THE JOURNAL OF
INVESTMENT CONSULTING

A reprinted article from Volume 21, Number 1, 2022

Applications of Machine Learning in Wealth Management

By John M. Mulvey, PhD, Junhan Gu, Margaret Holen, PhD, and Yuqi Nie

INVESTMENTS & WEALTH INSTITUTE®

VOLUME 21
NUMBER 1
2022

FEATURE

Applications of Machine Learning in Wealth Management

By John M. Mulvey, PhD, Junhan Gu, Margaret Holen, PhD, and Yuqi Nie

ABSTRACT

This paper provides a targeted review of machine learning methods that are impacting the field of wealth management. This is an area of great breadth and dynamic growth, and we focus on applications that in our experience are delivering benefits in practice to wealth management businesses by improving investment performance and enabling personalized services at scale. We highlight novel approaches to customized financial planning systems and deep learning algorithms with novel data sources linked to natural language processing concepts. We also discuss emerging challenges and opportunities that may shape the future path of this rapidly evolving area.

INTRODUCTION

Machine learning (ML) concepts are disrupting traditional ways of doing business across all sectors of today's economy. In finance, investment and wealth management firms have been deploying these approaches to improve the efficiency and the quality of their services and products. Global banks and asset managers have joined universities in launching research initiatives in a search for the latest breakthrough ML technologies (Huang et al. 2020; Veloso et al. 2021). The largest tech firms manage large suites of ML models that consume vast amounts of micro-level data that fuel growth and protect moats for their trillion-dollar businesses. Wealth management and fintech are seeing increasing opportunities for these technologies.

Rather than attempting a comprehensive survey of applications of ML in wealth management, we cover selected topics where we see wide applications to asset-only returns (alpha) and to scaling individualized services tailored to individual investors' goals in the context of their income and liabilities.

A critical matter for both these applications involves collecting and analyzing data on a wide scale. For personalized services, micro-level data allows investment firms to better understand and service their customers and promises to improve the engagement outcomes.[1] These data sources are expanding exponentially and are likely to continue growing at the same pace (Kolanonic and Krishnamachari 2017; Stanford University 2021; *The Economist* 2017).[2] For investment performance, potential opportunities afforded by alternative data motivate investors to collect information from a wide array of news sources and social media feeds, along with data related to underlying businesses, gleaned from the web, and from satellites. (Monk et al. 2019; Ekster and Kolm 2021).

The largest tech firms manage large suites of ML models that consume vast amounts of micro-level data that fuel growth and protect moats for their trillion-dollar businesses. Wealth management and fintech are seeing increasing opportunities for these technologies.

Compared with traditional statistics, ML algorithms reduce dependencies on structural models and parametric assumptions. Instead, there is a focus on performance of models across training, validation, and out-of-sample testing, requiring larger amounts of representative data.[3] The philosophy is closely aligned with classic frequentist approaches in statistics: "Let the data do the talking." Distinguishing features of recent ML algorithms are: (1) high dimensional parameter sets, (2) nonlinear relationships, and (3) a much-diminished role of structural assumptions. Another common feature involves constructing high-performing models from ensembles of relatively simple, low-performing models (Hastie et al. 2009). Israel et al. (2020) suggest that these concepts are evolutionary rather than revolutionary. Regardless, the emergence of a tidal wave of micro-level data and powerful ML algorithms enables numerous applications to wealth management and shifts research away from financial theory toward data-driven approaches.

Table 1 lists four of the interrelated domains in wealth management that are being impacted by ML concepts. We emphasize

Table 1

AREAS OF WEALTH MANAGEMENT EXPLOITING MACHINE LEARNING TECHNOLOGIES

Wealth Management Area	Uses of Machine Learning
Improve investment performance	• Improve forecasting from market and financial data • Integration of micro-level and alternative data with sentiment analysis and language embeddings
Address personalized wealth management at scale	• Goal-driven investing for individuals and institutions • Customize portfolios • Multi-period wealth management systems
Enhance investment implementation	• Optimized asset allocation and portfolio construction • Automated trading systems with lower transaction and market impact costs
Increase efficiency and profitability of investment management firms	• Fraud detection • Scoring to allocate high-touch services • Recommender systems to connect products and customers

the first three in this review. Our coverage provides a curated sample of the large number of studies and projects in this arena.[4]

For decades, quantitative methods have been applied to wealth management, such as the search for alpha or returns in excess of passive indexed portfolios. Recent ML technologies have been enabled by novel computational algorithms run on specialized computer hardware, e.g., graphical processing units (GPUs) and tensor processing units (TPUs) and by increasing data availability, including micro-level and alternative data. The interest in and the range of applications of machine learning in wealth management seems poised to expand.

OVERVIEW: MACHINE LEARNING CONCEPTS

TYPES OF LEARNING: SUPERVISED, UNSUPERVISED, AND REINFORCEMENT

Machine learning methods are often categorized according to their underlying processes and goals (Jordan and Mitchell 2015): (1) supervised learning, (2) unsupervised learning, and (3) reinforcement learning.

The objective of supervised learning is to discover relationships between an outcome assumed to be "ground truth" and input data, in order to forecast the future from current input data. Model specification includes parameters chosen based on a training/validation process; the quality of predictions is assessed via out-of-sample test data. One example of supervised learning is to consider equity returns as output data and financial reports and historical prices as input data. In another example of supervised learning, a firm may conduct research on identifying the features that indicate if a new client will purchase a specified security or service based on the data from existing clients' previous purchase patterns.

In contrast, the goal of unsupervised learning is to identify patterns that could be helpful in better understanding the data without labels. Examples of unsupervised learning include cluster analysis that identifies groups of customers and products they tend to buy, to inform product recommendation systems. Another example is separating massive numbers of news articles or customer communications into distinct segments.

In reinforcement learning, the goal is to render a sequence of decisions with feedback or rewards over time, and sometimes only at the conclusion of the sequence. Examples of reinforcement learning include game-playing systems for chess and Go, like the Google system AlphaZero—a famous successful example of reinforcement learning (Silver et al. 2016).

METHODS: MODEL COMPLEXITY TIERS

To summarize the current state of ML in wealth management, we separate widely used machine learning methods into three levels or generations,[5] which we describe in table 2 along with representative applications. This framework divides machine learning techniques according to the complexity of the models and algorithms and the required number of parameters. In the first level, the models are direct extensions of traditional statistical methods such as regression models. Here, the number of hyperparameters is small, models are easily interpretable, and implementations are easily managed even in legacy systems. In the second level, the methods are nonlinear and sometimes discrete, and often sit outside traditional statistics, with a larger number of hyperparameters than level 1, requiring more modern systems implementation. The third level takes up the most complex methods, with exceptionally large sets of parameters. Many of these are direct applications or extensions of deep neural networks (Li 2017; Heaton et al. 2018). Although neural networks have been known for decades, it is only recently that these concepts have shown full potential for solving a wide range of challenging application problems, including image processing. Today, researchers across the globe are working to determine categories of wealth management that might be amenable to significant breakthroughs using these approaches.[6] However, much of this research has not yet been widely deployed in the financial community and applications are rarer than in level 1 or 2.

Every application requires adequate, representative data in order to determine the best set of hyperparameters within a training and validation phase. Hyperparameters are commonly determined by a process called cross-validation, which requires subsampling from one portion of the historical training data, fitting the model using a range of hyperparameter values, then

Table 2

THREE LEVELS OF MACHINE LEARNING METHODS IN WEALTH MANAGEMENT

	Machine Learning Techniques		Hyperparameters	Applications Examples	Interpretability
First Generation: Traditional Statistical Methods	LASSO (L1)		Small number of regularization parameters	**Mainstream financial modelling:** Feature selection, shrinkage, factor models, sparse covariance matrix estimate, portfolio optimization	High
	Ridge (L2)				
	Elastic Net (L1 & L2)				
	Best Subset Selection (L0)				
	Nonconvex Penalization				
Second Generation: Nonlinear Methods (with a larger number of hyperparameters)	Supervised Learning (Regression and Classification)	Decision Trees	Maximum depth of tree, minimum samples required for a node, number of features to consider for the split	**Quantitatively oriented setting:** Forecast the macroeconomic trend and asset returns, fraud detection for personal financing	Medium
		Random Forest	Number of trees, number of samples to draw from original data for each tree		
		Gradient Boosting	Number of trees, learning rate, number of samples to draw from original data for each tree		
		Support Vector Machine	Depends on the kernel		
	Unsupervised Learning	Principal Component Analysis	Choose top k principle components based on variance explained	Dimension reduction	
		Hierarchical Clustering, K-means	May need to specify the number of clusters	Clustering: outlier detection, pattern recognition	
Third Generation: Deep Learning Methods	General Nonlinear Methods	Deep Neural Networks	Millions or even billions of hyperparameters	**More advanced applications:** NLP, sentiment analysis, multi-period financial planning systems	Low
		Deep Reinforcement Learning			

"validating" the model on the remaining training data, and then choosing hyperparameter values that perform best in that validation process. Once the hyperparameters are chosen, an out-of-sample "test" analysis is conducted with the withheld data—the test phase. It is critical for true validity to carry out the testing step only once (Hastie et al. 2009). Unfortunately, testing poses a particular challenge in many investment applications, where "out-of-training-and-validation" sample data may be limited, and any historical data may not be representative of future performance. The data requirements and hyperparameter tuning processes become more onerous as the number of parameters increase in level 3. Arnott et al. (2019), Giglio et al. (2021), and Harvey and Liu (2021) discuss related backtesting and data snooping issues.

GENERATIVE VERSUS DISCRIMINATIVE ML MODELS

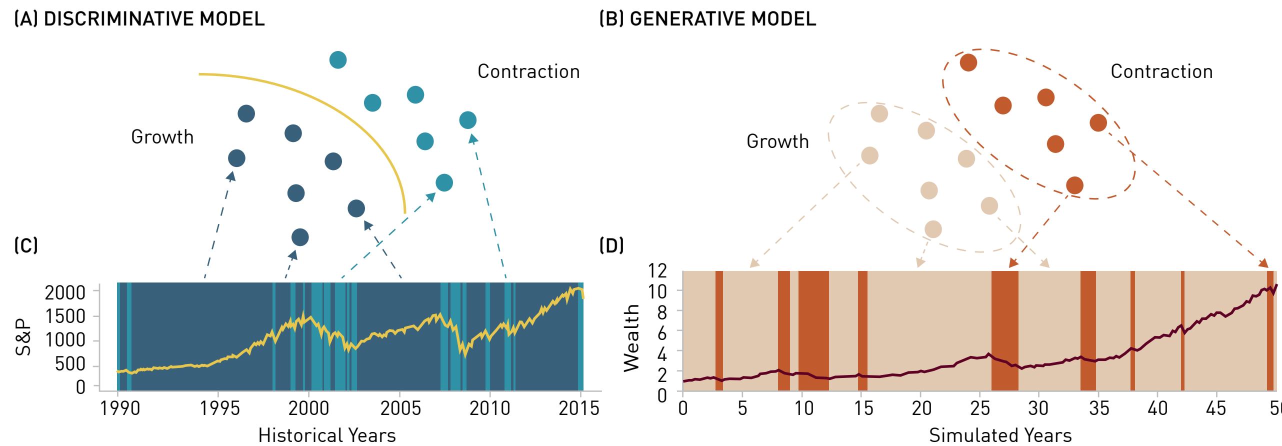
Another way to interpret machine learning methods is to distinguish (1) "generative methods" that enable the development of synthetic data, from (2) "discriminative methods" that give distributions of possible outcomes based on a fixed dataset, and (3) "decision machine" models that generate only predictions of outcomes, but do not attempt to associate the probabilities of those predictions (Bishop 2006). Generative models tend to work well in applications with a relatively lack of data, as often occurs in wealth management, though they pose unique challenges relative to other model types.

Generative models first became widespread in communities using deep learning for text and image applications, which benefit from generative models' "creative" capabilities, such as intelligently enhancing a corrupted photo. More specifically, given a dataset $\{x_1, x_2, ..., x_i\}$ sampled from a distribution $p_{model}(x)$, we can direct a generative model to learn to construct augmented data with a distribution as close as possible to $p_{data}(x)$. The models can be implemented with structural assumption priors, as is the case with hidden Markov models, or via nonparametric approaches.

Classification tasks provide a relatively simple context to contrast between discriminative and generative models. The most common generative model is a Gaussian Naïve Bayes model, which assumes the population of each class is distributed normally around a class-specific mean and covariance. The model is fit by determining the mean and standard deviation parameters, thus learning a joint distribution $p(x,y) = p(x|y)p(y)$, where x is the input data and y is the class label. In contrast, the most common discriminative models are generalized linear models such as logistic regression, in which the probability of each

Figure 1

DISCRIMINATIVE MODELS VS. GENERATIVE MODELS WITH A FINANCIAL EXAMPLE



Source: A, B Authors' generated plot; C, D Mulvey and Liu (2016)

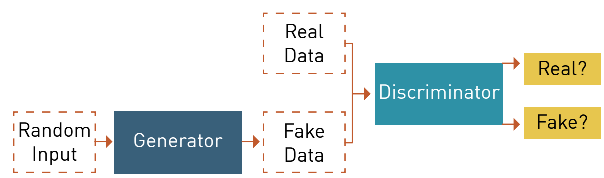
class conditional on the input variable *p(y|x)*, is a sigmoid function of a linear combination of those input variables. One can derive logistic regression parameters from Gaussian Naïve Bayes parameters by Bayes rules (Ng and Jordan 2001). Conversely, a decision tree model gives only a prediction of class membership given input data, not a probability of that class being valid.[7]

Figures 1A and 1B display an intuitive comparison between discriminative and generative models. Figures 1C and 1D depict a corresponding financial example from Mulvey and Liu (2016). Here, figure 1C employs the trend-filtering algorithm to identify regimes for the S&P 500 index, which determines the growth periods and contraction periods in history via a discriminative method; figure 1D indicates a potential future wealth path of the S&P 500 Index created by a generative method, with growth and contraction periods shown. A generative model is able to construct many future representative paths for financial planning studies.

We can categorize generative models into two main classes (Goodfellow 2016): One is "explicit" modeling of the density $p_{model}(x)$, for example, variational autoencoder (VAE) (Kingma and Welling 2013; Kingma et al. 2014). VAE is a generative variant of autoencoder (Kramer 1991), which uses an encoder to map data to a low-dimensional space, then employs a decoder to reconstruct it back to the original data. By adding noise between the process of encoder and decoder, VAE gains the ability to generate augmented data that has never been seen. This idea is implemented by introducing a latent random variable z, which represents the underlying structure of the data. By doing so, instead of maximizing the log-likelihood of $p_{model}(x)$, VAE maximizes a lower bound of it. This approach offers a way to generate expanded data from applications with limited sample data, a common problem in financial applications.

Figure 2

FRAMEWORK OF GENERATIVE ADVERSARIAL NETWORKS



The other class refers to an "implicit" way of estimating the probability distribution of data, in which the model doesn't calculate $p_{model}(x)$ but just directly generates data that matches the probability distribution of training data. The generative adversarial network (GAN) (Goodfellow et al. 2014) is a primary example in this class. Figure 2 displays its structure clearly. It consists of two distinct neural networks (generator and discriminator) with an adversarial relationship. The algorithm aims to pose data that looks "real" to cheat the discriminator, while the discriminator tries to distinguish fake data from the real data. Ideally, after iteratively training, the generator will create a database possessing a similar distribution to the original real data.

There are also alternative ways of building a generative model, e.g., PixelRNN (Oord et al. 2016), Boltzmann machines (Salakhutdinov and Hinton 2009), and an increasing number of methods being proposed in this active literature. However, practice tends to converge around a small number of models with demonstrated success in applications, and we have focused on those in this paper.

APPLICATIONS OF MACHINE LEARNING IN WEALTH MANAGEMENT

FORECASTING MARKETS AND THE SEARCH FOR ALPHA

This section discusses applications of ML to forecasting markets in the context of setting the parameters for portfolio models.[8]

Classical mean-variance optimization (MVO) chooses a portfolio in a single period after which all uncertainty is resolved (Markowitz 1952). Given its shortcomings including sensitivity to estimated input parameters and concentration of portfolio risk, this mean-variance trade-off approach led to more discussion on alternate and additional definitions of risk, asset-liability management, and multi-period financial models. Due to the comparative difficulty to estimate expected returns in MVO models, risk-based asset allocation also has gained popularity in both industry and academia over the past decade. Risk budgeting sets up portfolios in such a way that the different investment categories held in a portfolio make up a target risk contribution. Bridgewater's All-Weather Strategy is an early example of the risk-parity allocation strategy. Under certain circumstances, the risk-based optimization can achieve better downside risk management than a mean-variance portfolio.

One fundamental challenge for all those portfolio models arises from accurately estimating the parameters of these models, such as the expected returns for the assets and the covariance matrix. Estimation errors will adversely impact portfolio accuracy and investment performance, especially when the dimensions of the model are large.

A notable antecedent of machine learning methods to finance involves estimating returns using shrinkage estimates, which generally involve optimizing the sum of a standard "goodness of fit" metrics and a "penalty" term associated with the size of the model. Stein (1956, 1981) proved that the usual approach for estimating the mean (sample average) is inadmissible when estimating three or more variables. He showed that the standard loss function can be reduced by shrinking the sample means. Shrinkage has become a mainstay concept in financial applications and many other domains,[9] because it can help discover the most relevant features, increase sparsity, and allow for hyperparameters (regularization) to improve the out-of-sample testing.

ML is widely applied in estimation of covariance matrixes for portfolio models, with the goal of sparse matrixes that omit small, often highly unstable, correlations. This has been an especially active area of research. Regularization techniques are employed widely to consistently estimate large covariance matrixes. There are two general mainstream approaches: rank-based and factor-model based covariance estimation (Fan et al. 2016). The rank-based method is appealing in the analysis of financial data because the distribution of the data-generating process is typically non-Gaussian and heavy-tailed. The sparsity assumption needs to be applied with care in covariance estimation, as it is sometimes unrealistic because the returns of financial securities are strongly correlated with risk drivers shared across and within markets. In the cases where the sparsity property is not directly applicable, factor-based methods can model the conditional sparsity; namely, we decompose the financial data into a common risk component and an idiosyncratic component. Conditional on the common risk factors, the covariance of the idiosyncratic component is sparse. In high dimensions, the unknown factors and loadings can be estimated by methods, including principal component analysis, and more recently variational autoencoders.

Pedersen et al. (2021) propose a simple shrinkage method called enhanced portfolio optimization to improve the estimate. This method first shrinks all correlations toward zero by modifying the correlation matrix $\tilde{\Omega} = (I-\Theta)\Omega + \Theta I$, where $\Theta \in [0,1]$ and $\tilde{\Omega}, \Omega, I$ are modified correlation matrix, original correlation matrix, and identity matrix, respectively. Then it computes the standard MVO portfolio. Specifically, it sets the shrinkage hyperparameter via a validation step to maximize the portfolio's Sharpe ratio. This machine learning approach is different from other existing literature that chooses correlation shrinkage to maximize the fit of the covariance matrix. They demonstrate that tuning to maximize Sharpe ratio yields a much larger shrinkage parameter, which fixes estimation errors in both risk and expected return and yields a large performance improvement empirically.

Shrinkage methods also have been vastly employed because many of the factor loadings will be equal to zero or nearly so, thus pinpointing the most critical features. Penalized regression methods include LASSO (Tibshirani 1996), the smoothly clipped absolute deviation penalized regression (Fan and Li 2001), and the regression with a minimax concave penalty (Zhang 2010). For example, Harvard University Endowment applies LASSO shrinkage procedures by penalizing the absolute value of the betas in its flexible indeterminate factor-based asset allocation model (Blyth et al. 2016). The motivation for macro-factor models is the strong shift to alternative asset categories that possess multiple risks (Swensen 2009). For instance, private equity is exposed to illiquidity risks, equity risks, leverage risks, and sometimes interest-rate risks. Indeed, these methods have expanded the potential for asset owners to customize definition of asset categories to support their decision processes, ongoing monitoring, and presentation to constituents (Mulvey and Holen 2016). A similar situation occurs in the traditional long-short factors, such as size, value, momentum, and profitability, wherein investment firms choose distinct routes for translating the broad definitions into concrete portfolios.

PERSONALIZED WEALTH MANAGEMENT AT SCALE

Personal finance is a major growth area, partially because of the increase in concentrated wealth across the globe.[10] Here, we refer to customized investment recommendations and related software tools, tailored for a single investor or firm. To render targeted advice requires knowledge of an individual's financial circumstances and estimates of that individual's aspirational goals. In conjunction, there is a need for significant new forms of financial planning systems that address reward and risk measures that align with the investor's goals. The resulting targets and risks can be approximated by means of goal-risks and liability-driven investing over temporal settings (Brunel et al. 2017; Martellini et al. 2019; Mulvey et al. 2019; Nevins 2014; Nystrup et al. 2019).

Personalized service requires looking beyond the traditional risk measures of performance, such as Sharpe ratios and maximum drawdown, to the risks faced by individuals that future liabilities will be unmet and financial goals will not be achieved in a timely manner. Take the case of a university endowment, where the risks entail the probability that the capital will generate inadequate returns and contributions to pay future aspirational levels (Mulvey and Liu 2016), or an individual who is setting aside annual savings in a defined contribution plan to retire at a certain age and is evaluating alternative target-date funds.[11] These issues fall under the umbrella term "personal finance," analogous to "personal medicine" whereby a treatment or drug is crafted to match individual circumstances.

Where do ML and data science impact these issues? Ideally, the availability of micro-level data allows a wealth manager to better understand clients, leading to tailored investment recommendations. As an example, a saving recommendation can be linked to the median saving for groups of people in similar circumstances as the client. Asset allocation can be referenced to one's peer group. There is evidence that individuals are sensitive to these comparisons. Micro-level data allows for these types of recommendations; many people consider micro-level data to be a distinguishing characteristic of big-tech firms and fintech (*The Economist* 2017; Kolanonic and Krishnamachari 2017).

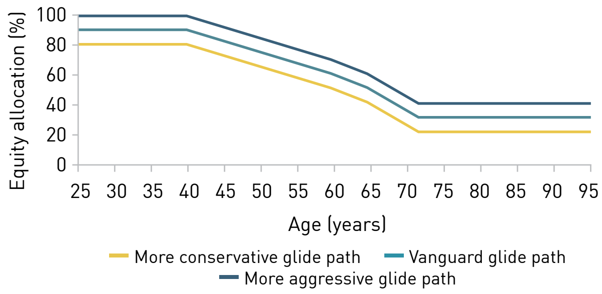
Given adequate, representative micro-level data and appropriate tools, investment managers have opportunities to develop better advice and also construct customized products with automated rebalancing rules. The rise of target-date funds for retirement is a straightforward case wherein the asset mix periodically is adjusted based on the investor's age and risk tolerance. Figure 3 shows the glidepath for a Vanguard fund relative to two others. Undoubtedly, greater customization will occur in the future, as individuals see the advantages.

There are related applications such as direct indexes for wealthy individuals and institutional investors, adaptive versions of target-date funds, and specialized securities such as tontines for retirees (Gemmo et al. 2020). Machine learning concepts, including recommender systems, may play a role by helping evaluate the pros and cons of these customized products. Many investors have inadequate time and expertise to carry out dynamic investment strategies, and customized products offer a ready path to implementation.

Similarly, enhancements in forecasting, such as those discussed previously, can have a positive impact on investment performance

Figure 3

GLIDE PATH FOR A VANGUARD TARGET-DATE FUND



Source: Donaldson et al. (2015)

for both asset returns and goal-driven measures by taking advantage of newer data sources and technical breakthroughs, including natural language processing (discussed below).

These improved forecasts can lead to dynamic strategies that are better able to adapt to changing circumstances and superior risk management, relative to both asset risks and goal-driven risks. Furthermore, the development of automated investment systems, referred to as robo-advisors, makes it possible to automatically implement various portfolio management tasks for retail investors, including tax-loss harvesting, asset location and goal-based investment (Grealish and Kolm 2021a, b).

APPLICATIONS OF NATURAL LANGUAGE PROCESSING

This section highlights advanced ML research that extends sentiment analysis in our quest to enhance investment performance. Much of this research is undertaken by experts in ML; advanced methods only recently have appeared in the investment management literature. The genesis of this research involves natural language processing (NLP) concepts. Historically, sentiment underpins many successful forecasts of the economy and markets such as the Conference Board's Consumer Confidence Index (2021) and the University of Michigan Index of Consumer Sentiment. Dashboards, such as Goldman Sachs Investment Research's Leading Indicators, provide forecasts of the future direction of markets and the economy (Hatzius et al. 2018).

The inevitable time lags and importance of sentiment motivates the need for improved methods for collecting and analyzing qualitative financial information. There have been many studies of sentiment by reference to lexicons, e.g., counting positive and negative words taken from financial and general interest publications and social media. These studies may account for the specialized language in financial markets (Loughran and McDonald 2011)—bull and bear markets—and ML concepts have proven beneficial (Agaian and Kolm 2017; Sohangir et al. 2018).

Big-tech firms have greatly improved NLP algorithms over the recent past. This domain is built on the idea of transforming words and written speech into vectors of real numbers—

language embedding. Once the qualitative data is embedded, the task of decoding the information can be deployed for tasks such as language translation, online search tasks, sentence extrapolation, understanding surveys, and so on. The goal is to convey the underlying meaning of words by reference to the context of the sentence or paragraph. Google was an early innovator, with an effort to "read" millions of publications on Wikipedia and other sources in order to pre-train a language. In theory and given specialized computer systems such as tensor processing units, once pre-training is accomplished, it is a relatively modest task to specialize the embeddings for particular applications. A number of pre-trainers are available, including the now famous Google's BERT, and OpenAI's GPT-2 and GPT-3 (Brown et al. 2020). In parallel, a few researchers have built specialized pre-trainers for financial analysis—finBERT (Liu, Huang et al. 2020). The BERT system has proven very effective at improving standardized ML tests (Devlin et al. 2019) across a wide variety of difficult NLP tasks. Mishev et al. (2020) provide details of this progression. This area is only beginning to become mainstream in wealth management, but we expect much growth in research along these directions (Dong et al. 2020).

A major breakthrough occurred in 2017 with the notion that informational content can be identified more precisely by selective analyses of the input written data (Vaswani et al. 2017). They introduce the transformer model, which utilizes the method called self-attention that focuses on various parts of the input by capturing the contextual relationship. It was a groundbreaking paper because the proposed model could take advantage of parallelism to significantly accelerate the training process. In addition, the transformer is capable of memorizing and capturing long-term contextual dependency from the input (Ramos-Perez et al. 2021).

The catchy title "attention" refers to the important attribute that human visual systems are able to focus on portions of a picture or video. Adding an attention layer to a deep neural network has shown to greatly decrease the computational resources required to pre-train a massive set of language data. Another benefit is that the training can be accomplished in a parallel computational mode—again increasing the NLP capabilities. Larger amounts of training data often lead to improved results on standardized tasks.

A direct extension of the written language research is to embed market-based time series data (Li et al. 2018), e.g., stock ticker data. NLP algorithms are custom made for deciphering and perhaps forecasting time series information (Zerveas et al. 2021). In a similar fashion, current research efforts are aimed at improvements in forecasting market events—including prices, bid-ask spreads, volume, and so on—over short time periods by "reading" historical time series and embedding these patterns. As before, the decoding phase can provide almost real-time matches with historical data in a quest to improve forecasting accuracy for trading systems.

A related research area entails alternative sources of sentiment and macroeconomic data. For example, many subindexes do not arrive in a timely fashion at a fixed date such as the first of the month. There is much research on alternative data sources to collect surrogate indicators (Rasekhschaffe and Jones 2019; Kolanonic and Krishnamachari 2017). Suppose that a market forecast indicates that inflation is a key feature. Then, there is incentive to collect inflation measures at a great granularity (micro-level) and higher frequency than monthly. Similarly, for investments in equities or debt, investors can aggregate publications, social media posts, or job postings, but they must use NLP topic models to connect the corporate issuer with the individual products or brands mentioned in an article (Ekster and Kolm 2019).

MULTI-PERIOD FINANCIAL PLANNING SYSTEMS

Since the early discoveries by Merton (1969) and Samuelson (1969), there have been research efforts on extending single period investment models such as that of Markowitz. Much of this work has focused on solving stylized model structures by dynamic programming and related algorithms so that, ideally, implementable policy prescriptions can be derived.[12] An example is the early study of transaction costs leading to the intuitive no-trade zones (Davis and Norman 1990). Recent ML research has shown that generic multi-period structures can be solved for realistic-size applications by employing third-level ML algorithms, especially versions of deep neural networks (LeCun et al. 2015). Bertsimas and Pachamanova (2008), Dempster et al. (2007), Li and Forsyth (2019), Li and Mulvey (2021), and Nystrup et al. (2019) provide a sample of research in solving multi-period financial planning models.

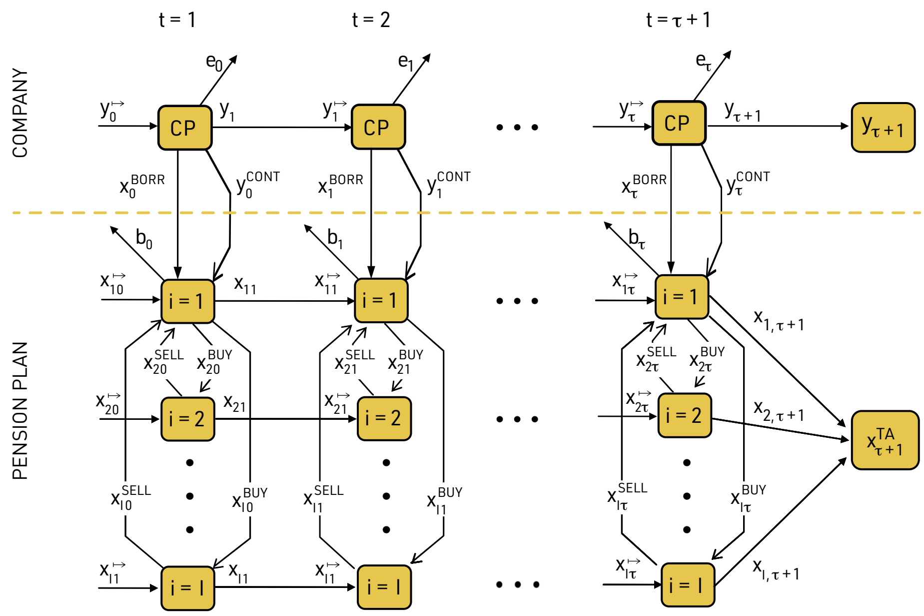
The single-period portfolio models, discussed earlier, are less useful than a multi-period framework because they weakly approximate practical issues. A full-scale multi-period model provides generality to address real-world considerations. Individual pension plan investors, for instance, should be keenly interested in ascertaining the likelihood of achieving a comfortable retirement rather than treating risks solely as short-term volatility. The optimal level of annual saving is onerous to address in a single-period setting. Investors can benefit by reference to risk measures dictated by goal-driven, temporal-linked objectives in addition to classical measures such as volatility. Applications include retirement plans, university endowments, nonprofit foundations, sovereign wealth funds, reinsurance companies, banks, and many others (see, e.g., Ziemba and Mulvey 1999).

A multi-period investment system can address transaction and market impact costs such as taxes, intermediate cash flows, and numerous other real-world considerations. As a generic

VOLUME 21
NUMBER 1
2022

Figure 4

MULTI-PERIOD INVESTMENT MODEL



Source: Mulvey et al. (2008)

example, figure 4 provides a graphical network of a financial planning model. Capital flows from left to right over time on the graph. Each column of nodes depicts the securities or asset categories at the beginning of each time period between the current time t=1 and the horizon t=τ+1. The vertical arcs represent rebalancing decisions and intermediate cash flows minus savings and spendings. The horizontal arcs represent the movement of capital across time periods in which the returns are indicated as stochastic multipliers along the horizontal arcs. At each period, asset performance is dictated by uncertain outcomes. Uncertainties are modeled via a discrete set of representative scenarios.[13] Each scenario is assigned its own graphical network. Dempster et al. (2007) and Mulvey et al. (2008) discuss these issues in detail. The time steps can be at any frequency depending upon the application, from real-time trading systems (Almgren and Chriss 2001) to financial planning for individuals and institutions with long horizons (Mulvey and Liu 2016; Li and Forsyth 2019). Many investment problems can be posed with a generic graphical optimization modeling framework.[14] Next, we highlight details for solving the resulting multi-period financial planning systems.

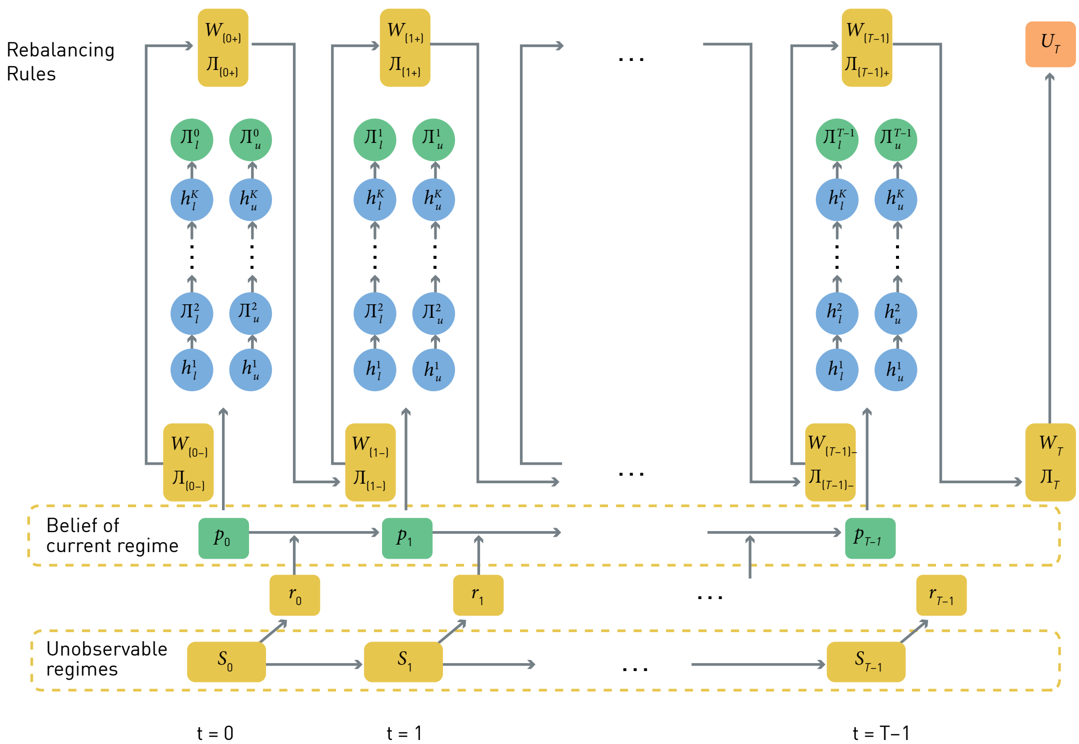
Discovering the optimal solution to a multi-period model is a complex task due to the curse of dimensionality, temporal uncertainties, and parameter estimation. To overcome computational constraints, researchers have implemented third-level ML algorithms such as feedforward neural networks or combinations of traditional algorithms such as dynamic programs and deep neural networks (Li and Forsyth 2019; Li and Mulvey 2021). Figure 5 shows the graphical structure of a neural network version of the previous multi-period investment model. On the left-hand side are the inputs to the neural network, such as prescribed by relevant economic and market conditions (features). On the right-hand side are the outputs—such as the expected wealth and risk measures at the horizon and intermediate time junctures. The intermediate parameters indicate the investment and related decisions that depend upon the inputs. In other words, depending upon the input picture, the deep neural network renders the missing investment decisions, along with the associated reward and risk measures.

The basic neural network algorithm strives to identify the best set of parameters that optimize the stated objective functions: Sharpe ratios, maximum drawdown values, probability of meeting a wealth goal at the horizon, and so on without reference to a particular model structure. This framework possesses generality to construct and solve numerous multi-period financial models. Importantly, computational time of employing neural networks can be close to linear in many cases. Today, we can address realistic-size financial planning models with multiple periods, given computational resources specialized for deep neural networks (GPUs and TPUs).

A promising extension of the deep neural network model includes an implicit layer within the deep neural network. Going back to von Neumann and the early days of decision-making under uncertainty, the longstanding traditional approach to decision models involves two steps. First, we model uncertainties by employing structural stochastic models and fit the parameters by minimizing a specified loss function.

Figure 5

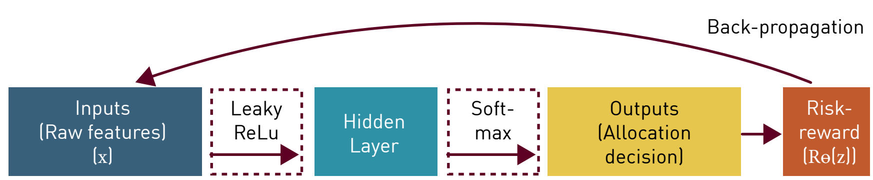
COMPUTATIONAL GRAPH OF THE DEEP NEURAL NETWORK



Source: Li and Mulvey (2021)

Figure 6

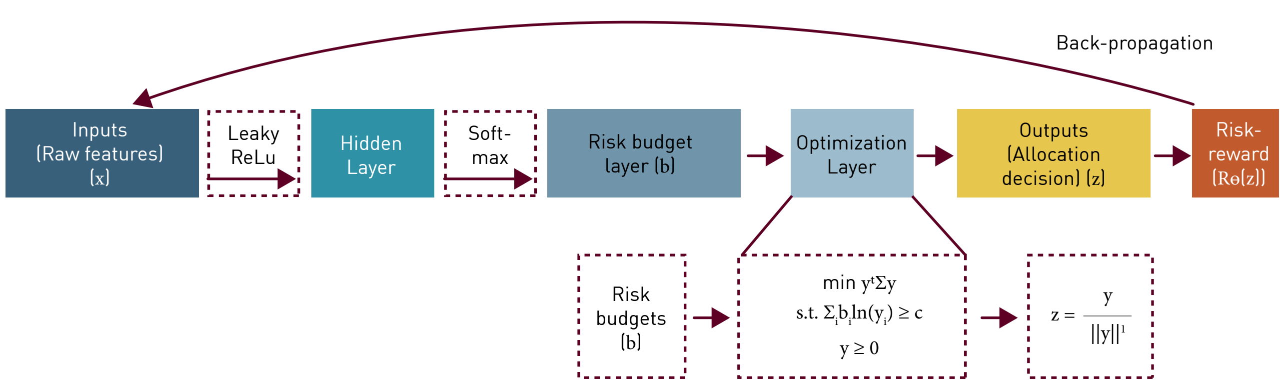
COMPUTATIONAL GRAPH OF MODEL-FREE APPROACH



Source: Uysal et al. (2021)

Figure 7

COMPUTATIONAL GRAPH OF MODEL-BASED APPROACH



Source: Uysal et al. (2021)

VOLUME 21
NUMBER 1
2022

Second, given the results of stage one, we construct and solve a stochastic optimization model to find the optimal decisions. However, the two optimization models (fitting parameters and maximizing an investor's utility function) possess differing objective functions.

To address this inconsistency and to include an optimization model, researchers are investigating the integration of the two parts into a single step via a deep neural network plus an implicit optimization layer[15]—called end-to-end modeling (Amos et al. 2018; Bertsimas and Kallus 2020; Donti et al. 2017). Here, rather than allowing the neural network algorithm to select any choice of parameters at each layer, the neural network architecture identifies input parameters for the optimization. There are possible advantages. First, the process may improve the solution algorithm, although the training takes longer than the neural network without an implicit layer. Second, there are ways to address the interpretability of the neural network due to the structure of the implicit layer, as has been recently proposed for image processing applications of neural networks (Chen et al. 2019). And the integrated model has consistency between the stochastic fitting and the decision models. Figure 5 shows a graph of a traditional deep neural network, whereas the end-to-end model with the implicit layer is depicted in figure 6.

OVERCOMING DATA CHALLENGES WITH GENERATIVE MODELS

Financial data poses a challenge for ML for several reasons. First, the signal-to-noise ratio is relatively low and not necessarily stationary over time. Second, representative data often is not freely and publicly available, thus limiting accessibility for researchers to explore methods and promulgate success. Third, the nature of representative data can be difficult to identify. This drives the demand for high-quality data augmentation, where generative models can play an essential role. Promising research has taken place in this quest (Wan et al. 2017; Wiese et al. 2020; Takahashi et al. 2019; Ratner et al. 2017). In these efforts, the generated data is evaluated through various metrics of distributional properties; herein, the generated models such as VAE (Wan et al. 2017) and GAN (Wiese et al. 2020; Takahashi et al. 2019) have shown benefits. For example, VAE can generate imputed data to complete a volatility surface (Bergeron et al. 2021). Further, data generation can be useful with imbalanced data (Wan et al. 2017), which is often the case in finance, such as merger-and-acquisition or crash predictions.

A primary goal when generating augmented data is to mitigate overfitting (Pardo and Lopez 2020). For instance, we can fine-tune the model to prevent backtesting overfitting (Koshiyama et al. 2021). As another example, when dealing with a large number of news articles, we can inversely feed the price movement as a condition into models to generate the most relevant news and thus avoid overfitting (Matsubara et al. 2018).

Apart from data generation, representation learning also can benefit from generative models. Because the generator learns how to generate data, it should identify some of the representation structure. GAN offers a tool to predict stock prices by capturing the underlying data distribution (Zhou et al. 2018; Zhang, Zhong et al. 2019). Modern machine learning technologies such as transformers have been shown to work well in learning representations from financial data (Sokolov et al. 2020). Recently, researchers have employed generative models to learn interpretable representations directly on time series (Fortuin et al. 2019); we expect similar analyses can be applied to financial data.

There are other applications regarding generative models in finance including fraud detection (Leangarun et al. 2018; Sethia et al. 2018), privacy (Assefa et al. 2020), and portfolio management (Mariani et al. 2019). In a nutshell, the generative models are naturally consistent with the goals in financial analysis, thus enriching the ways of applying machine learning methods to real-world data. However, the generation process depends upon the assumption that the method can derive synthetic data from historical data without strong biases and a naïve perspective. As they say, forecasting is difficult, especially when predicting future outcomes.

CHALLENGES AND FUTURE DIRECTIONS

Challenges that arise when applying machine learning algorithms to wealth management have common elements with other domains such as health sciences, where decisions can have long-term, high-stakes implications for businesses deploying them and consumers who are affected by them. Future implementations across fields will depend on technical breakthroughs and on behavioral, privacy, fairness, and regulatory considerations.

First, an important challenge involves backtesting and the appropriate selection of historical data for evaluating advanced investment strategies. Unfortunately, the nature of financial crises, recoveries, and mid-cycle periods can be difficult to project into the future as novel drivers of uncertainty arise. Recently, the coronavirus pandemic has been unique in its impact on markets and individuals across different sectors. Any business process that presumes continuity can be challenged by shifting patterns, and machine learning algorithms are no exception.

Overfitting of models becomes a greater issue as the wide array of machine learning techniques available, along with plentiful and inexpensive computing resources, allow investors to test myriad quantitative trading strategies and risk management models. Although we can use cross-validation to validate overfitting, that solution is imperfect. Cross-validation and out-of-sample testing require adequate data, and high-quality financial data are limited, and some data, like published audited financial metrics of extant companies, are only available on a monthly or quarterly basis for the past five to seven decades. Moreover, there is no true out-of-sample period when we perform historical backtests because investors are clearly aware of when and how the market crises and rallies unfolded in the past. Also, the academic literature on wealth management seldom considers realistic transaction costs and implementation shortfall in a comprehensive manner. Friction encountered in actual trading can consume a substantial fraction of expected returns of a dynamic portfolio strategy, especially high-turnover ones. Estimation of transaction costs is a major issue for financial institutions and an active area of application of machine learning (e.g., Zhang, Zohren et al. 2019). New paradigms are being developed to address these risks and opportunities, e.g., Arnott et al. (2019) develop a backtesting protocol for quantitative finance research.

The issue of limited data is especially acute in relatively new markets, e.g., cryptocurrencies such as bitcoin. In this case, history provides little to no evidence about the behavior of these assets under stressful time periods such as the 2008–2009 crash, and investors cannot take full advantage of the machine learning algorithms that depend upon representative data for training/validation and testing purposes (López de Prado 2020). Absent sufficient data for both the cross-validation process of hyperparameter selection and out-of-sample testing, ML algorithms do not reach their full capabilities.

Additional challenges involve the role of regulators, privacy issues, fairness, and intellectual property topics. For example, a fintech firm may have discovered a powerful algorithm for estimating credit risks. But privacy concerns, monopoly powers, fairness, and regulatory constraints may prevent the use of this technology, despite its superior performance.

Last, a number of machine learning algorithms possess opaque structures with millions of parameters and thus cannot be readily interpreted. Deep neural networks are especially prone to interpretability issues. Elsayed et al. (2021) argue that some deep learning models tend to be overly complex in comparison to other ML techniques. The authors propose a window-based input transformation that improves the performance of XGBoost to levels that outperform many state-of-the-art DNN models. Lack of interpretability can be a concern for executives overseeing algorithmic adoption, because they seek multifaceted understandings of mechanisms and risk. Business leaders often see great utility in applying their domain knowledge and business experience to high-impact decisions, and so they are wary of methods that are not amenable to intuitive understanding (Selbst and Barocas 2018). Rudin (2019) argues the importance of designing models that are inherently interpretable in the first place rather than using more complex "black box" methods in a high-stake decision context. Model complexity can pose difficulties for those closer to model deployment as well, obscuring the reasons that models do not perform as expected, or failing to provide frameworks for identifying when new observations diverge from those of the training datasets. There is general agreement that interpretability will be a future barrier and research is underway to address this challenge (Stanford University 2021). Interpretability may be less important in high-frequency decisions, such as transaction cost models, because shifts in performance could be observed readily in outcomes, allowing ex post performance to be evaluated quickly with limited risk. But even in these environments, unexpected events are arising, such as flash crashes, which can lead to unpredictable outcomes. The desire for transparency and interpretability has spurred a growth in probabilistic ML approaches (Ghahramani 2015).

CONCLUSION

We expect ongoing progress on applications of ML to many areas of wealth management. Investment performance applications will benefit from an increase in novel data sources, powerful algorithms for exploiting the data enabled by new computer architectures, and well-funded research teams at asset management firms. As these tools become more prevalent and powerful, there will be increasing numbers of ML implementations, such as fully automated trading systems (Burhani et al. 2020).

In conjunction, we envision a greater focus on personal finance, and by inference to reliance on information targeted to specific investors. Micro-level data will be critical. Machine learning algorithms offer opportunities to analyze data aimed at greater understanding of the needs and desires of individual and institutional investors.

ML technologies enable new opportunities to deploy customized recommendations and securities/products—e.g., direct indexing, which can complement current procedures and may in some cases replace human advisors. We foresee the deployment of automated dynamic strategies tailored to individual investors, rather than simplistic approaches such as current target-date funds. These products aim to reduce the investor's effort to implement dynamic rebalancing decisions—faster and easier. If carefully designed, these securities can improve financial well-being, e.g., by tax-loss harvesting. Improving the financial education of investors and advisors will be a significant related activity due to the complexity of the ML technologies.

New applications of these modern ML methods, such as extensions to sentiment analysis and NLP, are emerging. The availability of micro data provides two benefits. First, the focus shifts from theory to the data. Second, the algorithms can take advantage of a large number of parameters. For instance, the first-level elastic net regularization algorithm requires two hyperparameters—one for each norm: L1 and L2. The most recent ML algorithms (third level) take these ideas to the limit

VOLUME 21
NUMBER 1
2022

with millions or even billions of parameters (GPT-3 has almost 2 billion parameters). This complexity offers potential for addressing wealth management problems with new and powerful approaches.

Several challenges will hinder applications of ML. When algorithms are applied to individuals, privacy and fairness concerns need to be addressed. Regulations and compliance risks may evolve with the growth of automated investment systems, i.e., robo-advisors, and other customized and automated dynamic strategies for investors. Overfitting and the lack of interpretability may inhibit success in wealth management applications, especially with the limited data and non-stationarity typical of financial markets. The temptation to overfit is even greater when the approach or model itself is hard to justify by a priori reasoning. These issues are exacerbated by the massive number of parameters within advanced machine learning algorithms such as deep neural networks. Wealth managers need to be extra vigilant when the ML algorithms are opaque.

Despite these barriers, progressive improvements are occurring in wealth management based on machine learning concepts and the relevant micro-data sources. Machine learning has made dramatic progress over the past decade in areas that were thought to be inaccessible for decades, with transformational impact in many areas. Will wealth management be another success story? Time will tell. We are cautiously optimistic about the intersection of these domains, especially if greater attention is paid to the non-technical challenges.

John M. Mulvey, PhD, is professor of operations research and financial engineering and founding member of the Bendheim Center for Finance at Princeton University. Contact him at mulvey@princeton.edu.

Junhan Gu is a master in finance student in the Bendheim Center for Finance at Princeton University. Contact her at junhang@alumni.princeton.edu.

Margaret Holen, PhD, is a lecturer in the operations research and financial engineering department at Princeton University. Contact her at holen@princeton.edu or through LinkedIn.

Yuqi Nie is a graduate student in the electrical and computer engineering department at Princeton University. Contact him at ynie@princeton.edu.

ENDNOTES

1. There is debate about the so-called improvements that big-tech firms offer. For example, in many communities, competition from big-tech has been blamed for the loss of local businesses lamented by many people.
2. There is much attention from business and policy spheres on the management of the massive micro-level data, with a focus on efficiency along with fairness, ethics and privacy.
3. There are challenges in applying machine learning to investment problems, such as identifying alpha opportunities due to the changing nature of markets and other issues as discussed later in the paper.
4. Numerous related investment areas are seeing similar developments such as fully automated loans and fraud-detection systems. We leave these applications for others to discuss.
5. These levels (or generations) are less dependent upon chronology of their introduction than upon the nature and complexity of the methods.
6. Research is underway to improve the interpretability of the advanced methods and for robustness and inference purposes. We do not cover these concepts in this paper due to space limitations.
7. Determining probabilities from decision tree models is done with an additional step called calibration (see e.g., Zadrozny and Elkan 2002).
8. Additional related applications and methods are described in the appendix.
9. Shrinkage is useful when comparing performance of specialists. For example, MacKenzie et al. (2015) provide a short primer in evaluating hospital mortality across centers in the U.S. Northeast region.
10. Total wealth in the United States increased to an all-time high of $141 trillion in the fourth quarter 2021 (U.S. Federal Reserve 2021: Burgess 2021).
11. In the literature, there are competing terms: asset and liability management, dynamic financial planning, enterprise risk management (Dempster et al. 2007; Ziemba and Mulvey 1999).
12. Perhaps the most famous policy is the fixed-proportional rule wherein the asset proportions are rebalanced to a target asset allocation mix each period. This policy rule is optimal under a strict set of assumptions including temporal independence, no transaction costs nor intermediate cash flows, maximizing CRRA expected utility, and so on.
13. The graphical network is posed for each scenario. In addition, there is a need for non-anticipatory conditions. Additional constraints can be readily incorporated within a multi-period model.
14. There are numerous approaches for generating scenarios for the uncertain parameters. The traditional approach is to construct a set of stochastic processes, calibrated by means of capital market assumptions. Another approach is to employ bootstrap sampling of historical performance. Reinforcement learning provides another framework that might be exploited (Burhani et al. 2020; Jordan and Mitchell 2015; Li 2017).
15. A closely related framework is reinforcement learning, whereby the decisions are made in a progressive fashion and adjusted with regard to a reward system. This notion can be placed within the end-to-end framework.

REFERENCES

Agaian, S., and P. Kolm. 2017. Financial Sentiment Analysis Using Machine Learning Techniques. *International Journal of Management and Financial Innovations* 3, no. 1 (August): 1–9.

Almgren, R., and N. Chriss. 2001. Optimal Execution of Portfolio Transactions. *Journal of Risk* 3, no. 2 (winter): 5–40.

Amos, B., I. Rodriguez, J. Sacks, B. Boots, and J. Kolter. 2018. Differentiable MPC for End-to-End Planning and Control. *Advances in Neural Information Processing Systems* 31 https://papers.nips.cc/paper/2018/file/ba6d843eb4251a4526ce65d1807a9309-Paper.pdf.

Arnott, R., C. Harvey, and H. Markowitz. 2019. A Backtesting Protocol in the Era of Machine Learning. *Journal of Financial Data Science* 1, no. 1 (winter): 64–74.

Assefa, S., D. Dervovic, M. Mahfouz, T. Balch, P. Reddy, and M. Veloso. 2020, October. Generating Synthetic Data in Finance: Opportunities, Challenges and Pitfalls. In *Proceedings of the First ACM International Conference on AI in Finance*: 1–8. https://dl.acm.org/doi/abs/10.1145/3383455.3422554.

Bergeron, M., N. Fung, J. Hull, and Z. Poulos. 2021. Variational Autoencoders: A Hands-Off Approach to Volatility. arXiv preprint arXiv:2102.03945.

Bertsimas, D., and N. Kallus. 2020. From Predictive to Prescriptive Analytics. *Management Science* 66, no. 3: 1,025–1,044.

Bertsimas, D., and D. Pachamanova. 2008. Robust Multiperiod Portfolio Management in the Presence of Transaction Costs. *Computers and Operations Research* 35, no. 1 (January): 3–17.

Bishop, C. M., and N. M. Nasrabadi. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Blyth, S., M. C. Szigety, and J. Xia. 2016. Flexible Indeterminate Factor-Based Asset Allocation. *Journal of Portfolio Management* 42, no. 5: 79–93.

Brown, T., et al. 2020. Language Models are Few-shot Learners. Advances in Neural Information Processing Systems 33. https://arxiv.org/abs/2005.14165.

Brunel, J., T. Idzorek, and J. Mulvey. 2017. Asset Allocation in Practice 1, CFA Institute, Report for CFA Exam 3.

Burgess, R. 2021. The Most Important Number of the Week is $142 Trillion. *Bloomberg* (September 25). https://www.bloombergquint.com/gadfly/record-u-s-household-net-worth-of-142-trillion-is-double-edged-sword.

Burhani, H., G. Ding, P. Hernandez-Leal, S. Prince, D. Shi, and S. Szeto. 2020. Aiden: Reinforcement Learning for Order Execution. Borealis AI Research. https://www.borealisai.com/en/blog/aiden-reinforcement-learning-for-order-execution/.

Calomiris, C. W., and H. Mamaysky. 2019. How News and Its Context Drive Risk and Returns Around the World. *Journal of Financial Economics* 133, no. 2 (August): 299–336.

Chauvet, M., and S. Potter. 2013. Forecasting Output. Chapter 3 in *Handbook of Economic Forecasting* (Vol. 2, Part A): 141–194.

Chen, C., O. Li, C. Tao, A. Barnett, C. Rudin, and J. K. Su. 2019. This Looks Like That: Deep Learning for Interpretable Image Recognition. *Advances in Neural Information Processing Systems* 32. https://papers.nips.cc/paper/2019/hash/adf7ee2dcf142b0e11888e72b43fcb75-Abstract.html.

Conference Board. 2021. Consumer Confidence Survey and Index. https://conference-board.org/data/consumerconfidence.cfm.

Cont, R., and A. De Larrard. 2013. Price Dynamics in a Markovian Limit Order Market. *SIAM Journal on Financial Mathematics* 4, no. 1: 1–25.

Coulombe, P. G., M. Leroux, D. Stevanovic, and S. Surprenant. 2020. How is Machine Learning Useful for Macroeconomic Forecasting? arXiv preprint arXiv:2008.12477.

Davis, M., and A. Norman. 1990. Portfolio Selection with Transaction Costs. *Mathematics of Operations Research* 15, no. 4 (November): 676–713.

Dempster, M., G. Mitra, and G. Pflug. 2007. Introduction to Special Issue on Financial Planning in a Dynamic Setting. *Quantitative Finance* 7, no. 2 (April): 111–112. doi:10.1080/14697680701287144.

Deveikyte, J., H. Geman, C. Piccari. and A. Provetti. 2020. A Sentiment Analysis Approach to the Prediction of Market Volatility (December). arXiv preprint arXiv:2012.05906.

Devlin, J., M. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, June 2–7, 2019, Minneapolis, Minnesota: 4,171–4,186.

Donaldson, S., F. Kinniry, Jr., V. Maciulis, A. Patterson, and M. DiJoseph. 2015. Vanguard's Approach to Target-Date Funds. Vanguard Research. https://www.vanguard.com/pdf/s167.pdf.

Dong, Y., D. Yan, A. Almudaifer, S. Yan, Z. Jiang, and Y. Zhou. 2020. BELT: A Pipeline for Stock Price Prediction Using News. *2020 IEEE International Conference on Big Data (Big Data)*, 2020: 1,137–1,146. doi: 10.1109/BigData50022.2020.9378345.

Donti, P., B. Amos, and J. Kolter. 2017. Task-based End-to-end Model Learning in Stochastic Optimization. *Advances in Neural Information Processing Systems* 30: 5,484–5,494. https://arxiv.org/abs/1703.04529.

Ekster, G., and P. Kolm, 2021. Alternative Data in Investment Management: Usage, Challenges, and Valuation. *Journal of Financial Data Science* 3, no. 4 (fall): 10–32.

Elsayed, S., D. Thyssens, A. Rashed, H. S. Jomaa, and L. Schmidt-Thieme, 2021. Do We Really Need Deep Learning Models for Time Series Forecasting? arXiv preprint arXiv:2101.02118.

Erlwein-Sayer, C., S. Grimm, A. Pieper, and R. Alsac. 2021. Forecasting Corporate Credit Spreads: Regime-Switching in LSTM. https://ssrn.com/abstract=4003338.

Fan, J., and R. Li. 2001. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* 96, no. 456: 1,348–1,359.

Fan, J., Y. Liao, and H. Liu. 2016. An Overview of the Estimation of Large Covariance and Precision Matrices. *The Econometrics Journal* 19, no. 1: C1–C32.

Fan, X., X. Guo, Q. Chen, Y. Chen, T. Wang, and Y. Zhang, 2022. Data Augmentation of Credit Default Swap Transactions Based on Sequential GAN. *Information Processing and Management* 59, no. 3 (May). https://www.sciencedirect.com/science/article/pii/S0306457322000188.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the Factor Zoo: A Test of New Factors. *Journal of Finance* 75, no. 3: 1,327–1,370.

Fortuin, V., M. Hüser, F. Locatello, H. Strathmann, and G. Rätsch. 2019. SOM-VAE: Interpretable Discrete Representation Learning on Time Series. *International Conference on Learning Representation*, May 6–9, 2019, New Orleans, Louisiana.

Gemmo, I., R. Rogalla, and J. Weinert. 2020. Optimal Portfolio Choice with Tontines Under Systematic Longevity Risk. *Annuals of Actuarial Science* 14, no. 2: 302–315. https://EconPapers.repec.org/RePEc:cup:anacsi:v:14:y:2020:i:2:p:302-315_4.

Ghahramani, Z. 2015. Probabilistic Machine Learning and Artificial Intelligence. *Nature* 521, no. 7553: 452–459.

Giglio, S., Y. Liao, and D. Xiu. 2021. Thousands of Alpha Tests. *Review of Financial Studies* 34, no. 7: 3,456–3,496.

Glasserman, P., and H. Mamaysky. 2019. Does Unusual News Forecast Market Stress? *Journal of Financial and Quantitative Analysis* 54, no. 5: 1,937–1,974.

Glasserman, P., K. Krstovski, P. Laliberte, and H. Mamaysky. 2020. Choosing News Topics to Explain Stock Market Returns. arXiv preprint arXiv:2010.07289.

Goodfellow, I. 2016. NIPS 2016 Tutorial: Generative Adversarial Networks. arXiv preprint arXiv:1701.00160.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27. https://papers.nips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Grealish, A., and P. N. Kolm. 2021a. Robo-Advisors Today and Tomorrow: Investment Advice Is Just an App Away. *Journal of Wealth Management* 24, no. 3 (winter): 144–155.

——. 2021b. Robo-Advisory: From Investing Principles and Algorithms to Future Developments (January 31, 2021). https://ssrn.com/abstract=3776826.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33, no. 3: 2,223–2,273.

Hamilton, J. D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica: Journal of the Econometric Society* 57, no. 2: 357–384.

Harvey, C. R., and Y. Liu. 2021. Uncovering the Iceberg from Its Tip: A Model of Publication Bias and p-Hacking. https://ssrn.com/abstract=3865813.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed. New York: Springer.

Hatzius, J., S. Stehn, N. Fawcett, and M. Chaudhary. 2018. Can Our Leading CAI Help Predict Asset Prices? Goldman Sachs Economics Research Report.

Heaton, J., N. Polson, and J. Witte. 2018. Deep Learning in Finance. arXiv preprint arXiv:1602.06561.

Huang, J., J. Chai and S. Cho. 2020. Deep Learning in Finance and Banking: A Literature Review and Classification. *Frontiers of Business Research in China* 14, no. 13. https://doi.org/10.1186/s11782-020-00082-6.

Huang, W., C. A. Lehalle, and M. Rosenbaum. 2015. Simulating and Analyzing Order Book Data: The Queue-Reactive Model. *Journal of the American Statistical Association* 110, no. 509: 107–122.

Israel, R., B. Kelly, and T. Moskowitz. 2020. Can Machines "Learn" Finance? *Journal of Investment Management* 18, no. 2: 23–36.

Jordan, M., and T. Mitchell. 2015. Machine Learning: Trends, Perspectives, and Prospects. *Science* 349: 255–260.

Kingma, D., and M. Welling. 2013. Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114. https://arxiv.org/abs/1406.5298.

Kingma, D., D. Rezende, S. Mohamed, and M. Welling. 2014. Semi-supervised Learning with Deep Generative Models. Advances in Neural Information Processing Systems 27. arXiv:1406.5298.

Kolanonic, M., and R. Krishnamachari. 2017. Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing. J.P. Morgan Report. https://www.cfasociety.org/cleveland/Lists/Events%20Calendar/Attachments/1045/BIG-Data_AI-JPMmay2017.pdf.

Kolm, P. N., J. Turiel, and N. Westray. 2021. Deep Order Flow Imbalance: Extracting Alpha at Multiple Horizons from the Limit Order Book. https://ssrn.com/abstract=3900141.

Koshiyama, A., N. Firoozye, and P. Treleaven. 2021. Generative Adversarial Networks for Financial Trading Strategies Fine-Tuning and Combination. *Quantitative Finance* 21, no. 5: 797–813.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the Cross-Section. *Journal of Financial Economics* 135, no. 2: 271–292.

Kramer, M. 1991. Nonlinear Principal Component Analysis Using Auto-Associative Neural Networks. American Institute of Chemical Engineers, *AIChE Journal* 37, no. 2: 233–243.

Leangarun, T., P. Tangamchit, and S. Thajchayapong. 2018. Stock Price Manipulation Detection Using Generative Adversarial Networks.

In *IEEE Symposium Series on Computational Intelligence (SSCI).* Bengaluru, India, November 18–21, 2018.
LeCun, Y., Y. Bengio, and G. Hinton, 2015. Deep Learning. *Nature* 521: 436–444.
Li, H., Y. Shen, and Y. Zhu. 2018. Stock Price Prediction Using Attention-based Multi-Input LSTM. In *Proceedings of Machine Learning Research* 95: 454–469.
Li, X., A. S. Uysal, and J. Mulvey. 2022. Multi Period Portfolio Optimization Using Model Predictive Control with Mean-Variance and Risk Parity Frameworks. *European Journal of Operational Research* 299, no. 3: 1,158–1,176.
Li, X., and J. Mulvey. 2021. Portfolio Optimization Under Regime Switching and Transaction Costs: Combining Neural Networks and Dynamic Programs. *INFORMS Journal on Optimization* 3, no. 4 (fall): 315–443, C3.
Li, Y. 2017. Deep Reinforcement Learning: An Overview. arXiv: 1701.07274v2.
Li, Y., and P. Forsyth. 2019. A Data-driven Neural Network Approach to Optimal Asset Allocation for Target Based Defined Contribution Pension Plans. *Insurance: Mathematics and Econometrics* 86: 189–204.
Liu, Y., Q. Liu, H. Zhao, Z. Pan, and C. Liu. 2020, April. Adaptive Quantitative Trading: An Imitative Deep Reinforcement Learning Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence* 34, no. 2: 2,128–2,135.
Liu, Z., D. Huang, K. Huang, Z. Li, and J. Zhao. 2020. FinBERT: A Pre-trained Financial Language Representative Model for Financial Text Mining. In *Proceedings of 29th International Conference on Artificial Intelligence, Special Track on AI in FinTech*. https://www.ijcai.org/proceedings/2020/0622.pdf.
López de Prado, M. 2020. *Machine Learning for Asset Managers.* Cambridge, UK: Cambridge University Press.
Loughran, T., and W. McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10Ks. *Journal of Finance* 66, no. 1: 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.
MacKenzie, T., G. Grunkemeier, G. Gurnwald, J. O'Malley, C. Bohn, Y. Wu, and D. Malenka. 2015. A Primer on Using Shrinkage to Compare In-Hospital Mortality Between Center. *Annals of Thoracic Surgery* 99, no. 3 (March): 757–761.
Mariani, G., Y. Zhu, J. Li, F. Scheidegger, R. Istrate, C. Bekas, A. Cristiano, and I. Malossi. 2019. PAGAN: Portfolio Analysis with Generative Adversarial Networks. arXiv preprint arXiv:1909.10578.
Markowitz, H. 1952. The Utility of Wealth. *Journal of Political Economy* 60, no. 2: 151–158.
Martellini, L., V. Milhau, and J. Mulvey. 2019. Securing Replacement Income with Goal-Based Retirement Investing Strategies. *Journal of Retirement* 7, no. 4: 8–26.
Matsubara, T., R. Akita, and K. Uehara. 2018. Stock Price Prediction by Deep Neural Generative Model of News Articles. Institute of Electronics, Information and Communication Engineers. *IEICE Transactions on Information and Systems* E101–D, no. 4.
Merton, R. 1969. Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *Review of Economics and Statistics* 51, no. 3: 247–257.
Mishev, K., A. Giorgjevikj, I. Vodenska, L. Chitkushev, and D. Trajanov. 2020. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access* 8: 131,662–131,682. https://ieeexplore.ieee.org/document/9142175.
Monk, A., M. Prins, and D. Rook. 2019. Rethinking Alternative Data in Institutional Investment. *Journal of Financial Data Science* 1, no. 1 (winter): 14–31.
Mulvey, J., and M. Holen. 2016. The Evolution of Asset Categories: Lessons from University Endowments. *Journal of Investment Consulting* 17, no. 2: 48–58.
Mulvey, J., and H. Liu. 2016. Identifying Economic Regimes: Reducing Downside Risks for University Endowments and Foundations. *Journal of Portfolio Management* 43, no. 1: 100–108.
Mulvey, J., L. Martellini, H. Hao, and N. Li. 2019. A Factor- and Goal-Driven Model for Defined Benefit Pensions: Setting Realistic Benefits. *Journal of Portfolio Management* 45, no. 3: 165–177.
Mulvey, J., K. Simsek, Z. Zhang, F. Fabozzi, and W. Pauling, 2008. Assisting Defined-Benefit Pension Plans. *Operations Research* 56, no. 5: 1,066–1,078.
Nevins, D. 2014. Goals-based Investing: Integrating Traditional and Behavioral Finance. *Journal of Wealth Management* 6, no. 4 (spring): 8–23.
Nevmyvaka, Y., Y. Feng, and M. Kearns. 2006. Reinforcement Learning for Optimized Trade Execution. In *Proceedings of the 23rd International Conference on Machine Learning*, June 25–29, 2006, Pittsburgh, Pennsylvania: 673–680.
Ng, A., and M. Jordan. 2001. On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes. *Advances in Neural Information Processing Systems* 14. https://papers.nips.cc/paper/2001/file/7b7a53e239400a13bd6be6c91c4f6c4e-Paper.pdf.
Ning, B., F. H. T. Lin, and S. Jaimungal. 2018. Double Deep Q-Learning for Optimal Execution. arXiv/1812.06600.
Nystrup, P., S. Boyd, E. Lindström, and H. Madsen. 2019. Multi-period Portfolio Selection with Drawdown Control. *Annals of Operations Research* 282, no. 1–2: 245–271.
Nystrup, P., P. Kolm, and E. Lindström. 2020. Greedy Online Classification of Persistent Market States Using Realized Intraday Volatility Features. *Journal of Financial Data Science* 2, no. 3 (summer): 25–39.
Oord, A., N. Kalchbrenner, and K. Kavukcuoglu. 2016. Pixel Recurrent Neural Networks. arXiv:1601.06759.
Pardo, F., and R. Lopez. 2020. Mitigating Overfitting on Financial Datasets with Generative Adversarial Networks. *Journal of Financial Data Science* 2, no. 1 (winter): 76–85.
Pedersen, L. H., A. Babu, and A. Levine. 2021. Enhanced Portfolio Optimization. *Financial Analysts Journal* 77, no. 2: 124–151.
Ramos-Perez, E., P. Alonso-Gonzalez, and J. Nunez-Valazquez. 2021. Multi-Transformer: A New Neural Network-Based Architecture for Forecasting S&P 500 Volatility. *Mathematics* (Multidisciplinary Digital Publishing Institute) 9, no. 15. http://doi.org/10.3390/math9151794.
Rasekhschaffe, K., and R. Jones. 2019. Machine Learning for Stock Selection. *Financial Analysts Journal* 75, no 3: 70–88.
Ratner, A., S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré. 2017. Snorkel: Rapid Training Data Creation with Weak Supervision. In *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases* 11, no. 3: 269–282. doi: 10.1007/s00778-019-00552-1.
Rudin, C. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1, no. 5: 206–215.
Salakhutdinov, R., and G. Hinton. 2009. Deep Boltzmann Machines. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*. April 16–18, 2009. Clearwater Beach, Florida.
Samuelson, P. 1969. Lifetime Portfolio Selection by Dynamic Stochastic Programming. *Review of Economic Statistics* 5, no. 3: 239–246.
Selbst, A. T., and S. Barocas, 2018, The Intuitive Appeal of Explainable Machines, 87 *Fordham Law Review* 1085. https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5569&context=flr.
Sethia, A., R. Patel, and P. Raut. 2018. Data Augmentation using Generative Models for Credit Card Fraud Detection. In *International Conference on Computing Communication and Automation.* December 14–15, 2018.Greater Noida, India.
Silver, D., A. Huang, C. Maddison, A. Guez, L. Sifre, G. V. D. Driessche, and J. Schrittwieser. 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529 (January): 484–489.
Sohangir, S., D. Wang, A. Pomeranets, and T. Khoshgoftaar. 2018. Big Data: Deep Learning for Financial Sentiment Analysis. *Journal of Big Data* 5, no. 3: 1–25. https://journalofbigdata.springeropen.com/track/pdf/10.1186/s40537-017-0111-6.pdf.
Sokolov, A., J. Mostovoy, B. Parker, and L. Seco. 2020. Neural Embeddings of Financial Time-Series Data. *Journal of Financial Data Science* 2, no. 4 (fall): 33–43. doi: jfds.2020.1.041.
Stanford University. 2021. Gathering Strength, Gathering Storm: The One Hundred Year Study on Artificial Intelligence. Study Panel Report. https://ai100.stanford.edu/2021-report/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence.
Stein, C. 1956. Inadmissibility of the Usual Estimator of the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium* 1: 197–206.
———. 1981. Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics* 9, no. 6: 1,135–1,151.
Stock, J., and M. Watson. 2019. *Introduction to Econometrics*, 4th ed. Pearson.
Swensen, D. 2009. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*, 2nd edition. New York: Free Press.
Takahashi, S., Y. Chen, and K. Tanaka-Ishii. 2019. Modeling Financial Time-Series with Generative Adversarial Networks. *Physica A: Statistical Mechanics and Its Applications* 527 (August): 121261. https://doi.org/10.1016/j.physa.2019.121261.

The Economist. 2017. The World's Most Valuable Resource is No Longer Oil but Data. *The Economist* (May 6). https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Théate, T., and D. Ernst. 2021. An Application of Deep Reinforcement Learning to Algorithmic Trading. *Expert Systems with Applications* 173: 114632.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1: 267–288.

——. 2014. Adaptive Piecewise Polynomial Estimation Via Trend Filtering. *Annals of Statistics* 42, no. 1: 285–323.

Uematsu, Y., and S. Tanaka. 2019. High-dimensional Macroeconomic Forecasting and Variable Selection via Penalized Regression. *Econometrics Journal* 2, no. 1: 34–56.

U.S. Federal Reserve System. 2021. Derivation of U.S. Net Worth. https://www.federalreserve.gov/releases/z1/20210923/html/b1.htm.

Uysal, A., and J. Mulvey. 2021. A Machine Learning Approach in Regime-Switching Risk Parity Portfolios. *Journal of Financial Data Science* 3, no. 2: 87–108.

Uysal, A., X. Li, and J. Mulvey. 2021. End-to-End Risk Budgeting Portfolio Optimization with Neural Networks. arXiv preprint arXiv:2107.04636.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention is All You Need. *Advances in Neural Information Processing Systems* 30. https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Veloso M., T. Balch, D. Borrajo, P. Reddy, and S. Shah. 2021. Artificial Intelligence Research in Finance: Discussion and Examples. *Oxford Review of Economic Policy* 37: 564–584.

Wan, Z., Y. Zhang, and H. He. 2017. Variational Autoencoder Based Synthetic Data Generation for Imbalanced Learning. In *IEEE Symposium Series on Computational Intelligence (SSCI)*. Honolulu, Hawaii, November 27–December 1, 2017.

Wiese, M., R. Knobloch, R. Korn, and P. Kretschmer. 2020. Quant GANs: Deep Generation of Financial Time Series. *Quantitative Finance* 20, no. 9: 1,419–1,440.

Zadrozny, B., and C. Elkan, 2002, July. Transforming Classifier Scores into Accurate Multiclass Probability Estimates. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining:* 694–699.

Zerveas, G., S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff. 2021. A Transformer-based Framework for Multivariate Time Series Representative Learning. In *KDD '21, Virtual Event*. http://doi.org/10.1145/3447548.3467401.

Zhang, C. 2010. Nearly Unbiased Variable Selection under Minimax Concave Penalty. *Annals of Statistics* 38, no. 2: 894–942. doi: 10.1214/09-AOS729.

Zhang, K., G. Zhong, J. Dong, S. Wang, and Y. Wang. 2019. Stock Market Prediction Based on Generative Adversarial Network. *Procedia Computer Science* 147: 400–406.

Zhang, Z., S. Zohren, and S. Roberts. 2019. DeepLOB: Deep Convolutional Neural Networks for Limit Order Books. In *IEEE Transactions on Signal Processing* 67, no. 11: 3,001–3,012. DOI: 10.1109/TSP.2019.2907260.

Zhou, X., Z. Pan, G. Hu, S. Tang, and C. Zhao. 2018. Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets. *Mathematical Problems in Engineering* 2018, article ID 4907423. https://www.hindawi.com/journals/mpe/2018/4907423/.

Ziemba, W., and J. Mulvey (editors). 1999. *Worldwide Asset and Liability Modeling*. Cambridge, UK: Cambridge University Press.

APPENDIX
HOW ML ENHANCES INVESTMENT PERFORMANCE

The appendix discusses how machine learning techniques can enhance investment performance. We start from machine learning applications in macroeconomic forecasting. Next, we discuss tactical and operational decisions such as short-term trading strategies using accurate forecasts of market trends and asset returns, and lower transaction and market impact costs. We provide a selective review of several machine learning applications and implementations.

MACROECONOMIC FORECASTING

Occasional dramatic changes in macroeconomic and financial time series are usually associated with financial crises and recessions, followed by heavy tails, volatility clustering, and co-movements among asset returns. Heavy tails and financial contagion risks necessitate the forecast of market trend, or more specifically, detection and prediction of regimes, which can be defined as different time periods with varying patterns and feature loadings.

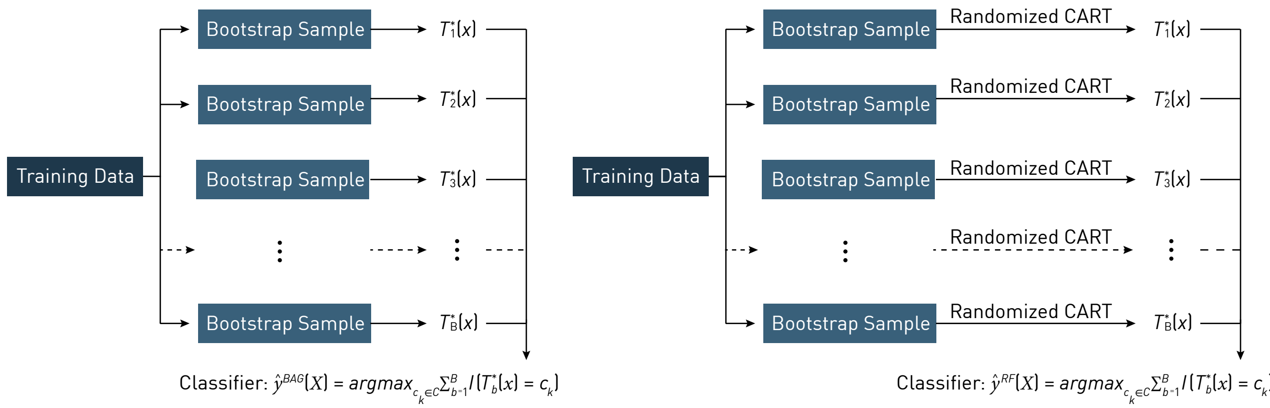
Traditional approaches collect macroeconomic data and survey market sentiments in order to forecast the future economy. Leading economic indicators can be useful to forecast the timing, probability, and magnitude of the economy or market condition. As an example, Goldman Sachs introduced its leading current activity indicator (LCAI) in 2015, an alternative summary measure of economic growth constructed from major real activity indicators, including surveys on consumer confidence and sentiment (Hatzius et al. 2018). LCAI provides an early signal on the short-term global growth momentum: The indicator led the 2016 sell-off in U.S. Treasury yields, because it picked up a few months ahead of long-term rates. It also moved ahead of equity prices, weakening ahead of the 2015 equity soft-patch and leading the 2017 S&P rally. However, Goldman does not disclose the details about how it selected these features—likely by expert judgment.

Extending traditional approaches such as Goldman's LCAI, a growing number of studies have applied recent machine learning models in macroeconomic forecasting. In the data-rich environment, theoretical and empirical research generally follows two directions to handle a large macroeconomic dataset effectively: one is the factor model approach, where we assume the common variation among many observed economic variables can be represented by unobserved factors (Stock and Watson 2019); the other is the sparse modeling approach, a visible alternative to the factor models, which imposes parsimonious structure by different feature selection methods to forecast macroeconomic outcomes (Uematsu and Tanaka 2019). Progress has also been made in understanding the properties of ML algorithms such as nonlinearity, regularization, cross-validation and alternative loss function when they are applied to predict macroeconomic outcome (Coulombe et al. 2020).

Modeling the varying patterns in financial time series by regime-switching models has become a common practice to hedge against tail risks. Regimes can be defined by various methods, including the trend filtering algorithm (Tibshirani 2014). After the regime identification, the next period's regime can be predicted using different supervised learning approaches. Hamilton (1989) introduces a Markov-switching framework to model the business cycle using U.S. real gross national

Figure A1

ILLUSTRATION OF BAGGING AND RANDOM FOREST ALGORITHM



product. Hidden Markov model classification is a popular approach in regime modeling and prediction, partly due to the interpretability of the resulting hidden states in terms of business cycle, risk-on and risk-off, etc. Chauvet and Potter (2013) reproduce the real-time forecasting problem of U.S. output growth using solely the data available at the forecast time, without looking-forward bias. They find that most traditional econometrics models, such as autoregressive models, perform poorly during recessions, whereas the dynamic factor Markov switching model has better accuracy. Nystrup et al. (2020) introduce a greedy online classifier that contemporaneously determines which hidden state a new observation belongs to, with possible applications to moderate to high frequency trading. Erlwein-Sayer et al. (2021) estimate an HMM model and feed the regime-switching information as features into an LSTM, in order to forecast the corporate credit spread. Their findings indicate that the LSTM performance can be improved when regime information is added.

With the regime forecast information, investors can better diversify risks and protect capital during crash periods. For example, the probability of the next period being a crash can be estimated using machine learning techniques such as random forest and gradient boosting (Uysal and Mulvey 2021). Then, a heuristic or a separate optimization algorithm can be employed to form a portfolio based on these probability predictions.

FORECAST OF ASSET RETURNS

Empirical asset pricing research aims to explain the cross-section of expected returns and measure the risk premium of different assets. Over the decades, academic researchers and practitioners have accumulated hundreds of factors that are significantly predictive and generate positive risk-adjusted returns beyond traditional risk factors. Traditional statistical tools are no longer an adequate choice if a researcher would like to test potentially hundreds of hypotheses, or to compute conditional expectation of a future excess return based on high dimensional factors, the number of which even exceeds the number of observations. Mainly in response to the high correlation among factors, investors and researchers have utilized variable selection and dimension reduction techniques, such as principal component analysis, to identify prevailing factors driving the market.

Kozak et al. (2020) tackle the challenge in this high-dimensional setting by imposing an economically motivated prior on stochastic discount factor coefficients that shrinks the contributions of low-variance principal components of the candidate factors, and thus summarizes the joint explanatory power of a large number of cross-sectional return predictors. Feng et al. (2020) provide an approach for improving the robustness of selecting new factors relative to a set of benchmark factors. Gu et al. (2020) demonstrate large economic gains to investors using machine learning forecasts with linear regression, generalized linear models with penalization, dimension reduction via principal components regression and partial least squares, regression trees (boosted trees and random forests), and neural networks.

Those modern techniques accommodate nonlinearities and improve the return predictions. For example, regression trees are known to be interpretable but highly unstable, because a small change in the data may drastically change the model. We can promote stability by using bagging (bootstrap aggregating) algorithms. More specifically, instead of fitting the tree on all the features as in the original bagging algorithm, we can use the random forest method by fitting the trees on randomly selected subsets out of all the features (see figure A1). The rationale is that the correlation among samples hinders variance reduction, so we use random dropouts to decorrelate. This can solve the instability problem, where we simply estimate the desired regression tree on many bootstrap samples with random dropouts and make the final prediction as the average of the predictions across the trees.

TRADING STRATEGIES USING ALTERNATIVE DATA

In addition to traditional structured datasets that originate from securities exchanges or regulatory disclosures, quantitative and fundamental institutional investors also leverage alternative data to uncover investment opportunities. Alternative data is less structured than the traditional sets of data, hence it is more challenging to extract useful information from those alternative sources such as financial transactions, mobile devices, satellites, and social media. Issues include a shortage of historical data, a small coverage universe, and various irregularities specific to each alternative dataset limit their applicability for algorithmic-only funds (Ekster and Kolm 2021). Nonetheless, as the alternative data vendors become more sophisticated, the deployment of alternative-data-driven investment becomes more promising in creating innovative alphas for the investors' portfolios.

The institutional investors will first evaluate the cost of acquiring a dataset (both the direct purchase cost, but also the opportunity cost of time invested in analyzing a dataset), and then use appropriate technologies and methods to process and analyze those data. Social media sentiment analysis and web search trends are two popular types of alternative data from individual activities. Sentiment analysis is applicable not only to individual stock names; it can also be used to trade broad market indexes. See Calomiris and Mamaysky (2019), Deveikyte et al. (2020), Glasserman and Mamaysky (2019), and Glasserman et al. (2020) for recent studies. Further ML concepts based on language embeddings are discussed in the "Applications of Natural Language Processing" section.

It is noteworthy that the development in the alternative data market likely will change the investment landscape. As more investors obtain insights from alternative datasets, the market will start reacting faster. The traditional datasets may gradually lose the ability to generate predictive signals and more standardized and high-quality alternative datasets will become available (Kolanonic and Krishnamachari 2017; Ekster and Kolm 2021).

HIGH FREQUENCY TRADING

In recent years, increasingly widespread high-frequency data has enabled investors to study the financial ecosystem through a different lens—market participants' actions and interactions. Many of the price prediction ideas emerge from the identification of informative trading behaviors: reversion from sudden moves, less liquid asset following more liquid asset, or large traders with directional goals. Some well-known strategies using high-frequency data include bid-ask imbalance, trade imbalance, "sweep" signal for rapid reversion, and pairs trading. In the trading signal development and evaluation process, researchers need to backtest strategies using either actual market data or a market simulator to show improvement in slippage or profit and loss. There is a trade-off when choosing between the two: Actual market data reflects accurate market dynamics but cannot incorporate price impact; stochastic simulation can do arbitrarily many sample paths but is only partially realistic. Though complete reproduction of the market is impossible, some simulation models calibrate Markov queuing systems to the order book dynamics and provide insight into the relation between order flow and price dynamics in limit order markets (Cont and De Larrard 2013; Huang et al. 2015).

After developing a promising trading signal, the next question that naturally arises is how to design an execution strategy to optimize the result, namely, how to optimally set up and unwind the positions to capture the alphas. Trading execution always has a cost because of the bid-ask spread or the impact of the trade on the market. Market microstructure studies both micro impact (such as the response to individual actions) and macro impact (such as the cost of trading meta-orders).

Many artificial intelligence-based electronic trading platforms, such as Aiden developed by RBC Capital Markets, use the computational power of deep reinforcement learning (RL) to improve trading results. The black box of those real-time trading systems remains mysterious to the public, but academic researchers have also started employing deep learning in high frequency trading. Nevmyvaka et al. (2006) is an early large-scale application of RL to optimize trading execution in the financial markets, which experimentally demonstrates that RL approaches are well-suited for optimized execution. Deep Q-learning networks also have been applied to optimal trade execution (Ning et al. 2018; Théate and Ernst 2021). Imitative learning deep RL techniques, which are taught by an intelligent trading agent, can help in quantitative trading with balance between exploration and exploitation (Liu, Liu et al. 2020). Kolm et al. (2021) forecast high-frequency returns by training off-the-shelf artificial neural networks on order flow at the most granular level.

5619 DTC Parkway, Suite 600
Greenwood Village, CO 80111
Phone: +1 303-770-3377
www.investmentsandwealth.org