign is.

Nguyen, Yosinski, and Clune (2015) demonstrated something even more surprising. In addition to making nothing out of something (like a computer not recognizing a stop sign), computers can make something out of nothing by misinterpreting meaningless images as real objects. For example, state-of-the-art DNNs misidentified a series of black and yellow lines as a school bus, completely ignoring the fact that there were no wheels, no door, and no windshield in the picture, because computer algorithms do not know in any relevant sense what a school bus is.

Sharif, Bhagavatula, Bauer, and Reiter (2016) reported that the state-of-the-art deep neural network programs used in facial biometric systems can be fooled by persons wearing colorful eyeglass frames. One of the authors, a white male, was misidentified as Milla Jovovich, a white female, 88 percent of the time, and another author, a 24-year-old Middle Eastern male, was misidentified as Carson Daly, a 43-year-old white male, 100 percent of the time—all because the eyeglass frame colors led the computer program astray. Humans do not make such mistakes because we know what eyeglass frames are, and we know that we should look past the frames to identify the person we see. Computers know none of this; they just match pixels as best they can.

A Knowledge Discovery

AI stock programs are susceptible to analogous mistakes because the algorithms do not understand in any real sense the data that they manipulate and torture. Numbers are just numbers and labels are just words.

To demonstrate this concretely, I analyzed daily observations on 100 potential explanatory variables in 2016 to see if a data-mining algorithm could uncover a simple model for predicting the level of the S&P 500 the next day. Considering all 100 possible explanatory variables, I used a multiple regression algorithm to estimate 9,900 models with two explanatory variables.

It would have taken me many months to estimate the parameters of these 9,900 models using a pencil and paper. It took my computer a few seconds. The best of these 9,900 estimated models used variables 58 and 94:

$$P = 1{,}640.64+1.83X_{58} + 3.62X_{94}$$

The correlation between the predicted and actual values of the S&P in shown in Figure 1 is an impressive and highly statistically significant 0.93. I should evidently let my algorithm buy stocks when it predicts an increase in the S&P 500 the next day and sell when it predicts a decrease.

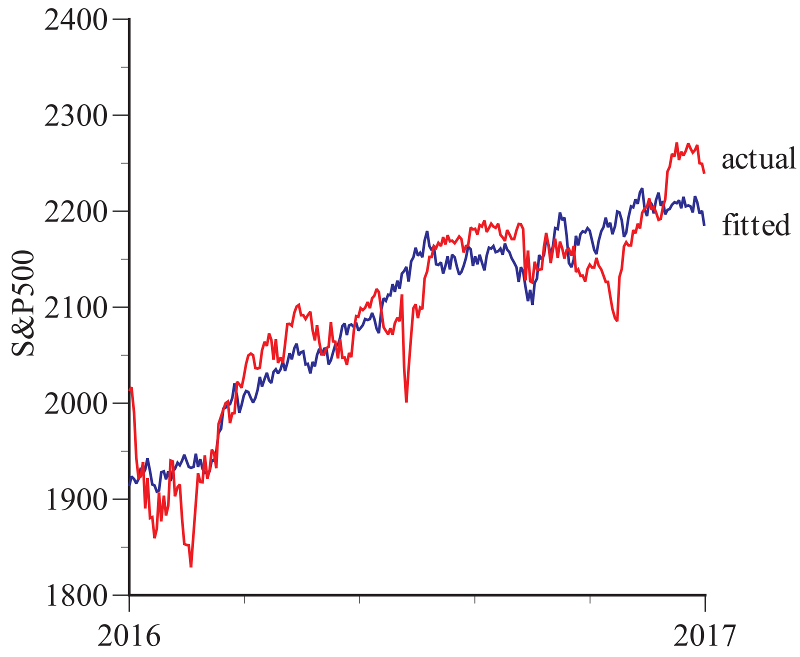

Figure 1 Some Knowledge Discovery For Stock Prices

What are these two predictors for the S&P 500? Suppose that they are the daily high temperature in Curtin, Australia, and the daily low temperature in Antelope, Montana. A human would know that this is nonsense. An algorithm would not. Humans know what stock prices and temperatures are. They know U. S. stocks are not made more or less valuable by the high or low temperatures in these two small cities, one of which is in Australia. An algorithm would not know this, because a computer cannot comprehend what these data are.

A computer does not know what a stock is. It could retrieve a definition of *stock*, though it might be a different kind of stock; perhaps merchandise, animals, or bouillon. Even if the computer program found the correct definition of *stock*, it would not know what the words in the definition mean, though it could retrieve definitions of these words, too, and then definitions of the words in those definitions. Beyond retrieving definitions, a computer does not know, in any real sense, what a stock is, what a stock price is, or why stock prices go up and down. Nor does it know what the high and low temperatures in Curtin and Omak are or whether they might plausibly be related to U. S. stock prices.

A computer search for the words *stock prices* and *Australian temperatures* is unlikely to turn up anything that the computer would interpret as supporting or contradicting the statistical pattern it discovered, and, if it did find anything, the computer would be hard-pressed to assess the reliability of what it found. In addition, the whole claim of “knowledge discovery” is that computers will discover new, previously unknown patterns and relationships. By definition, a knowledge discovery is not something that has already been reported. How can a computer program that does not understand words tell whether its knowledge discovery makes sense? It cannot.

This trading model was selected after estimating 9,900 models with 2016 data and identifying the most accurate model. Because it was based on data, rather than logic, we shouldn’t expect it to work very well in predicting stock prices in 2017. Figure 2 shows that the accuracy in 2017 is –0.54. Yes, that is a negative sign. When the model predicted an uptick or downtick in stock prices, the opposite was likely to occur.

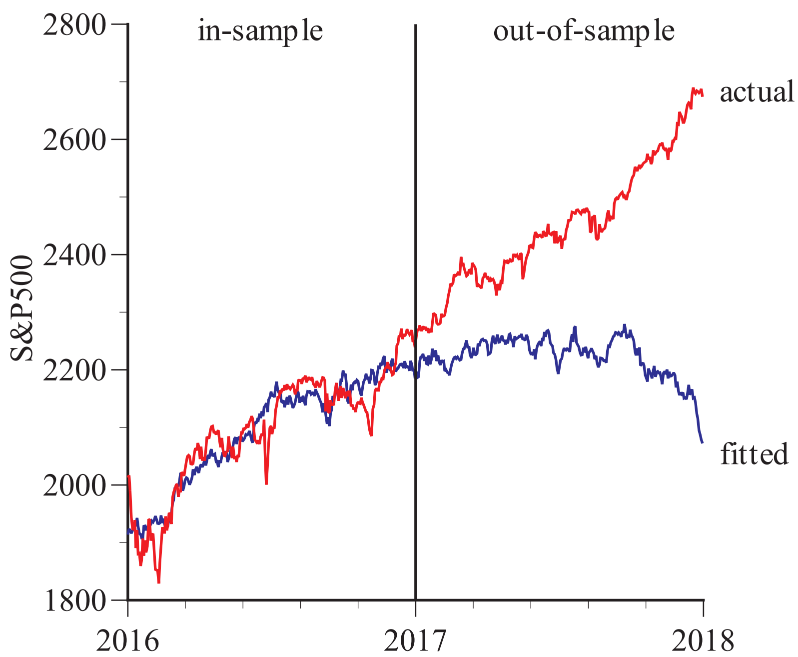

Figure 2 Coincidence, not Knowledge

What happened? How can a model work so well one year and so badly the next? That is the inescapable nature of data mining. Choosing a model simply because it fits a particular set of data closely virtually guarantees that it won't do nearly as well with fresh data. For a model to work with fresh data, it needs a theoretical foundation. It has to make sense. Correlation does not supersede causation.

It might be tempting to think that perhaps this discovered statistical relationship between stock prices and temperatures in these two cities is real—that my algorithm discovered a previously unknown relationship. Anticipating such a temptation, I did not actually use daily temperatures or any other real variables. Each of the 100 candidate explanatory variables was a randomly generated variable that—because it random—is guaranteed to have no systematic relationship to the S&P 500. By chance alone, variables 58 and 94 happened to be statistically correlated with the S&P 500, The algorithm predicted stock prices based on data that have nothing at all to do with stock prices, but happened to have been temporarily correlated with them during the in-sample estimation period.

That is the point. Even though all the data were generated randomly and have nothing whatsoever to do with stock prices, my data-mining algorithm found some variables that were fortuitously correlated with the S&P 500. The trading rule uncovered by my data mining algorithm was not knowledge discovery. It was coincidence discovery.

The data analyzed by a trading algorithm might include some variables that really do matter; however, when an algorithm ransacks data looking for statistical relationships, the more variables it considers, the more likely it is that the variables it chooses are coincidental rather than causal (Calude and Longo 2016).

Out-of-Sample Validation

Since a data-mined model's weaknesses can be exposed by the deterioration of the model's fit using fresh data, it is reasonable to hold out part of the available data for testing after the algorithm has identified a promising trading model. Data-mine part of the data for knowledge discovery and then validate the results by testing the discovered model with data that were set aside for this purpose (Mayers and Forgy 1963; Mark and Goldberg 2001).

It is always a good idea to test a model with fresh data. However, choosing a data-mined model via a repetitive cycle of in-sample estimation and out-of-sample testing does not ensure that a useful model will be chosen. Just as some models are certain to fit the in-sample data by luck alone, so some models are certain to fit both the in-sample and out-of-sample data. Uncovering a model that fits all the data is just another form of data mining, and doesn't solve the fundamental problem, which is that models chosen to fit the data, either part of the data or all of the data, cannot be expected to fit new data nearly as well.

To illustrate this point, Figure 3 shows the in-sample and out-of-sample fits for all 9,900 models that use 2 of the 100 randomly generated variables. For the 2016 data used to estimate the models, the correlation between the predicted and actual values cannot be less than zero, because the best-fit model can always ignore the word variables completely and have a correlation of zero. The average correlation for the in-sample models was 0.73.

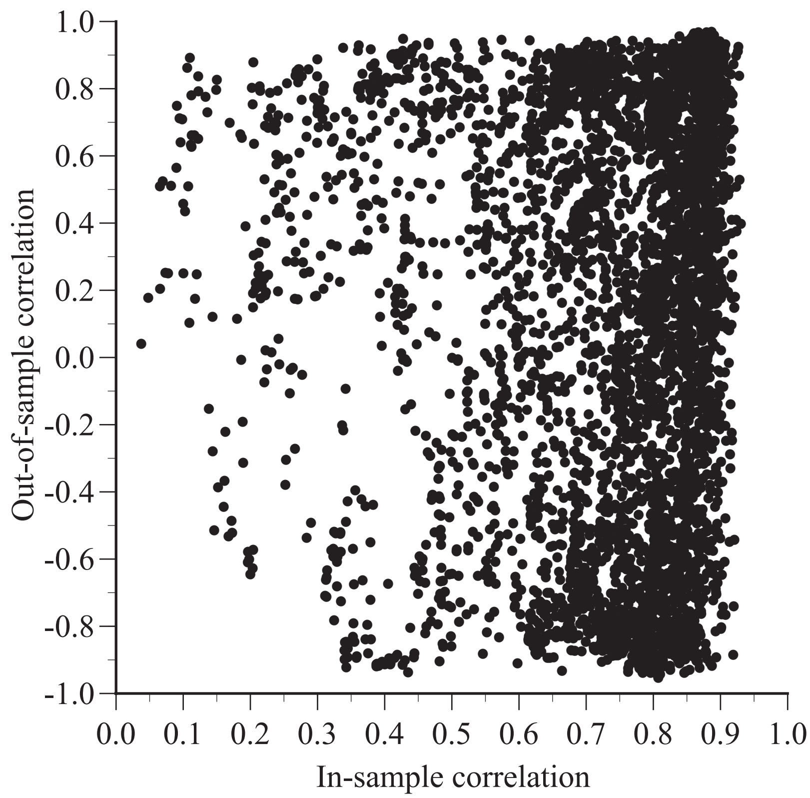

Figure 3 All 9,900 In-Sample and Out-of-Sample Correlations

For the 2017 out-of-sample data that were set aside to test the models, the correlation is equally likely to be positive or negative because the words are, after all, random numbers that have nothing at all to do with stock prices. We expect the average correlation to be close to zero. For these particular data, the average out-of-sample correlation happened to be 0.02.

Nonetheless, some out-of-sample correlations were, by chance, strongly positive and others were strongly negative. Figure 3 shows that several models fit the in-sample data well and also do well out-of-sample, sometimes even better than in-sample. That is the nature of chance, and these are chance variables.

In-sample, where the models are fit to the data, 129 models have a correlation above 0.90. Out-of-sample, where the models are accurate only by chance, six of the models with correlations above 0.090 in-sample have even higher correlations out-of-sample. These six models pass the validation test with flying colors even though it was pure luck. They are still useless for predicting stock prices. If we didn't know better, we might think that we discovered something important. But, of course, we didn’t. All we really discovered is that it is always possible to find models that do well in-sample and out-of-sample, even if the data are just random noise.

Models should be tested with set-aside data, but set-aside data are not a cure for energetic data-miners. A trading algorithm would have no trouble finding a model that performs almost as well (or even better) out-of-sample as in-sample, even though the variable being predicted is only coincidentally related to the explanatory variables. A trading algorithm cannot evaluate the plausibility of a discovered pattern because it does not understand in any meaningful sense what the data are and whether they might reasonably be related to stock prices.

Conclusion

Black-box trading algorithms are appealing because computers do so many things so well. However, the inarguable fact that computers do many difficult things much better than humans does not mean that computers are better investors. When it comes to investing, computers are much more efficient than humans at using data-mining to discover patterns, but completely incapable of gauging whether the unearthed patterns are potentially useful, or are merely coincidental and therefore fleeting and useless. Only humans can make that assessment.

If a trading algorithm is hidden inside a black box, then no one—neither computers or humans—can tell if a discovered patterns is useful or useless. Precluding human judgment is a flaw, not a feature.

References

Allan, John H. 1976, *The Winds of Wall Street*, *The New York Times*, 1976.

Anderson, Chris, 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*, June 23.

Association of National Advertisers, 2017, "AI" Voted ANA Marketing Word of The Year For 2017," press release, December 6

Begoli, E., and J. Horsey, 2012. Design Principles for Effective Knowledge Discovery from Big Data, Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference.

Bolen, Johan, Mao, Huina, and Zeng, Xiaojun, 2011. Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, 2 (1), 1-8.

Calude, Cristian S, Longo, Giuseppe. 2016. The deluge of spurious correlations in big data, *Foundations of Science,* 22, 595–612 https://doi.org/10.1007/s10699-016-9489-4.

Cios, Krzysztof J., Pedrycz, Witold, Swiniarski, Roman W., and Lukasz Andrzej Kurgan. 2007, *Data Mining: A Knowledge Discovery Approach*, New York, Springer.

Coase R, 1988. How should economists choose?, in *Ideas, Their Origins and Their Consequences: Lectures to Commemorate the Life and Work of G. Warren Nutter*, American Enterprise Institute for Public Policy Research.

Cooper, Michael J., Dimitrov, Orlin, and P. Raghavendra Rau. 2001. A Rose.com by Any Other Name. *The Journal of Finance*. 56. 2371-2388.

Davis, Ernest, 2018, Collection of Winograd Schemas, https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WSCollection.html

Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A. et al. (2017). Robust Physical-World Attacks on Deep Learning Models. <https://arxiv.org/abs/1707.08945>

Fayyad, Usama, Piatetsky-Shapiro, Gregory, and Padhraic Smyth. 1996. From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, 17 (3), 37-54.

Hofstadter, Douglas. 2018. The Shallowness of Google Translate, *The Atlantic*, January 30

Kecman, Vojislav. 2007. Forward, Cios, Krzysztof J., Pedrycz, Witold, Swiniarski, Roman W., and Lukasz Andrzej Kurgan, Data Mining: A Knowledge Discovery Approach, New York, Springer, xi.

Knight, Will, 2016, Will AI-Powered Hedge Funds Outsmart the Market?, *MIT Technology Review*, February 4.

Koppett, Leonard, Carrying Statistics to Extremes, S*porting News*, Feb 11, 1978.

Krueger, Thomas M., and William F. Kennedy. 1990. An examination of the Super Bowl Stock Market Predictor, *The Journal of Finance*, 45 (2), 691-697.

Levesque, H., Davis, E., and Morgenstern, L. 2012, The Winograd Schema Challenge. *Principles of Knowledge Representation and Reasoning (KR).* Piatetsky-Shapiro, G. 1991. Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop. *AI Magazine* 11(5): 68–70.

Mark, J. and Goldberg, M. A. 2001. Multiple regression analysis and mass assessment: A review of the issues, *The Appraisal Journal*, 56, 89–109.

Mayers, J. H. and Forgy, E.W. 1963. The Development of numerical credit evaluation systems. Journal of the American Statistical Association, 58 (303), 799–806.

Nguyen, Anh, Yosinski, Jason, and Jeff Clune, 2015. Deep Neural Networks are Easily Fooled:

High Confidence Predictions for Unrecognizable Images, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

RK Capital Partners, presentation, March 8, 2017.

Sagiroglu, Seref, and Duygu Sinanc, 2013, Big Data: A Review, Collaboration Technologies and Systems (CTS), 2013 International Conference.

Sharif, Mahmood, Bhagavatula, Sruti, Bauer, Lujo, and Michael K. Reiter, 2016, Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 1528-1540.

Su, Jiawei, Vargas, Danilo Vasconcellos, and Sakurai Kouichi, One pixel attack for fooling deep neural networks, November 2017. <https://arxiv.org/abs/1710.08864>

Tullock, Gordon, 2001, A Comment on Daniel Klein's "A Plea to Economists Who Favor Liberty," *Eastern Economic Journal*, 27 (2), 203-207.

Winograd, Terry, 1972. Understanding Natural Language. *Cognitive Psychology*, 1–191

Zweig, Jason, Super Bowl Indicator: The secret history, *The Wall Street Journal*, January 28, 2011.