3RD QUARTER - 2019 INVESTMENT PUBLICATIONS HIGHLIGHTS

Third quarter's edition summarizes four articles on machine learning, an exciting field within quantitative finance:

- the first article considers whether machine learning techniques that have yielded significant breakthroughs across several scientific fields can produce similar results in finance;
- the second argues that when machine learning techniques are properly applied, they can help us better understand the relationship between factors and stock returns;
- the third assesses whether machine learning can improve recession prediction in real time, and recommends a dynamic asset allocation strategy based on the machine learning model's recession signal; and
- the last article provides a protocol for quantitative financial research designed to limit common pitfalls and maximize the number of investment strategies that produce real world results.

CAN MACHINES "LEARN" FINANCE?

Ronen Israel, Bryan Kelly, and Tobias Moskowitz, AQR Capital Management, June 7, 2019

Technological advances have made it possible for machines to uncover new, complex relationships between variables in several scientific fields. In this article, the authors begin by explaining machine learning techniques and why they have facilitated several breakthroughs across various fields before turning to its application in finance. They conclude that nuances specific to finance have prevented current machine learning techniques from making similar breakthroughs in predicting asset returns, but it is in its early days and the authors see significant potential for its use in other areas of finance.

Whereas traditional computer programing techniques rely on humans to feed a computer a set of rules to relate input and output variables, the machine learning approach relies on the computer to uncover relationships between input and output variables with little or no human direction. Machine learning techniques rely on longstanding statistical principles, but significant improvements in data storage and processing power have enabled researchers to quickly analyze large datasets and uncover complex, new relationships between variables that were previously unknowable under the traditional computer programming approach.



Researchers have successfully applied machine learning techniques in many scientific fields. Although machine learning appears well-suited for financial research, there are several real-world challenges that make it more difficult for machine learning to master financial tasks, such as return prediction. Perhaps most importantly, machine learning works best when there is a high success rate of predicting outcomes, or a high signal-to-noise ratio, within a system. Unfortunately, financial markets are extremely noisy, which makes for a weak signal-to-noise ratio. This is due to the general efficiency of financial markets, which leads to volatile swings in asset prices following unanticipated news (noise) and successful investment strategies (signals) being short-lived, as they are quickly adopted by market participants. Additionally, machine learning techniques perform best in areas with large amounts of data. Financial data is typically time-based, which limits the number of available observations. Finally, machine learning models can be complex, making it a challenge for asset managers that are not data scientists to communicate their investment strategy and its risks to their clients.

Applying machine learning to finance may be more difficult than in other scientific fields, but some early research models have shown that machine learning techniques have the potential to improve return prediction. The success stories are usually the result of sophisticated models that have uncovered complex, non-linear relationships that simpler methods, such as linear regression, may have previously missed. Machine learning is still in its early stages in the field of finance, and the authors believe the potential upside is still high, even if the gains so far have only been incremental. There are areas of finance that exist outside of return prediction (i.e., risk management, transaction cost modeling, and factor construction) that have stronger signal-to-noise ratios and may be more well-suited for machine learning. To date, however, the authors see the application of machine learning in finance as an evolutionary process and one that has yet to become revolutionary.

MACHINE LEARNING FOR STOCK SELECTION

Robert C. Jones and Keywan Christian Rasekhschaffe, *Financial Analysts Journal*, Vol 75, no 3 (Third Quarter 2019): 70–88

When machine learning is properly applied in finance, it has the potential to improve upon traditional statistical techniques in predicting stock returns. Yet, overfitting remains a major concern. The authors discuss machine learning techniques, including forecast combinations and feature engineering that can reduce overfitting and still produce superior results. They conduct a case study to show how these techniques can help us better understand the relationship between factors and future stock returns.

Traditional quantitative models that use company factors to forecast stock returns have struggled to generate alpha since the global financial crisis. As a result, practitioners have turned to more dynamic quantitative models that attempt to develop more flexible factor-timing strategies to predict stock returns. However, issues such as noisy data, multicollinearity, and the nonlinear relationship between factors and returns make traditional simple linear regression techniques ill-suited for this task. To help address this problem, investors are looking toward a growing investment field: machine learning. Machine learning uses computationally complex algorithms to derive meaningful relationships between variables without explicit human programming instructions. Machine learning models have two important properties relative to traditional linear regression models: (1) they can uncover complex, nonlinear patterns that were previously hidden, and (2) they are more effective in the presence of multicollinearity (when two or more input variables are highly correlated). However, a key issue for most quantitative financial research—overfitting—remains a major challenge for machine learning.

Overfitting occurs when a statistical model picks up the noise surrounding a signal instead of the signal itself. In such cases, the model will have good in-sample performance, but will produce poor results when applied to out-of-sample data. This poses a problem when using factors to forecast stock returns. Stock returns are noisy and have a low signal-to-noise ratio, making it difficult to distinguish an actual signal from irrelevant noise. In the presence of overfitting, favorable in-sample results of several machine learning financial models are not as promising as in the real world.

To overcome this issue, the authors suggest using two different approaches: (1) forecast combinations (combining forecasts from multiple outperforming models and removing forecasts from underperforming models) and (2) feature engineering (using institutional knowledge to structure the problems the models solve in a way amenable to machine learning). The former increases the robustness of the model by testing the results across multiple forecasting techniques, training sets, and factors to confirm if it produces similar patterns and results. The latter is one of the most effective ways to overcome overfitting; it improves the signal-to-noise ratio and forecasting accuracy by limiting the degree to which irrelevant noise overwhelms a model.

The authors conduct a case study to demonstrate the general power of incorporating these machine learning techniques in models for stock selection. They analyze the relationship between stock returns and a total of 194 company characteristics for thousands of stocks across 22 developed markets from 1994 to 2016. To limit overfitting, they reproduce their results using four separate machine learning models and three separate training windows. While each of the individual machine learning models and training windows exhibited strong results, the best forecasts of future stock returns came from a composite of the models and training windows. The authors conclude that their exercise shows that machine learning techniques—if applied correctly—can produce stock return forecasts that dramatically exceed those of models based on simple linear regression techniques, while also reducing the risk of overfitting.

MACHINE LEARNING FOR RECESSION PREDICTION AND DYNAMIC ASSET ALLOCATION

Yaser S. Abu-Mostafa, Alexander James, and Xiao Qiao, Journal of Financial Data Science, December 31, 2018

There is a vast body of literature dedicated to identifying turning points in the business cycle. In this paper, the authors use machine learning to try to more accurately identify the beginning and end of US recessions in real time. Not only do the authors find that machine learning techniques can improve the precision of recession prediction in real time, but they also show that investors can profit off of a dynamic asset allocation strategy based on the signals produced by their machine learning model.

The body in charge of providing the official dating of US expansions and recessions, the National Bureau of Economic Research (NBER), has historically announced turning points in the business cycle with a four- to 21-month delay. Due to the significant lag in reporting the beginning and end of business cycles, nowcasting—forecasting a condition in the present time because the full information will not be available until later—is key for recession prediction. Building on the existing body of work about nowcasting recessions, the authors use a common and flexible machine learning method to uncover complex relationships between four distinct components of the economy—labor markets, stock markets, goods markets, and bond markets—and the current state of the macroeconomy to forecast recessions.

The authors find that for the six US recessions that occurred between 1973 and 2018, their model typically identified the turning point in the business cycle within one to three months of the official NBER definition for expansions and recessions. In general, the results from the machine learning model were as accurate as other recession prediction models, but the machine learning model had a lag that was significantly shorter than other models.

The authors then attempt to create a profitable dynamic asset allocation strategy based on the real-time recession forecasts their model produced. Their dynamic asset allocation strategy altered the portfolio risk contribution from stocks and bonds based on the model's forecast of the state of the macroeconomy. Starting from an equal risk contribution from stocks and bonds (50/50), the strategy increased the risk contribution from stocks relative to bonds when the model forecasted an expansion (75/25) and vice versa when the model forecasted a recession (25/75).

The authors found that their dynamic asset allocation strategy outperformed an equal risk contribution strategy by 85 basis points per year. Risk-adjusted performance metrics, such as the Sharpe ratio and Calmar ratio, were also higher for the dynamic asset allocation strategy. The authors believe their findings illustrate how investors can use machine learning techniques to improve portfolio performance over the course of a business cycle.

A BACKTESTING PROTOCOL IN THE ERA OF MACHINE LEARNING

Rob Arnott, Campbell R. Harvey, and Harry Markowitz, Journal of Economic Literature, November 21, 2018

Machine learning tools hold considerable promise for financial research, but like other quantitative applications in finance, there is a significant risk that researchers will misapply these tools. Often, machine learning strategies do not perform as advertised in the real world. As a result, the authors believe the time is right to reflect on the finance industry's research process. They propose a seven-step research protocol to help researchers avoid common pitfalls, limit false discoveries, and identify winners when using machine learning tools, as well as traditional quantitative methods, to develop and back test investment strategies.

Advancements in data storage, processing power, and data analysis have made it more practical to apply machine learning techniques in financial research. Although machine learning represents a new and exciting area of quantitative finance, issues specific to financial data—limited data availability and noisy data—make it vulnerable to many of the same setbacks that have plagued traditional quantitative financial research methods for years, such as data mining. Moreover, the availability of open source software has reduced the barriers to entry to data analysis, increasing the potential for researchers to misapply advanced quantitative methods, such as machine learning, and produce investment strategies that fail to perform in live trading.

Given these considerations, the authors propose a comprehensive research protocol for quantitative finance. Their protocol includes seven key areas for consideration: (1) research motivation, (2) multiple testing and statistical methods, (3) sample choice and data, (4) cross-validation, (5) model dynamics, (6) model complexity, and (7) research culture. Each step serves a key role in the research process. For example, research motivated by an *ex-ante* hypothesis based on sound economic foundations will help limit data mining and increase the likelihood that the results of a model hold up in live trading. This is particularly important in machine learning, since the probability of a false positive remains high for strategies based on inputs with no economic logic, even after they are thoroughly cross-validated. Additionally, extensive cross validation can add a degree of complexity to a model. The in-sample results might improve because of the additional complexity, but the live trading results could disappoint if the model suffers from overfitting, especially if it is not based on sound economic theory.

Machine learning techniques have the potential to uncover new, more complex relationships within finance, but investment strategies based on these techniques are still susceptible to data mining and overfitting. A protocol is a simple step that can help minimize the number of false positives without eliminating the possibility of identifying successful strategies altogether. Ultimately, the authors believe their seven-step research protocol should help investment managers identify more investment strategies that are successful in live trading.

Copyright © 2019 by Cambridge Associates LLC. All rights reserved.

This report may not be displayed, reproduced, distributed, transmitted, or used to create derivative works in any form, in whole or in portion, by any means, without written permission from Cambridge Associates LLC ("CA"). Copying of this publication is a violation of US and global copyright laws (e.g., 17 U.S.C.101 et seq.). Violators of this copyright may be subject to liability for substantial monetary damages.

This report is provided for informational purposes only. The information does not represent investment advice or recommendations, nor does it constitute an offer to sell or a solicitation of an offer to buy any securities. Any references to specific investments are for illustrative purposes only. The information herein does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual clients. Information in this report or on which the information is based may be based on publicly available data. CA considers such data reliable but does not represent it as accurate, complete, or independently verified, and it should not be relied on as such. Nothing contained in this report should be construed as the provision of tax, accounting, or legal advice. Past performance is not indicative of future performance. Broad-based securities indexes are unmanaged and are not subject to fees and expenses typically associated with managed accounts or investment funds. Investments cannot be made directly in an index. Any information or opinions provided in this report are as of the date of the report, and CA is under no obligation to update the information or communicate that any updates have been made. Information contained herein may have been provided by third parties, including investment firms providing information on returns and assets under management, and may not have been independently verified.

The terms "CA" or "Cambridge Associates" may refer to any one or more CA entity including: Cambridge Associates, LLC (a registered investment adviser with the US Securities and Exchange Commission, a Commodity Trading Adviser registered with the US Commodity Futures Trading Commission and National Futures Association, and a Massachusetts limited liability company with offices in Arlington, VA; Boston, MA; Dallas, TX; Menlo Park, CA, New York, NY; and San Francisco, CA), Cambridge Associates Limited (a registered limited company in England and Wales, No. 06135829, that is authorised and regulated by the UK Financial Conduct Authority in the conduct of Investment Business, reference number: 474331); Cambridge Associates Limited, LLC (a registered investment adviser with the US Securities and Exchange Commission, an Exempt Market Dealer and Portfolio Manager in the Canadian provinces of Alberta, British Columbia, Manitoba, Newfoundland and Labrador, Nova Scotia, Ontario, Québec, and Saskatchewan, and a Massachusetts limited liability company with a branch office in Sydney, Australia, ARBN 109 366 654), Cambridge Associates Investment Consultancy (Beijing) Ltd (a wholly owned subsidiary of Cambridge Associates, LLC which is registered with the Beijing Administration for Industry and Commerce, registration No. 110000450174972), and Cambridge Associates Asia Pte Ltd (a Singapore corporation, registration No. 200101063G, which holds a Capital Market Services License to conduct Fund Management for Accredited and/or Institutional Investors only by the Monetary Authority of Singapore).