

Machine Learning in Risk Forecasting and its Application in Low Volatility Strategies

By Yuriy Bodjov

December 2018

Artificial intelligence and machine learning are two terms that have gained increased popularity within the investment industry in recent years. This area of computer science has successfully been used for tasks such as algorithmic trading, credit rating evaluation, 'robo-advising' or for generating superior alpha forecasts. However, relatively fewer efforts have been made to use it for risk forecasting. This paper will discuss the application of machine learning algorithms in providing independent risk forecasts, the benefits of combining these forecasts with traditional risk models and how to potentially use them in practice to construct better low volatility equity strategies.

Artificial intelligence or machine learning

Although these two terms are often used interchangeably, they do not mean the same thing. Artificial intelligence, in existence since 1956, is a more general term emphasizing the building of machines that think and react like humans. Machine learning is only a subset of artificial intelligence, its

focus being the use of algorithms to find patterns in data and then using models that recognize these patterns, in order to make forecasts on new data. In this article, we will focus on machine learning as it provides the tools we need to make successful risk predictions.

Traditional risk modelling vs. machine learning

Traditional risk models appeared with the Modern Portfolio Theory put forth by Harry Markowitz in 1952.¹ Over the years they went from one factor (Capital Asset Pricing Model or 'CAPM') to multi-factor models (Arbitrage Pricing Theory or 'APT'), adding layers of complexity. The vast majority of the equity risk models used within the financial industry presently are multi-factor models; given that they have proven to be an effective tool for risk management. There are two common approaches to model risk; the first is based on the assumption that the risk factors are known. These factors could be macroeconomic, such as inflation, industrial production and interest rates or they could be based on the company fundamentals like size, valuations, profitability or industry. The second approach used by the statistical risk models, doesn't use explicit risk factors, but instead infer these from panel data of security returns.

Each approach has its own advantages and disadvantages. For example, fundamental risk models are better suited for explaining the sources of risk as they are known in advance, but some important risk factors could be left out. Statistical models usually do a better job in portfolio construction, especially in the low volatility context, but lack explanatory power. Regardless of the approach used for choosing the risk factors, traditional risk models have a lot of common. First of all, they share the same basic quadratic format consisting of factors, a covariance matrix and exposures to these factors; plus the stock's specific risk:

$Risk = XFX^T + \Delta$, where:

X is the matrix of assets exposures to factors
 F is the factors covariance matrix, and
 Δ is a diagonal matrix of asset specific variances

This basic format is used by virtually all risk models, due to its adaptability and ease of use within the quantitative portfolio construction process. Traditional risk models use a fixed number of risk factors, ensuring the model's stability and lessening its sensitivity to outliers.

That being said, the risk models currently in use by most quantitative analysts do have some disadvantages. First, there is no supporting evidence to prove that the covariance matrix is the one and only way to describe and manage the risk. Second, traditional risk models don't capture all sources of risk, because they rely on a fixed number of risk factors. Adding new factors can marginally improve the performance, but there is a limit on the numbers of factors we could add before we run into other problems. In addition, the model's rigidity precludes all new emerging risk factors from being quickly captured and reflected. Finally, most traditional risk models are based on assumptions about the input data, such as it having a normal distribution, a linear relationship with the outputs, non-collinearity etc. Every time we make a transformation of the data to fit the model restrictions, there is reduction in the explanatory power of the outputs.

The machine learning approach to risk forecasting is very different. Instead of relying on pre-defined statistical measures of risk or risk models, it is based on a methodical learning of new concepts and then applying the acquired knowledge in a decision-making scenario. Analysts and managers with an in-depth understanding of the market environment and the fundamental characteristics of the companies they cover can select the stocks, which in their opinion, would represent low, medium or high risk. The criteria may vary from stock to stock and could be based on characteristics such as company size and corporate structure, dividend yields, liquidity, analyst coverage, quality of the management (based on personal assessment), valuations etc. Moreover, the fundamental metrics that analysts use can change over time, as they constantly look for new ideas and learn from past mistakes. Similar to the human brain, machine learning algorithms can be trained over time to discover patterns and data dependences; then from their observations, make forecasts about the expected future volatility of a stock. Hence, machine learning algorithms utilize a foundational tenet of learning, trial and error, to develop its knowledge base and achieve its objective.

A more formal way to describe the machine learning process is to find a target function f that best maps input variables (X) to an output variable (Y):

$$Y = f(X) + e$$

Machine learning algorithms are well adapted to classification problems, and from the point of view of risk forecasting, the output variable (Y) could be simply the level of future volatility e.g. low, medium or high. The input data (X) could come from different sources: fundamental stocks characteristics, industry and country information, macroeconomic variables, market, political, environmental indicators, or any predictive indicator a good financial analyst would use.

In stark contrast with traditional risk models, utmost importance is not placed on the details of the function we are learning; only that it makes the most accurate predictions about future volatility. In a similar fashion to the human brain, capable of acquiring knowledge from disparate sources, importance is not placed on how the knowledge is acquired; but the efficacy of its use. A second major difference from the traditional risk models is that the algorithms used in the machine learning process can change frequently by adding or removing factors “on-the-fly” to instantaneously reflect the new information. This removes the restraints imposed by the rigidity of the fixed form traditional risk models and is related to the possibility of using nonparametric algorithms such as *k-Nearest Neighbors*, *Decision Trees* or *Support Vector Machines*. Although these algorithms require more data, they remove obstacles related to non-linearity, sparse or missing data. Moreover, we are able to combine the forecasts from two or more algorithms to improve the forecasts.

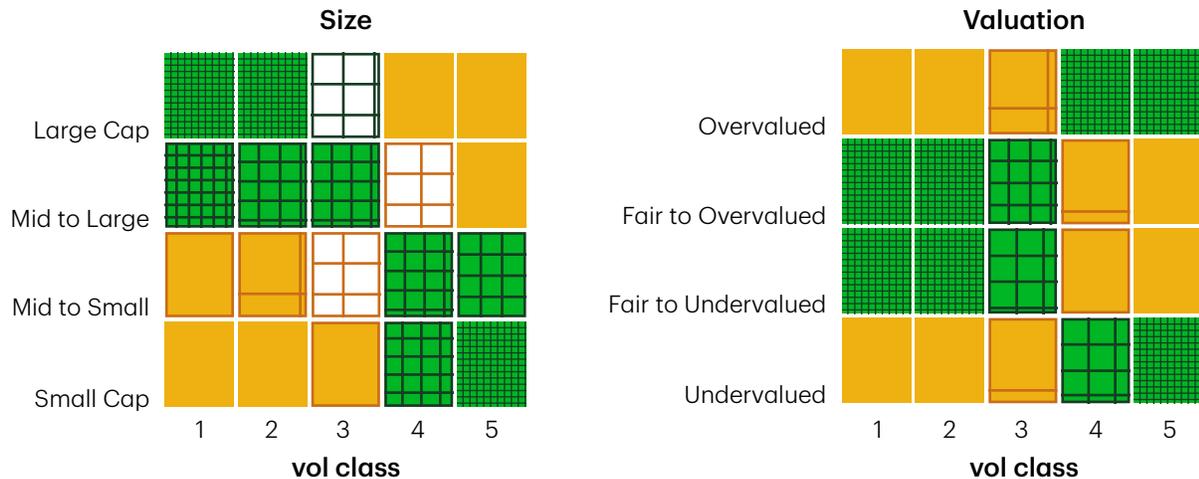
Equity risk forecasts with machine learning

This is essentially a classification problem where the goal is to group the stocks by expected future volatility based on all available information. We could use five groups (quintiles) or even three: low medium or high risk. In the machine learning vocabulary, these groups are called classes. In order to successfully separate the stocks into classes we need some information about the stocks, called attributes. The attributes could be:

- **Fundamental: financial ratios, style exposures (value, growth, size) etc.**
- **Market: price momentum, short interest**
- **Qualitative: corporate management, earnings quality**
- **Macro-economic: sensitivity to interest rates, economic cycles, changes in FX**
- **Industry or sector of activity**
- **Country of registration or activity**
- **ESG factors, etc.**

Before exploring the algorithms we need to examine the data. Many of the attributes are known risk factors: valuations, size, momentum or short interest. Some have linear relationship with the expected volatility, while others are clearly non-linear. For example, companies with larger market capitalizations are usually associated with lower

volatility and small-cap stocks are considered more risky. Hence, the size attribute would be easily used in a risk classification problem. Other attributes such as valuations, even if considered as a source of risk, cannot be used directly to map the stocks into volatility buckets.



Source: TDAM

* Sieve Diagrams:² The area of each rectangle is proportional to the expected frequency, while the observed frequency is shown by the number of squares in each rectangle. The difference between observed and expected frequency (proportional to the standard Pearson residual) appears as the density of shading, using color to indicate whether the deviation from independence is positive (green) or negative (orange).

Both stocks with high and low valuations are associated with a higher than average future volatility. To use it successfully we need additional information, such as the quality of the company's earnings, the stock's momentum, the amount of short interest, the company's leverage, where we are in the market cycle etc. Traditional risk models try to process this information simultaneously. However, one might argue that it would be better if some of the decisions are made sequentially. For example, a skillful analyst may first split the stocks into groups based on the quality of their earnings, before making a call based on valuations. Some machine learning algorithms work in a similar fashion, making decisions based on the knowledge acquired from the results of previous analysis.

Before proceeding with the algorithm selection, the data used would need to be separated into training and testing sets to avoid overfitting. The algorithms are first trained with the training data and then the accuracy of the forecasts is measured against the previously unseen testing data. A machine learning algorithm is then judiciously assessed and selected from a set suitable for classification problems. Examples of such algorithms are: *Logistic Regression*, *Linear Discriminant Analysis (LDA)*, *k-Nearest Neighbors (KNN)*, *Naïve Bayes*, *Classification and Regression Trees (CART)* and *Support Vector Machines (SVM)*. There are various techniques to improve the performance and reduce the overfitting of the volatility forecasts, the most promising being *Random Forest*³ and *Extra Trees*⁴ as well as boosting algorithms such as *AdaBoost*.⁵

An Example

Let's illustrate the process with the following example. We want to classify the stocks from S&P500 index in 3 groups in terms of expected future volatility: low (lowest 40%), medium (mid 20%) and high (highest 40%). For each stock in the universe we have a

very short list of attributes such as the sector of activity, value, profitability, growth, size, corporate management, short interest and earnings sentiment, all going back to 2000. We have also the forward volatility for each stock over rolling one year.

Example of training data as of Dec. 31, 2007

Ticker	Attributes								Class
	Value	Profit	Growth	Size	Corporate Management	Short Interest	Earnings Sentiment	Sector	
CPB	-0.149	0.179	-0.014	-0.104	1.288	0.267	0.056	Cons. Staples	low
CAH	0.256	0.102	-0.010	0.175	0.547	0.412	-0.496	Health Care	low
CCL	-0.078	-0.055	0.024	0.143	-0.293	0.001	-0.941	Cons. Discretion.	high
CAT	0.345	0.104	0.035	0.719	0.148	0.051	-0.578	Industrials	mid
JPM	-0.127	-0.133	-0.011	2.500	-0.008	0.519	0.129	Financials	high
CB	0.784	0.021	0.032	0.176	-0.235	0.644	1.309	Financials	mid

Source: TDAM and Bloomberg, September 2018

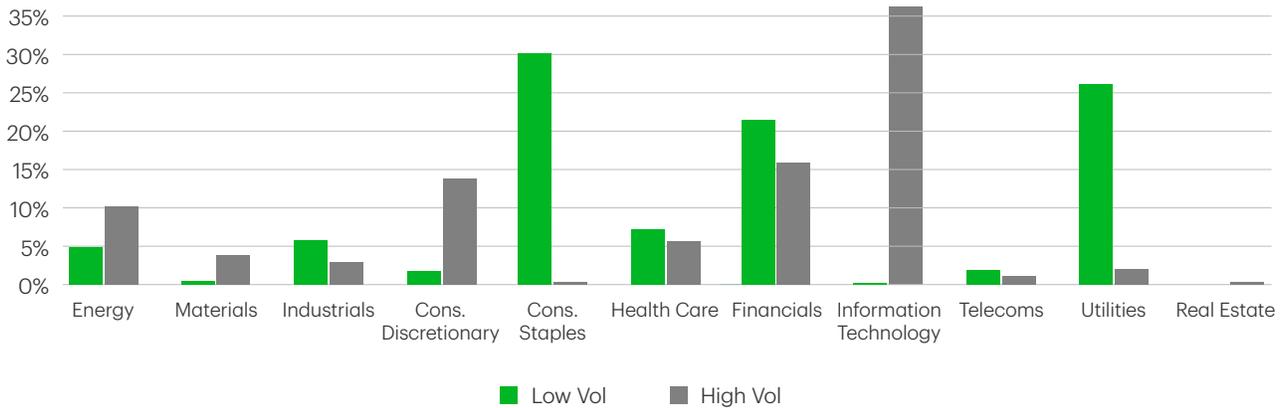
The 'tickers' represented above have been selected from the S&P500 Index to illustrate the training data process and were selected from the index randomly. This is not a recommendation to purchase or sell a specific stock or to adopt an investment strategy based on the selection shown. If the stocks were profitable in the past there is no guarantee they would be in the future.

The first training period is 3 years and the algorithm is evaluated over the subsequent year with the test data. Then the training period is sequentially increased and the algorithm is tested over the next year.

Because of the nature of the training data, we would expect that some of the intuition from the inputs will be reflected in the volatility forecasts. If our task is to choose the least volatile stocks, we would probably

choose stocks from non-cyclical sectors such as Financials, Utilities and Consumer Staples. If the task is to identify riskier stocks, the choice would probably go to Technology, Energy and other so-called cyclical sectors. The results from running the data through the machine learning algorithm show that our intuition is followed quite closely: the low volatility stocks come mostly from the least cyclical sectors and the opposite is true for the stocks forecasted to have high volatility.

Average Sector Distribution of Volatility Forecasts S&P500 (Feb 2004 – Feb 2017)

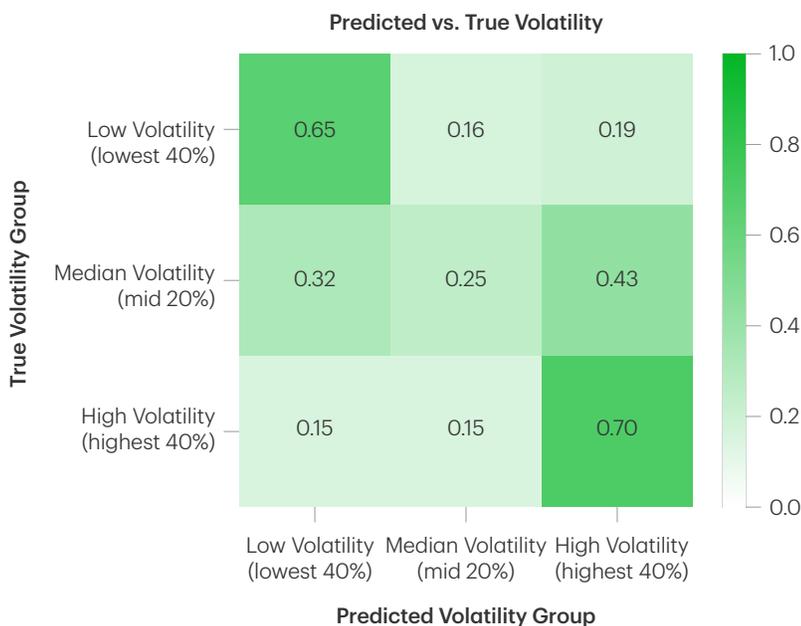


Source: Standard & Poor's and TDAM

Further analysis shows that the computer assigns to the low volatility group stocks which are not over-valued, but growing slowly. Stocks predicted to be riskier are on average of smaller sizes, trade at higher multiples and are managed in less shareholder friendly manners. All these arguments make a lot of intuitive sense.

The machine learning algorithms could easily be evaluated based on their classification accuracy. The algorithm measures the percentage of correct forecasts over the total number of forecasts. As we increase the training period, the accuracy of the forecasts increase, as new information becomes available.

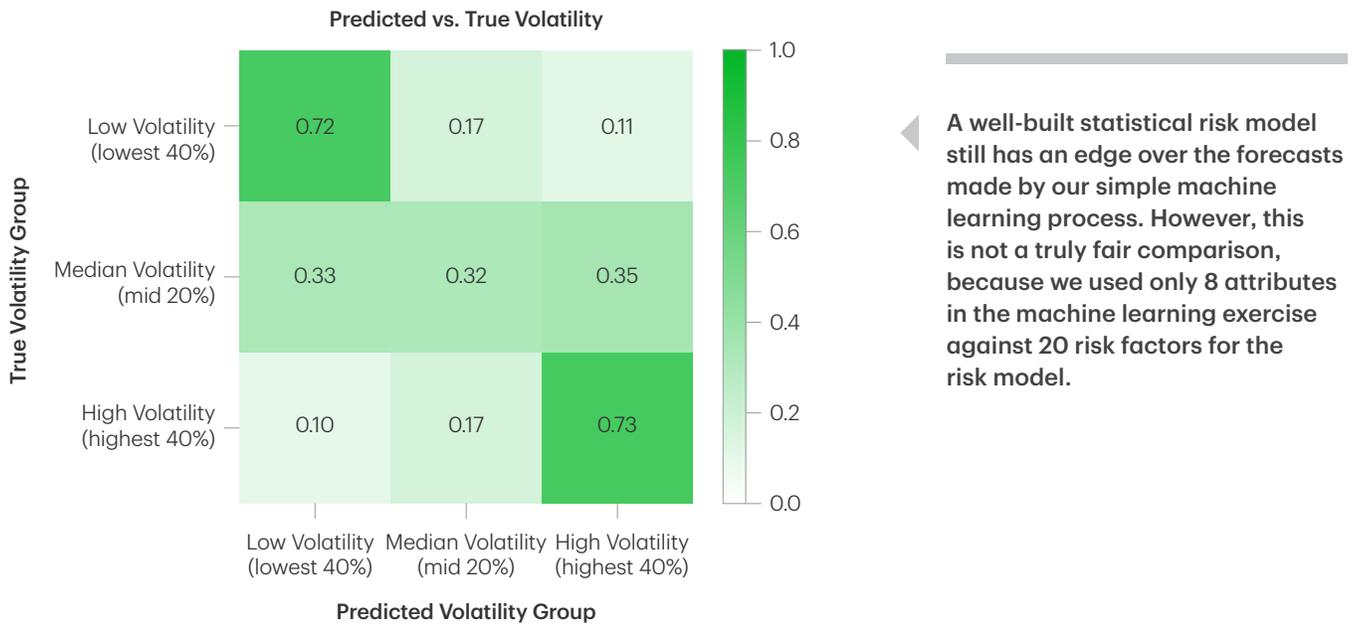
Confusion Matrix of Volatility Forecasts made by Machine Learning (Feb 2004 – Feb 2017)



Another insightful way to analyze the accuracy of the forecasts is by using the so-called confusion matrix. It shows the risk forecasts on the x-axis and the accuracy outcomes on the y-axis. The results from our simple exercise based on out-of-sample data between February 2004 and February 2017 are shown on the left.

The classification accuracy for the low volatility stocks is 65% compared to 70% for the high volatility stocks. This is a good score, but how does it compare to a traditional risk model over the same period?

Source: TDAM

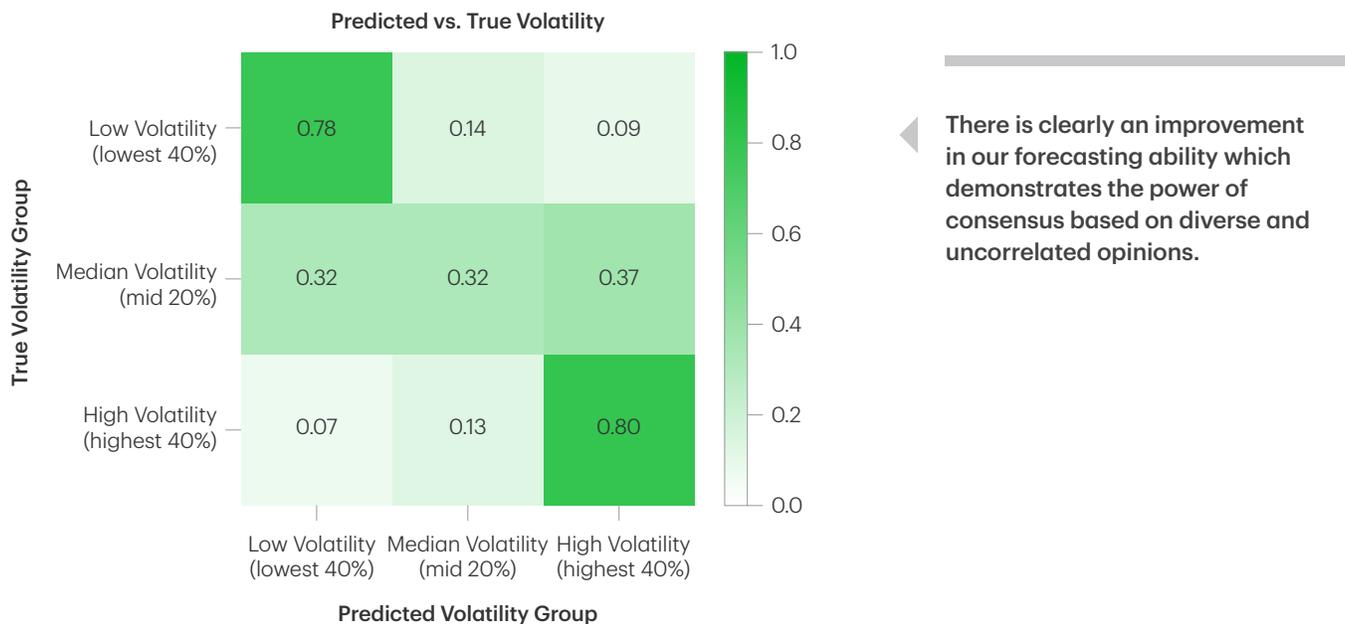


Source: TDAM

Combining volatility forecasts

The benefits derived from utilizing objective and independent sources of information, while making investment decisions, is inherently understood by investors. Having a (good) second opinion is always welcomed, even only to confirm an existing signal. We saw that traditional risk models and machine learning risk forecasts

have individually, a relatively good accuracy at predicting future volatility. If we compare the forecasts we see that these two approaches agree only about half of the time. It is interesting to see what happens with the classification accuracy for the stocks for which there is consensus:



Source: TDAM

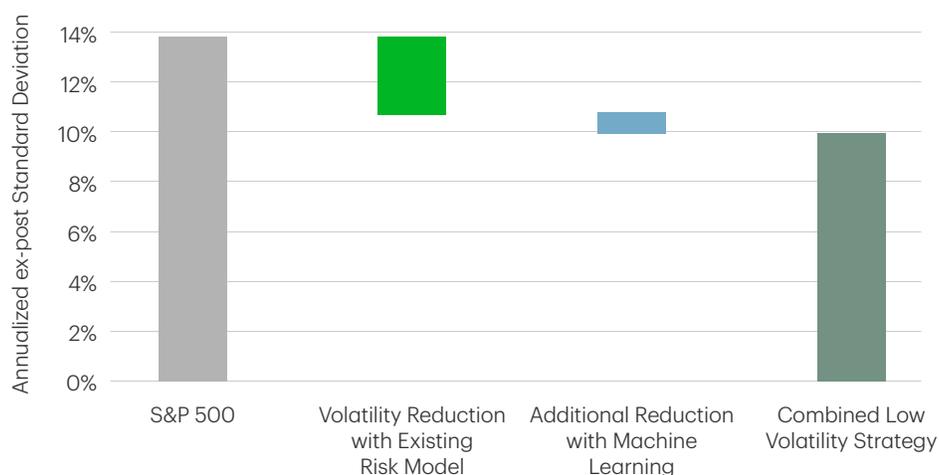
Using machine learning to seek to build better low volatility portfolios

There are various ways to combine the machine learning forecasts with the risk models within the investment process of the low volatility strategies.

One of the easiest approaches would be to filter out the stocks with the highest forecasted volatility

by the machine learning algorithm and then use an optimizer with the traditional risk model to create low volatility portfolios. The graph on the following page shows the results from using this method with the stocks from the S&P 500 universe.

Combining Volatility Forecasts by Filtering (Feb 2004 – Feb 2017)



Source: Standard & Poor's and TDAM

Combining both forecasts reduces the ex-post volatility beyond the results achieved by even the best performing risk models. A downside of this method is that it limits the universe of investable stocks, which could be problematic for an already small equity universe, like Canada. Another approach that we are currently exploring is to use the machine learning algorithms to forecast the stock specific volatility as part of the risk models. Currently, most of the risk models compute the stock specific risk as the difference

between the total return variance and the risk coming from the common risk factors. By definition, this idiosyncratic (firm-specific) risk is independent and uncorrelated with the factors from the risk model, but many outside factors could affect the stock's behavior. Machine learning algorithms are very well positioned to uncover and exploit these factors. The specific risk forecasts then could be used with the existing framework to create low volatility portfolios through a quadratic optimization.

In this paper, we have demonstrated that the use of new and innovative techniques could markedly improve the success rate of selecting low volatility stocks. We showed that combining the existing traditional risk models with the machine learning forecasts can lead to significantly higher predictive power. Portfolio solutions based on reducing volatility, such as the low volatility funds, can greatly benefit from the use of machine learning as an important additional input in the portfolio construction process.



¹Press release. NobelPrize.org. Nobel Media AB 2018. Thu, 22 Nov 2018. <https://www.nobelprize.org/prizes/economic-sciences/1990/press-release/>
²A sieve diagram shows the frequencies in a two-way contingency table in relation to expected frequencies under independence, and highlights the pattern of association between the row and column variables. The sieve diagram was proposed by Riedwyl and Schüpbach in a technical report in 1983 and later called a parquet diagram. ³Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. ⁴An extra-trees regressor. This class implements a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. ⁵AdaBoost, short for Adaptive Boosting, is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire, who won the 2003 Gödel Prize for their work. It can be used in conjunction with many other types of learning algorithms to improve performance.

The information contained herein has been provided by TD Asset Management, and is for information purposes only. The information has been drawn from sources believed to be reliable. Graphs and charts are used for illustrative purposes only and do not reflect future values or future performance of any investment. The information does not provide financial, legal, tax or investment advice. Particular investment, tax, or trading strategies should be evaluated relative to each individual's objectives and risk tolerance. TD Asset Management operates through TD Asset Management Inc. in Canada and through TDAM USA Inc. in the United States. TD Asset Management Inc., TDAM USA Inc., Greystone Managed Investments Inc. and Epoch Investment Partners, Inc. are wholly-owned subsidiaries of The Toronto-Dominion Bank and comprise the asset management arm of the TD Bank Group. The TD Bank Group means The Toronto-Dominion Bank and its affiliates, who provide deposit, investment, loan, securities, trust, insurance and other products or services. The TD logo and other trademarks are the property of The Toronto-Dominion Bank.