October 30, 2018

MAN VERSUS MACHINE

MALTA (Machine Adaptive Learning Tactical Alpha) Global Stock Selection Model

- **Re-Adapting Machine Learning to Finance.** We begin with a comprehensive review of how Machine Learning (ML) techniques can be applied to stock selection. In particular, we demonstrate how to transform a traditional linear model (where we attempt to forecast a point estimate of future stock return) to a classification model of outperformers/underperformers. We illustrate the pros and cons of existing classification algorithms – covering classification tree, random forest, and boosting techniques (i.e., AdaBoost and Gradient Boosting Machine). Lastly, we introduce our next generation boosting algorithm MBBT (Multi-Branch Boosted Tree), a more effective and efficient technique compared to existing ML techniques especially in financial applications.
- **MBBT Machine Learning Algorithm.** Our MBBT algorithm is uniquely designed to capture the U- or inverse U-shaped payoff profile frequently observed in financial data. The MBBT is exceptionally robust to overfitting and powerful in dealing with high-dimensional problems consisting of hundreds or thousands of potential input factors, with full transparency. Furthermore, we develop a suite of approaches to handle the time series aspect of financial data that is not well covered in current ML literature.
- **Introducing MALTA – Another Global Stock Selection Model.** Our MALTA (Machine Adaptive Learning Tactical Alpha) model is our second flagship global stock selection model. Compared to our LEAP (L-Economic Alpha Processing) model, MALTA is completely driven by ML, emphasizes nonlinear patterns, and is more robust than other ML techniques. It is also highly suitable for fundamental investors with concentrated positions. The MALTA has a Sharpe ratio above 3.0x in most regions, including the US, Europe, Asia, and Japan. Given the low correlation between LEAP and MALTA, a combined model can boost performance even further.
- **Aligning Investment Horizon with Model Development.** For investors with large AUM, low turnover, long investment horizons, and/or facing high transaction costs, we construct our MALTA-HC (High Capacity) model. We also have a high frequency version – MALTA-StatArb. The MALTA-StatArb model is designed for either daily or intraday rebalance, with an approximate holding horizon of a week. The MALTA-StatArb delivers superior and uncorrelated alpha to traditional multifactor models.

Source: pixabay.com

Sheng Wang
SWang@wolferesearch.com

Yin Luo, CFA, CPA
YLuo@wolferesearch.com

Miguel Alvarez
MAlvarez@wolferesearch.com

Javed Jussa
JJussa@wolferesearch.com

Gaurav Rohal, CFA
GRohal@wolferesearch.com

Jason Zhong, PhD
JZhong@wolferesearch.com

QES Desk Phone: 1.646.582.9230
Luo.QES@wolferesearch.com

Table of Contents

A LETTER TO OUR READERS

Introducing a new Global Stock Selection Model – MALTA

Since our first pure Machine Learning (ML) based stock selection model[1] published in 2005, in the past decade, we have been gradually improving our suite of algorithms in our research. Before ML/AI became the buzz words in finance, even just a few years ago, ML was still being perceived as data mining and overfitting by mainstream practitioners and finance academics. Even until today, many managers are still not convinced whether ML can truly generate out-of-sample alpha above and beyond traditional fundamental and quantitative processes.

The vast majority of ML algorithms were originally developed in a cross-sectional context. For example, in medical research, insurance fraud detection, or a marketing campaign, the goal is primarily to identify patterns among a large number of subjects at a given point-in-time. The time dimension is not taken into account. Even today, there is limited research in time series ML. For investment management, however, time is as important as the cross-sectional aspect. For stock selection models, we at least have a large number of cross-sectional observations, i.e., number of stocks. For global macro modeling, time is often the dominant or only dimension.

In this research, we first offer a comprehensive review of how ML techniques can be applied in stock selection. In particular, we demonstrate how to transform a traditional linear model (where we attempt to get a point estimate of stock return) to a classification setting. Instead of forecasting stock returns, we attempt to identify future outperformers and underperformers.

Getting a point estimate of future stock return is extremely challenging, if not impossible. Models may overfit the data, which will deter out-of-sample prediction. On the other hand, classification problems tend to be more robust. In addition, some investors are concerned more about hit rate (e.g., percentage of stocks with above or below average returns) than average return. In particular, if a manager's portfolio is highly concentrated, she can't take advantage of a model with a great average return dominated by a few outliers. Fundamental managers naturally prefer models with higher hit rates; therefore, a classification setting may fit their needs better. More importantly, a classification-based model is naturally different from a regression framework; therefore, providing diversification benefit.

We illustrate the pros and cons of existing classification algorithms covering classification trees, random forests and boosting (i.e., AdaBoost and Gradient Boosting Machine). Next, we introduce our next generation boosting algorithm – MBBT (Multi-Branch Boosted Tree) – more effective and efficient than existing ML techniques, especially for financial applications

Our MBBT algorithm is uniquely designed to capture and exploit the often observed U- or inverse U-shaped payoff profile exhibited by considerable data in finance. The MBBT is exceptionally robust to overfitting and powerful in high-dimension problems with hundreds or thousands of potential input factors, with full transparency. Furthermore, we develop a suite of approaches to handle the time series aspect of finance data that is not well covered in the current ML literature.

The MALTA model has delivered exceptional performance in all nine regions of the world, especially in the four largest markets – the US, Europe, Asia, and Japan, with a Sharpe ratio over 3.0x. Even in

[1] The first version of our Machine Learning (ML) stock selection model was named QED (Quantitative Equity Dynamic) model.

recent years, when traditional multifactor models have struggled in some of the most difficult markets such as the US and Japan, the MALTA model is able to retain its efficacy.

Our MALTA model is different from and more importantly, complementary to our first flagship global stock selection model LEAP (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017]). MALTA is completely driven by ML, emphasizes nonlinear patterns, is more robust to changes in the market environment, and suits fundamental investors with concentrated positions. As a result, the combined LEAP+MALTA model boosts performance even further in all nine regions of the world.

In this research, we also discuss two of the three most important alignment issues in portfolio construction[2]. The first one is horizon and decay mis-alignment. Alpha model could have a much longer (or shorter) investment horizon than (therefore, different signal decay from) the risk model, which generates insufficient trades (or unnecessary turnover). For investors with large AUM, low turnover, long investment horizon, and/or facing high transaction costs, we construct our MALTA-HC (High Capacity) model. Similarly, we also have a high frequency version – MALTA-StatArb. The MALTA-StatArb model is designed for either daily or intraday rebalance, with a holding horizon around a week. The MALTA-StatArb delivers superior uncorrelated alpha to traditional multifactor models. In our statistical arbitrage model, we also address the third mis-alignment issue – the alignment between model development and trade execution.

How Clients can use our Research

For fundamental investors, our MALTA, LEAP, and other specialized models such as SMEC (NLP on management presentations) and SPEC (NLP on corporate regulatory filings) offer a complete solution using alternative data and machine learning. In our experience, our systematic stock selection models are materially different from traditional fundamental analysis; therefore, should provide for broader diversification benefit.

For quantitative investors, the complete results of our backtesting are available upon request, which should assist managers in their own research. More importantly, we offer daily data feeds that investors can plug into their own investment process directly. Given the heavy usage of alternative data and machine learning techniques, we expect our models to be sufficiently different from most standard approaches on the market.

Regards,

Yin, Sheng, and Luo's QES team

[2] The third misalignment issue is factor misalignment. Alpha models are often developed in-house, while risk models are mostly from commercial risk vendors. Therefore, the factors used in the alpha and risk models may have some overlap, but most likely are different. The optimizer may emphasize the difference between the alpha and risk models and therefore, produce a sub-optimal portfolio. Factor alignment is covered in our previous research. Please contact us for details.

A Machine Learning Evolution

Artificial Intelligence (AI) and Machine Learning (ML) have gained tremendous popularity in many fields of science, technology, and increasingly in our daily lives, including finance and investment. We have always been a true believer and a proponent of applying machine learning techniques in investment research.

Machine learning is also known as data mining, pattern recognition, predictive modeling, etc. This is a rapid evolving field, where new algorithms are constantly being developed. Most people refer to machine learning as the various algorithms such as random forest, classification trees, and neural networks. However, machine learning is much more than the tools and techniques for uncovering patterns in data. Machine learning is a new philosophy and the entire process of developing models in a way that we can understand and quantify the model's prediction accuracy on future live data.

Our Past Work on Machine Learning

Since we launched our research at Wolfe Research in early 2017, we have been actively applying machine learning in a number of our models:

- **LEAP** or L-Economic Alpha Processing (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017c]) is our first flagship stock selection model. We introduced a suite of feature selection tools (e.g., random forest and modified Fama-MacBeth), nonlinear and linear components, along with a macro overlay in the LEAP model.
- **BALI** (Bank Alpha LEAP Insight, see *Multi-Dimensional Alpha: Banking on the Banks – Welcome to BALI*, Luo, et al [2018]) and **LEAP-China** (see *Multi-Dimensional Alpha: The Silk Road to China*, Wang, et al [2017c]) extended the LEAP methodology to banks and China domestic equity markets, respectively.
- In Quant **CSI** (see *Multi-Dimensional Alpha: Quant CSI*, Jussa, et al [2017a]), we applied our MBBT (Multi-Branch Boosted Tree) algorithm in accounting fraud detection.
- **SPEC** (Systematic Profiling of EDGAR Composite, see *Multi-Dimensional Alpha: Text Mining Unstructured Corporate Filing Data*, Rohal, et al [2017]) and **SMEC** (Systematic Mining of Earnings Calls, see *Multi-Dimensional Alpha: Tone at the Top? Quantifying Management Presentation*, Rohal, et al [2018a]) models use NLP (Natural Language Processing) and other machine learning algorithms to analyze textual documents to derive investment insights.
- **SMAP** (Systematic M&A Prediction) model (see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a]) combines a suite of different machine learning methods to predict the probability that a company becomes a takeover target in the near term.
- **MRM** or Margin Revision Model (see *Multi-Dimensional Alpha: Unlocking Hidden Value*, Jussa, et al [2017b]) uses our proprietary Boosted Logit algorithm to identify the likelihood of turnaround (proxied by margin improvement).
- In the **SIRC** (Sensitivity to Interest Rate Changes) model (see *Portfolio Compass: Stocks Most Exposed to Rising Interest Rates*, Jussa, et al [2018b]), we attempt to systematically measure a company's interest rate sensitivity, using the elastic net regularization technique.

- **Capri** (Compensation, Ability, and Performance Ranking Indicator) model is our answer to ESG (see *Multi-Dimensional Alpha: Pay for Performance*, Rohal, et al [2018b]). Given the highly nonlinear and complex interactions often observed among corporate governance factors, we implemented the xgBoost algorithm in the Capri model.
- The vast majority of risk arbitrage and event-driven strategies are managed by discretionary portfolio managers. In our **SARA** (Systematic Alpha form Risk Arbitrage) model (see *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)*, Wang et al [2018a] and *Global Systematic Risk Arbitrage – SARA Global*, Wang et al [2018b]), we studied a wide range of machine learning methods to predict the deal success probability, remaining deal duration, and deal premium for both US and global M&A transactions.

Machine Learning Algorithms Toolbox

As we have demonstrated in previous research, sophisticated machine learning algorithms can often improve prediction accuracy substantially. However, there are hundreds of machine learning techniques available and new algorithms are still being introduced constantly (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017c] for a general overview). More importantly, there is no theoretical guidance as to what algorithms are best suited for a given scenario. Therefore, algorithm selection is part-art and part-science. In practice, it is often determined by empirical performance and researchers' personal preference. Below is a partial list of the algorithms that we have tested in our published research.

- **LASSO, Ridge, and Elastic Net.** OLS (Ordinary Least Squares) regression, despite its wide usage in both academic and practitioner research, often suffers from estimation errors. Robust regression or regularization techniques typically shrink the estimated coefficients, which often improves out-of-sample performance. We applied the LASSO technique in our research on short interest (see *Multi-Dimensional Alpha: New Insights in Short Interest – A Global Perspective*, Rohal, et al [2017b]). Elastic net combines the L1 (LASSO) and L2 (Ridge) penalties to handle the overfitting issue from linear regression. LASSO, ridge, and elastic net are just more elegant linear regressions; therefore, are easy to implement and interpret. in *Portfolio Compass: Stocks Most Exposed to Rising Interest Rates* (see Jussa, et al [2018b]), we applied the elastic net technique to measure stock sensitivity to interest rate. In *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)* (see Wang et al [2018a]), we predict deal premium by combining the MBBT's feature selection and the elastic net's regularized regression to overcome the small sample, overfitting, and estimation error problems.
- The **CART** (Classification and Regression Tree) algorithm was originally developed by Breiman, et al [1984] and has since seen widespread applications in multiple fields. The CART model is transparent, typically fast to train and implement. However, performance by itself is not particularly strong, especially in complex context (see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a]). We typically use the CART to identify some high level nonlinear patterns and then combine it with other models (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP* Luo, et al [2017c]).
- **RF** (Random Forest). RF is an improved version of CART, by assembling hundreds or thousands of CART models. The RF model often offers significant improvements over CART (see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a]). However, similar to

SVM, it is difficult to interpret and suffers from the transparency issue. In our LEAP model, we applied the RF algorithm as a nonlinear feature selection tool (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017c]).

- The performance of **ANN** (Artificial Neural Networks) is typically stronger linear regression (see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a]). However, overfitting and interpretation remain as the top two challenges of applying the ANN model in practice.
- **Deep Learning** (typically perceived as the next generation of ANN) has gained great popularity in recent years. In finance, however, as we demonstrated in *Multi-Dimensional Alpha: Tone at the Top? Quantifying Management Presentation* (see Rohal, et al [2018a]), it does not necessarily provide much improvement from a traditional ANN. In finance, sample size is typically smaller, while signal-to-noise ratio is often much higher than in other fields, such as playing the GO game, which may contribute to the suboptimal performance of the algorithm.
- The **SVM** (Support Vector Machine) algorithm has been widely used as an effective classification technique for a long time. In our experience, see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a], SVM does offer decent performance, but model transparency is often cited as a major weakness. In addition, the SVM does not scale very well – as the size of training data, especially, as the number of input factors (or features) increases, the computational needs spikes up quickly.
- In our proprietary **Boosted Logit** model, we combined the ability of our MBBT algorithm as a feature selection tool and the easy-to-implement nature of the logit model. As explained in *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)* (see Wang, et al [2018a]), the Boosted Logit model preserves the strength of the MBBT technique and the simplicity of the logistic regression in predicting deal premium.
- The **xgBoost** algorithm is a powerful implementation of the gradient boosting framework. In *Multi-Dimensional Alpha: Pay for Performance* (see Rohal et al [2018a]), we applied the xgBoost feature in our corporate governance model called Capri.
- **MBBT** (Multi-Branch Boosted Tree). Our proprietary MBBT algorithm is the next generation of the AdaBoost and GBM (Gradient Boosting Machine) algorithms. The AdaBoost model was first introduced by Schapire [1998], which has shown robust performance in facial detection (see Wu, et al [2004]) and human pose estimation (see Wang, et al [2010]). We extend the AdaBoost/GBM algorithms by allowing a more flexible and powerful tree splitting. While AdaBoost/GBM is almost always based on binary-split trees, our MBBT can trigger a decision tree in any number of branches. Although in theory, binary trees, with multiple layers, can proxy almost any complex features. In practice, it often struggles to fit the U- and inverse U-shaped data in finance. In our experience, even a three-branch tree can beat the traditional binary trees considerably, especially in finance. More importantly, we also incorporate the time series properties often observed in financial data. While the vast majority of machine learning algorithms were developed and designed to process cross-section data, our MBBT algorithm can handle the time series dimension more effectively. As shown in *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)* (see Wang, et al [2018a]), our MBBT model can outperform other machine learning algorithms at a much faster computational speed.

In this research, we are introducing our second global stock selection model called MALTA (Machine Adaptive Learning Tactical Alpha), using our proprietary MBBT algorithm to achieve the following goals:

- **Uncorrelated Performance.** A strong out-of-sample performance is always the first priority for any stock-selection model. Furthermore, we hope our MALTA model is sufficiently different from our existing LEAP.
- **Transparency.** Although asset owners have become increasingly comfortable with more complex models in recent years, we strive to develop transparent models. As we will explain in the remainder of this paper, the MALTA model is fairly transparent and the patterns identified are mostly intuitive.
- **Scalability**. An effective model needs to be scalable, i.e., it is easy to train and update, especially as we expend the model globally with over 14,000 securities and to higher frequency (e.g., daily).
- **Adaptiveness**. The market environment changes constantly, due to both micro (e.g., investors constantly improve their investment process and market becomes more competitive over time) and macro (e.g., unprecedented monetary, fiscal, and international trade policy interventions) factors. A real AI-based model should be adaptive to potential paradigm shifts on its own without human override; while at the same time, it should also strive to avoid too many false positives, which triggers unnecessary turnover.

FROM REGRESSION TO CLASSIFICATION

Traditional asset pricing research (e.g., Fama French [1993, 1996], Carhart [1997]) and quantitative equity investors use primarily linear multifactor models (see *Multi-Dimensional Alpha: Signal Research and Multifactor Models*, Luo, et al [2017b] for a comprehensive review). Factors are chosen often manually by analysts, based on economic intuition and backtesting. Factor weights are determined by either optimization (e.g., Grinold and Kahn [1999]) or linear regression (e.g., Luo, et al [2017b]).

Regression Models

In a regression setting, the output is a continuous variable, e.g., stock return or volatility. Regression models can be linear or nonlinear. Ordinary least squares (OLS), partial least squares (PLS), penalized models (e.g., ridge regression, the LASSO, and the elastic net) all belong to the linear regression family. OLS regressions tend to have low bias, whereas penalized models have low variance. There is no theory to suggest the best model for a given scenario; therefore, it is mostly an empirical question and the optimal choice depends on the data. Linear models can be written in the following form; they are generally interpretable, as the coefficients have intuitive meanings; and it is also easy to make statistical inference with linear models.

$$r_{i,t} = \beta_{0,t} + \sum_{k=1}^{K} \beta_{k,t} f_{i,k,t-1} + \varepsilon_{i,t}$$

Where, $r_{i,t}$ is the return of stock i at time t; $\beta_{k,t}$ is the estimated coefficient (commonly known as factor return) for factor k at time t; K is the number of factors; $f_{i,k,t-1}$ is the score (also known as factor exposure) of factor k for of stock i at time $t-1$; and $\varepsilon_{i,t}$ is the regression residual.

The above regression is often performed at each given point-in-time and repeated over multiple periods, i.e., the so-called Fama-MacBeth [1973] procedure.

Linear models can be augmented by including higher order and interaction terms to account for nonlinear relationship, but it tends to be arbitrary and remains restrictive.

In addition to linear regression, there is also a wide selection of nonlinear regression models, e.g., artificial neural networks (ANNs), multivariate adaptive regression splines (MARS), support vector machines (SVMs), and K-nearest neighbors (KNNs).

Classification Models

Some problems in finance are naturally fit for classification models. For example, our SMAP (Systematic M&A Prediction) model (see *Multi-Dimensional Alpha: Machine Learning Takeovers*, Wang, et al [2017a]) combines a suite of different machine learning methods to predict the probability that a company becomes a takeover target in the near term. In these applications, the output variable is categorical, e.g., yes and no.

In this research, the natural output – future stock return that we try to predict – is still continuous. However, there are good reasons to translate the problem from continuous to discrete. Instead of forecasting stock returns, we convert the problem to predicting outperformers versus underperformers. For example, we can classify the top 30% stocks as outperformers and similarly the bottom 30% as underperformers.

There are a number of reasons why such a transformation may be more desirable. The first motivation is that getting a point estimate of future stock return is extremely challenging, if not impossible. Models may overfit the data and may not be able to generalize to predict returns out-of-sample. On the other hand, classification problems tend to be more robust. In addition, some investors concern about hit rate (e.g., percentage of stocks with above or below average returns) more than the average. In particular, if a manager's portfolio is highly concentrated, she can't take advantage of a model with a great average return, but dominated by a few outliers. Fundamental managers naturally prefer models with higher hit rates; therefore a classification setting may fit their needs better. More importantly, a classification-based model is naturally different from a regression framework; therefore, provides diversification benefit.

Let's use a concrete example to show the difference between regression- and classification-based models. In *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP* (see Luo, et al [2017c]), we introduced a simple eight-factor benchmark model (BM), as a proxy for conventional multifactor quantitative models. The eight common stock selection factors[3] are chosen for their simplicity and popularity, from each style category:

- Value: Trailing Earnings Yield – we prefer companies with high earnings yield
- Value: Book-to-Market – we buy companies with high book-to-market, i.e. cheap stocks on valuation
- Growth: Consensus FY1/FY0 EPS Growth – we prefer companies with high earning growth
- Price Momentum: 12M Total Return excluding the most recent month – we prefer companies with positive price momentum
- Analyst Sentiment: 3M EPS Revision –we buy companies with positive earnings revisions

[3] These are the same eight factors used in our Benchmark models, see *Multi-Dimensional Alpha: Signal Research and Multifactor Models* (Luo, et al [2017b]).

- Quality – Profitability: Return on Equity (ROE)- we like firms with high ROEs
- Quality – Leverage: Debt/Equity Ratio – we prefer companies with low financial leverage
- Quality – Earning Quality: Sloan's Accruals – we buy companies with low accruals.

As a comparison, we develop a classification-based model using our MBBT algorithm on the same set of eight factors. We will elaborate the MBBT algorithm shortly, but for the time being, you can think of it as an advanced classification algorithm.

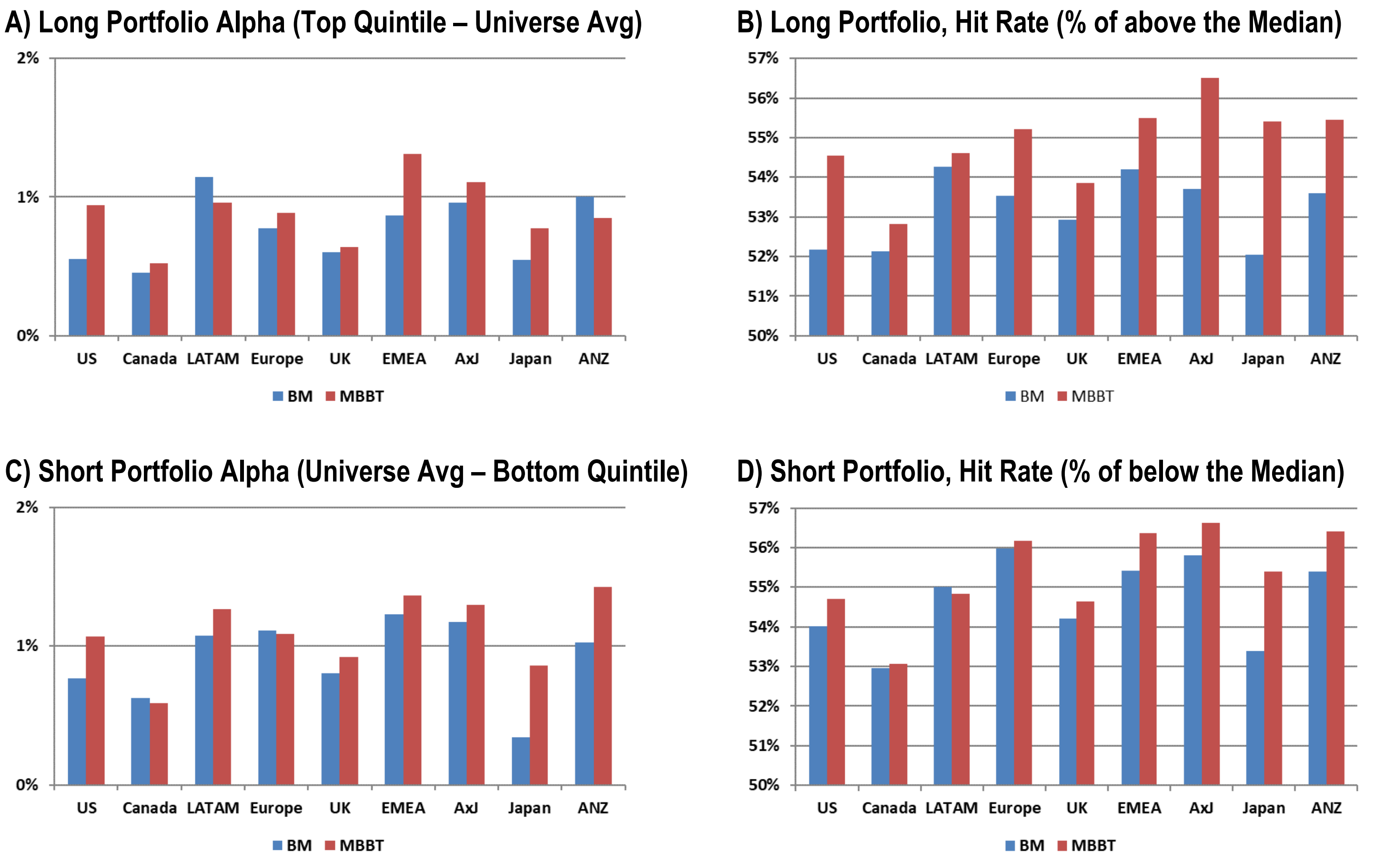
As shown in Figure 1(A), on the long-side (i.e., top 20% of stocks or top quintile), both BM and MBBT models deliver higher significantly alpha[4]. MBBT beats BM in eight of the nine regions. Similarly, the MBBT model also achieves a higher hit rate, i.e., percentage of stocks in the long portfolio with higher returns than the median of the universe, than the BM (see Figure 1), in all nine regions.

Both models also deliver strong alpha on the short side[5] (see Figure 1(C). The MBBT model outperforms the BM in eight regions. Again, the MBBT model earns a higher hit rate (i.e., percentage of stocks in the short portfolio with lower returns than the median of the universe) than the BM., especially in the US and Japan (see Figure 1D).

[4] The long-side alpha is computed as the excess return of holding the top quintile stocks (equally weighted) above and beyond the average return of all stocks in our investment universe.

[5] Similar to how we compute the alpha on the long-side, the short-side alpha is calculated as the difference between the average return of the investment universe and the bottom decile stocks (equally weighted).

Figure 1 Average Return versus Hit Rate



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Let's Grow a Real Tree

Tree-based models follow the same process as how we make decisions; therefore, they are fairly intuitive. Tree-based models also tend to be simple, computationally fast, and reasonably accurate. They conduct automatic feature selection (factor selection), are robust to outliers, and can handle highly nonlinear relationships. Single tree models suffer from model stability issues (i.e., a small change in data can drastically change the structure of the tree) and weak performance. Ensemble[6] method can significantly improve the performance of single trees.

CART and other Simple Tree Models

A typical Classification and Regression Tree (CART) has an upside-down decision tree structure that divides the universe into sub-regions. For single trees, the key is to ensure data points in each divided region are as homogenous as possible. There are three key decisions to make:

[6] In machine learning, ensemble methods use multiple machine learning algorithms to improve predictive performance that could be obtained from any of the constituent learning algorithm alone. We will discuss bagging, boosting, and random forest as examples of ensemble models shortly.

- The predictor to split on and the value of the split,
- The depth or complexity of the tree, and
- The prediction equation in the terminal nodes

In the CART model (see Breiman, et al [1984]), the algorithm starts with the entire data set (S) and searches every distinct value of every predictor to find the predictor and the split value that partitions the data into two groups (S_1 and S_2) such that the overall SSE (Sums of Squares Error) are minimized:

$$SSE = \sum_{i \in S_1} (y_i - \overline{y_1})^2 + \sum_{i \in S_1} (y_i - \overline{y_2})^2$$

Where, $\bar{y}$ and $\bar{y}$ are the averages of the training set outcomes within groups S_1 and S_2, respectively.

We can keep growing the tree to improve the in-sample accuracy (called the tree growing step). However, we will also quickly overfit the tree, with little out-of-sample predictive power. A typical way to cut back from the full tree to a smaller but more robust structure is called pruning, by adding a penalty term to the SSE:

$$SSE_{C_p} = SSE + c_p \times (\#of\ Terminal\ Nodes)$$

Where c_p is the complexity parameter, to be tuned using, for example, cross validation.

Tree-based models can handle missing data well. Tree algorithms are also fairly fast, unless there are a large number of categorical predictors – in that case, tree models have to try every combination of categories. Lastly, tree models automatically conduct feature selection. Intuitively, predictors that appear at the top levels or appear multiple times in the tree are more important. They suffer from model instability and suboptimal prediction accuracy

Classification Trees can be constructed with the same philosophy as regression trees. The classification tree algorithm partitions the data into binary splits, by maximizing homogeneity within each sub-group, where homogeneity is defined by either Gini index or cross entropy. Another popular algorithm is the C4.5 model (see Quinlan [1993]), where the splitting criteria is based on information theory. Quinlan later on introduced C5.0, by incorporating boosting and other features.

To demonstrate how the CART model works, let's use a simple example. Each month, for stocks in our investment universe, we classify them into Outperformers (top 30%) and Underperformers (bottom 30%), based on each stock's country/sector adjusted returns. Then, we use the ~400 factors in our factor library to build a CART model. Figure 2 (A) shows a fitted CART model for the US market as of September 2018, using the previous 10 years of data. The first split is based on EBIT/EV, where expensive stocks fall into the right branch and are categorized as Underperformers. For cheaper stocks (based on EBIT/EV), we further divide them on realized volatility – stocks that are cheap and have high realized volatility are put into the middle branch and classified as Underperformers[7]. The left branch is further split on target price implied return – an intrinsic value metric – to further refine our value measurement.

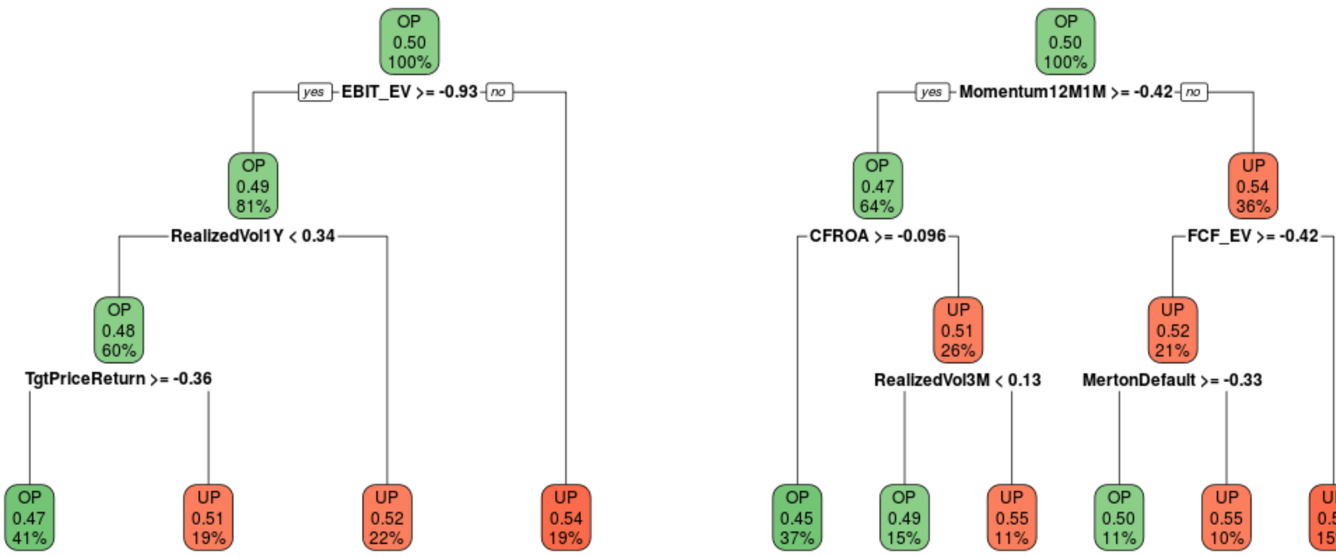
The CART model for US stocks is rather small – with three factors and four terminal nodes. On the other hand, the model for Europe is more complex, with five factors and six terminal nodes (see Figure 2 B). The CART model for Europe is more balanced with value (free cash flow/EV), quality (cash flow

[7] These stocks are likely to have high default risk; therefore, they are cheap for a reason.

return on assets and Merton's distance to default), price momentum (12-month return excluding most recent month), and volatility.

Figure 2 CART Models for US and Europe

A) US **B) Europe**



Note: **Green** color indicates predicted Outperformers (OP), while **red** means Underperformers (UP). At each spliting point, the left branch shows the stocks that meet the condition (i.e., yes), while the right branch includes samples that do not meet the condition (i.e., no). The second number in each box shows the actual percentage of Underperformers, while the last number denotes the percentage of observations fell in the specific node.

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Bagging

Many machine learning techniques suffer from the overfitting problem, i.e., high variance. Tree-based models, in particular, tend to be unstable. Breiman [1996] proposes a highly effective model averaging approach called Bagging (Bootstrap Aggregation), which is often used with tree models, but can also be applied to other machine learning algorithms. Bagging essentially involves the following steps:

- Draw a random sample of data from the original data set, via bootstrap
- Construct a model (e.g., train an unpruned tree model)
- Make predictions
- Repeat the above process m times
- The final prediction is an average of all the above m trees

Bagging is quite effective in reducing model variance and has shown meaningful performance improvement in many applications.

Another useful byproduct from the bagged model is that it produces a useful measure of out-of-sample performance. For each bootstrapped sample, certain sample data points are not used for model training. These data points are called "out-of-bag", which can be used to assess the predictive performance. The average of all out-of-bag error rates tends to be a reliable measure of the model's real out-of-sample performance.

There is one tuning parameter for bagging, m or the number of bootstrap samples to average. Empirical research generally shows steep performance improvement even with a small number (e.g., $m < 10$) but it then quickly tails off.

Bagging tends to be computationally intensive, but speed can be improved with parallel computing. Bagged models are also less interpretable.

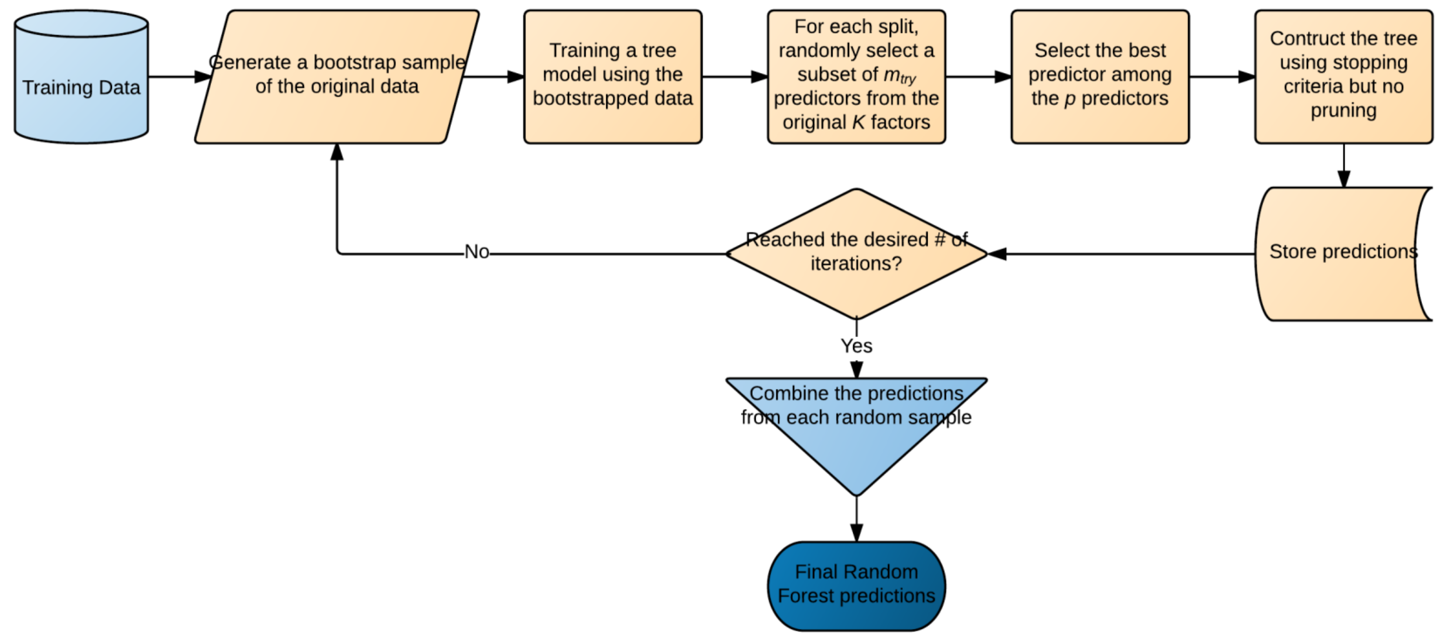
Random Forest (RF)

There are two potential issues with the bagging procedure. First, those factors with strong predictive power are likely to be at the top levels of most bagged trees; therefore the trees are correlated. Secondly, if some factors are highly correlated, they are likely to be in many trees. Breiman [2001] proposes a random forest algorithm, which treats these two problems effectively. The key difference between a random forest and bagged tree is, that each tree split, the algorithm randomly select a subset m_{try} factors from the K original factors. Breiman [2001] recommends to set $m_{try} = \sqrt{K}$.

Figure 3 shows the random forest model flowchart. In summary, the random forest algorithm uses bootstrap to select a sub-sample of data and a subset of factors, fit a CART model[8] with a non-pruned tree. Then it repeats the same procedure many times[9]. Finally, we take the average from each CART model's prediction to derive our final forecast.

[9] In practice, many researchers fit 1,000 trees, as a standard rule of thumb. However, the model performance is typically not very sensitive to the number of trees.

Figure 3 Random Forest Flowchart



Sources: Wolfe Research Luo's QES

The RF algorithm has two key tuning parameters (the number of random factors selected for each tree and the number of trees to grow), but the model is actually robust to both parameters. Between the two parameters, the number of random factors selected for each tree (called m_{try}) is more important and there are various discussions in the academic literature about the starting value.

In our experience, the random forest model tends to have a much stronger performance than CART. However, it has two significant hurdles to implement in practice:

- It is difficult to interpret the model, because it is not easy to show hundreds of trees, and
- It can be very slow[10]

One of the most useful tools to visualize the random forest model is the variable importance plot. Breiman [2001] describes the way to rank the factors based on their importance in explaining the model and reducing out-of-sample prediction error. To measure the importance of the ith variable, the values of the ith factor are permuted among the training and the out-of-bag error is again computed on this new dataset. The importance score for the factor is the average of the difference in out-of-bag error before and after the permutation over all trees.

Figure 4 (A) shows the top 10 factors selected by the RF algorithm, as of December 31, 1996, using 10-year trailing monthly data. Because we require all factors to have complete 10-year data, we only have about 100 factors at the end of 1996. It is somewhat surprising to see that the year-over-year earnings growth, abnormal volume, size (log market cap), and kurtosis are in the list, because these are not commonly used factors. As of the end of 2017, half of the top 10 factors remain the same as in

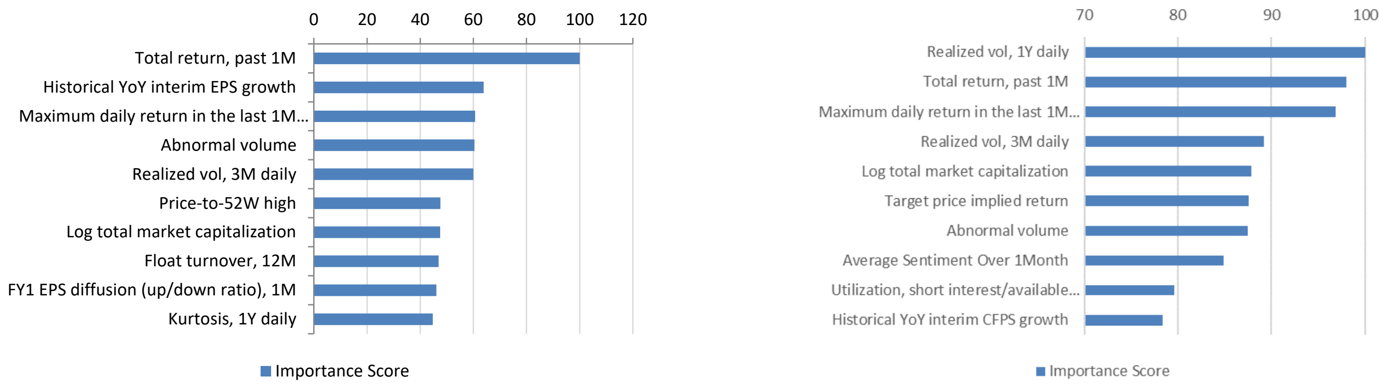
[10] In our backtesting of 3,000 stocks, 400 factors, and 10-year of monthly data, it takes about two hours on a 25-core Linux server to fit one single model.

1996, which highlights the stability of the random forest model. More interestingly, we note that two unconventional Big Data factors – Markit's securities lending short interest signal and Ravenpack's news sentiment factor enter the top 10 list.

Figure 4 The Top 10 Factors Selected by the Random Forest Algorithm

A) Data as of 12/31/1996, US

B) Data as of 12/31/2017, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Boosting – from AdaBoost to Gradient Boosting Machine

Boosting is a machine learning meta-algorithm, based on the philosophy of ensemble method (see Valiant [1984], Kearns and Valiant [1989]). Similar to RF, boosting is also in the ensemble method family, i.e., instead of using a single model, we construct multiple models and use the average prediction from all underlying models as our final forecast. Unlike RF, boosting is a sequential process, i.e., one model is added to the previous model at a time. In boosting, each component is often called a weak learner, meaning a classifier (or factor) that is only slightly correlated with the true classification. By sequentially summing up weak learners, we can build a single strong learner, i.e., a classifier (or model) that can well classify our problem.

Boosting can be applied to almost any machine learning algorithm. In practice, boosting is typically implemented on tree-based classification algorithms. In this section, we briefly introduce AdaBoost (see Freund and Schapire [1996]), Gradient Boosting Machine (see Friedman, et al [2010]), and then our MBBT.

AdaBoost

Schapire [1990] and Freund [1995] first introduced AdaBoost theory, while Freund and Schapire [1996] provided the first practical implementation. A simple description of the algorithm is shown in Figure 5.

To summarize, AdaBoost generates a sequence of weak classifiers. In our implementation, each weak classifier is a single factor. At each iteration, the algorithm finds the best classifier (factor) that can classify the current sample with the highest hit rate, based on the current sample weights. The sub-samples that are incorrectly classified in the kth iteration receive higher weight in the next (i.e., $(k+1)$th) iteration, while those sub-samples that are correctly classified receive lower weight. Therefore, the AdaBoost algorithm automatically shifts its focus on the sub-samples that are difficult to classify by the previous classifiers (factors). It guarantees to improve (or at least not to decrease) in-sample hit rate, while in practice, we often observe out-of-sample performance lift as well. At each iteration, the algorithm is required to learn a different aspect of the data, focusing on regions that contain difficult-to-classify samples. At each iteration, a stage weight is computed based on the error rate at that literature. The overall sequence of weighted classifiers (factors) is then combined into an ensemble and most likely has a much stronger predictive power than each of the underlying factor.

Figure 5 A Simple Representation of the AdaBoost Algorithm

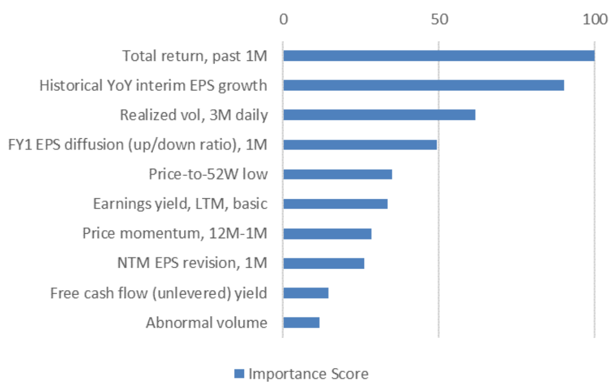
- Let's use a rolling sample, e.g., rolling one-year of monthly data
- At each month end, outperformers (e.g., top 30% of stocks based on forward one-month return) are represented with a value of $+1$, while underperformers (e.g., bottom 30% of stocks based on forward one-month return) have a value of -1
- Each observation (i.e., stock) have the same starting weight of $1/N$, where N is the total number of stocks in the rolling sample
- Start a for loop, from $k = 1 \ to \ K$
    - Fit a weak classifier (e.g., a CART model), using the weighted samples and compute the kth model's misclassification error (ε_k)
    - Compute the kth stage value as $\log(\frac{1-\varepsilon_k}{\varepsilon_k})$
    - Update the sample weights giving more weight to incorrectly predicted samples and less weight to correctly predicted samples
- End of the for loop
- Computed the boosted classifier's prediction for each stock by multiplying the kth stage value by the kth model prediction and adding these quantities across K. If this sum is positive then classify the stock as $+1$ (i.e., outperformer); otherwise, as -1 (i.e., underperformer)

Sources: Wolfe Research Luo's QES

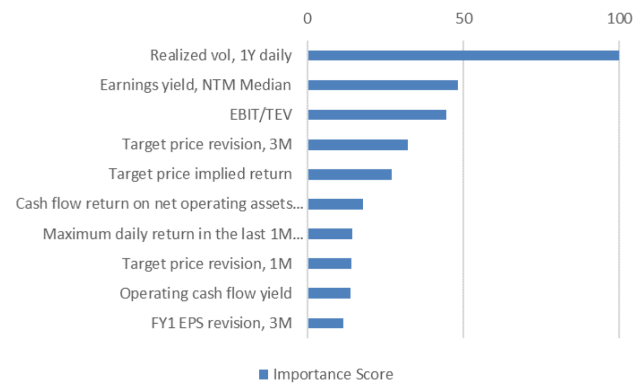
Figure 6 shows the top 10 factors selected for US at the end of 1996 versus 2017 by AdaBoost. Interestingly, the top factors selected by AdaBoost are similar as the ones chosen by RF (see Figure 4), which should not be a surprise, given both algorithms are ensembles of classification trees. It is important to reiterate that RF constructs a large number of trees independently; therefore, it can be easily paralleled. On the other hand, the AdaBoost algorithm builds classification trees sequentially, which typically delivers better in-sample fit.

Figure 6 The Top 10 Factors Selected by the AdaBoost Algorithm

A) Data as of 12/31/1996, US



B) Data as of 12/31/2017, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Gradient Boosting Machine

Friedman [1999] generalized the AdaBoost algorithm by allowing optimization of an arbitrary differentiable loss function. The resulting algorithm is often called Gradient Boosting Machine (GBM) of Stochastic Gradient Boosting.

For example, we can formulate our binary classification problem using a logistic regression function:

$$\widehat{p_i} = \frac{1}{1 + \exp[-f(x)]}$$

Where $f(x)$ is a model prediction in the range of $[-\infty, \infty]$.

Using the Bernoulli distribution, the algorithm for GBM can be written in Figure 7.

Figure 7 A Simple Representation of the GBM Algorithm

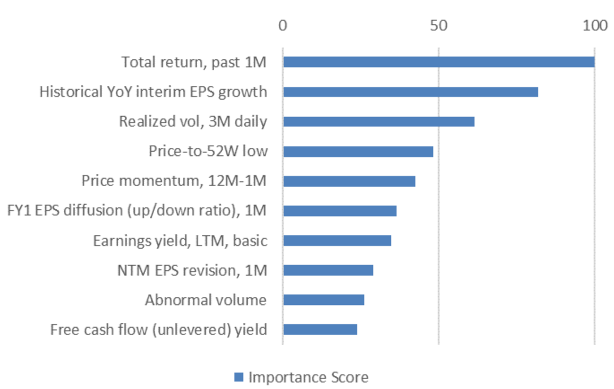
- Let's use a rolling sample, e.g., rolling one-year of monthly data
- At each month end, outperformers (e.g., top 30% of stocks based on forward one-month return) are represented with a value of $+1$, while underperformers (e.g., bottom 30% of stocks based on forward one-month return) have a value of 0
- Initialize all predictions to the sample log-odds $f_i^{(0)} = log\frac{\hat{p}}{1-\hat{p}}$
- Start a for loop, from $k = 1\ to\ K$
  - Compute the residual (i.e., gradient) $z_i = y_i - \widehat{p_i}$
  - Randomly sample the training data
  - Train a CART model on the random sample using the residuals as the outcome
  - Compute the terminal node estimates of the Pearson residuals: $\varepsilon_i = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{p_i})}{\frac{1}{n}\sum_{i=1}^{n}\widehat{p_i}(1-\widehat{p_i})}$
  - Update the current model using $f_i = f_i + \lambda f_i^{(k)}$
- End of the for loop
- Computed the boosted classifier's prediction for each stock by multiplying the kth stage value by the kth model prediction and adding these quantities across K. The final combined prediction is the probability that a stock will outperform in the next month.

Sources: Wolfe Research Luo's QES

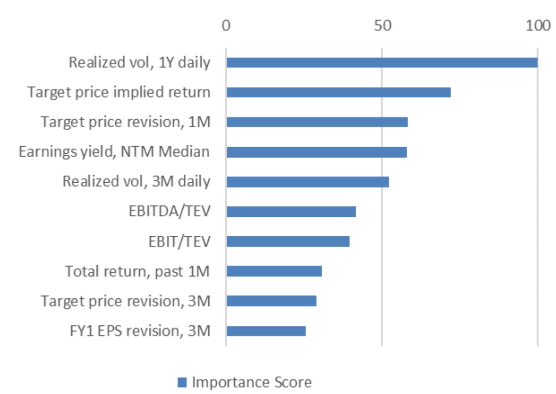
Figure 8 shows the top 10 factors selected for US at the end of 1996 versus 2017 by GBM. Interestingly, the factors selected by GBM are very similar as the factors selected by AdaBoost (see Figure 6), but sufficiently different from the RF model (see Figure 4). For example, at the end of 1996, the top 10 factors selected by GBM are exactly the same as those selected by AdaBoost, with a small difference in ordering. At the end of 2017, seven out of top 10 factors selected by GBM are the same as AdaBoost, while the overlap with the RF model is minimal. The results are intuitive, since GBM is a natural extension of the AdaBoost, while RF is a rather different type of machine learning algorithm.

Figure 8 The Top 10 Factors Selected by the GBM Algorithm

A) Data as of 12/31/1996, US



B) Data as of 12/31/2017, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

MBBT DESIGNED FOR FINANCE

In practice, boosting is typically implemented with tree-based base learners. Most existing tree-based models assume binary splits at each level. In theory, a highly nonlinear zigzag pattern can be learned and proxied by tree models with multiple binary splits. In finance, however, the most frequently observed nonlinear patterns are typically U- or inverse U-shaped. The payoff is also often asymmetric, i.e., one extreme is typically very different from the other side. In *Multi-Dimensional Alpha: Pay for Performance* (see Rohal, et al [2018b]), we find corporate governance data often exhibits a classic inverse U-shape curve. For example, companies with the highest paid CEOs often underperform, possibly because overpaid executives have less incentives to innovate. Similarly, firms with the poorest paid executives also significantly lag behind. The payoff pattern is also highly asymmetric, in that the lowest paid companies deliver substantially lower returns than the best paid firms, albeit both extremes underperform the average.

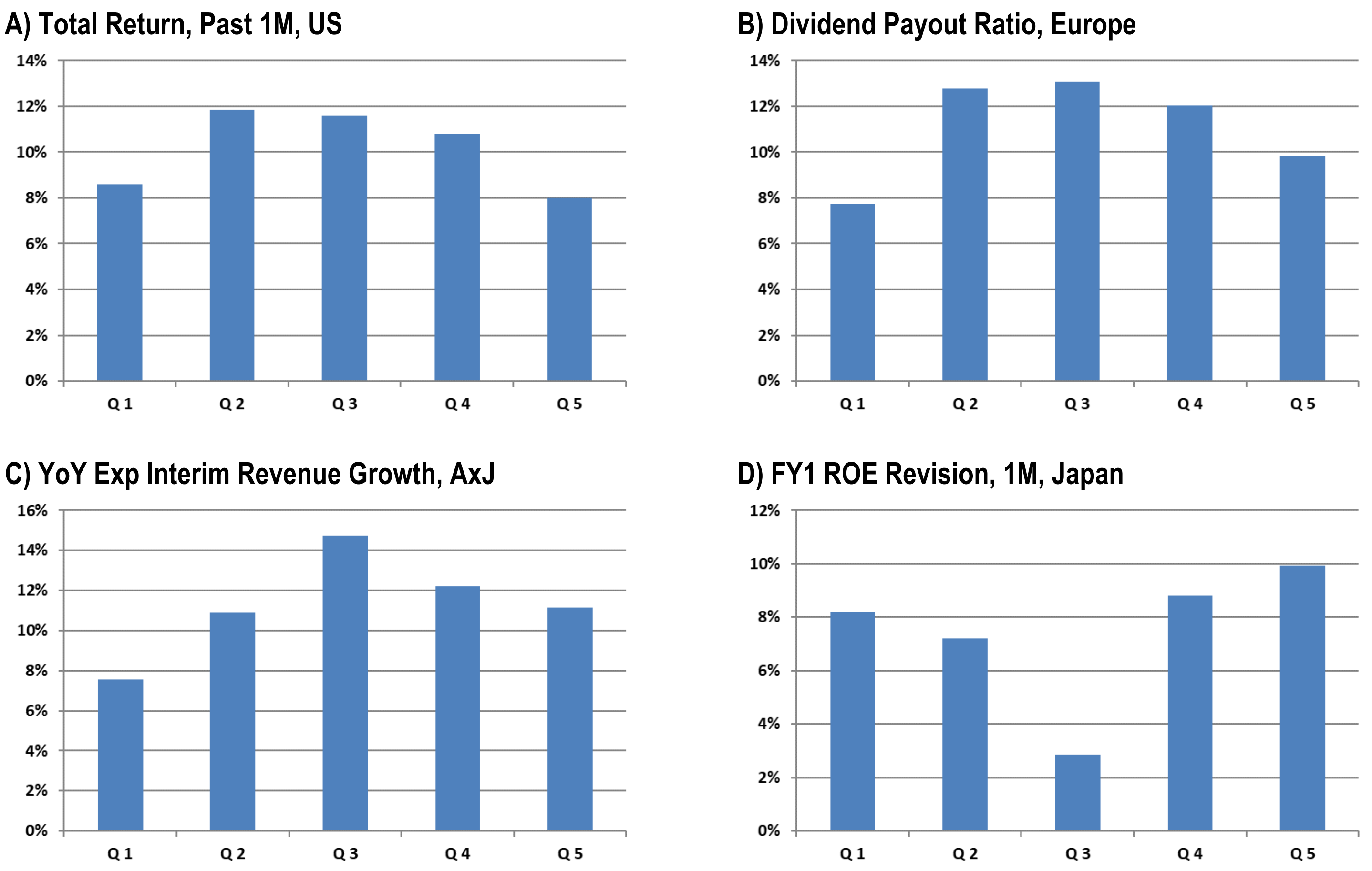
Figure 9 shows a few examples of nonlinear patterns that we often observe in finance. Short-term mean reversal signal (e.g., stocks with the highest return in the last month often deliver terrible returns in the subsequent month) used to be a strong factor in the US. However, possibly due to the popularity of statistical arbitrage hedge funds, the reversal factor has lost its appeal in recent years. As shown in Figure 9(A), although the spread between Q1 (worst performing stocks in the past month) and Q5 (best performing stocks in the last month) is minimal, the middle three quintiles outperform the two extremes meaningfully. Similarly, stocks with excessively high or low payout ratios in Europe both trail behind companies with modest payout policies (see Figure 9B) – not paying sufficient dividends resembles agency problems[11], while firms with unsustainably high payout ratios are likely to cut dividends.

In Asia, although the payoff to high growth companies has stalled, firms with modest growth expectations – possibly due to their reasonable valuation – actually produce the best future returns (see Figure 9C). Lastly, in Japan, it is strange to see that investors have been rewarding stocks with both high and low ROE revisions. It is intuitive to say that stocks with positive ROE revision to outperform. The similar outperformance of negative ROE revision names is possible due to investors' initial over-reaction to bad news, which leads to a mid-term reversal. Nonetheless, all four charts in Figure 9 demonstrate the classic U- or inverse U-shaped curves in finance. Using traditional linear regression, none of these four factors are statistically significant predictors of future stock returns. Similarly, adding them to convention multifactor models only blends in noises.

Nonlinear classification models can identify patterns that are irrelevant to most investors; therefore, is likely to lead to unexplored territory.

[11] In corporate finance theory, the interest of firm management and shareholders is not always aligned. Management has the tendency to build an "empire", i.e., not willing to pay out all free cash flows as dividends.

Figure 9 Common Nonlinear Patterns Observed in Finance



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

MBBT Algorithm – Uniquely Designed for Finance

Our MBBT algorithm is based on the AdaBoost and GBM philosophy, but extends the boosting to the next level. It is uniquely designed to process financial data. We find binary trees (e.g., CART) may not be able to capture the U-/inverse U-shaped data efficiently. As the name suggests, MBBT allows multi-branch tree splits. Although in theory, a true U-/inverse U-shape curve can still be captured by multiple layers of binary splits, in practice, most binary tree algorithms either stop too early (which fails to catch one-side of the curve) or grows too deep (which overfits the data). We divide the factors into quantiles and design the week classifier based on hit rate of the outperformers over underperforms for each quantile. Details of the MBBT algorithm is shown in Figure 10.

Figure 10 A Simple Representation of the MBBT Algorithm

- Let's use a rolling sample, e.g., rolling one-year of monthly data
- At each month end, outperformers (e.g., top 30% of stocks based on forward one-month return) are represented with a value of $+1$, while underperformers (e.g., bottom 30% of stocks based on forward one-month return) have a value of -1
- Each observation (i.e., stock) have the same starting weight of $1/N$, where N is the total number of stocks in the rolling sample
- Start a for loop, from $k = 1\ to\ K$
    - Find the best weak classifier w^k (i.e., a factor in our factor library)
        - Divide the training data into Q quantiles based on the chosen weak classifier (factor)
        - For each quantile q, calculate the total weights for outperformers W_+^q and total weights of underperformers W_-^q
        - Find the best weak classifier (factor) that minimizes the loss function $\sum_{q=1}^{Q} \sqrt{W_-^q W_+^q}$
        - For stocks in the quantile q, the output of the weak classifier w^k is computed as $\frac{1}{2}\log(\frac{W_+^q}{W_-^q})$
    - Update the sample weights, by giving higher weights to incorrectly predicted samples and lower weights to correctly predicted samples
        - For positive samples, the weight is updated by a scaler of exp(-w^k)
        - For negative samples, the weight is updated by a scaler of exp(w^k)
    - Normalize the weights so that they add up to 1
- End of the for loop
- Compute the boosted classifier's prediction for each stock by adding up all the weak classifiers. If this sum is positive, then classify the stock as $+1$ (i.e., outperformer); otherwise, as -1 (i.e., underperformer)

Sources: Wolfe Research Luo's QES

A Horserace of Machine Learning Algorithms

Now, let's conduct an interesting horserace of machine learning algorithms. We use the same eight factors in the BM. To make sure a fair comparison, we backtest the following five machine learning algorithms – CART, RF, AdaBoost, GBM, and MBBT, using the same procedure:

- At the end of each month, we train our model using a rolling one-year data.
- All five machine learning algorithms have built-in feature selection functions; therefore, we let each model to select the factors, identify the pattern, and make predictions.
- We then track the performance of a long/short tercile portfolio[12], based on a monthly rebalance frequency.
- The results from all five models are completely out-of-sample.

[12] Throughout this paper, we use predominately long/short quintile portfolios. However, in this section, we use tercile portfolios. The reason is that the CART algorithm typically only has a few terminal nodes; therefore, a few categorical predictions. Many of the time, the CART model does not give a refined enough prediction to form quintile portfolios.

First of all, we want to see which algorithm can accurately identify the true pattern. We cherry-pick two factors – earnings revision and book-to-market – the former factor has a true pattern of an inverse-U shape, while the latter demonstrates a more conventional linear monotonic configuration (see Figure 11A). Obviously, the true pattern is unknown *ex ante*. Now, it is up to each machine learning algorithm to search for these patterns, each month at a time, using data only available at that time. Therefore, we are comparing each technique's true ability to identify patterns *ex ante*.

First of all, although all five machine learning algorithms are nonlinear classification tools, they are all able to identify a classic linear pattern (book-to-market) perfectly, out of sample.

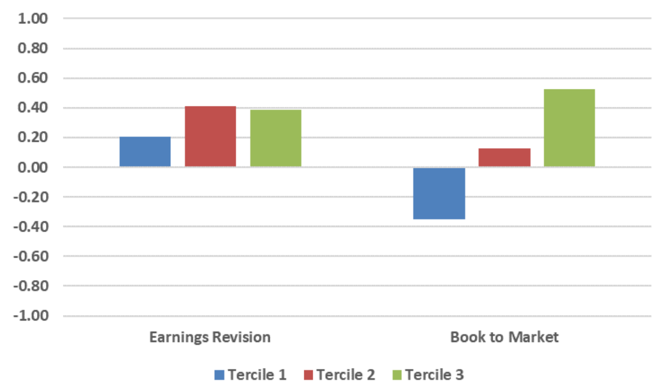
For a nonlinear pattern presented by the earnings revision factor, however, even though the first four existing algorithms (CART, RF, AdaBoost, and GMB) can the first four models can pinpoint a nonlinear shape, they are not able to correctly distinguish a convex from a concave curve (see Figure 11B-F). Our MBBT is the only algorithm that can properly diagnose the unusual inverse U-shape.

Figure 11 Patterns Identified by Different Machine Learning Algorithms for Japan

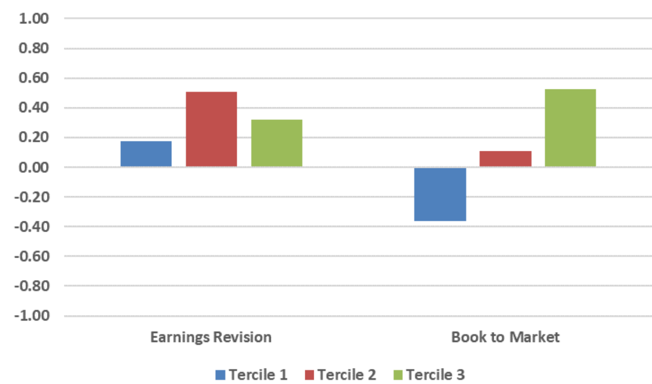
A) True Pattern (CAGR)

B) CART

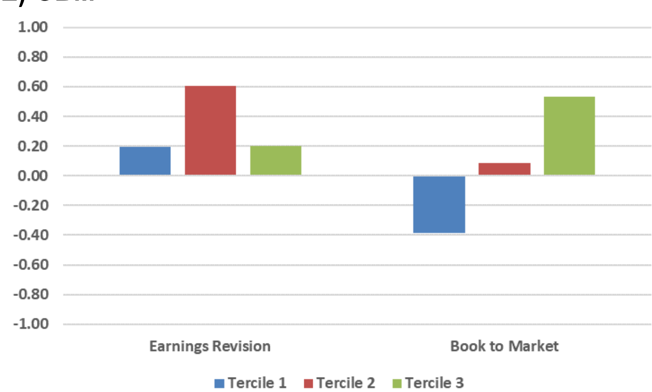
C) Random Forest



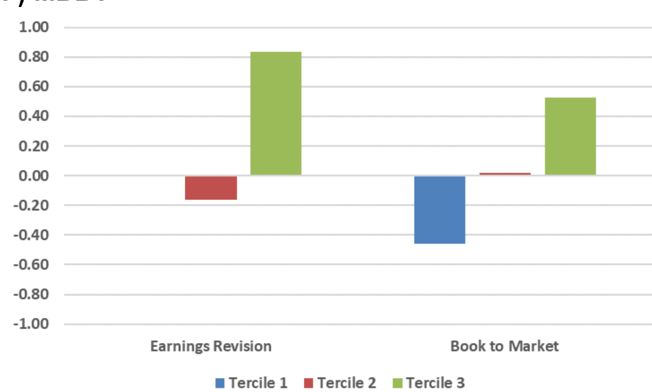
D) AdaBoost



E) GBM



F) MBBT



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

As discussed in *Multi-Dimensional Alpha: Machine Learning Takeovers* (see Wang, et al [2017a]), statistical accuracy, however, does not automatically leads to economic profit. Now, let's see which of the five algorithms delivers the highest out-of-sample return.

As shown in Figure 12, all five machine learning algorithms outperform the naïve BM, which equally weights the eight underlying factors. Therefore, machine learning models, if designed and applied properly, generally are able to identify patterns better than traditional handpicking factors.

Secondly, ensemble learning (RF, AdaBoost, GMB, and MBBT) all beats the CART model considerably, which confirms George Box's famous argument that "all models are wrong but some are useful" – model averaging generally improves performance.

Lastly, in this case, a more accurate pattern recognition algorithm – MBBT – does lead to stronger investment performance. The MBBT model outperforms all four baseline machine learning algorithms with a higher Sharpe ratio (see Figure 12A), better return (see Figure 12B), lower volatility (see Figure 12C), and lower drawdown (see Figure 12D).

Figure 12 A Horserace of Different Machine Learning Algorithms, Japan

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

MBBT Under the Hood

Despite the recent hype about AI and machine learning, the existing academic literature of using AI/ML in investing is rather limited. Fundamental investors worry about using a model that they do not fully understand. Quantitative investors who are trained from the classic finance angle often label ML as data mining/overfitting. In this section, we use our MBBT algorithm as an example to address the following questions people typically ask in terms of machine learning:

- Can machine learning improves performance with only a handful of factors?
- Why is it important to capture the nonlinear payoff pattern of factors?
- How can we interpret the output of machine learning model?
- Why our MBBT is robust and less likely to face the overfitting problem?
- Is it true that including more factors always improves performance?
- How long should the look-back window (rolling window) be? Are longer training periods, i.e. more training data improve the model performance?

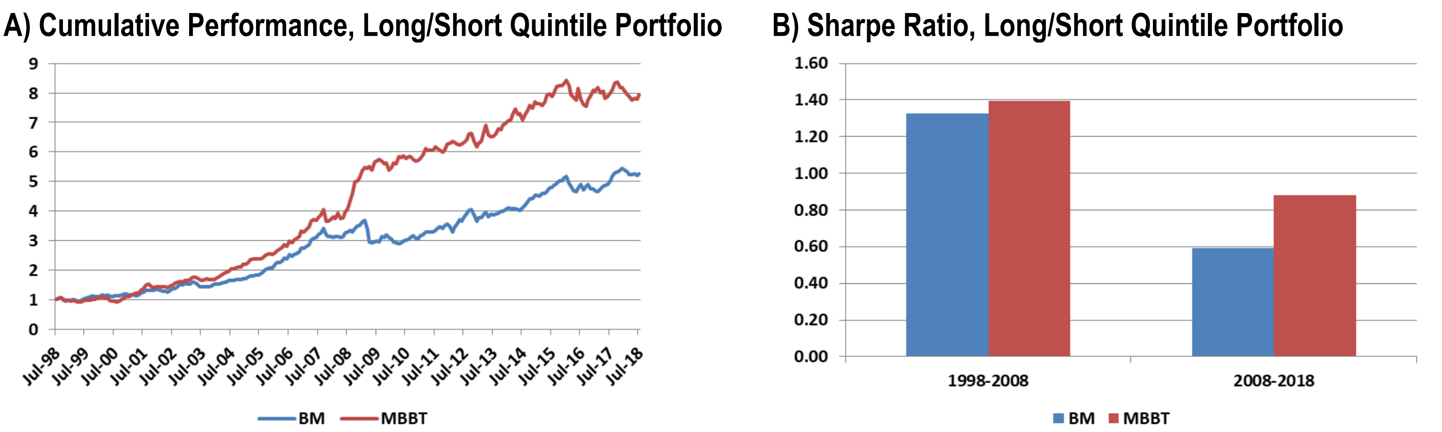
Improved Performance even with a Small Factor Pool

Among the many things that machine learning excels is feature selection, i.e., deciding what factors to use for a given context, at a given point-in-time. In today's Big Data world, portfolio managers often face too many (rather than too few) relevant data points. the performance improvement by combining hundreds of different factors. In this section, we want to address the question – if we only have a few factors, do we still need to use machine learning?

To show incremental contribution from machine learning, we compare the performance of an equally weighted BM (Benchmark) model with our MBBT trained model, using the same eight factors.

For simplicity, our MBBT model is trained at the end of every month, using a rolling 12-month window, for the Japanese equity market. As shown in Figure 13(A), our MBBT model outperforms the BM model considerably. The outperformance is especially significant in recent years (see Figure 13B). Indeed, in the post-2008 period, the MBBT model delivers a Sharpe ratio that is 50% higher than the BM.

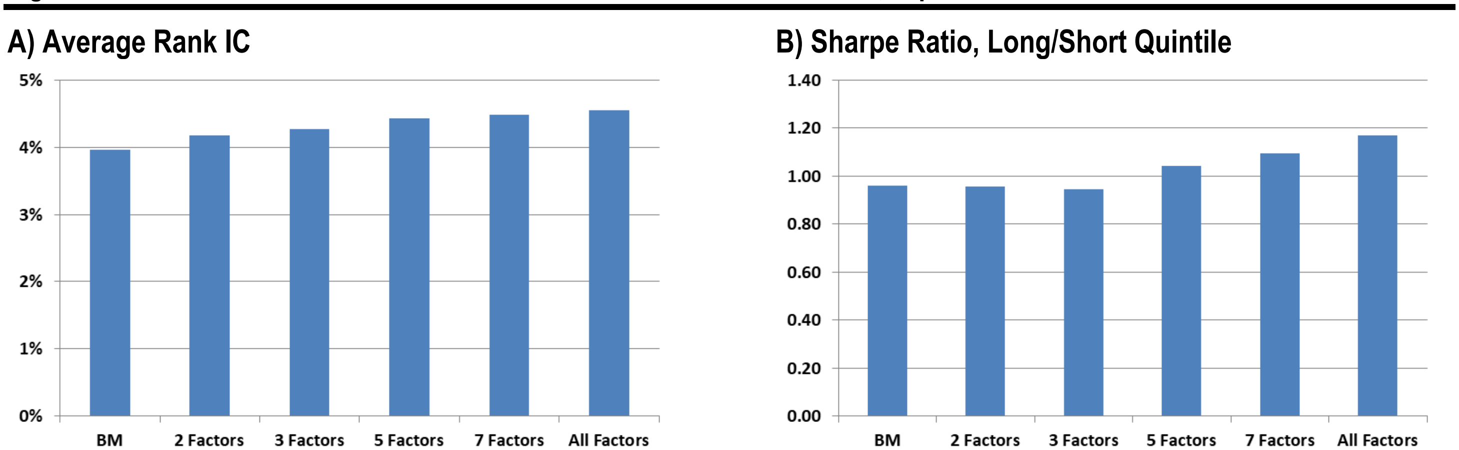
Figure 13 Machine Learning Using the Same Input Factors as the Benchmark Model, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As a reminder, the MBBT algorithm essentially "boosts" or adds factors sequentially. As shown in Figure 14, even with only two factors, the model already matches and even slightly outperforms the BM. As the number of factors increases, the model performance also improves, albeit at a slower pace, i.e., the law of diminishing marginal returns. Please note that the performance is completely out-of-sample.

Figure 14 Performance Increase with More Factors From Benchmark Models, Japan Model



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, HIS Markit, Wolfe Research Luo's QES

Capturing Nonlinear Payoff Pattern

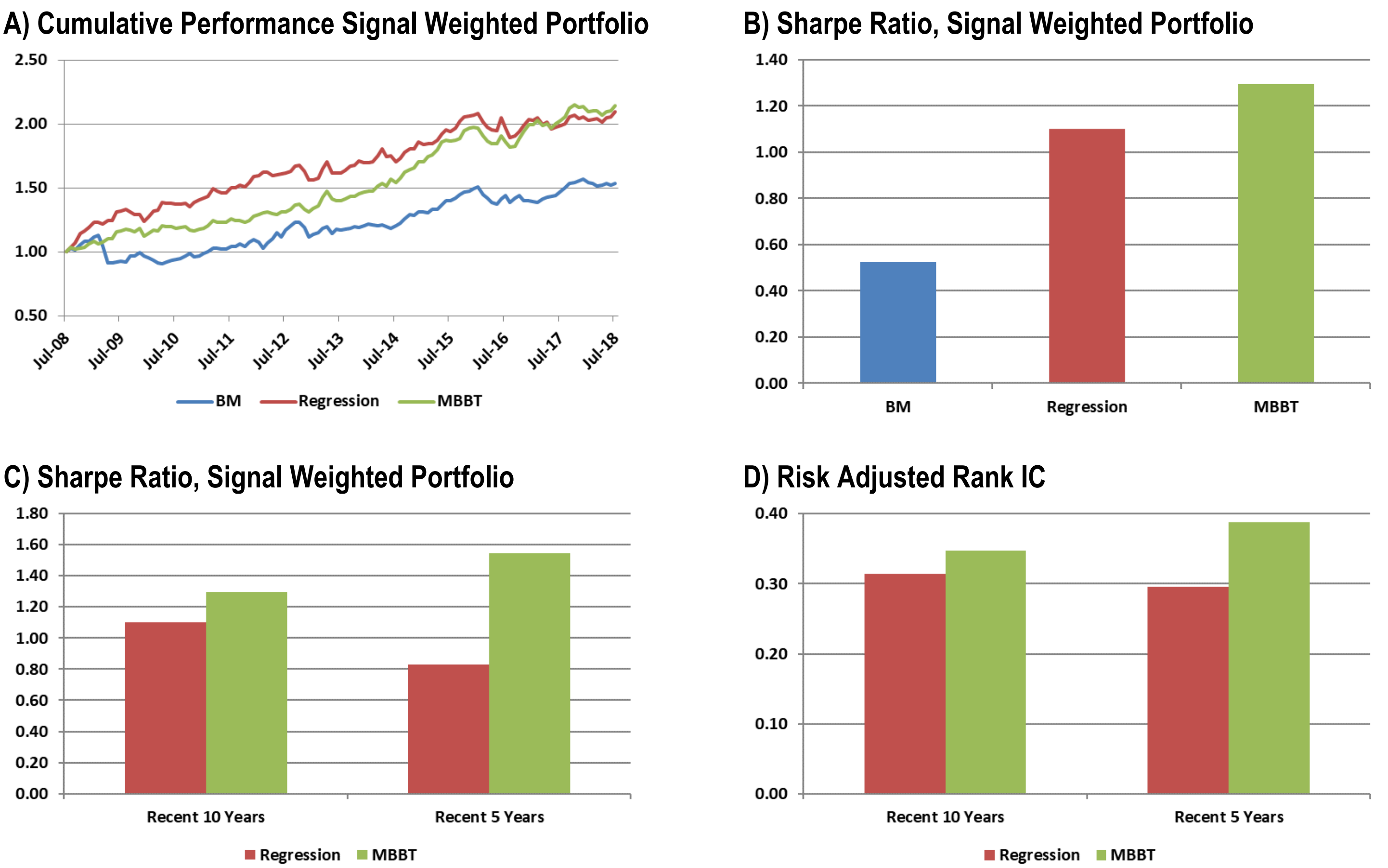
The improvement of MBBT model over BM primarily comes from two sources – dynamic factor selection and capturing nonlinear patterns. To illustrate the importance of capturing the nonlinear factor payoff pattern, we also compared the performance with linear dynamic regression model using the same 12 month rolling window as the training data.

To assess the incremental value added from nonlinear modeling, we construct two comparable models:

- **Regression.** We use a 10-year rolling window to build the model. Factor returns are estimated using linear regression. The model is re-fitted each month.
- **MBBT.** We use the same 10-year rolling window and fitting frequency (i.e., monthly), on the same set of eight factors. However, model is estimated using our MBBT algorithm.

Let's use one of the most difficult markets (from the alpha generation perspective) – Japan, as an example. As shown in Figure 15(A), the cumulative return of the Regression and MBBT models is similar, but the MBBT model's Sharpe ratio is almost 15% higher (see Figure 15B). In the past five years, when traditional factor models such as the BM are essentially flat, the MBBT model's Sharpe ratio is almost twice as high as the Regression model (see Figure 15C). As a verification, measured by risk-adjusted rank IC, the MBBT model also shows much stronger performance than the Regression model, especially in recent years (see Figure 15D).

Figure 15 Improvement from Capturing Factor Nonlinearity, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

MODEL TRANSPARENCY

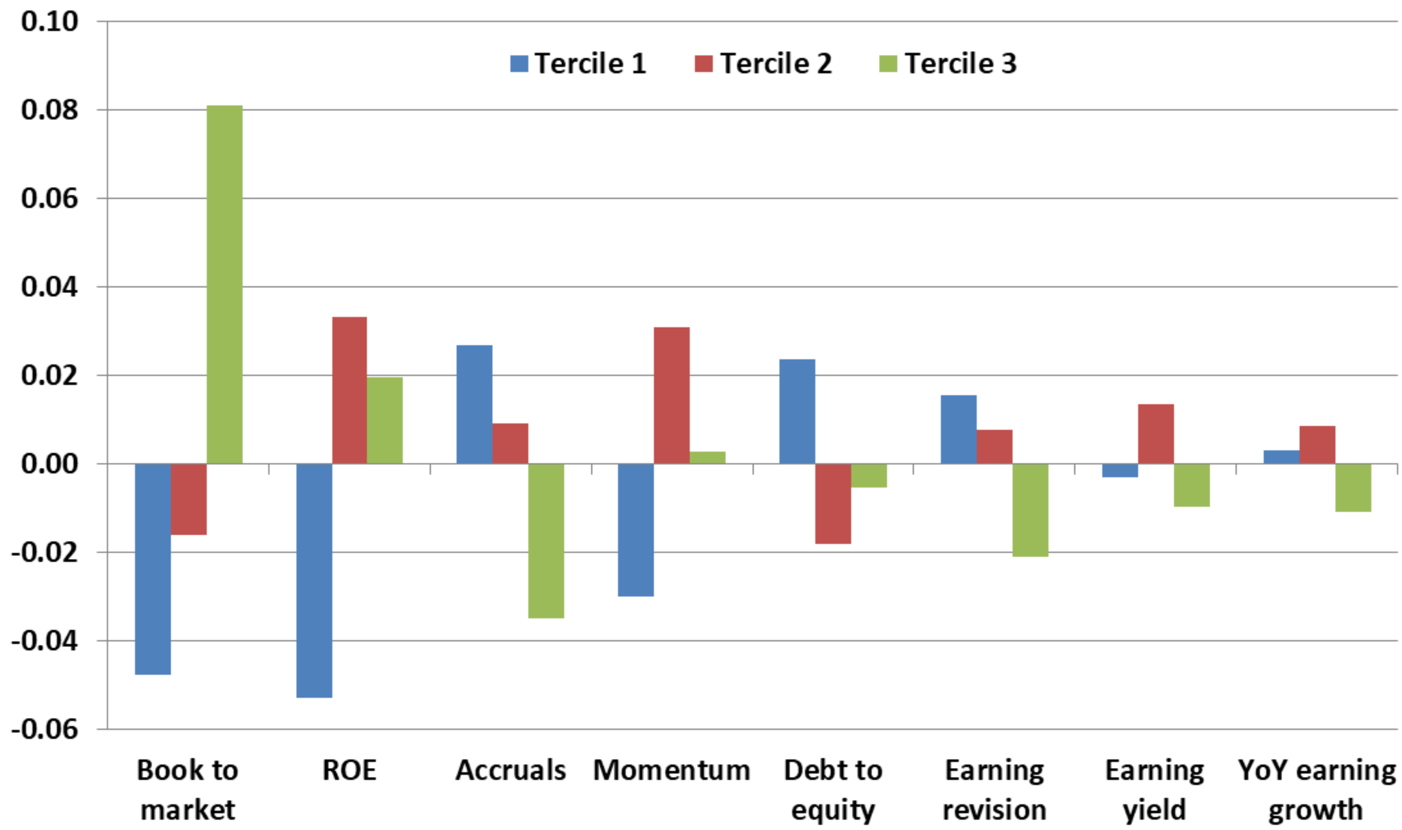
Essentially, the MBBT algorithm is a suite of multi-layer multi-branch decision trees. As a result, the model turns out to be much more transparent than other ensemble methods. For each branch (i.e., quantile) in the boosted tree, based on a given factor, the MBBT algorithm assigns a different value. The final prediction for a given stock is the weighted sum of all the assigned values based on each branch/factor that this stock travels along. The MBBT algorithm can capture nonlinear patterns, especially the U-/inverse U-shaped data often observed in finance surprisingly well.

As a demonstration, Figure 16 shows the order of the eight factors, along with the patterns captured, using the MBBT model for Japan[13].

- In this case, book-to-market is the first factor, which contributes the most weight in the final prediction, followed by ROE, accruals, and price momentum.
- The patterns for book-to-market, accruals, and earnings revision factors are not exactly linear.
- Most factors show strong nonlinear structure. For instance, both the cheapest and most expensive stocks (based on earnings yield) are expected to underperform companies with modest valuation.
- The MBBT algorithm also conducts implicit factor timing. For most factors, the direction and pattern identified by the model are reasonable intuitive. However, the MBBT algorithm can indeed deliver counter-intuitive predictions. For example, currently, the model bets against earnings revision – in particular, the stocks with the highest earnings momentum are to be avoided.

[13] The chart is based on the patterns identified as of July 2018. Please note that our models are fully dynamic; therefore, the factors/patterns also change over time. In this demonstration, we use a rolling 10-year data to build the model.

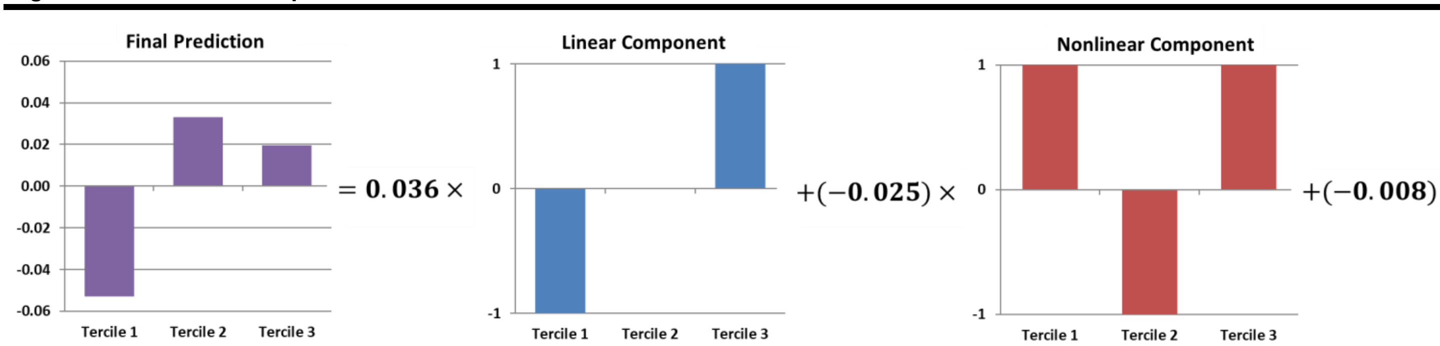
Figure 16 Nonlinear Payoff Pattern Identified, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

We can further decompose each factor into a linear component and a nonlinear element, which may help investors to understand the intuitive behind the MBBT algorithm. For example, the nonlinear inverse U-shape curve for the ROE factor (see Figure 16) comprises a linear upward sloping part (see Figure 17) and a nonlinear V-pattern. The linear component of ROE has the intuitive direction, i.e., more profitable firms are preferred. The two components can be treated as two separate models.

Figure 17 Pattern Decomposition, ROE Factor



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

CURSE OF DIMENSIONALITY?

Many investors associate model overfitting with machine learning. If applied naively, given the speed of computation and processing, machine learning algorithms can handle (almost) unlimited number of factors. Furthermore, as the number of factors increases, in-sample model fit almost always goes up and performance generally improves. Obviously, there is no guarantee that out-of-sample predictive power also progresses the same way. As dimensionality increase, out-of-sample performance using traditional variable selection techniques such as stepwise regression often falls sharply. As a result, critics suggest using economic intuitive to select factors rather than an automated procedure.

As we discussed in *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP* (see Luo, et al [2017c]), machine learning techniques, if used properly, can actually help researchers to greatly reduce the risk of model overfitting. For example, simple 10-folder cross validation typically produces far more honest true out-of-sample performance than traditional variable selection techniques. In machine learning, there is a wealth of algorithms and research on feature selection (i.e., factor selection). In *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)* (see Wang et al [2018a]), we demonstrated that our MBBT algorithm is particularly effective in high-dimension variable selection problems.

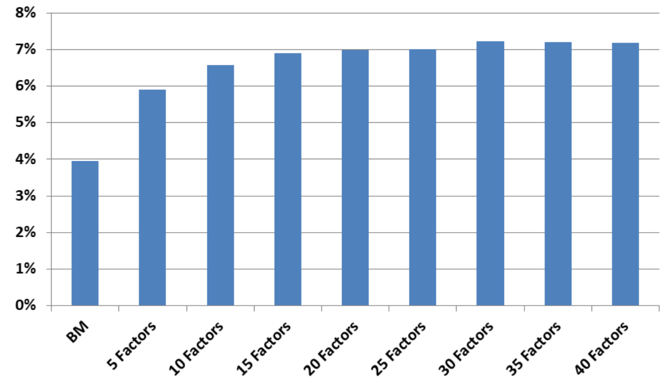
Optimal Number of Factors

In a conventional linear framework, a two-factor model is almost always better than a single factor form, but a model with hundreds of factors would is probably a overkill. Most portfolio managers would probably stick to 10-20 factors.

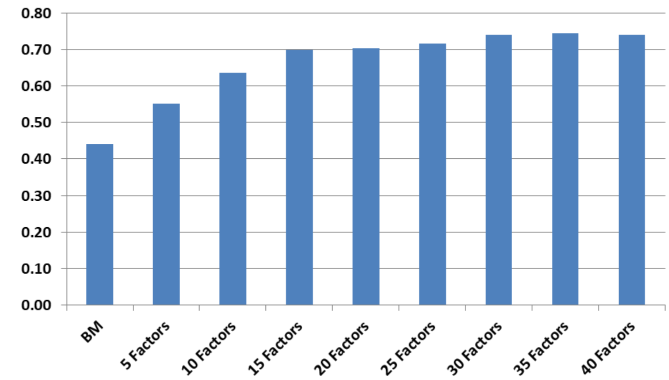
As an example, let's train our MBBT model using a suite of around 200 common factors as input variables. Figure 18 shows the out-of-sample performance of our MBBT model, as the number of chosen factors in the model increases from five to 40. Initially, the model performance jumps rapidly, as we lift the factors from five to 10, due to diversification benefit. However, after 15 factors, the incremental improvement quickly levels off, i.e., the law of diminishing marginal contribution. More importantly, as we overfit the model with more factors than necessary, e.g., 40 factors, the out-of-sample performance does not decrease. Therefore, the MBBT algorithm is extremely robust to large dimension problems and model overfitting. Including a large number of factors in the model does add up to computational speed; however, it does not hurt the model's out-of-sample performance.

Figure 18 The Relationship between Out-of-sample Performance and Model Complexity, Japan

A) Average Rank IC



B) Risk Adjusted Rank IC



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

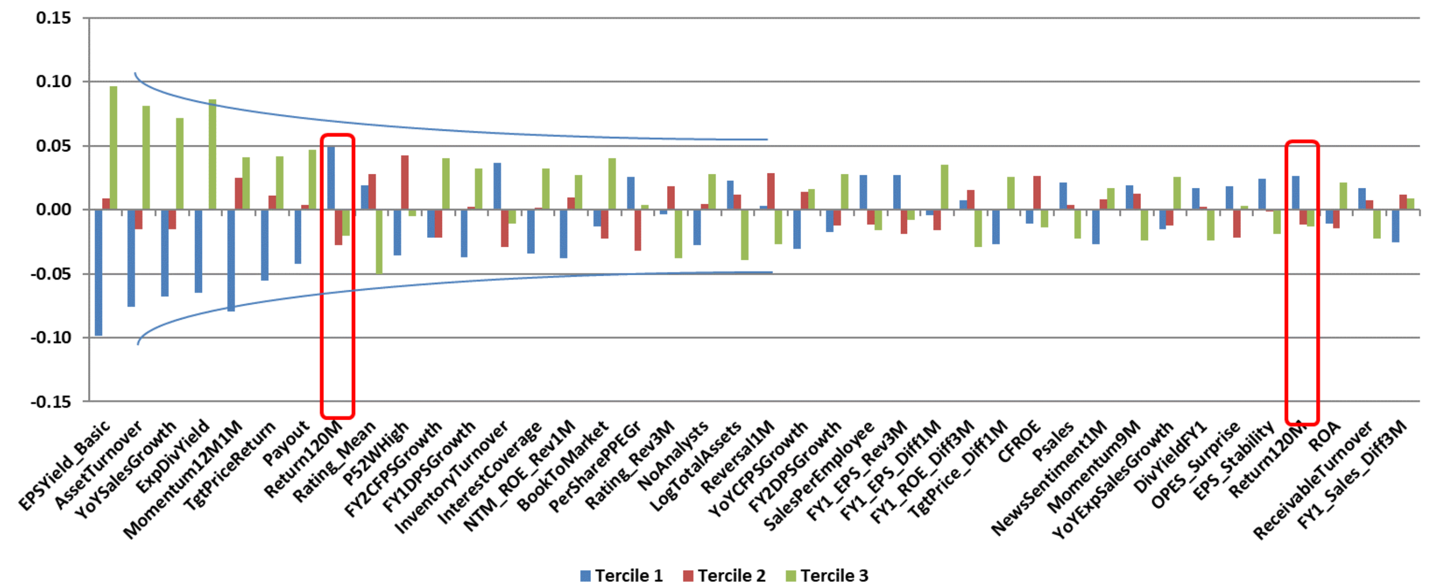
The reason why our MBBT model, and other boosting-type of models, is robust to overfitting is mainly due to the very design philosophy behind boosting. By construction, boosting is a sequential model building process, in that each new weak learner (i.e., factor) is added, after we overweight the sub-sample that the previous model does not fit well. Therefore, by construction, each new factor adds incremental explanatory power to the model.

Figure 19 shows the 40-factor MBBT model for Japan, in the order as the factors are selected. We can see the first few factors have outsized magnitude than the remaining factors. As more factors are added to the model, variable importance[14] decreases exponentially. Therefore, even if we overfit the model, after the first 15-20 factors, the remaining factors have little impact on the final score (i.e., predicted probability of outperformance).

In addition, although we do not explicitly restrict the model to re-select the same factor at a later stage, in practice, the MBBT algorithm rarely chooses the same factor more than once. In our example, among the top 40 factors selected (see Figure 19), only one factor (120-month return) is initially selected as #8 (to capture the long-term reversal effect on the bottom tercile) and then at step #37 with a minimal weight (with slight difference for the top two terciles). In theory, if the pattern is highly nonlinear zigzag type, it might require the MBBT algorithm more than once to fit the data. In practice, as we discussed in the earlier section, finance data often demonstrate U- or inverse U-shape curves; therefore, in most cases, it only takes the MBBT model once to identify the main feature in the data.

[14] Since the MBBT algorithm is nonlinear in nature, factor importance is not the same as factor weight.

Figure 19 First 40 Factors Selected by MBBT for Japan (As of December 31, 2017)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Lastly, boosting-type of models are also robust to multicollinearity. Highly correlated factors (either linearly or nonlinearly) are highly unlikely to be re-selected again in a later stage, once one of the factors from the cluster is used in an earlier step, because similar factors are unlikely to add incremental predictive power.

For example, in our factor library, there are many earnings yield related factors (see Figure 20). As expected, most of these factors are highly correlated. For the MBBT algorithm, the basic trailing earnings yield is considered as very important. However, once it is being selected, none of the other earnings yield factors enter the model, even though they may be just as predictive by themselves.

Figure 20 Earnings Yield Related Factors, Japan

	Trailing, Basic	Trailing, Fully Diluted	FY1	FY2	Growth flow yield	EBIT / TEV
Trailing, Basic	100%					
Trailing, Fully Diluted	99%	100%				
FY1	83%	82%	100%			
FY2	67%	66%	94%	100%		
Growth flow yield	95%	95%	86%	76%	100%	
EBIT / TEV	90%	91%	64%	45%	81%	100%

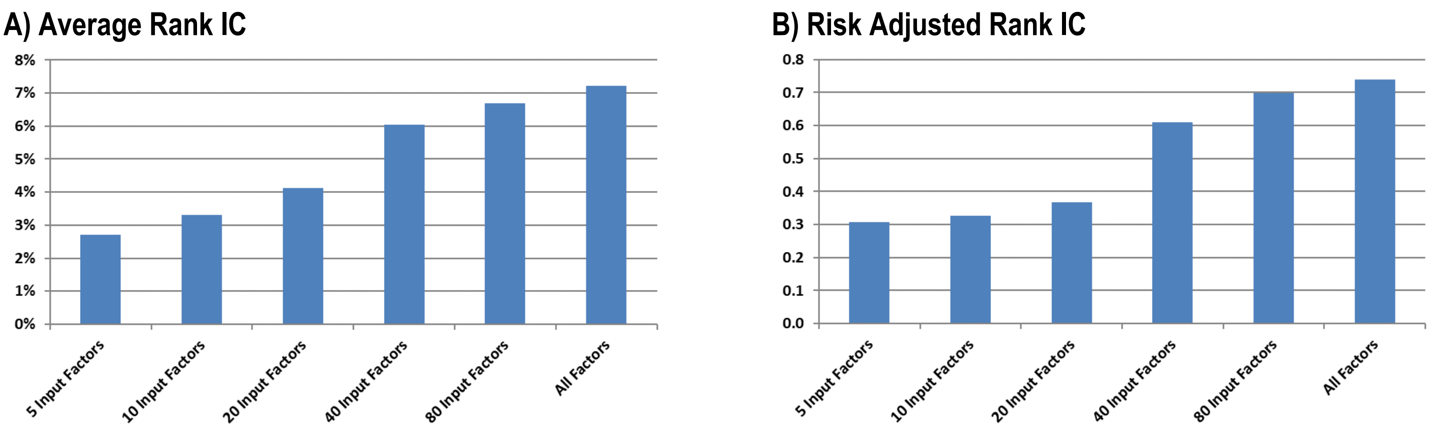
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Factor Library – The More, the Merrier

The previous section discussed the optimal number of factors that we want to use in our model. A closely related question is how many input factors we should use to build our model. We could either start from our entire factor library, or some preliminary feature selection (i.e., factor selection) algorithm can be applied first. Some machine learning algorithms have built-in feature selection features, e.g., random forest, AdaBoost, MBBT. On the other hand, many traditional model building techniques (e.g., linear regression) are not robust to high dimensionality. We have discussed the feature selection issue extensively (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017c] and *Multi-Dimensional Alpha: Systematic Alpha from Risk Arbitrage (SARA)*, Wang et al [2018a]).

Since the MBBT algorithm is fairly effective as a feature selection tool, we do not need a separate filtering process. Figure 21 shows the out-of-sample performance of our MBBT model, as we increase the number of input factors from five to the entire factor library. Initially, as the number of factors increases, performance rises rapidly. The first 80 factors capture the majority of the available information. The incremental value from 80 to 200 factors is only modest.

Figure 21 The Relationship between Out-of-sample Performance and the Size of Information Set (# of Input Factors), Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Sample Size – How Much is Enough?

Most machine learning algorithms tend to be data hungry; therefore, require a large enough data sample to accurately identify the true patterns. In stock selection context, there are numerous ways to expand our training sample, but mostly on two key dimensions:

- **Enlarge our sample cross sectionally.** We do not necessarily have to train our model on the same sample of stocks as our investment universe. For example, when we try to model risk arbitrage transactions in Canadian merger and acquisition deals, we know our data sample is relatively small, with only a handful transactions as of any given point in time. We could broaden our training sample by including US deals. Obviously, as we increase our sample considerably, we also introduce significant noises. US equity market is similar to, but certainly different from Canadian market. It is a trade-off between sample size and sample homogeneity. As discussed in *Global Systematic Risk Arbitrage – SARA Global* (see Wang, et al [2018]), we find strong empirical evidence that sample size outweighs sample homogeneity. Even if we are only interested in trading M&A arbitrage in Canada, there is a significant benefit by studying US transactions.
- **Stack up cross sectional sample over time.** Few researchers would build a model using data only from the most recent period. Instead, most of the time, we would pool cross-sectional observations from multiple periods together. There is again, no theory to help us select the

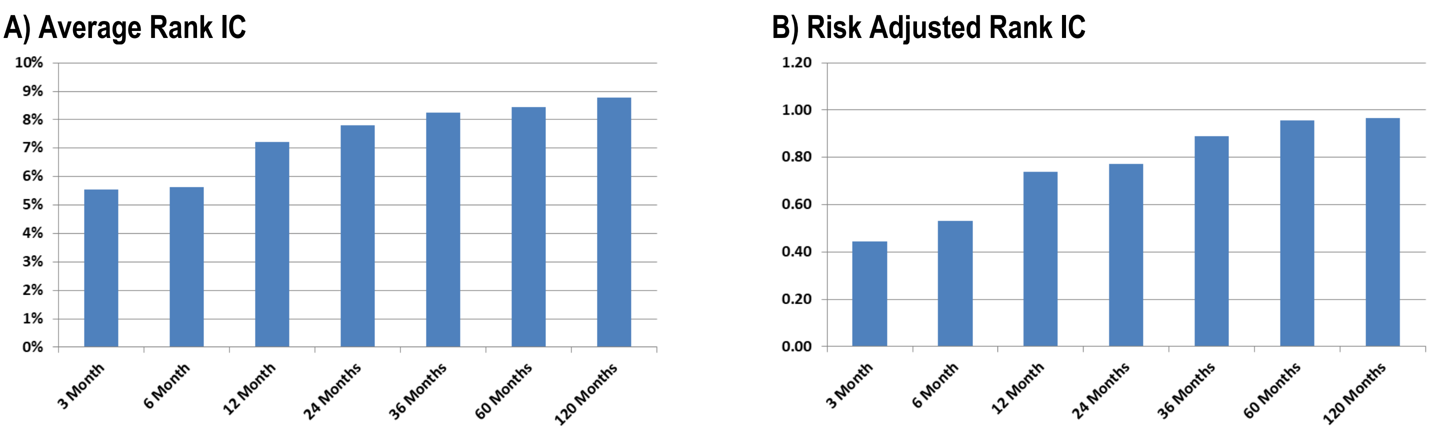
"optimal" size of rolling window (or look-back window); therefore, it is another empirical question. In our model design, the rolling window is another hyper parameter that we need to calibrate.

Increasing the size of the rolling window amplifies our sample size. On the other hand, a long training window also suffers from the following problems:

- We introduce serial correlation in data structure. Although stock returns (i.e., the dependent variable or decision variable) have almost not autocorrelation[15], many factors (e.g., price momentum, and accounting-data related) are highly correlated over time, by construction.
- As we expand our rolling window, we bring in data from distant past. As the underlying economic and market environment changes over time, the data from multiple periods ago may be irrelevant in today's market.
- As we lengthen our look-back window, our model becomes more stable and less adaptive to the changes in market environment.
- If we use a very long training window, e.g., 10 years of data, we would have much fewer alternative data-based factors in the model. Alternative data tend to have shorter history than our traditional factors.

As shown in Figure 22, initially, as we expand our rolling window, i.e., from three months to a year, out-of-sample performance improves considerably. However, after 12-month of data, incremental performance improvement plateaus.

Figure 22 The Relative Between Size of Rolling Window and Out-of-Sample Performance, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

[15] The "weak" form of EMH (Efficient Market Hypothesis) essentially assumes that there is no serial correlation in stock returns, i.e., you can't predict future stock prices on the basis of past stock prices.

MALTA – THE BUILDING BLOCKS

Most of our existing models are constructed with a fixed rolling window. The size of the look-back window is typically chosen to balance economic relevance, sample size, available factor data, and data calibration. For example, our first flagship global stock selection model – LEAP – uses a 10-year rolling window (see *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP*, Luo, et al [2017c]).

In this research, we explore the idea of combining different moving windows into a composite model, to capture the potential diversification benefit. Our MALTA (Machine Adaptive Learning Tactical Alpha) model is developed on four sets of training windows:

- A **Medium-Term Model** is trained using a rolling 12 month window
- A **Seasonal Model** is constructed using a rolling 10-year window of the same calendar months
- A **Hedge Model** is developed to provide a downside insurance to our baseline scenario, using those periods that are different from our current environment
- A **Short-Term Model** uses only the previous month data

SEASONAL MODEL

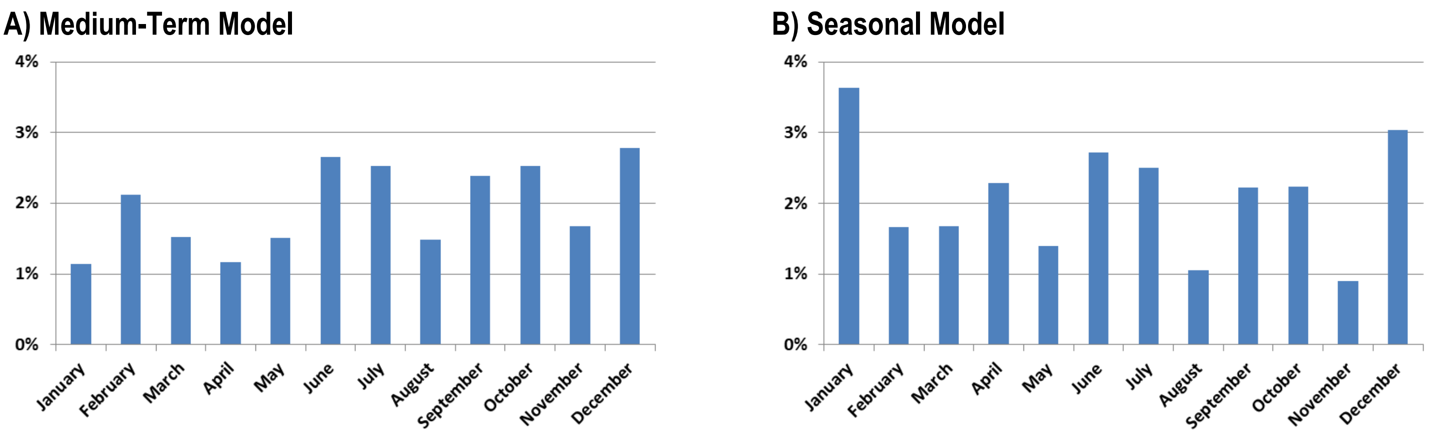
Seasonality in asset pricing has long been well documented (e.g., Rozeff and Kinney [1976], Bouman and Jacobson [2002]). We have also demonstrated how to incorporate factor return seasonality in dynamic stock selection models:

- In *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP* (see Luo, et al [2017c]), we find three strong seasonal patterns in the US and most developed markets (e.g., Canada, Europe, UK, and Australia) – the January effect, the sell in May and go away phenomenon, and the December tax loss selling and window dressing hypothesis. For example, in January, investors tend to be more optimistic about investment outlook; therefore, they are more likely to take on risky bets. As a result, small cap, high beta, and low quality factors tend to perform well in January.
- In *Multi-Dimensional Alpha: The Silk Road to China* (see Wang, et al [2017b]), we observe a different January effect for China's domestic equity market. Because the most important holiday in China is the Chinese New Year (also known as the Spring Festival), which typically falls in late January or early February, the phycological reference point is not the same Gregorian calendar January. As a result, the most important seasonality pattern in Chinese equity market is typically around the Chinese New Year or February by the solar calendar.

To fully capture the seasonal effect – not only the well-known linear seasonal patterns identified in academia and our previous research, but also any other nonlinear relationship, we introduce our Seasonal Model, which is trained using the past 10 years of monthly data, from the same calendar month. For example, at the end of December 2017, we would use the December data from the past 10 years (e.g., December 2016, December 2015, all the way to December 2007) to develop our Seasonal Model, which is then used to predict stock returns of January 2018.

As shown in Figure 23(A), our Medium-Term Model, built on a rolling 12-month data, performs well on average, but exhibits noticeable seasonal variations. In particular, the Medium-Term Model performs the best in December and worst in January. On the other hand, our Seasonal Model paints a quite different picture, with the best performance in January (see Figure 23B). The two models are complementary in many other months too. For instance, the Medium-Term model holds up well in November, while the Seasonal Model struggles.

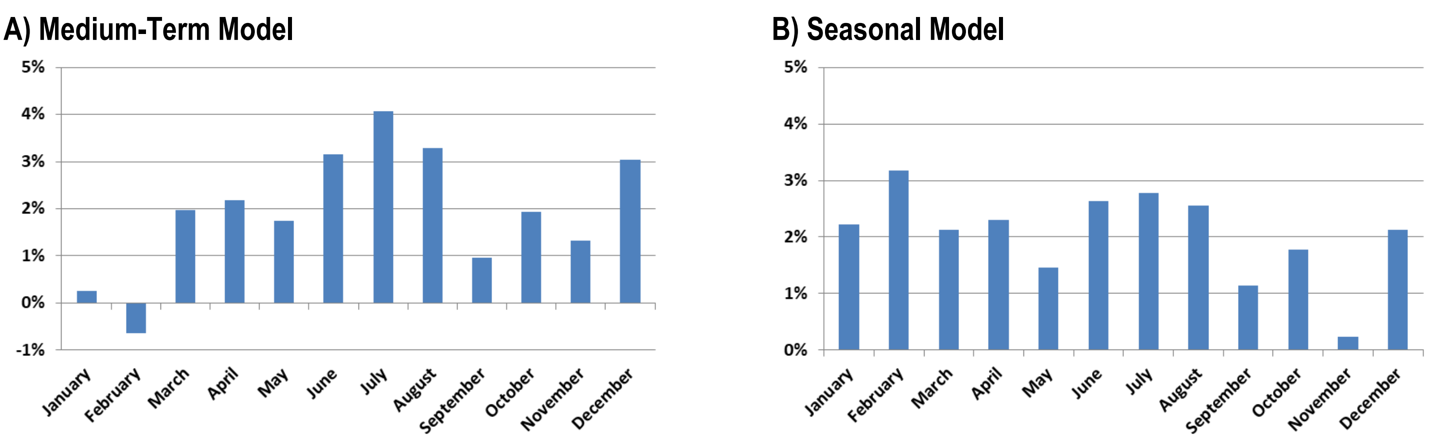
Figure 23 Average Monthly Alpha, Long/Short Quintile Portfolio, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

In Asia ex Japan, we observe not only strong January effect, but also the Chinese New Year phenomenon in February. As shown in Figure 24(A), our Medium-Term model struggles in both January and February, while our Seasonal Model delivers exceptional performance (see Figure 24B).

Figure 24 Average Monthly Alpha, Long/Short Quintile Portfolio, Asia ex Japan

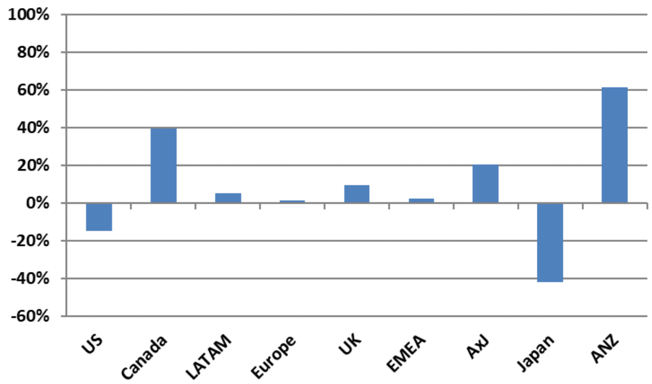


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

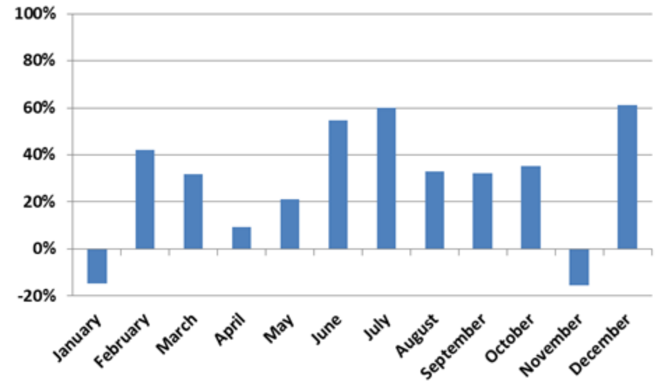
Not only just in the US and Asia, in all nine regions of the world, the Medium-Term and Seasonal models are either uncorrelated or only modestly correlated in January (see Figure 25A). In the US and Japan, the correlation is even negative. In addition to January, we also see sizable diversification benefit/low correlation in other months (see Figure 25B).

Figure 25 Correlation between the Medium-Term Model and the Seasonal Model

A) Correlation in January



B) Monthly Correlation, US

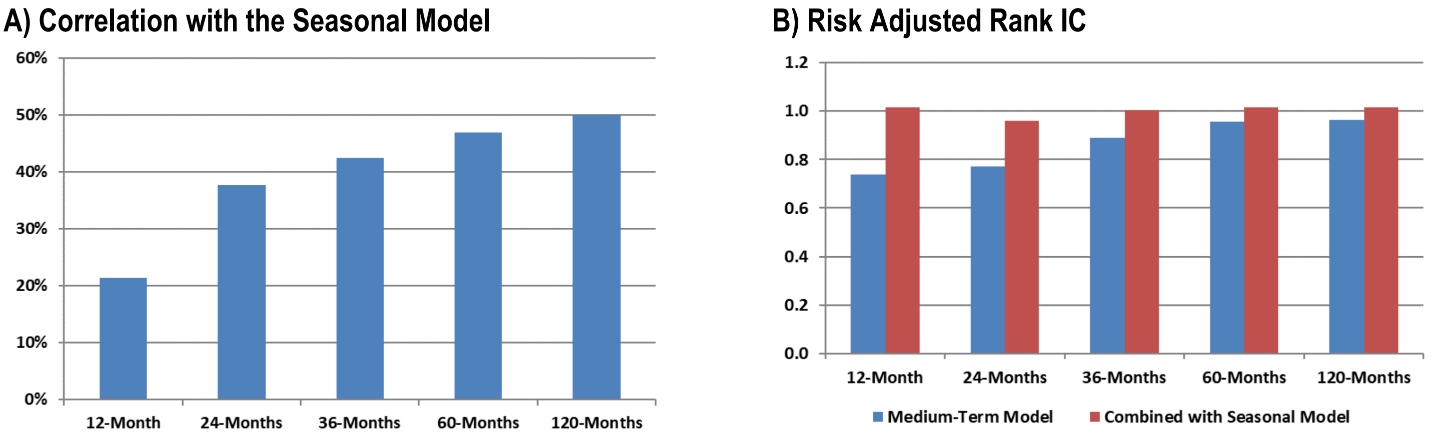


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As shown earlier in Figure 22, as we expand our trailing training window, out-of-sample performance improves, albeit at a slower speed. On the other hand, as the size of the rolling window increases, the overlap between the Medium-Term and Seasonal models also grows. For example, a 12-month Medium-Term model has only one common month with the Seasonal model, while a 120-month Medium-Term model would include all 10 months used in the Seasonal model. Therefore, the correlation between the Seasonal model and Medium-Term model increases, as we broaden the rolling window (see Figure 26A). Our current Medium-Term model, which uses a 12-month rolling window, is only 20% correlated to the Seasonal model. If we bolster the training window to 120 months, the correlation between the two models spikes to 50%.

As shown in Figure 26(B), when we blend the Medium-Term and Seasonal models together, the combined model is robust to the choice of rolling window for the Medium-Term model. As we expand the training window, the performance of the Medium-Term model increases, but the correlation with the Seasonal model also swells. Setting a shorter rolling window (e.g., 12 months) also meaningfully reduces the computational need. Therefore, we set the size of the rolling window to be 12 months for our Medium-Term model throughout the rest of the research. The combined Medium-Term and Seasonal model is called the Baseline model.

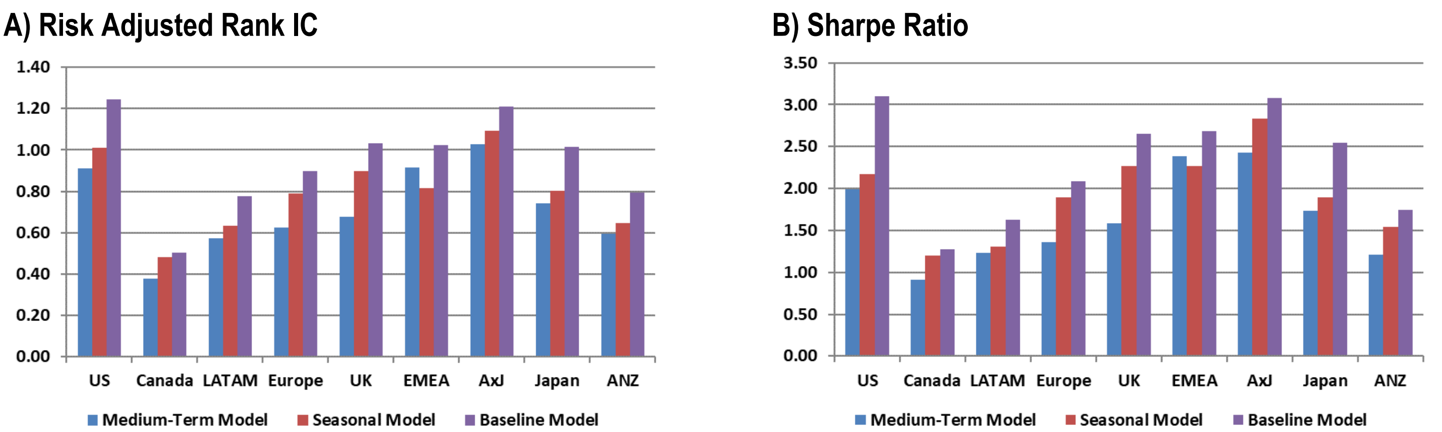
Figure 26 Combining the Seasonality Model with the Medium-Term Model, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Due to diversification benefit, combining the Medium-Term model and the Seasonal models boosts performance substantially in all nine regions (see Figure 27).

Figure 27 Combining the Medium-Term and Seasonal Models



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

HEDGE MODEL

One common criticism about quantitative models is that all models are backward looking; therefore, they almost always miss the turning point. We would argue that not only just quantitative models, but also essentially almost all human knowledge is based on historical experience and data. We either explicitly (using models) or implicitly (using heuristics) learn from our experience and make adjustment to our reasoning process. As a result, most investors are likely to be wrong at the market turning point[16].

[16] There are certainly exceptions. For example, at every market crisis, there were commentators and investors claiming that they had anticipated such event well ahead of time. In reality, market turning points – by definition – are rare event. We probably never have

Given that almost all models are trained on historical data to optimize average fit, they are most likely to be wrong when the market environment does change. In *Multi-Dimensional Alpha: Style Rotation, Machine Learning, and the Quantum LEAP* (see Luo, et al [2017c]), we discussed the first approach to tackle regime shifts. In the LEAP global stock selection model, we propose to use exogenous macro variables to time factor performance. We do find style rotation (also known as factor timing) can add diversification benefit and reduce model drawdown.

In this research, we take a rather different approach. We argue that market environment changes are probably unobservable or at least difficult to identify in real time. Instead of predicting a regime shift, we would like to buy an "insurance" or a "hedge". Essentially, we want to add a Hedge model to our Baseline model (equally weighting the Medium-Term and Seasonal models) introduced in the previous section. The "insurance" policy pays off when we have a catastrophic event – in this case, when our Baseline model underperforms. Adding a Hedge model to the mix might marginally reduce average excess return, because we pay for an insurance premium. However, it should lower our strategy volatility and especially downside risk even more. Therefore, we expect our risk-adjusted performance to improve.

The Hedge model is constructed as follows:

- At each month end, we backtest the performance of our Baseline model (i.e., the combined Medium-Term and Seasonal models) for the previous 120 months; and
- The bottom half of our sample, i.e., the 60 months when the Baseline model's rank IC is behind the median, is used to train the Hedge model.

Accordingly, we essentially treat the bottom half of the past 10 years as an unobserved different regime; develop a Hedge model using data from that regime; and then combine the two models together to capture the diversification benefit. We would certainly define the hedge period using other percentiles (e.g., bottom 10% months). We take the bottom half for the following reason:

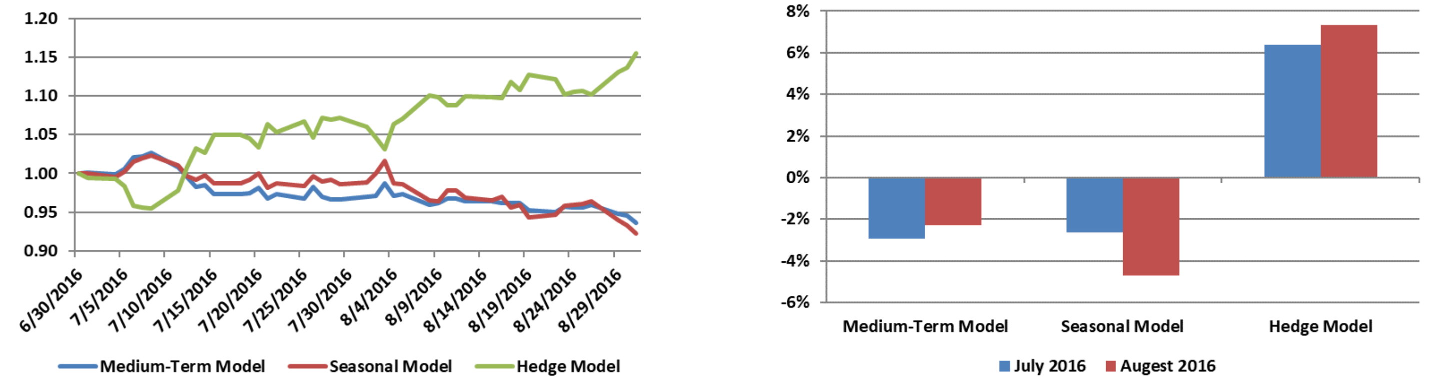
- We need a large enough sample to build our Hedge model; and
- Using a large enough sample also ensures that our Hedge model has reasonable performance on its own, i.e., we want to make sure the insurance premium to be as low as possible.

As an example, in the summer of 2016, both Medium-Term and Seasonal models (and therefore, the Baseline model) suffered from a modest drawdown of -5% in Japan (see Figure 28A). On the other hand, the Hedge model was almost a mirror image, up by 15% during the same period. As shown in Figure 28(B), the Medium-Term and Seasonal models both had negative IC in July and August 2016, while the Hedge model demonstrated superior predictive power for both months.

enough data to prove whether these commentators/investors are really skillful or just lucky. More importantly, if they were truly able to identify the turning events well ahead of time, they would have been wrong most of the time leading to the events. Therefore, even if they did make a large profit in the end, they would probably have a poor risk adjusted performance and a very low hit rate (i.e., losing money most of the time) – highly undesirable from an institutional investor's viewpoint.

Figure 28 Long/Short Quintile Portfolio, Japan, Summer 2016

A) Perfect Hedge by the Hedge Model **B) Rank IC**



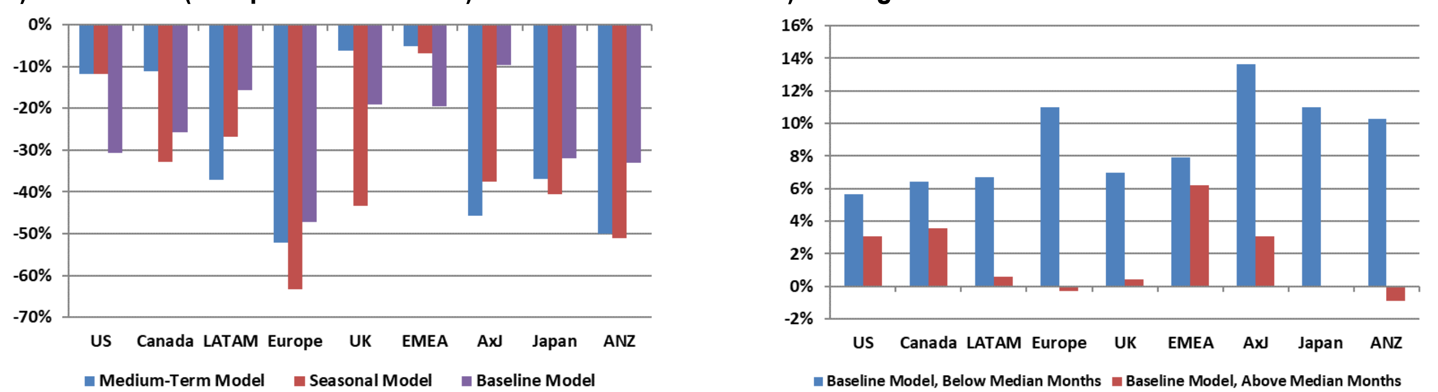
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As expected, the Hedge model is negatively correlated to the Medium-Term, Seasonal, and Baseline models in all nine regions (see Figure 29A). As shown in Figure 29(B), the Hedge model really shines when the Baseline model struggles[17]. Conversely, during the periods when the Baseline model delivers strong performance[18], the Hedge model is either flat or up marginally.

A traditional insurance policy requires a small premium during normal times and covers the losses in stressful times. Our Hedge model delivers modest or slightly negative performance when our Baseline model works well, while generates strong performance when our Baseline model struggles. To capture the diversification benefit from the Hedge model, we now introduce our Enhanced Model, which equally weights the Medium-Term, Seasonal, and Hedge models together.

Figure 29 Diversification Benefit by the Hedge Model

A) Correlation (Computed on Rank IC) **B) Average Rank IC**



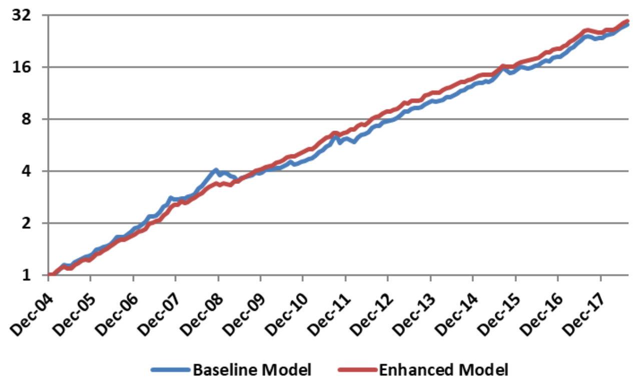
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

17 Defined as those periods when the Baseline model's IC is below the median.
18 Defined as those periods when the Baseline model's IC is above the median.

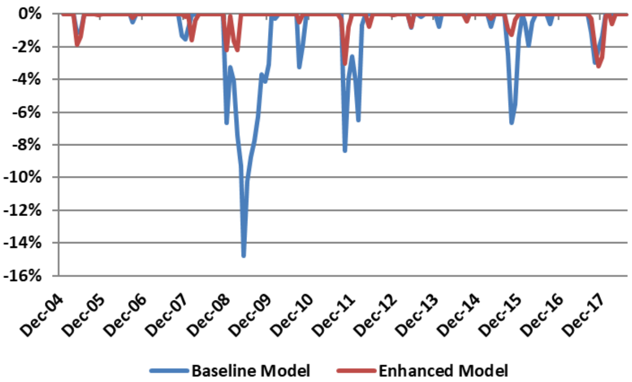
As shown in Figure 30(A), the Enhanced and Baseline models deliver similar total performance (as measured by a long/short quintile portfolio) in Asia ex Japan in the past 14 years. However, the Enhanced model reduces the maximum drawdown considerably (see Figure 30B). In particular, the downside risk suffered in March-May 2009 was scaled down from -15% to -2%.

Figure 30 Asia ex Japan Performance for Enhanced Model and Hedged Enhanced Model

A) Cumulative Performance



B) Maximum Drawdown

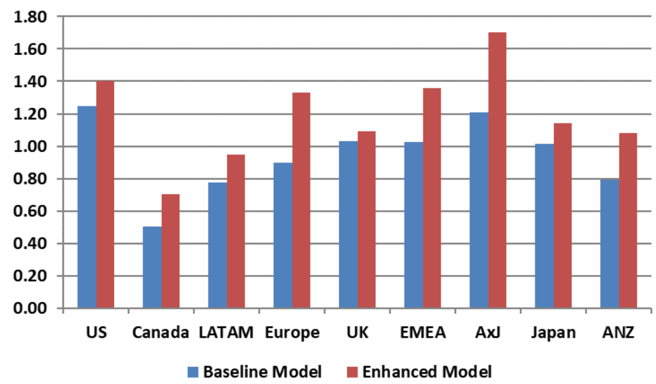


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

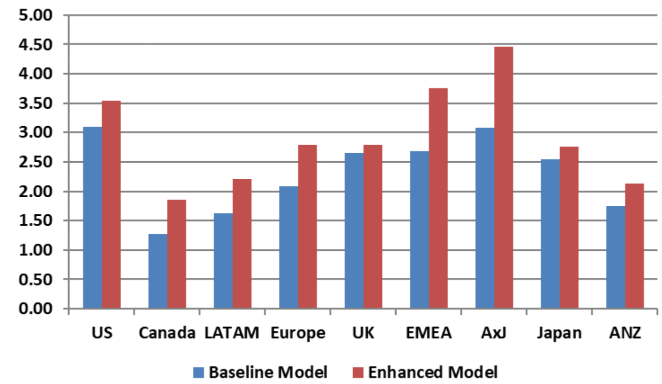
As shown in Figure 31(A) and (B), adding the Hedge model to the Enhanced strategy improves our predictive power and performance significantly across the nine regions of the world. More importantly, due to the diversification benefit offered by the Hedge model, the combined model also reduces volatility (see Figure 31C) and downside risk (see Figure 31D) considerably.

Figure 31 Performance, Baseline Model versus Enhanced Model

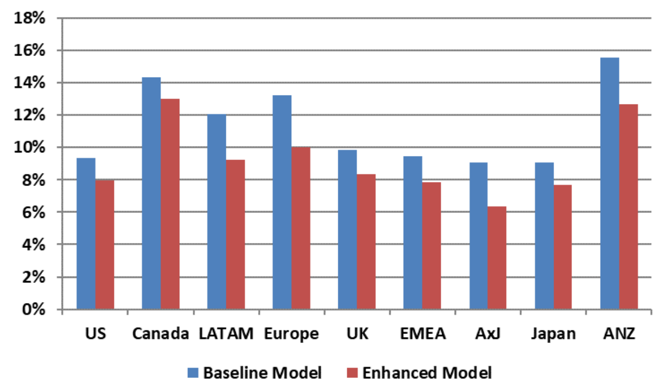
A) Risk Adjusted Rank IC



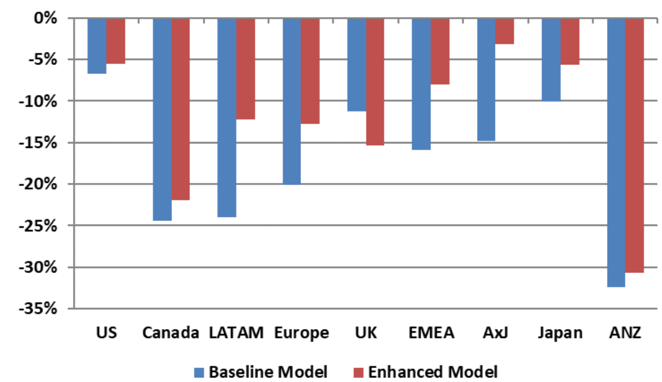
B) Sharpe Ratio, Long/Short Quintile Portfolio



C) Annualized Volatility, Quintile Portfolio



B) Maximum Drawdown, Quintile Portfolio



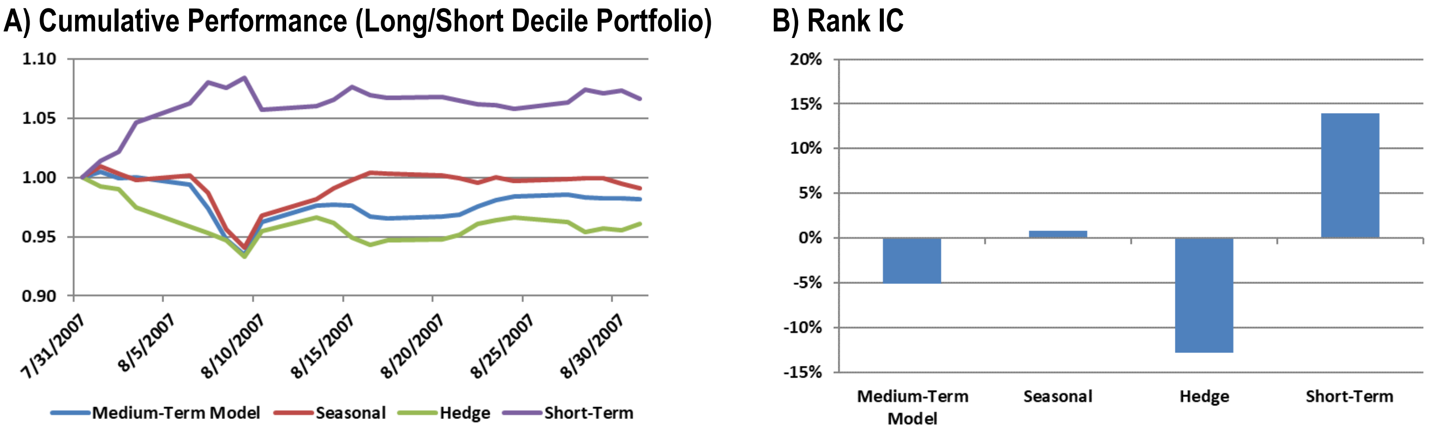
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

SHORT-TERM MODEL

Given that most machine learning algorithms are designed primarily for cross-sectional data, we need to explicitly account for the time series dimension. Each of the three components behind the Enhanced model offers some time aspect. The Medium-Term model strikes a balance between sample size and time relevance. The Seasonal model captures the long-term seasonal patterns, while the Hedge model takes advantage of the observations from an unobserved different regime. In this section, we introduce a Short-Term model, which is trained only using the last month's data. The Short-Term model can adapt to rapid changes in market environment.

There are a few cases when none of the three sub-components behind the Enhanced model performs well. One such extreme example is the quant melt down in August 2007, when all three models fell by more than -5% in the first two weeks (see Figure 32A). Interestingly, Short-Term model generated strong performance during this period, up by 8%. As shown in Figure 32(B), the Medium-Term and Hedge models had negative predictive power, while the Seasonal model was flat during August 2007. On the other hand, the Short-Term model had a monthly IC of nearly 15, greatly offsetting the losses from the other three models.

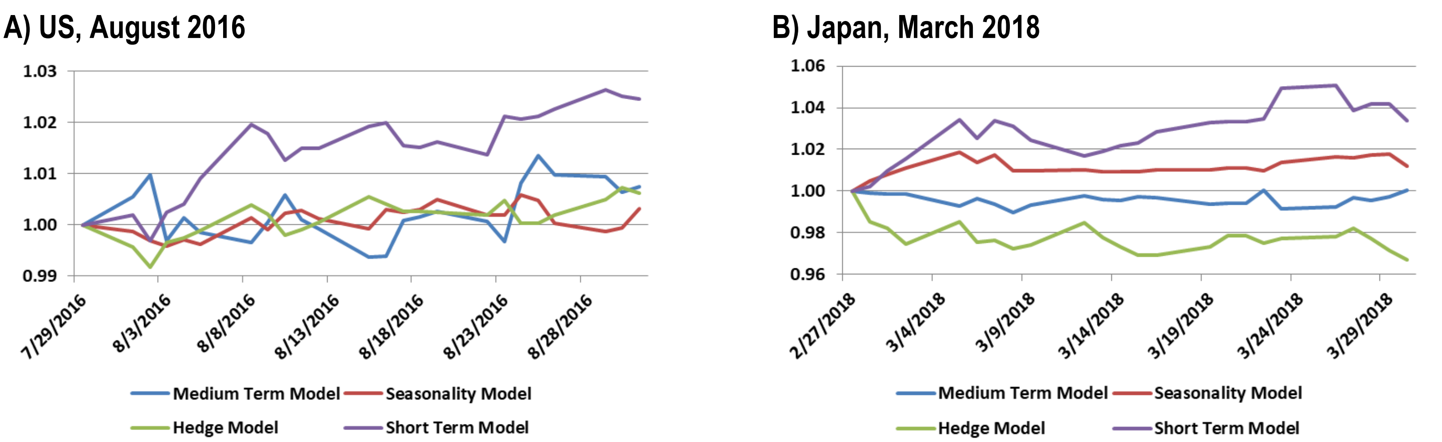
Figure 32 A Case Study, Quant Meltdown, August 2007, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

More recently, in August 2016, while the three models were flat in the US, the Short-Term model delivered an alpha of 3% (as shown in Figure 33A). Similarly, in March 2018, when the Seasonal and Medium-Term models were flat, the Hedge model plunged in Japan. During the same period, the Short-Term model came to the rescue (see Figure 33B).

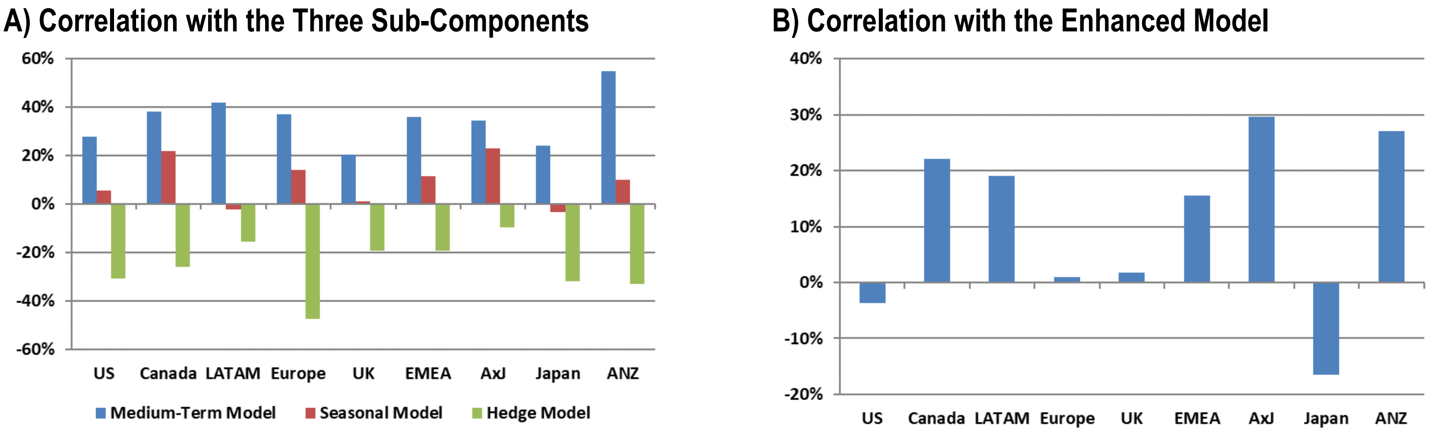
Figure 33 Two Recent Examples, Long/Short Decile Portfolio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As shown in Figure 34A, the Short-Term model is positively correlated to the Medium-Term and Seasonal components, but negatively correlated to the Hedge model. As a result, the correlation between the Short-Term model and the Enhanced model is fairly low across all regions (see Figure 34B).

Figure 34 Short-Term Model's Correlation with Other Models



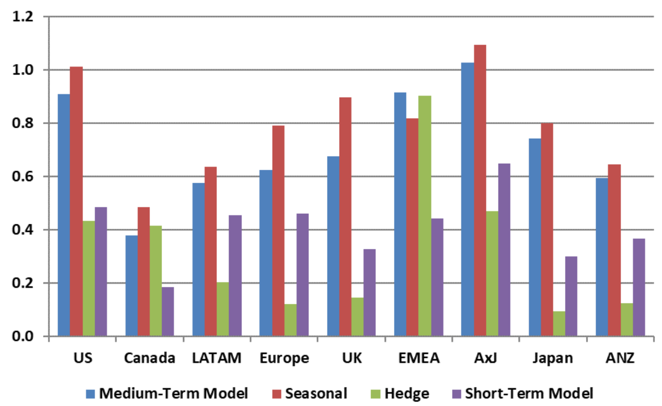
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

WELCOME TO MALTA

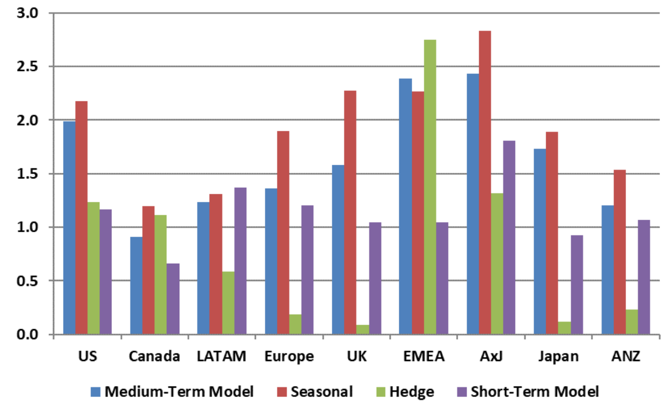
As shown in Figure 35(A) and (B), all our sub-models deliver strong performance across the nine regions of the world. On average, the Seasonal and the Medium-Term perform the best. However, as shown in Figure 35(C) and (D), the Hedge and Short-Term models also provide great diversification benefit.

Figure 35 Performance, Each of Sub-Model

A) Risk Adjusted Rank IC

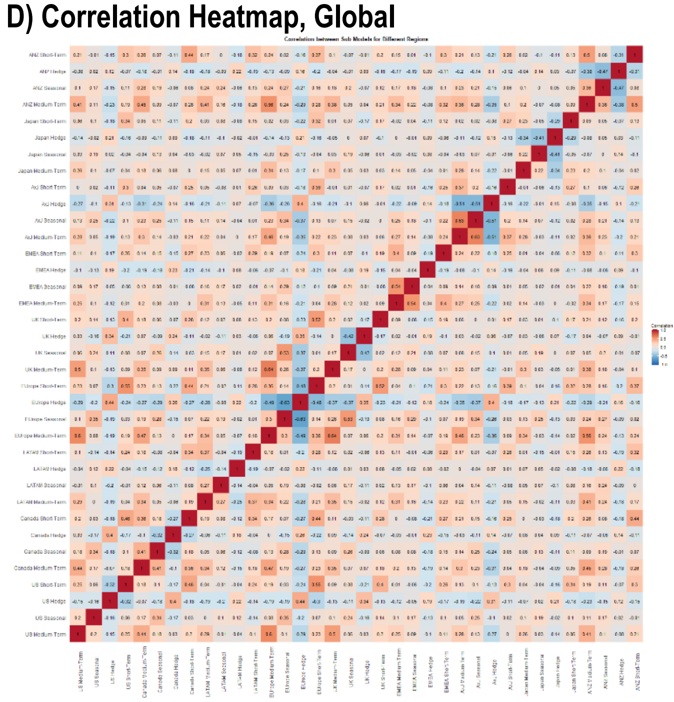


B) Sharpe Ratio, Quintile Portfolio



C) Correlation Heatmap, US

D) Correlation Heatmap, Global



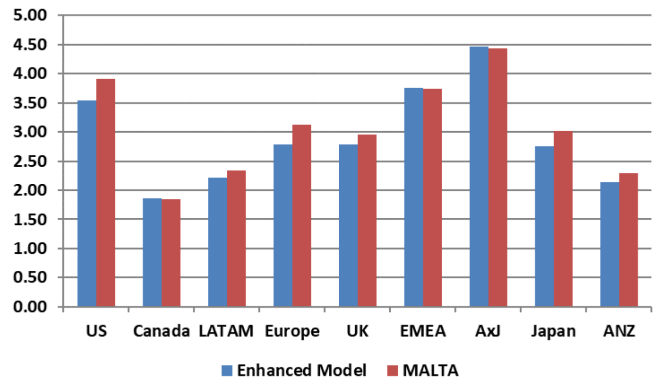
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Our final MALTA (Machine Adaptive Learning Tactical Alpha) model is a combination of the above mentioned four sub-components. Similar to our global macro models (see *Current Affairs: Systematic Global Macro Investing – Launching our Global Macro Forecasts*, Luo, et al [2017f]), the MALTA model takes advantage of the Bayesian model averaging philosophy.

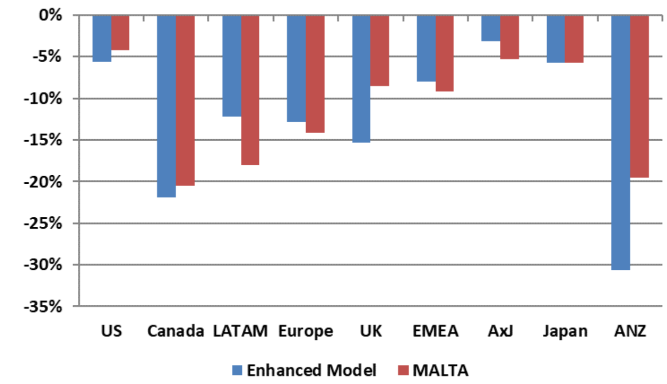
In particular, adding the Short-Term model to the Enhanced model boosts performance by 5%-15% (see Figure 36A) and lowers downside risk by 10%-45% for most of the developed markets (see Figure 36B).

Figure 36 MALTA Model Performance

A) Sharpe Ratio, Long/Short Quintile Portfolio



B) Maximum Drawdown, Long/Short Quintile Portfolio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

MALTA Global Stock Selection Model

In addition to our LEAP global stock selection model, the MALTA model offers another attractive alternative. The LEAP model is primarily a linear multifactor model, with a nonlinear component and a macro factor timing overlay. Conversely, the MALTA model is a suite of cutting-edge nonlinear machine learning algorithms. As we will demonstrate later, the two models are uncorrelated and complementary.

Global Coverage Universe

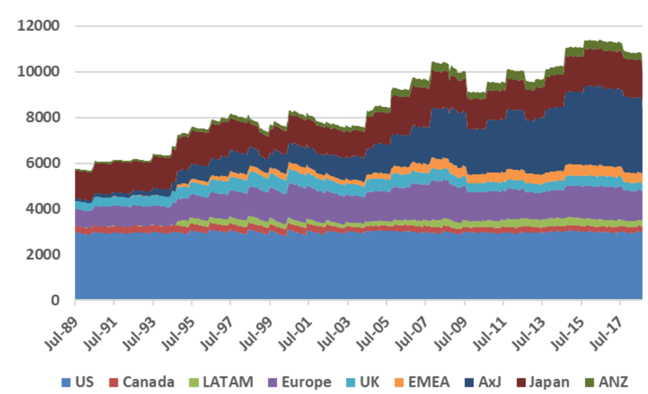
Similar to LEAP, we also have nine regional MALTA models:

- US (Russell 3000 universe)
- Canada (S&P/TSX Composite universe)
- Europe ex UK (S&P BMI universe)
- UK (S&P BMI universe)
- Asia ex Japan (S&P BMI universe)
- Japan (S&P BMI universe)
- Australia and New Zealand (S&P BMI universe)
- LATAM (S&P BMI universe)
- Emerging EMEA (S&P BMI universe)

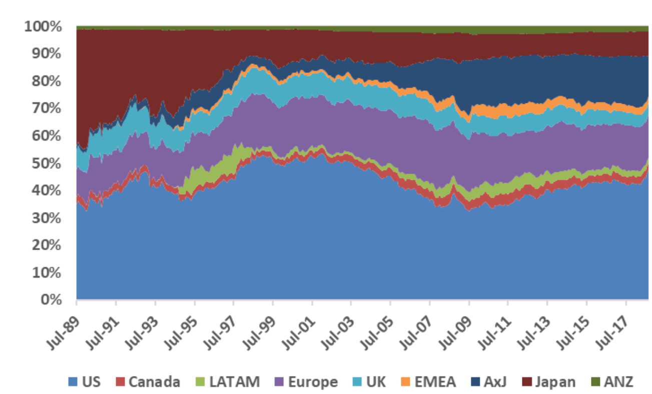
As shown in Figure 37 (A), US, Europe, AxJ, and Japan are the four largest regions by number of stocks, while US market counts for half of the global equity market by market capitalization (see Figure 37 B).

Figure 37 The Global Investment Universe, by Geographic Regions

A) # of Stocks, by Region



B) % of Market Capitalization, by Region



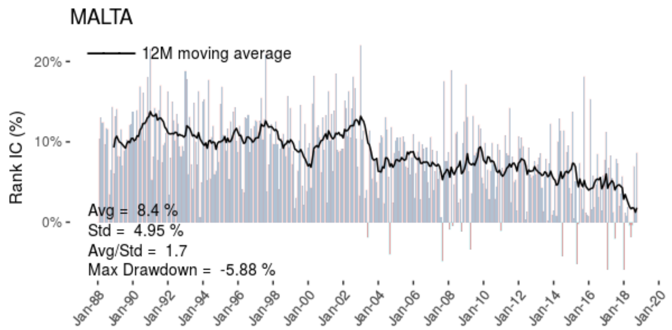
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Consistent Performance Over Time

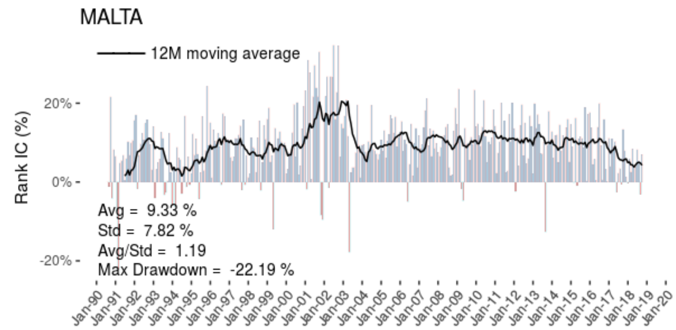
The MALTA model delivers consistent performance in all regions (see Figure 38).

Figure 38 MALTA Performance, Rank IC

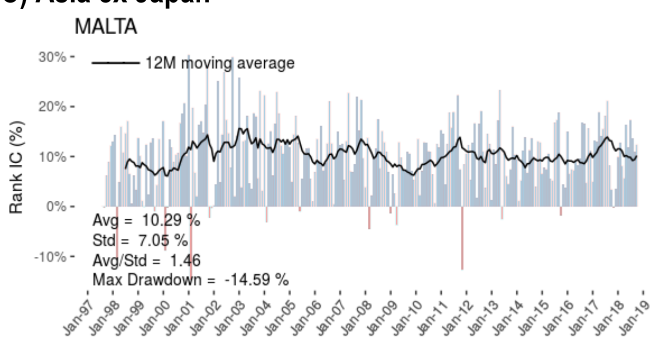
A) US



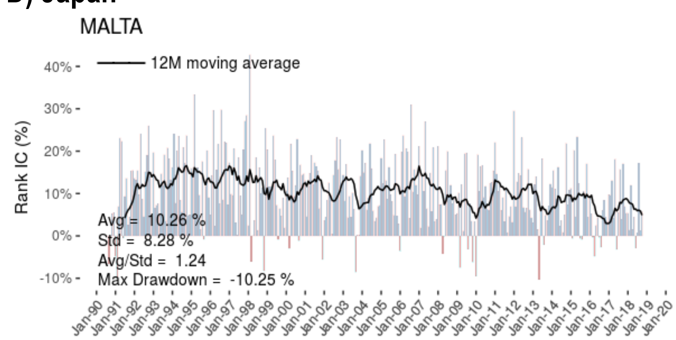
B) Europe



C) Asia ex Japan



D) Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

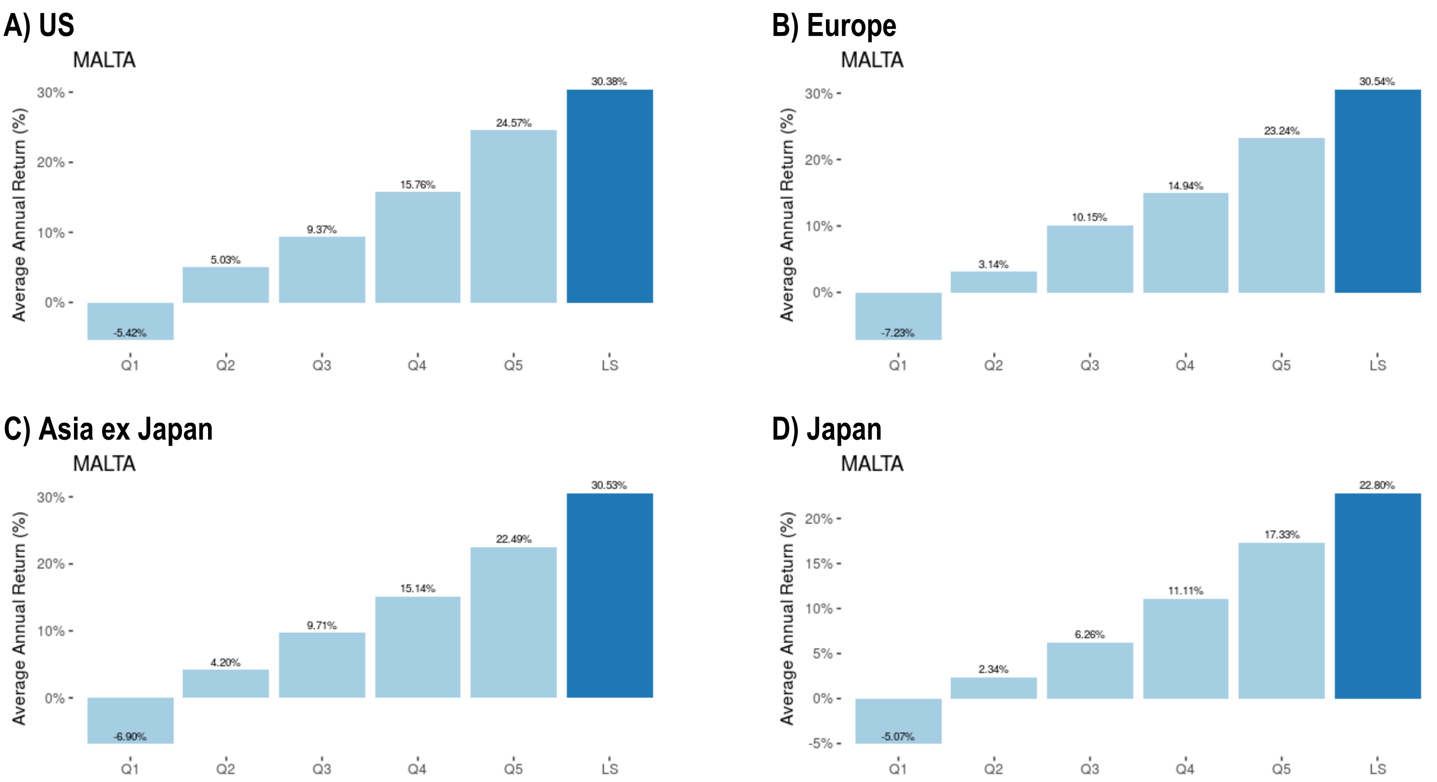
Effective on both the Long and the Short Sides

Many quantitative models suffer from asymmetric and nonlinear payoff patterns. For example, it is generally easier to find alpha on the short side, because:

- Many investors are long only and do not short. Therefore, the short side is relatively inefficient compared to the long side.
- There are incremental cost and compliance burden to short stocks, which also creates market inefficiencies, i.e., limited arbitrage.

Our MALTA model is highly effective on both the long and the short sides for all regions, with a relatively symmetric pattern (see Figure 39). In particular, the bottom quintile portfolio (i.e., the stocks to avoid or short) for all regions has a negative return, despite the overall market generally goes up in the long term.

Figure 39 MALTA Quintile Returns For Different Regions

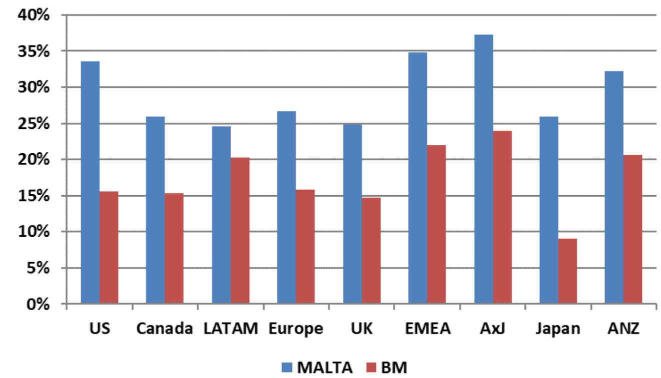


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

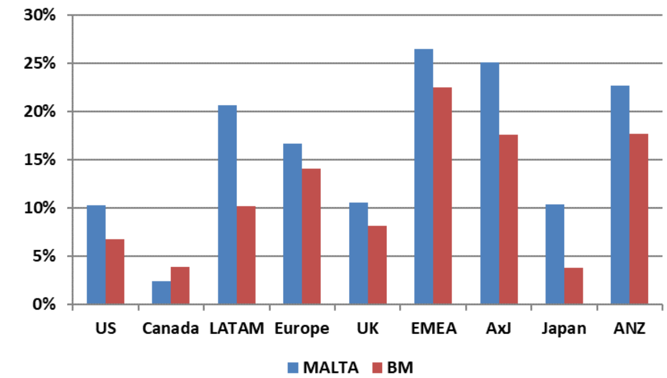
Compared to a conventional multifactor model – BM, the MALTA model produces much higher returns (see Figure 40A), especially in recent years (see Figure 40B). The Sharpe ratio of the MALTA is considerably better than the benchmark model, especially in the more challenging markets such as the US and Japan (see Figure 40C), which is consistent with the evidence based on risk-adjusted rank IC (see Figure 40D). Furthermore, the MALTA model is more likely to stay positive than the BM (see Figure 40E), with a substantial lower downside risk (see Figure 40F).

Figure 40 MALTA Outperforms the BM Significantly

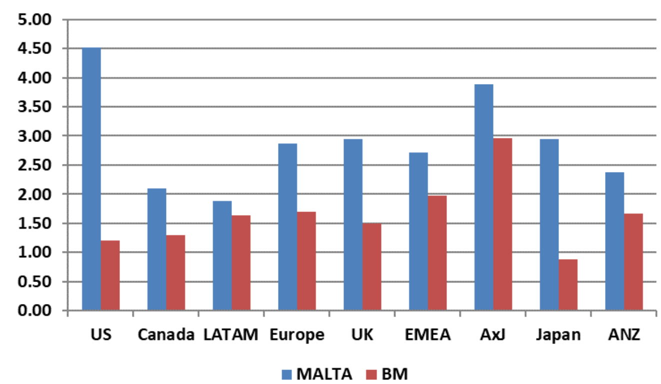
A) Annual Return (L/S Quintile Portfolio), Long Term



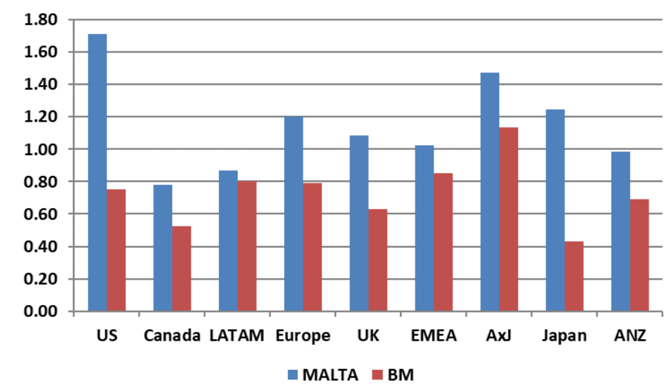
B) Annual Return (L/S Portfolio), 2015-Present



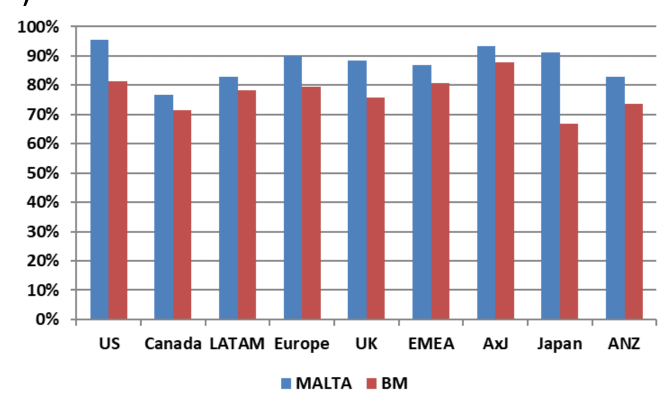
C) Sharpe Ratio (L/S Quintile Portfolio)



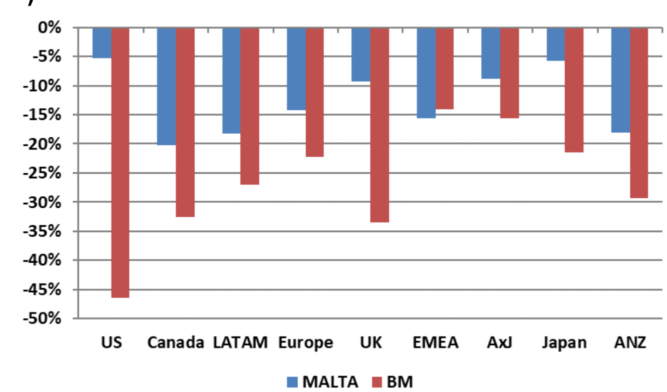
D) Risk Adjusted Rank IC



E) Hit Rate



F) Max Drawdown



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

MALTA AND LEAP ARE HIGHLY COMPLEMENTARY

Given that we have two competing global stock selection models – LEAP and MALTA, a natural question that investors might ask is which one to choose. Can one model replace the other? The short answer is no. The two models have a rather different design philosophy. As a result, we would argue that the two models are complementary and investors can use them both.

- **Linear versus Nonlinear.** The LEAP model is primarily built on a set of linear machine learning algorithms, with a minor nonlinear overlay. On the other hand, the MALTA is a complete nonlinear model.
- **Model Transparency.** Both models are fully transparent. However, given the LEAP's linear structure, it might be easier for investors to comprehend.
- **Regime Change.** The LEAP model uses exogenous macro variables to conduct explicit factor timing. On the one hand, it does take advantage of macro information. On the other hand, we also acknowledge the difficulties of macro timing. The MALTA model is constructed on a Bayesian model average approach, accounting for the time series dimension using four different training windows. Therefore, the MALTA model does not attempt to predict factor returns; rather it strives to have robust performance across different economic regimes.

A SIMPLE HORSERACE

As shown in Figure 41(A), both LEAP and MALTA models outperform the benchmark model BM considerably. Furthermore, MALTA model outperforms the LEAP in the US, Canada, LATAM, UK, EMEA, Japan, and ANZ, but the LEAP wins in Europe and Asia ex Japan. Furthermore, both LEAP and MALTA have much lower downside risk than the BM (Figure 41B).

Figure 41 A Horserace between LEAP and MALTA

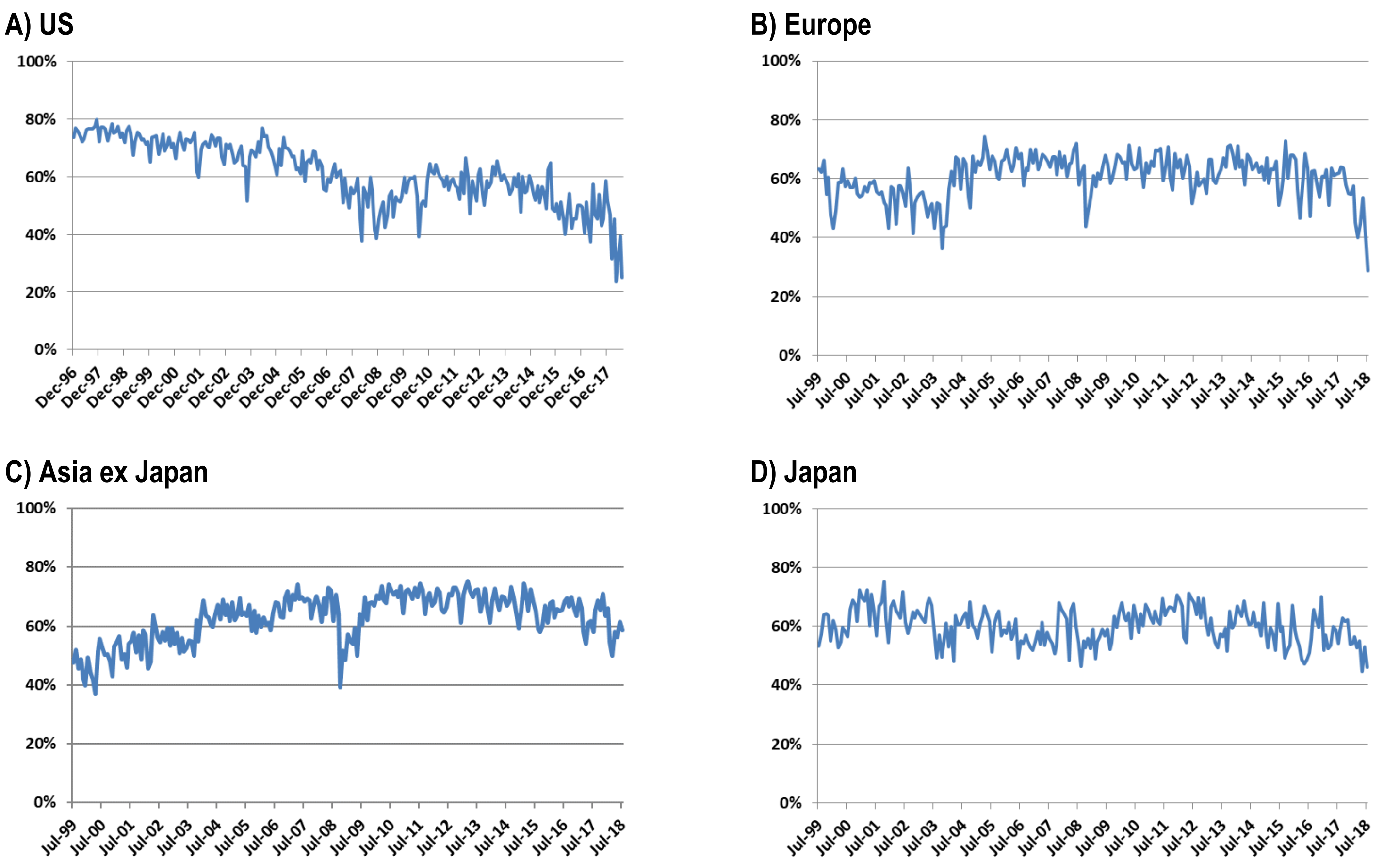
A) Risk Adjusted Rank IC **B) Maximum Drawdown**

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

CORRELATION BETWEEN LEAP AND MALTA

Figure 42 shows the cross-sectional signal correlation between LEAP and MALTA. Although the average correlation is around 50% for most regions, the two models have become increasingly decoupled in recent years.

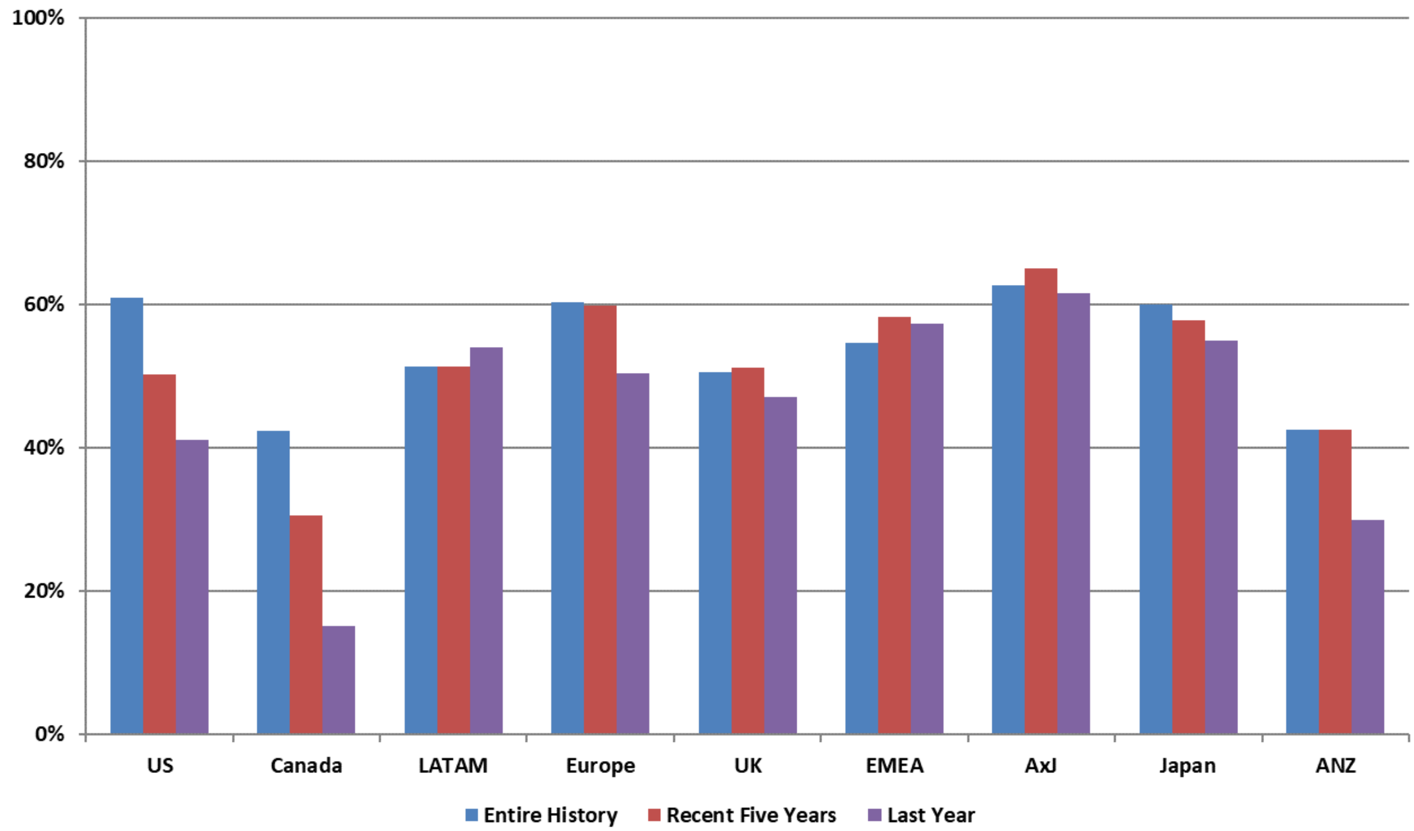
Figure 42 Cross-Sectional Signal Correlation between LEAP and MALTA



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Similarly, as shown in Figure 43, the long-term average correlation between LEAP and MALTA is around 50% for most regions. The correlation has dropped significantly in recent years, especially in developed markets. As quantitative investing has re-gained attractions in recent years, more conventional linear models face more competitions. On the other hand, completely nonlinear machine learning models such as the MALTA offers a fresh and different view.

Figure 43 Average Correlation between LEAP and MALTA



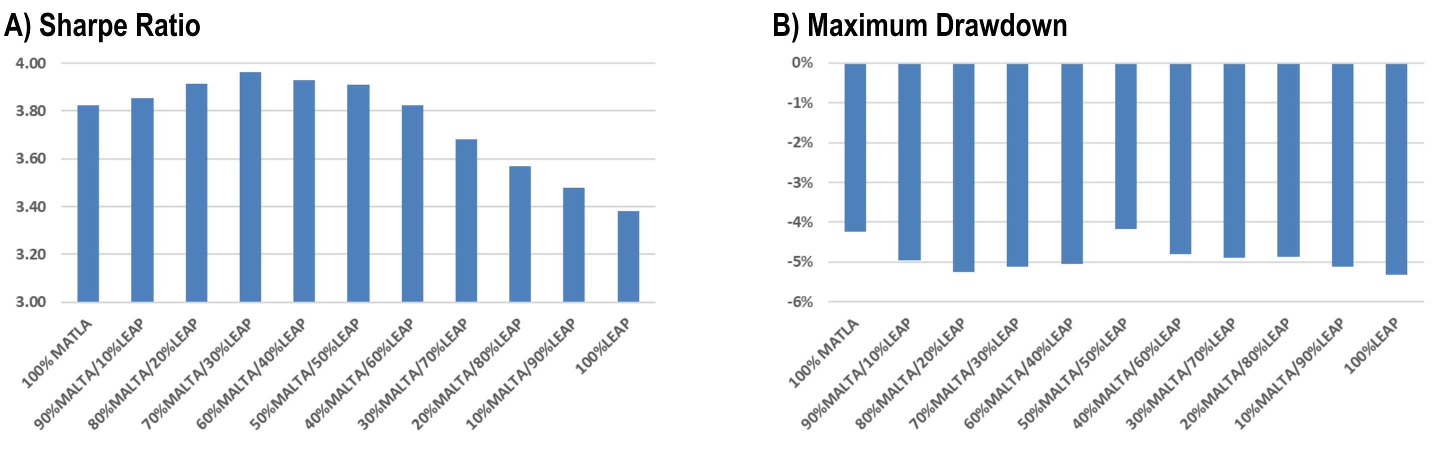
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

COMBINING MALTA WITH LEAP

Given that the MALTA and LEAP models are based on completely different machine learning algorithms and development philosophies, the two models are uncorrelated most of the time in all regions. We expect further diversification benefit by combining the two models together.

As shown in Figure 44(A), as we blend in the MALTA and LEAP together, the performance (as measured by Sharpe ratio) of combined model improves. In the US, the "optimal" weights appear to be 70% MALTA and 30% LEAP, with the highest Sharpe ratio and similar drawdown (see Figure 44B).

Figure 44 Combining the MALTA and LEAP Models, US



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

There are many sophisticated approaches to combine multiple models (see *Multi-Dimensional Alpha: Signal Research and Multifactor Models*, Luo, et al [2017b] for a comprehensive review). As a demonstration, we use the Grinold & Kahn weighting scheme [see Grinold and Kahn [1999] and Qian, Hua, and Sorensen [2007]), which is essentially a mean-variance optimization on the factor space. Mathematically, it is to maximize the expected IR (Information Ratio):

$$argmax_{\omega} IR = \frac{\omega' \widetilde{IC}}{\sqrt{\omega' \Sigma_{IC} \omega}}$$

Where,

ω is a $(K \times 1)$ vector of factor weights;

$\widetilde{IC}$ is a $(K \times 1)$ vector of expected factor IC; and

Σ_{IC} is a $(K \times K)$ covariance matrix of factor IC.

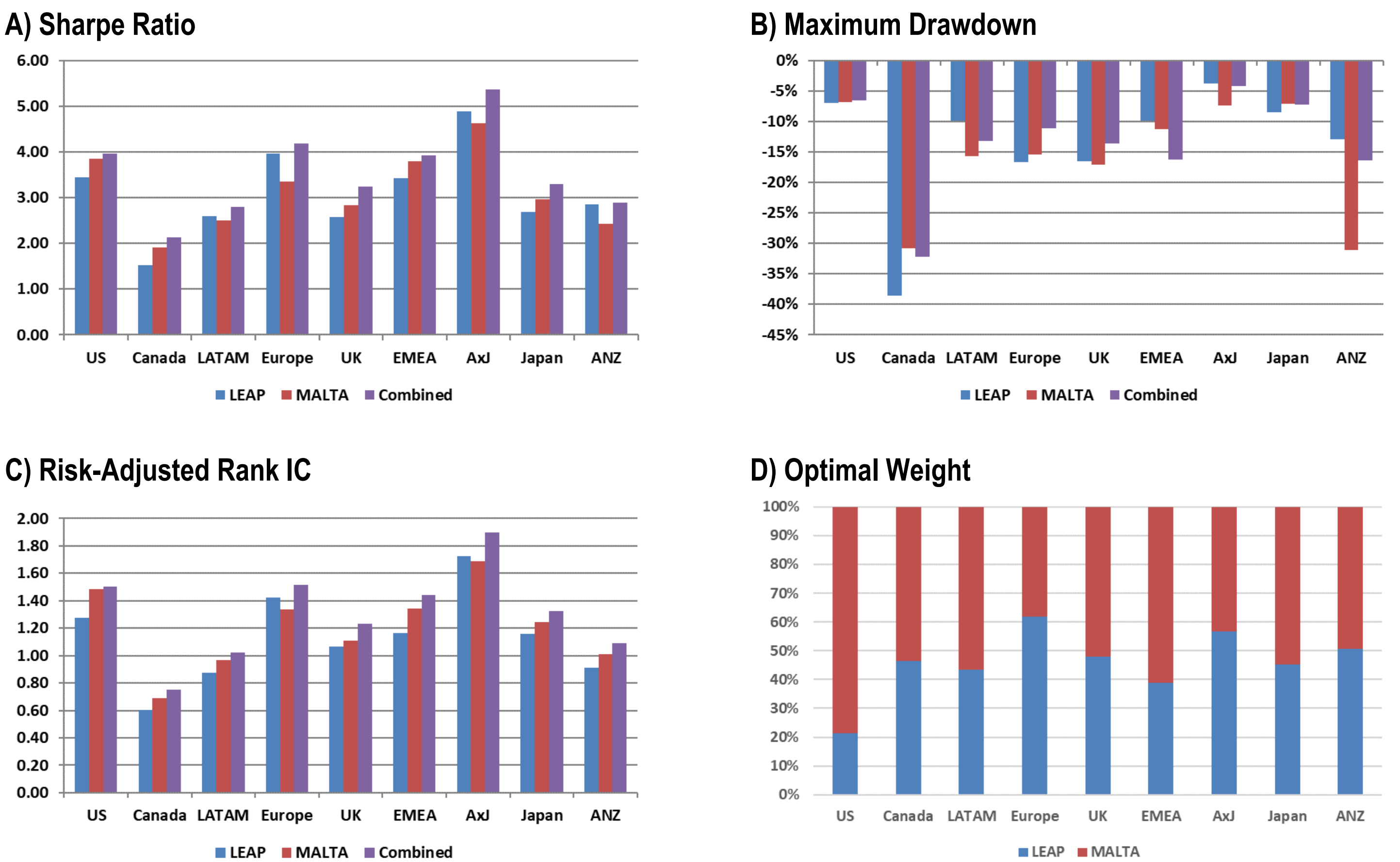
The above optimization problem has a closed-form solution, if we do not have any constraints:

$$\widehat{\omega} = \Sigma_{IC}^{-1} \widetilde{IC}$$

$\widehat{\omega}$ can be further normalized as $\widetilde{\omega} = \frac{\widehat{\omega}}{i' \cdot \widehat{\omega}}$, where i is a $(K \times 1)$ vector of 1's; then the weights add up to one.

As shown in Figure 45(A) and (C), the combined model boosts performance considerably from the two underlying MALTA and LEAP models, in all nine regions. Similarly, the combined model is also able to reduce downside risk in most regions (see Figure 45B). Lastly, in the more competitive markets (e.g., US and Japan), the combined strategy allocates more weights toward an unconventional nonlinear MALTA model, while in the markets where trading cost tends to be high (e.g., AxJ) or the factor tend to be more linear (e.g., Europe), LEAP model is more likely to dominate (see Figure 45D).

Figure 45 Combined MALTA/LEAP, Global



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

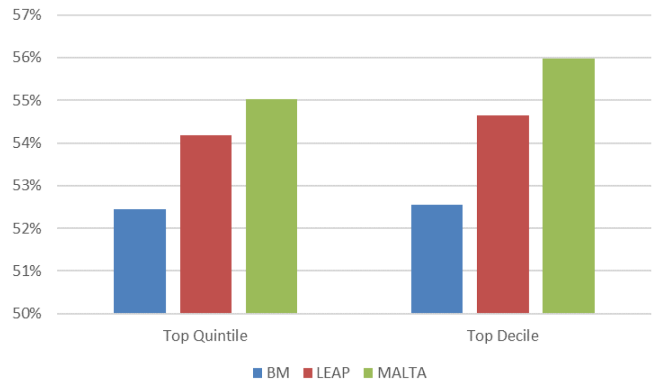
MALTA IS PERFECTLY DESIGNED FOR FUNDAMENTAL MANAGERS

As we explained earlier, we would expect a classification-based model (e.g., MBBT) to be different from and more robust than a regression-based model (e.g., LEAP). Furthermore, for fundamental investors who typically hold more concentrated portfolios than quantitative managers, the hit rate is more relevant. We measure hit rate as the percentage of stocks in our long (short) portfolio that outperform (underperform) the median of our investment universe.

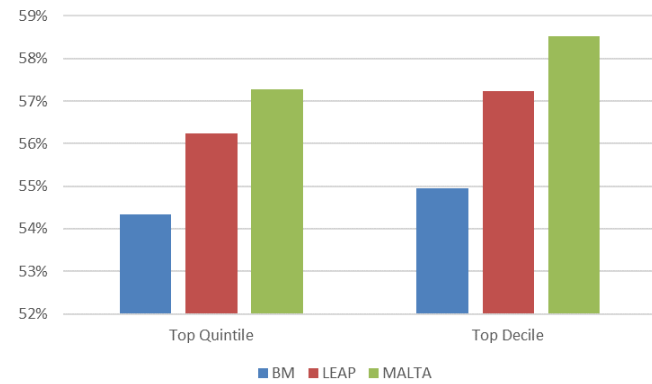
As shown in Figure 46(A), first of all, the MALTA model has a significantly higher hit rate than the LEAP, which beats the BM in the US equity market. Moreover, as we increase our conviction level from the top quintile (20%) to the top decile (10%), i.e., as our portfolio becomes more concentrated, the hit rate of all three models also rises. However, the MALTA model sees a substantially higher boost in hit rate than the BM and LEAP. We observe a similar pattern in Europe (see Figure 46B) and Asia (see Figure 46C). Globally, the MALTA model shows a considerably higher hit rate than the LEAP and BM in all nine regions.

Figure 46 Hit Rate for the Long Portfolio

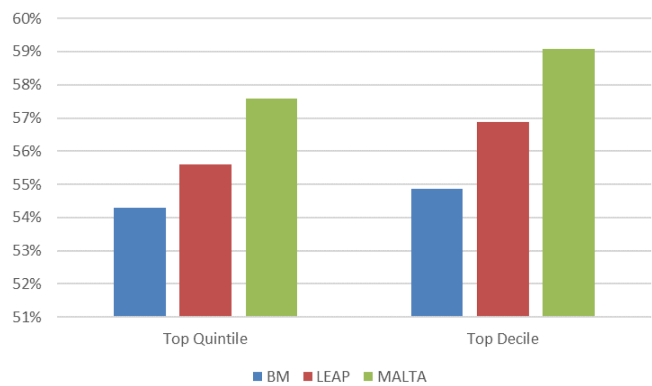
A) US



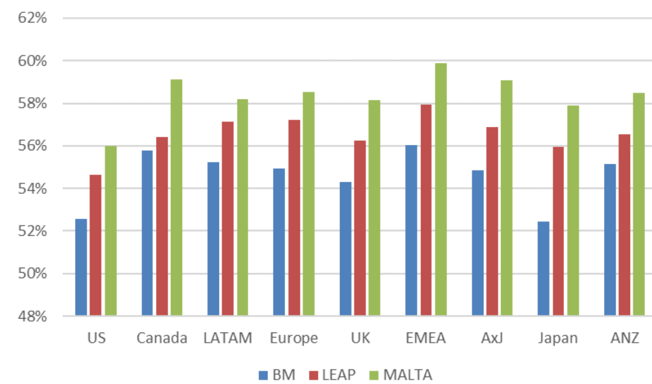
B) Europe



C) Asia ex Japan



D) Global

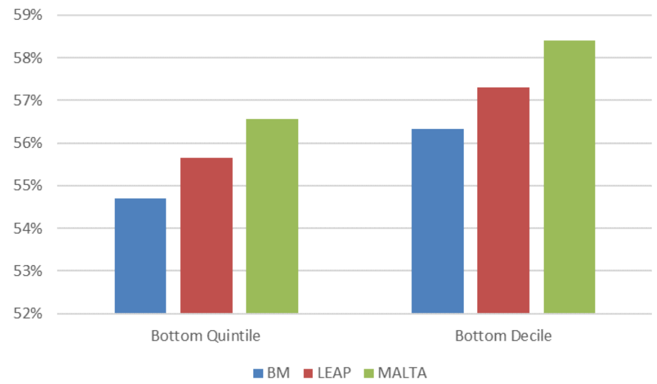


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

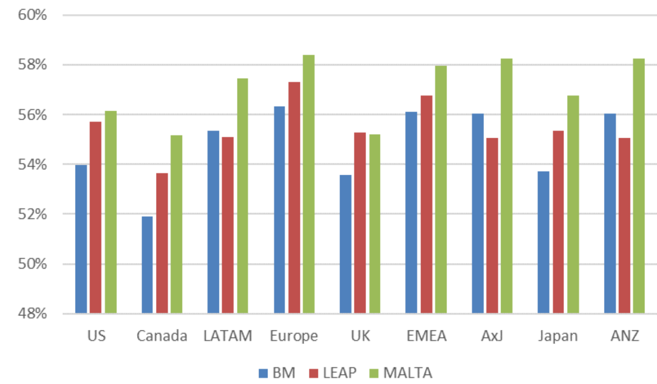
Similarly, on the short side, as our portfolio becomes more concentrated, from the bottom quintile (20%) to decile (10%), the hit rate improves for all three models, especially the MALTA in Europe (see Figure 47A). Globally, as shown in Figure 47(B), LEAP has a higher hit rate than BM in most regions, while MALTA has the highest hit rate in all regions.

Figure 47 Hit Rate for the Short Portfolio

A) Europe



B) Global



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

A LONG HORIZON, HIGH CAPACITY, LOW TURNOVER MODEL – MALTA-HC

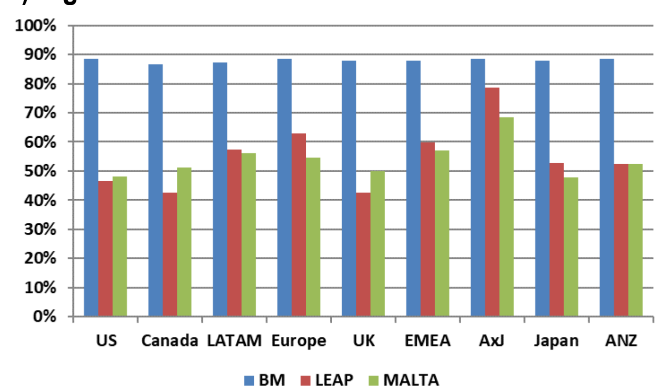
Machine learning based dynamic models generally have much higher turnover than traditional static multifactor models. By itself, high turnover is not necessarily good or bad. Because stock returns have almost not autocorrelation[19], a model that can perfectly predict next period's return should also have a serial correlation close to zero, i.e., a turnover of 100%. On the other extreme, a completely static model that never changes, should have an autocorrelation of 100% and a turnover of 0%. Therefore, generally speaking, a more accurate model should have a lower signal autocorrelation/higher turnover. The reverse statement, i.e., a higher turnover model is always better, however, is not true.

Therefore, investors should accept or reject a model solely based on its turnover. Instead, turnover, along with other model performance metrics (e.g., rank IC, long/short portfolio Sharpe ratio, downside risk, signal decay) should be considered in a unified manner. Details can be found in our previous research (see *Multi-Dimensional Alpha: Risk, Portfolio Construction, and Performance Attribution*, Luo, et al [2017d]).

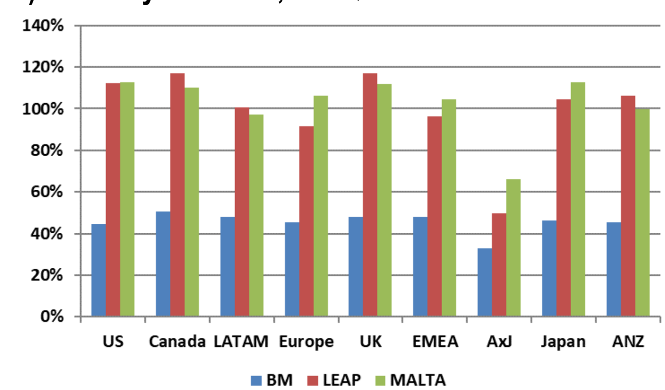
As expected , both LEAP and MALTA models have much lower signal autocorrelation (see Figure 48A) and considerable higher turnover than the benchmark model BM (see Figure 48B), precisely because both models have far stronger predictive power of future stock returns (see Figure 41A).

Figure 48 Signal Autocorrelation and Portfolio Turnover

A) Signal Autocorrelation



B) One Way Turnover, L/S Quintile Portfolio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

INTRODUCING A LONG HORIZON, HIGH CAPACITY, AND LOW TURNOVER MODEL – MALTA-HC

We have done extensive research on how to best align a model's forecasting horizon, signal decay, and turnover, with a target portfolio's holding horizon. In many occasions, investors have to manage their portfolios under certain constraints:

- **Holding Horizon.** In many institutional mandates, managers are required to hold their positions for multiple years.

[19] This is suggested by the weak form of EMH (Efficient Market Hypothesis).

- **Capacity.** Successful managers often attract multiple billions or even hundreds of billions of AUM. All models and fundamental investment processes have their own capacity. As the size of AUM increases, performance often suffers.
- **Transaction Cost.** How to model and incorporate transaction cost in model development and portfolio construction is rather complex. Interested readers please refer to *Multi-Dimensional Alpha: Risk, Portfolio Construction, and Performance Attribution* (see Luo, et al [2017d]).

For investors with long holding horizon, large capacity, low turnover, or facing high transaction cost, we are introducing our MALTA-HC (High Capacity) model. There are many potential ways that researchers can develop a low turnover model. In this section, we propose a simple yet highly effective approach. In essence, we line up our model's forecasting horizon with our portfolio holding horizon.

As we will elaborate in the following sections, the MALTA-HC model has the following features, compared to the base case MALTA structure:

- Higher signal autocorrelation/lower turnover
- Stronger predictive power for longer holding horizon
- Better after cost performance, when trading is expensive
- Better performance for investors with low/tight turnover constraint

The MALTA-HC is similar to our base case MALTA model, with four sub-components:

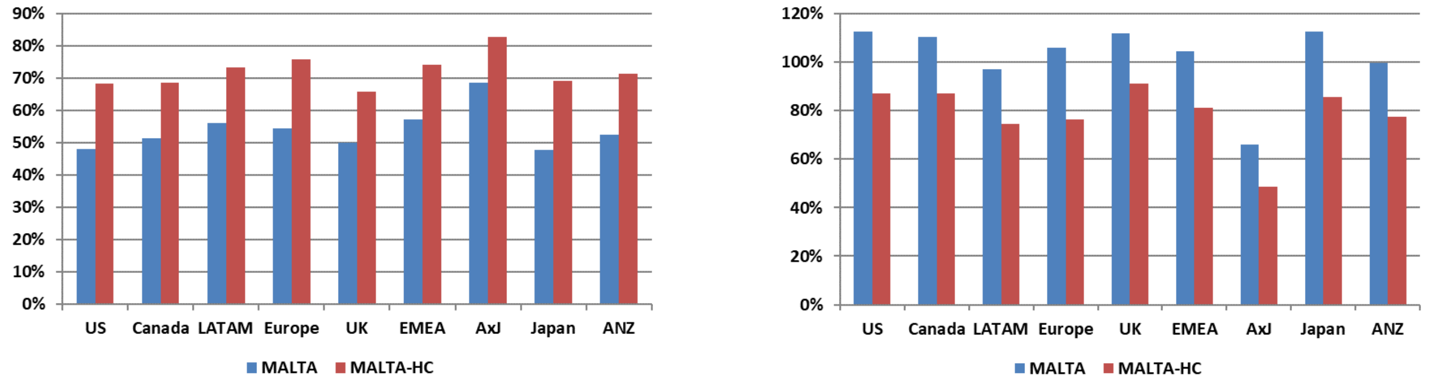
- The Medium-Term Model is trained using a rolling 12-month *overlapping* window;
- The Seasonal Model is developed using the same calendar quarter in the past 10 years;
- The Hedge Model is constructed using the bottom half of quarters in the past 10 years, based on the combined Medium-Term and Seasonal models; and
- The Short-Term Model uses only the previous quarter's data.

Lower Turnover

As shown in Figure 49A, the MALTA-HC model is able to boost the signal autocorrelation by about 20% in most regions. Similarly, the MALTA-HC model enjoys a turnover around 30% lower than the MALTA framework (see Figure 49B).

Figure 49 Higher Signal Autocorrelation and Lower Turnover

A) Signal Autocorrelation **B) One Way Turnover, L/S Quintile Portfolio**



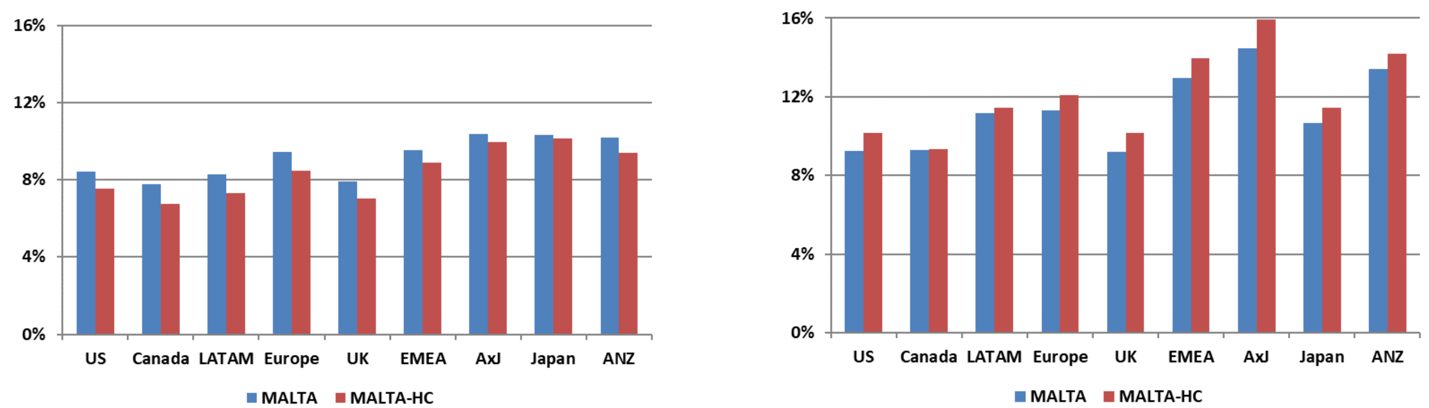
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Stronger Predictive Power for Longer Investment Horizon

As shown in Figure 50(A), for near-term return forecast (i.e., the subsequent month), the MALTA model outperforms the MALTA-HC. However, as we extend our holding period to six months, the MALTA-LC model shines (see Figure 50B), as the MALTA-HC is specifically calibrated to a longer forecasting horizon.

Figure 50 Stronger Predictive Power for Longer Holding Horizon

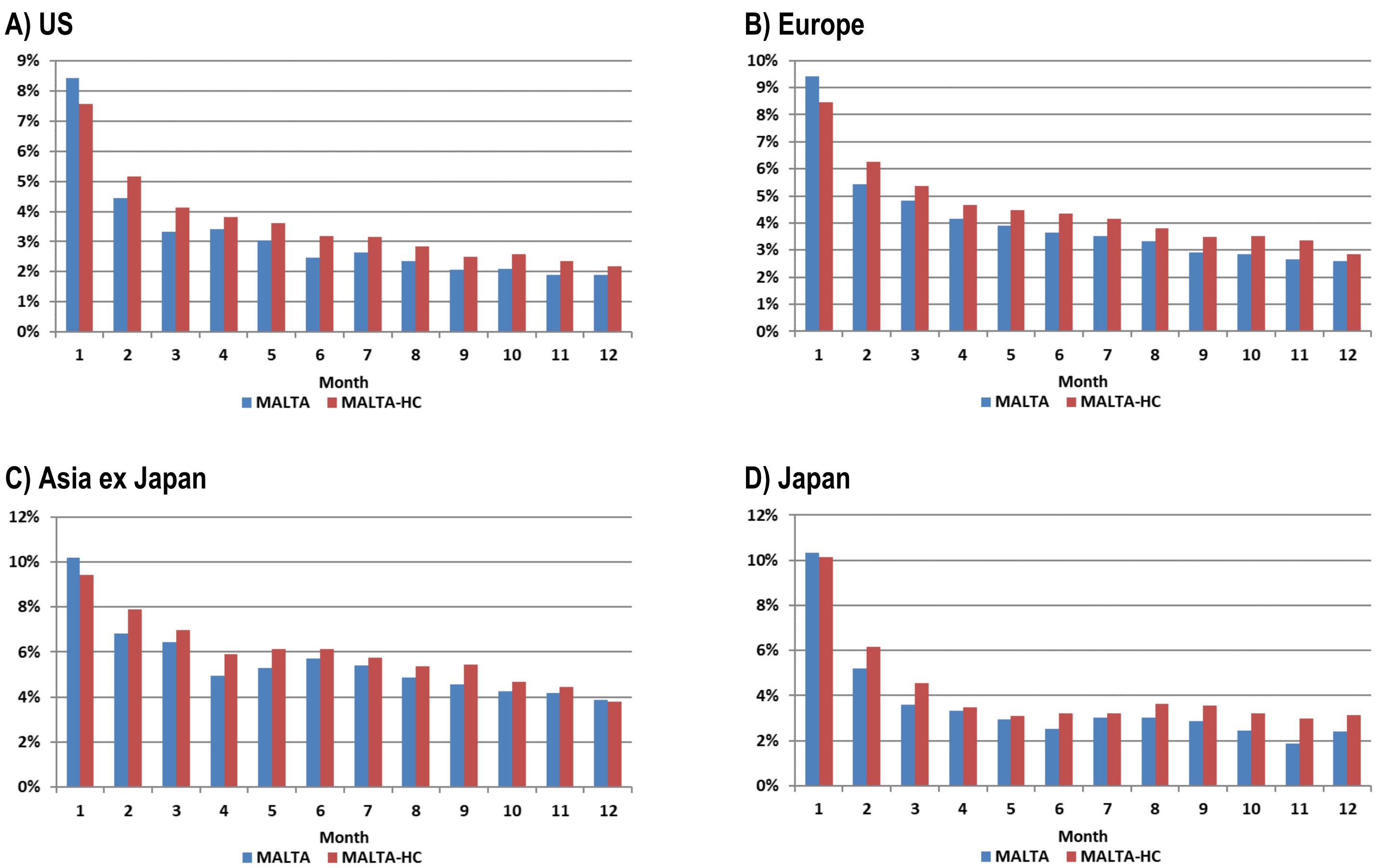
A) Correlation with Next Month's Stock Return (IC) **B) Correlation with Next 6M Stock Return (IC)**



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As shown in Figure 51, in each of the four largest regions (US, Europe, Asia, and Japan), the MALTA has a higher IC (predictive power) than the MALTA-HC at the nearest month. However, after one month, the MALTA-HC model clearly shows not only stronger efficacy, but also slower decay.

Figure 51 IC Decay



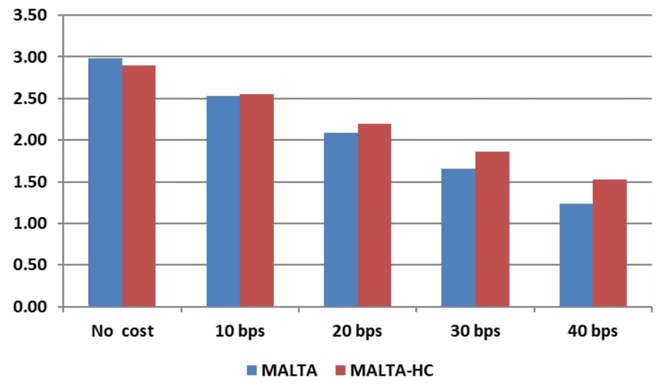
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Portfolio Performance and Transaction Cost

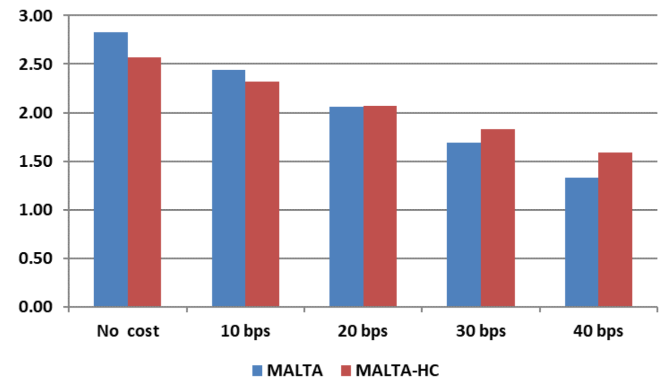
As shown in Figure 52, for an unconstrained portfolio, after cost performance declines, as we raise our transaction cost assumption. The MALTA-HC model has a slower decay profile than the base case MALTA model. For the UK market, the breakeven cost is around 10bps per trade (see Figure 52A), while for Japan, the two models are on par at a 20bps cost function. The market impact cost is a direct function of trade size, which depends on a fund's size. Therefore, for managers with large AUMs, the MALTA-HC model offers a more attractive choice.

Figure 52 After Cost Performance (Long/Short Quintile Portfolio)

A) Sharpe Ratio, UK



B) Sharpe Ratio, Japan



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

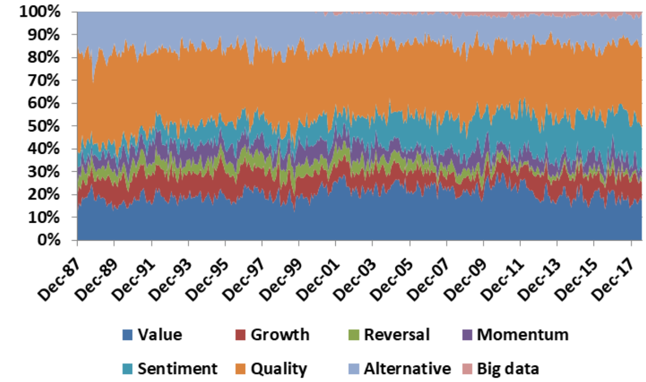
Factor Importance

Although the MALTA model is constructed using a complex suite of nonlinear machine learning models, the MBBT and the associated algorithms are fully transparent. We can transform the factor importance scores into conventional factor weighting. Compared to the base case MALTA (see Figure 53A), the MALTA-HC model tends to load more value and quality factors and less likely to bet on short-term reversal signals in the US. We also observe similar patterns in Europe (see Figure 53C and D) and Japan (see Figure 53E and F).

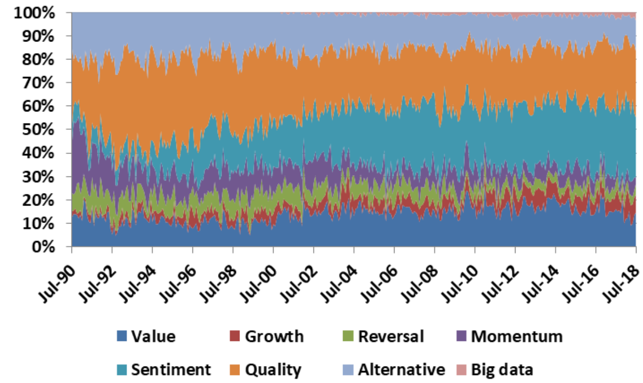
Figure 53 Factor Weights Over Time

A) US, MALTA

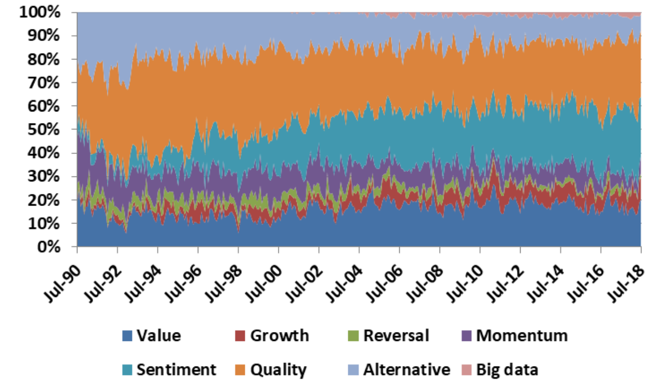
B) US, MALTA-HC



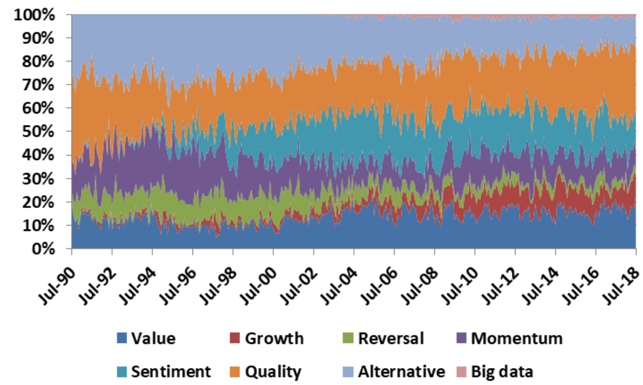
C) Europe, MALTA



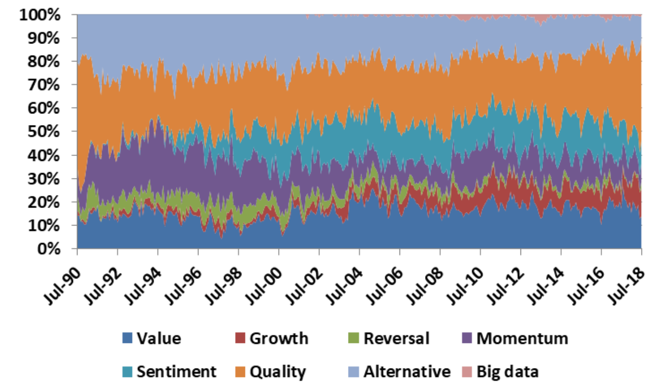
D) Europe, MALTA-HC



E) Japan, MALTA



F) Japan, MALTA-HC



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

PORTFOLIO CONSTRUCTION

Not all institutional portfolios are optimized. However, an optimizer does make it much easier to incorporate transaction cost assumption, portfolio turnover, benchmark tracking, and handle other constraints much easier. In this case, we use a simulated monthly rebalanced optimized portfolio to demonstrate the complex interaction of signal decay, portfolio turnover, and trading cost.

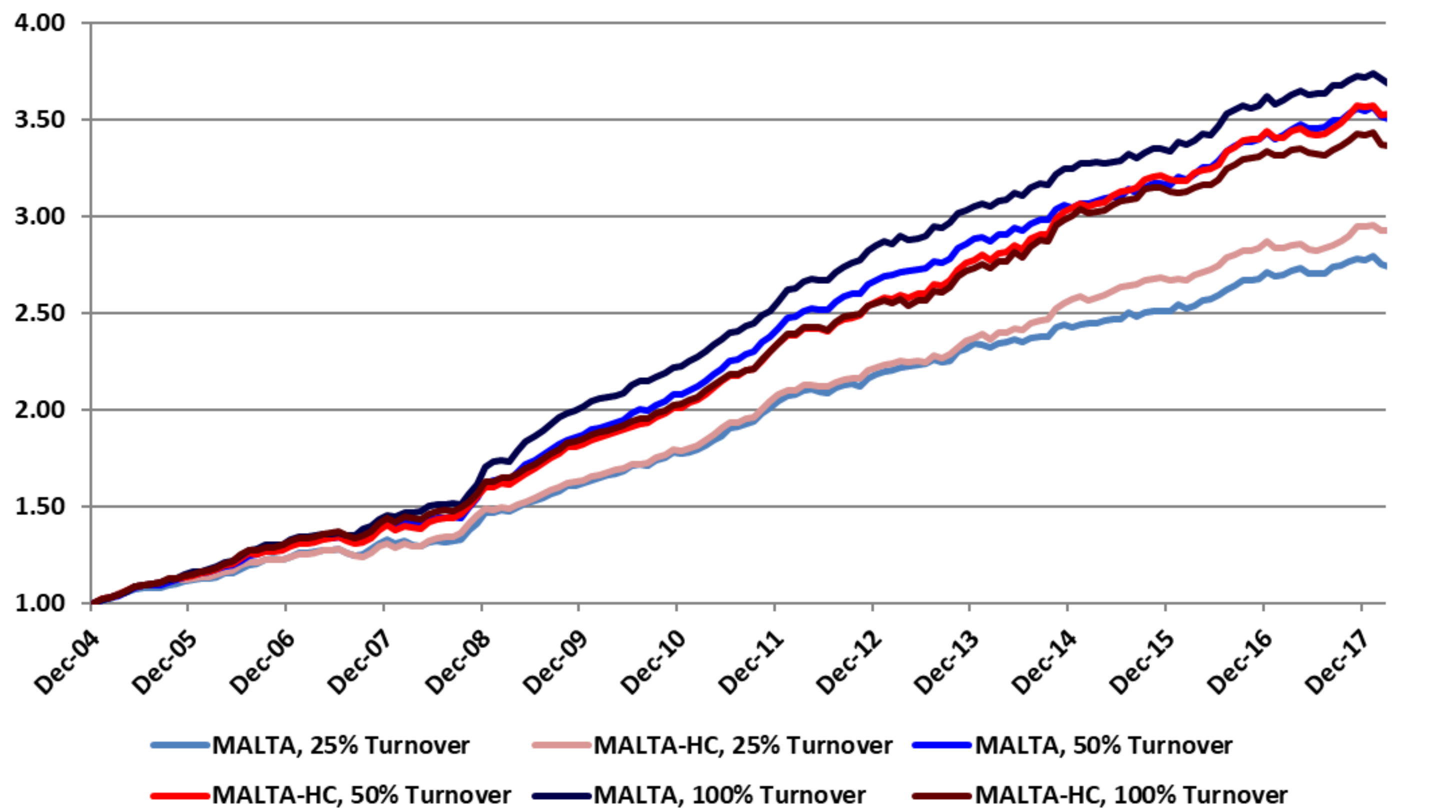
At the end of each month since December 2004, we construct two comparable portfolios:

- Target return/alpha is based on our MLLTA and MALTA-HC, respectively
- Rebalancing Frequency: Monthly
- Investment Universe: Russell 3000
- Market neutral, dollar neutral, 2x leverage (i.e., for $1 capital, we have $1 on the long side and $1 on the short side)
- Target Risk: 3% annual volatility
- Liquidity Constraint: at each rebalance, we can't trade more than 10% ADV (Average Daily Volume) of any single stock
- Notional AUM: US$100 million
- We repeat the simulation with three monthly two-way turnover constraint, 25%, 50%, and 100%, respectively.

As shown in Figure 54, we can draw the following conclusions, assuming a 10bps transaction cost per trade:

- For a highly predictive model such as the MALTA suite, the higher the portfolio turnover, the better is the *after-cost* performance.
- If we have a tight turnover constraint, i.e., we can't trade too much into our signal, the slow decay MALTA-HC model is a better choice.
- For portfolios with high turnovers, the fast moving MALTA model performs better.

Figure 54 Cumulative Performance with a 10bps Average Transaction Cost Assumption, US



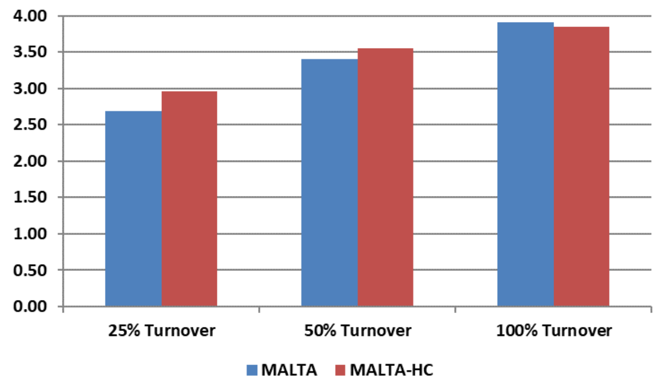
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Arguably, the 10bps transaction cost assumption is reasonable. However, as elaborated in *Multi-Dimensional Alpha: Risk, Portfolio Construction, and Performance Attribution* (see Luo, et al [2017d]), transaction cost is also a function of market volatility. It is obviously more expensive to trade in a volatile market when liquidity is low. Figure 55 shows a sensitivity analysis on portfolio *after-cost* performance as a function of different transaction cost and turnover scenarios:

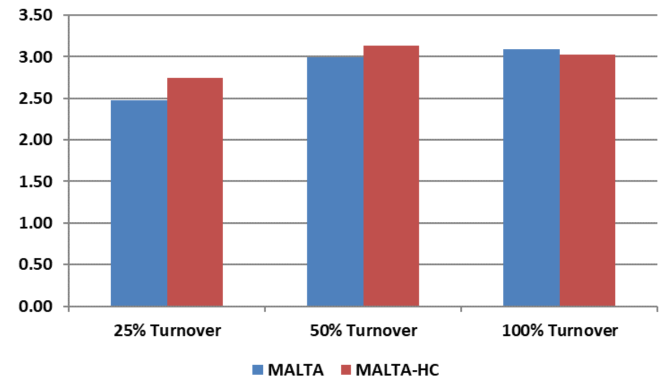
- MALTA model performs better with high turnover and/or low transaction cost. On the other than, MALTA-HC model is preferred when turnover is low and/or trading is expensive.
- When trading cost is low (e.g., no cost or 10bps), performance tends to improve as turnover increases. MALTA model is generally a better choice.
- When transaction cost is high (i.e., when market is volatile, or trading liquidity is low, or AUM is large), performance is more likely to deteriorate as turnover goes up. MALTA-HC model is typically preferred.

Figure 55 Optimized Portfolio Performance with Different Cost and Turnover (US)

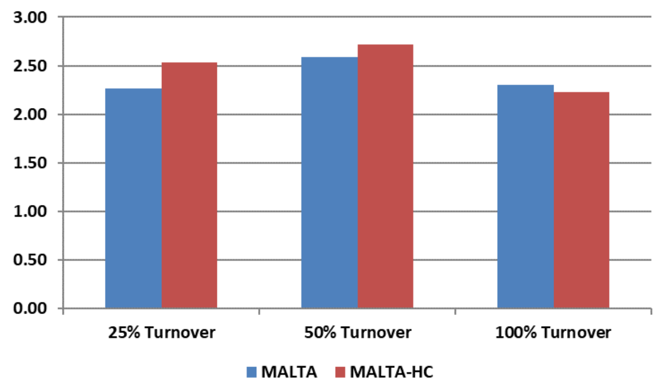
A) Sharpe Ratio, Assuming No Trading Cost



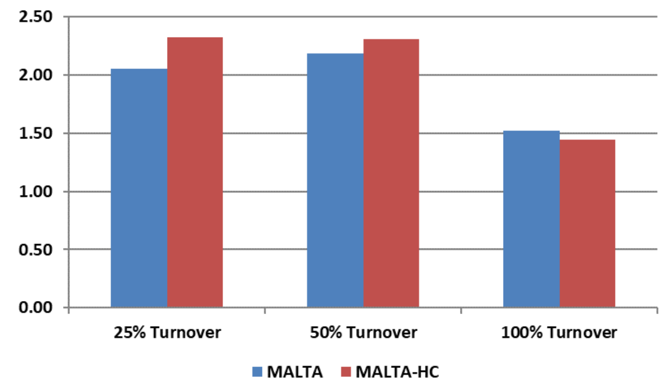
B) Sharpe Ratio, 10bps Cost Assumption



C) Sharpe Ratio, 20bps Cost Assumption



D) Sharpe Ratio, 30bps Cost Assumption



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

A FAST MOVING, UNCORRELATED MALTA STATARB MODEL

For some managers, an average holding of one year is still too short, while for other PMs, even one week is too long. In this section, we are introducing a high frequency statistical arbitrage version – MALTA-StatArb model. The MALTA-StatArb model is calibrated to predict the following week's stock return. The MALTA-StatArb can be implemented with intraday trading, daily rebalance, or weekly execution – generally speaking, the higher the frequency of execution, the better is the after-cost performance.

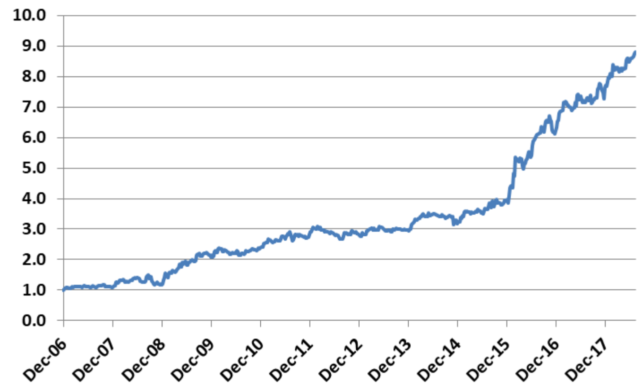
On a weekly horizon, most fundamental factors based on financial statement data become less useful, while technical signals (primarily driven by price and volume) and alternative data (e.g., news sentiment, short interest) turn out to be more effective. Therefore, in addition to our standard factor library, we could add hundreds of technical factors as potential features for the machine learning algorithms to choose.

As discussed extensively in our previous research, we can't use technical factors the same way as done by traditional technical analysts. Technical analysis is commonly applied to a single security, one at a time, in a time series fashion. For example, if GE is oversold, would it be a good entry point to buy the stock? We first compute our technical factors as in the traditional way, i.e., using past price/volume data for each stock, at each given point-in-time. Then, we normalize each stock's computed score over its own past history. Lastly, we compare all stocks in a cross-sectional fashion, similar to fundamental factors. Since most technical factors are calculated using the closing price, the earliest time we can trade these signals is on the next day's open. Therefore, we use Friday's closing price/volume and other information to compute our factors. For the time being, we assume that we can execute on the next business day's open. We will discuss other more realistic and sophisticated trade execution strategies shortly.

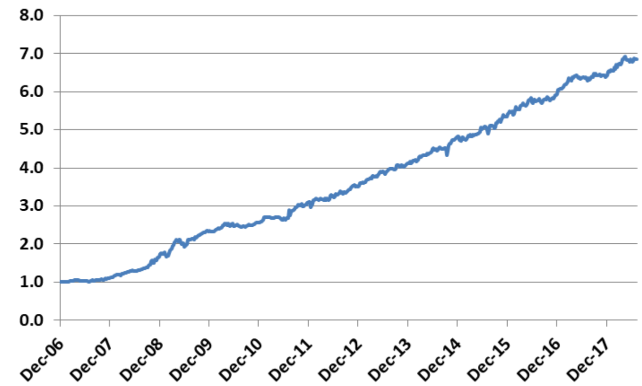
Figure 56 shows the performance of a few sample technical factors (the long/short decile portfolio, weekly rebalanced), pre-trading cost, in the US, Europe, Asia, and Japan. Despite the recent challenging market for active management, these technical factors remain highly profitable.

Figure 56 Long/short Decile Performance for Technical Factors

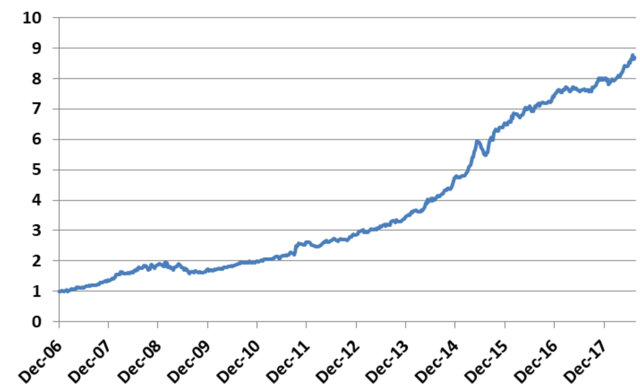
A) Moving Average Convergence Divergence (US)



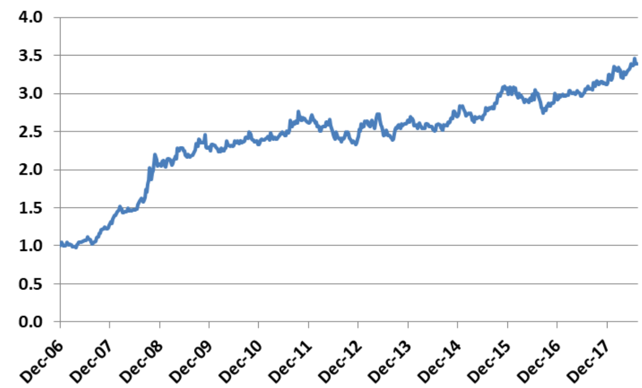
B) Percentage Volume Oscillator (Europe)



A) Average Directional Index (Asia ex Japan)



B) William % R (Japan)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

MALTA-StatArb

The MALTA-StatArb model follows a similar design as our base case MALTA model, with the same four sub-components. There are a number of key differences for our short-term statistical arbitrage model:

- The MALTA-StatArb model is re-fitted every week, after the market close on Friday.
- For the current version, we use primarily technical factors for the MALTA-StatArb model, mostly due to computing needs[20].
- All factors, including all technical factors, are computed using Friday's closing price/volume and other information.

[20] As we expand from monthly to weekly, the size of our data grows exponentially. We do plan to incorporate other factors in our factor library in the next release of our StatArb model.

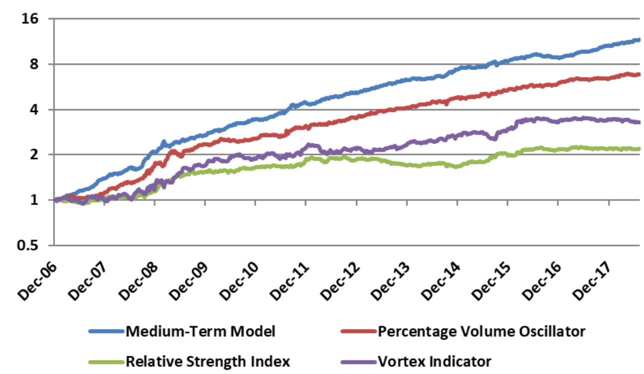
- Our target variable, i.e., stock return, is calculated as the return from the following Monday's open price[21] to Friday's closing price.
- The Medium-Term model is trained using trailing one-year of weekly data; the Seasonal model is based on a trailing 10-year window of the same calendar week; the Hedge model is constructed on the bottom half of weekly data, based on the performance of the combined Medium-Term and Seasonal models; and lastly, the Short-Term model uses only the most recent week's data.

Some Simple Motivation

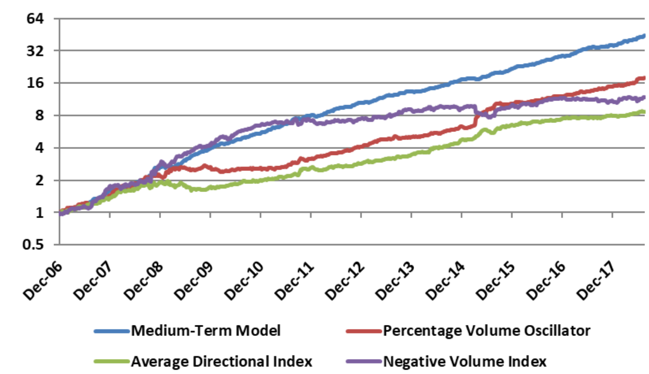
To help our readers to gain some additional insights and intuition behind the MALTA-StatArb model, Figure 57 shows the cumulative performance of the model, along with some of the best performing technical factors for Europe and Asia. Not surprisingly, our MALTA-StatArb model easily beats the best performing technical factors – a feature that is not necessarily true for most other statistical arbitrage models. Although a naïve investor may expect a composite model using multiple factors to outperform any underlying single variables, it is not that easy in practice. The reason is that we would not know the best performing single factor *ex ante*. Therefore, the MALTA-StatArb is pure out-of-sample, while the best technical factors are based on their ex post performance with significant look-ahead bias.

Figure 57 Cumulative Performance for Best Performing Technical Factors versus Machine Learning Model

A) Europe



B) Asia ex Japan

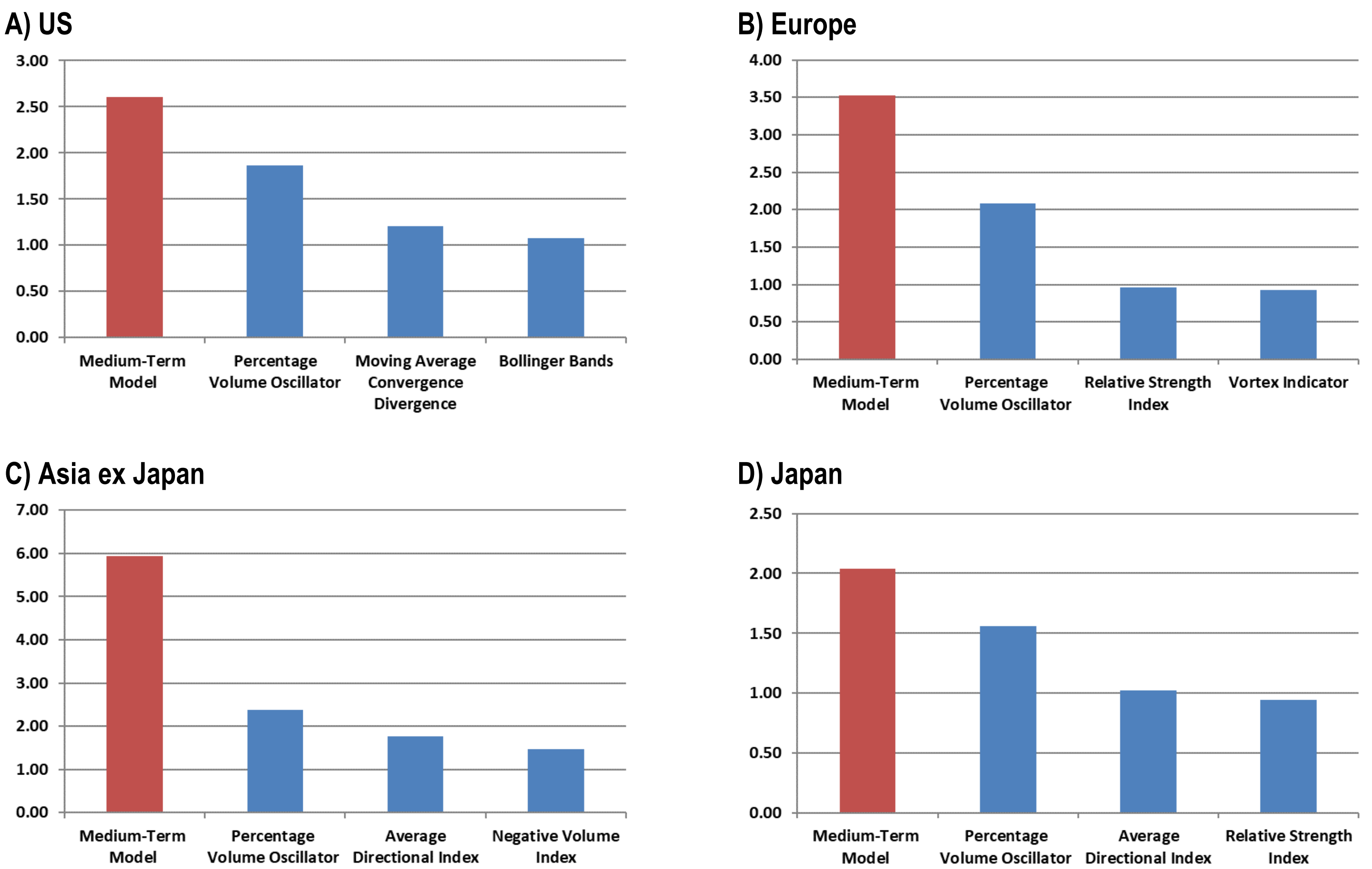


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

As shown in Figure 58, in the US, Europe, Asia, and Japan, the MALTA-StatArb outperforms the best underlying single factors considerable, especially in Asia ex Japan.

[21] If Monday is a market holiday, we will use the next business day's open price. We will also discuss other ways to compute our target returns in a later section.

Figure 58 Best Performing Technical Factors versus Machine Learning, Sharpe Ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Diversification Benefit from the Four Sub-Models

As shown in Figure 59, the four sub-models are almost uncorrelated, in almost all regions. In particular, the Hedge model, by design, offers tremendous diversification benefit.

Figure 59 Correlation Matrix

A) US

	Medium-Term	Seasonal	Hedge	Short-Term
Medium-Term	100%			
Seasonal	45%	100%		
Hedge	25%	8%	100%	
Short-Term	33%	11%	-6%	100%

B) Japan

	Medium-Term	Seasonal	Hedge	Short-Term
Medium-Term	100%			
Seasonal	28%	100%		
Hedge	27%	15%	100%	
Short-Term	24%	11%	-2%	100%

C) UK

	Medium-Term	Seasonal	Hedge	Short-Term
Medium-Term	100%			
Seasonal	23%	100%		
Hedge	35%	22%	100%	
Short-Term	28%	17%	14%	100%

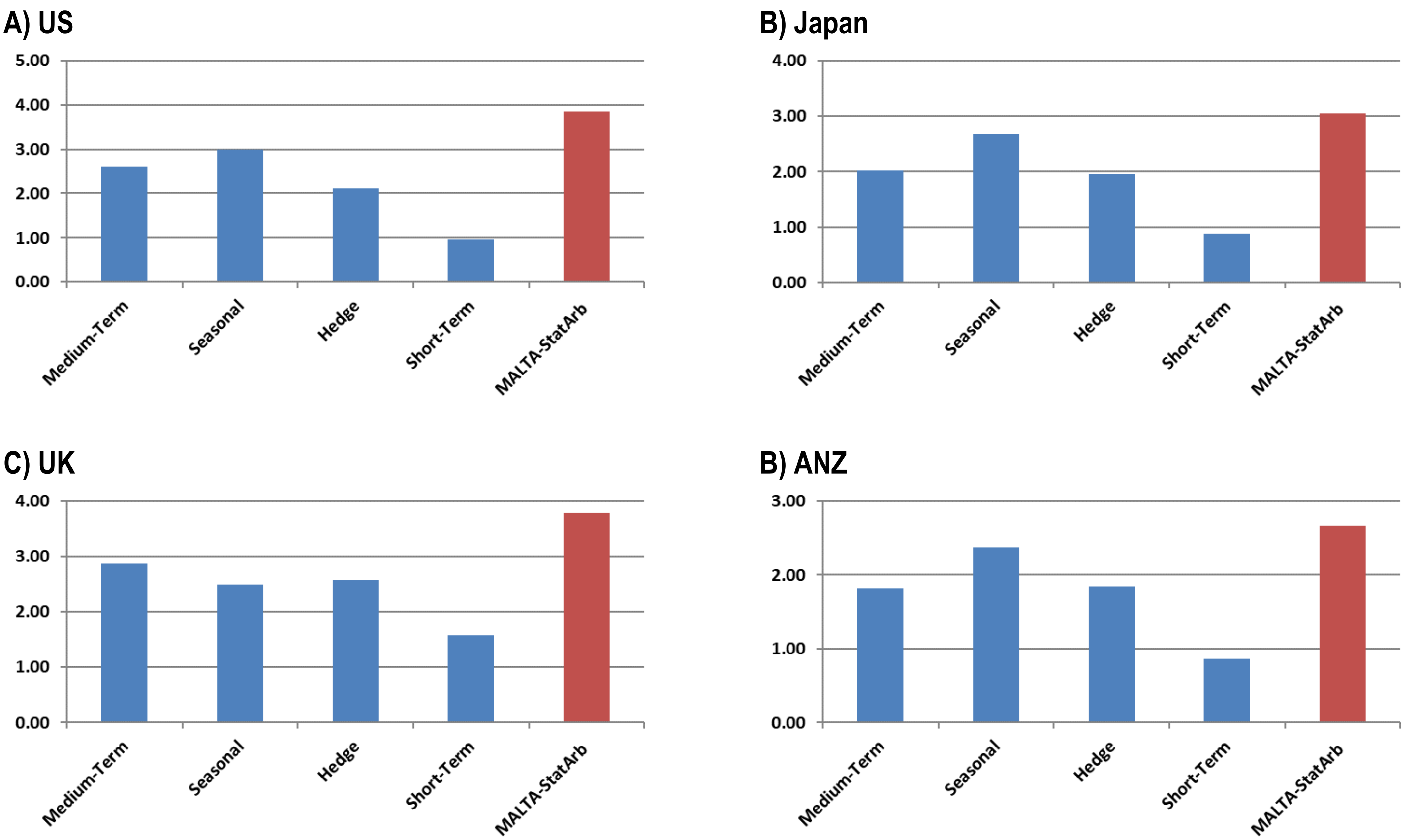
B) ANZ

	Medium-Term	Seasonal	Hedge	Short-Term
Medium-Term	100%			
Seasonal	19%	100%		
Hedge	49%	12%	100%	
Short-Term	35%	3%	18%	100%

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Because of the low correlated among the four sub-components, the combined MALTA-StatArb boosts performance further (see Figure 60).

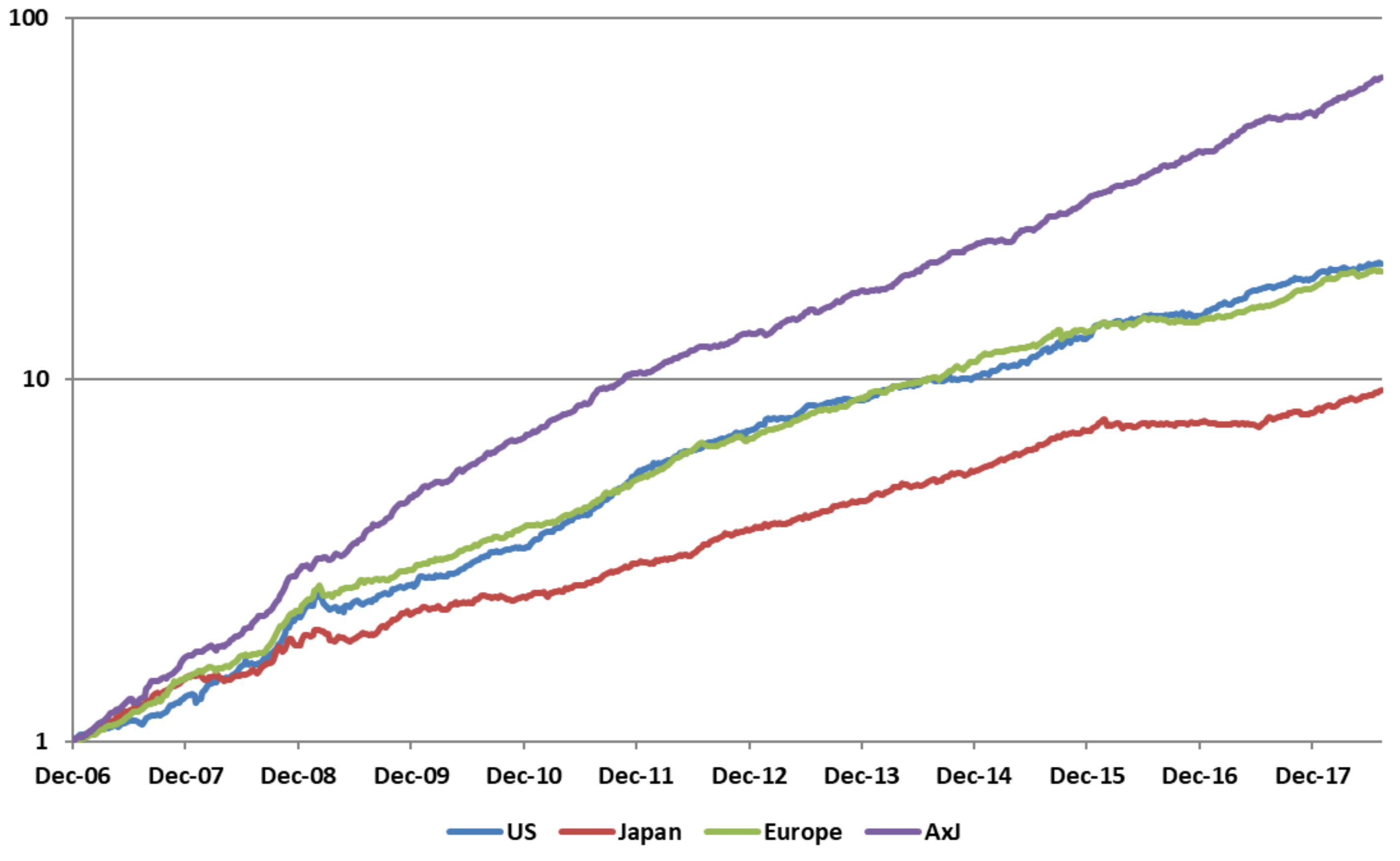
Figure 60 Sharpe Ratio of MATLA-StatArb and Sub-Models, Long/Short Decile Portfolio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Figure 61 shows the cumulative performance for a long/short decile performance, weekly rebalanced, using our MALTA-StatArb model, before transaction cost. The short horizon statistical arbitrage model demonstrates superior performance in all regions, especially in Asia.

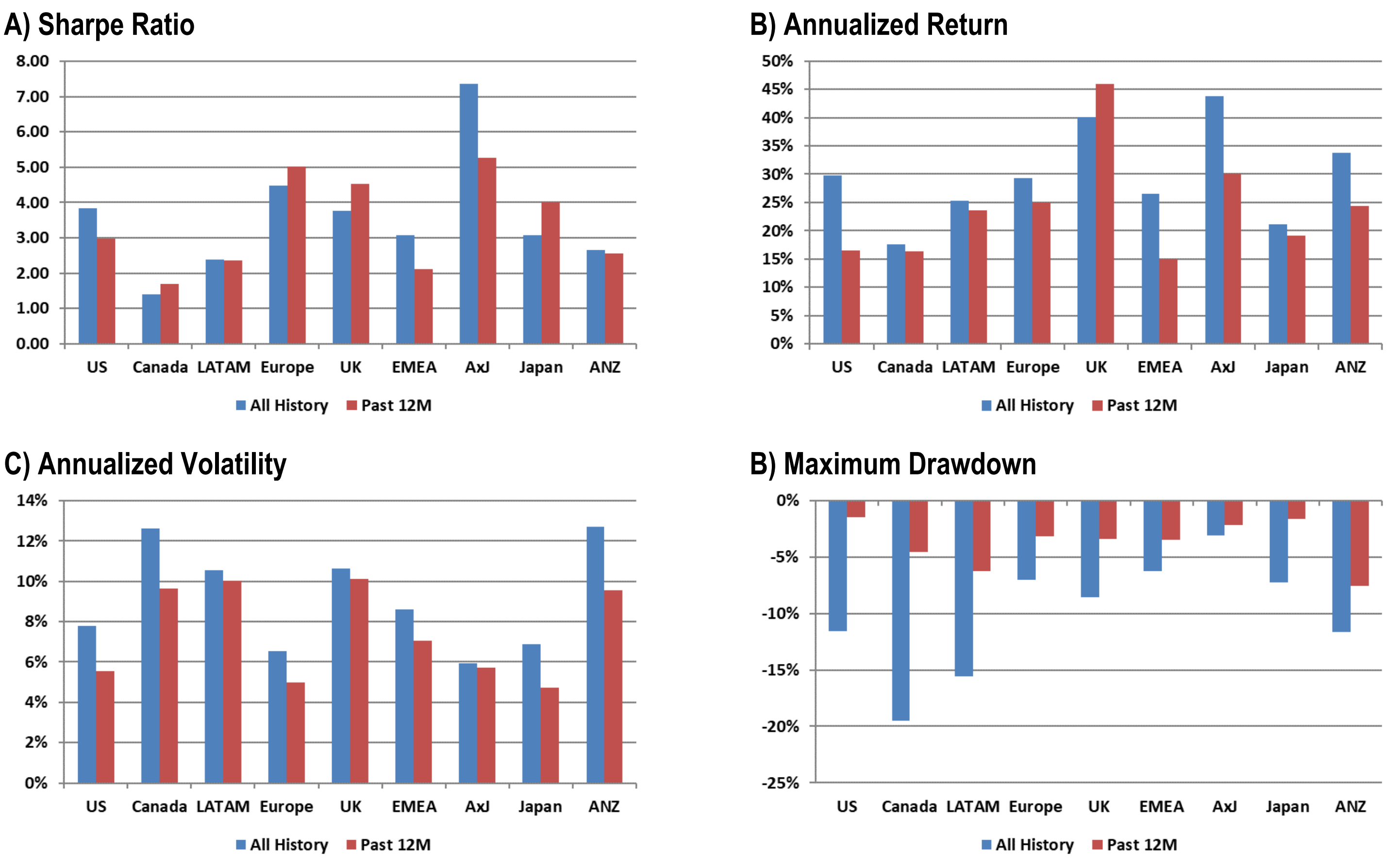
Figure 61 Cumulative Performance, MALTA-StatArb, Long/Short Decile Portfolio, Weekly Rebalance



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

As shown in Figure 62(A), the MALTA-StatArb model delivers superior performance in all nine regions of the world, including the most recent year, when many statistical arbitrage hedge funds suffered from negative performance and redemptions. As we discussed in *Multi-Dimensional Alpha: Why the Market Volatility has been so Low?* (see Luo, et al [2018a]), despite of a few temporary corrections, the equity market has been relatively calm in the past year. As a result, our MALTA-StatArb portfolio also produces lower volatility than the long-term historical average (see Figure 62C). Similarly, we also observe a lower downside risk across the board (see Figure 62D).

Figure 62 MALTA-StatArb Model Performance, Long/Short Decile Portfolio



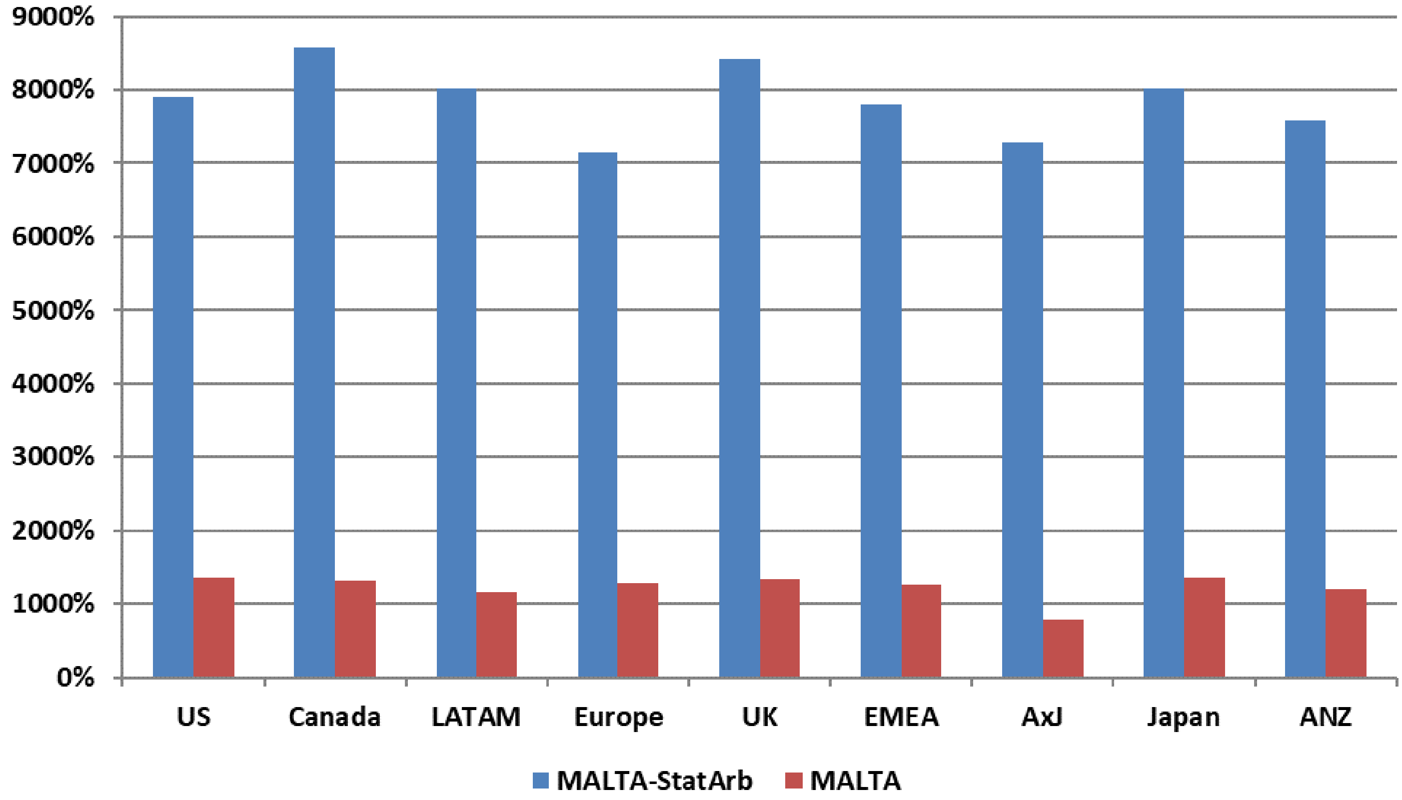
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

TURNOVER, TRANSACTION COST, AND SIGNAL DECAY

By design, the MALTA-StatArb model should have a much higher turnover than our base case MALTA. Furthermore, we trade the StatArb model weekly, while MALTA is based on a monthly rebalance. As shown in Figure 63, the annualized one-way turnover of MALTA-StatArb is about 7x-8x higher than the MALTA, based on a long/short decile portfolio[22]. As a result, the MALTA-StatArb model's performance is more sensitive to transaction cost.

[22] Please note that the turnover is computed on a completely unconstrained long/short decile portfolio. In practice, many managers may want to slow down the trade by adding a turnover constraint. We will discuss a few ways to reduce turnover shortly. The natural maximum one-way turnover for MALTA is 200% per month or 2400% per year. The maximum one-way weekly turnover for the MALTA-StatArb is 200%; therefore the maximum annual turnover is 10,400%.

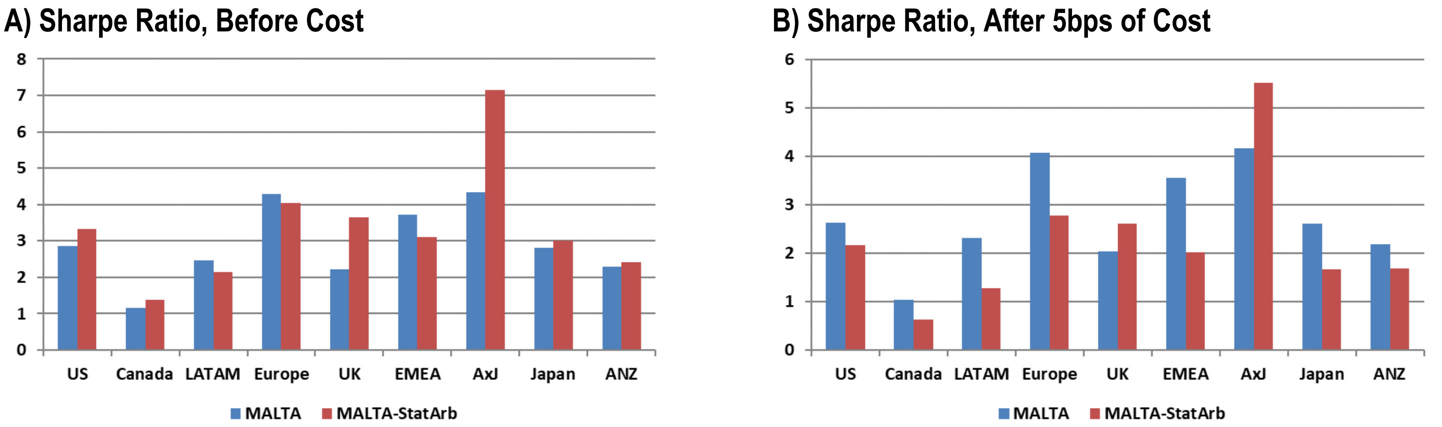
Figure 63 Annualized One-Way Turnover



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

As shown in Figure 64(A), MALTA-StatArb outperforms the MALTA model in most regions pre-transaction costs. However, due to the much higher turnover, even a modest 5bps trading cost can eat up much of the performance. On an after cost basis, MALTA-StatArb lags behind in most regions (see Figure 64B). Therefore, how to slow down our signal and/or how to reduce transaction cost is critical to the success of a statistical arbitrage strategy. In the next few sections, we will focus on the first aspect and propose a few improvements.

Figure 64 Performance Comparison, Long/Short Decile Portfolio, 2007-Present



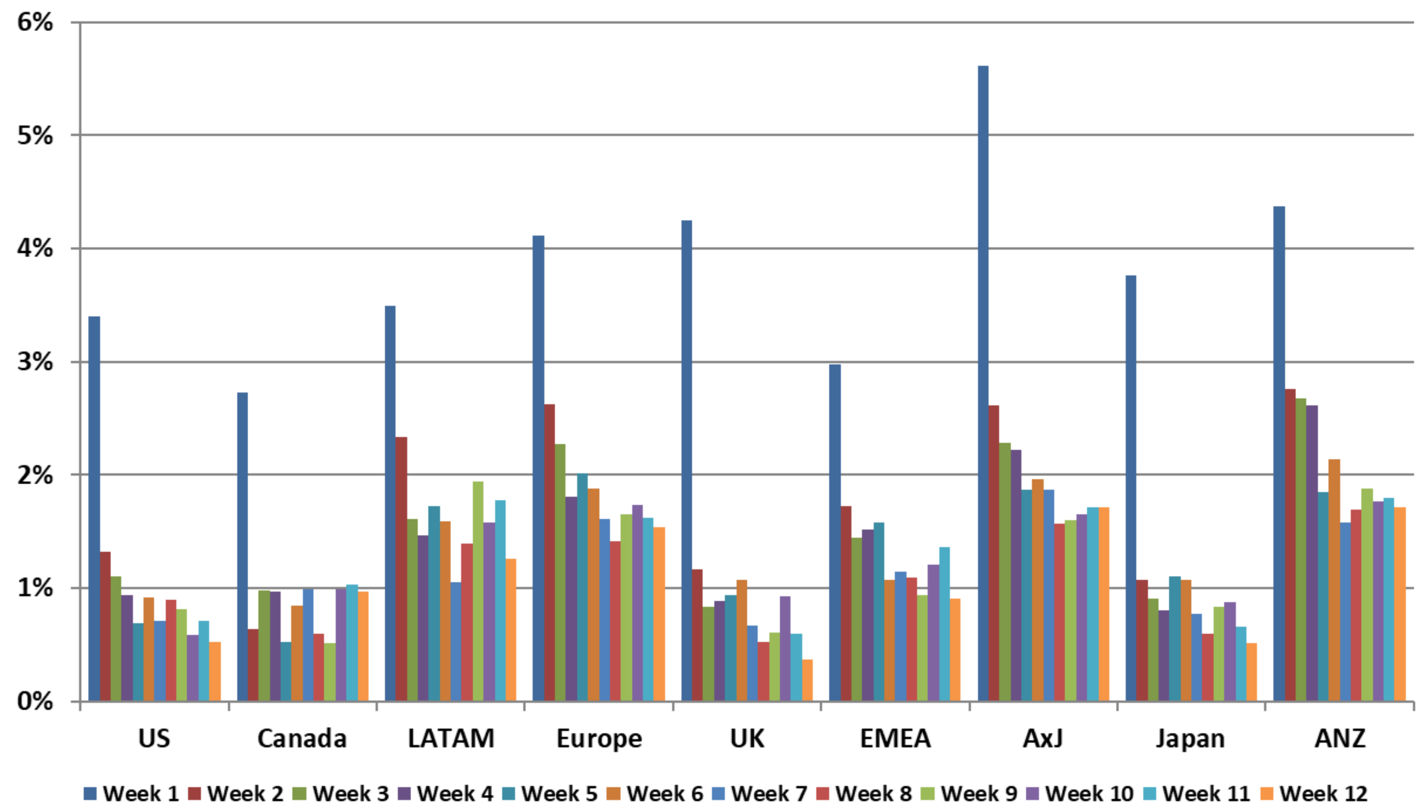
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Signal Decay

Figure 65 shows MALTA-StatArb model's information decay. Basically, we track the predictive power (i.e., Rank IC or the correlation between our model prediction and the next period's actual stock return) on a weekly basis. There are a few interesting patterns:

- The model is highly predictive for the first subsequent week's stock return. The model's predictive power falls sharply after the first week, but it remains reasonable even after three months.
- The MALTA-StatArb model's decay varies by regions. For example, the model retains much of its forecasting ability in Europe even after a month, but it loses much of its excess return in the UK after a week.
- Given that the model preserves decent performance for weeks, we could smooth our signal by combining the model's past predictions.

Figure 65 Signal Decay (Rank IC)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

To reduce turnover, we can smooth our signal by mixing the model's new prediction with scores from previous weeks. As a simple demonstration, we could use the following formula to smooth our signal:

$$\alpha_{i,t}^{Smooth} = (1-\lambda)\alpha_{i,t} + \lambda\alpha_{i,t-1}$$

Where,

$\alpha_{i,t}^{Smooth}$ is the smoothed signal for stock i at week t,

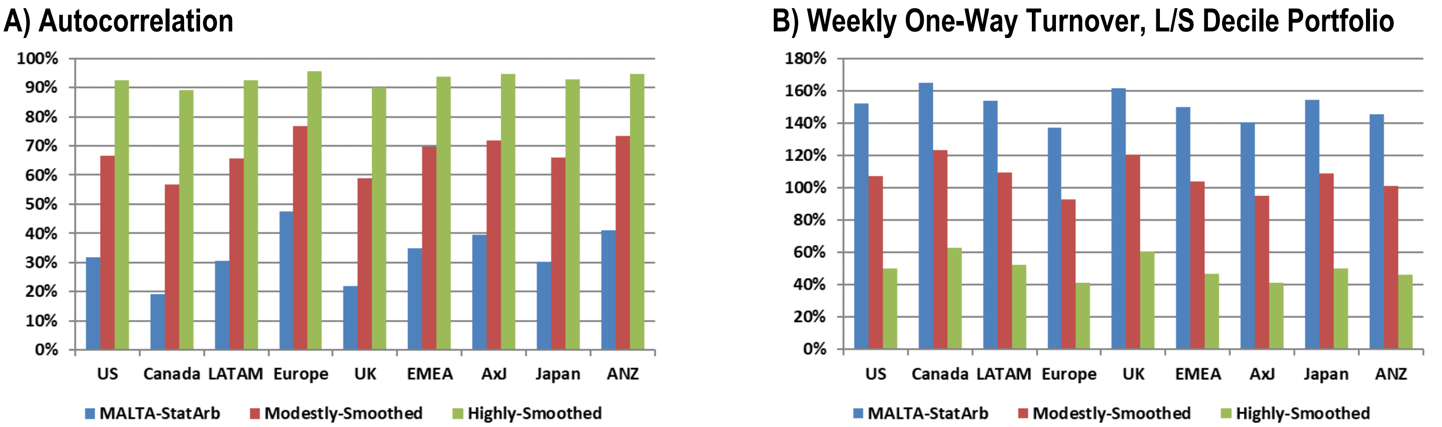
$\alpha_{i,t}$ is our original alpha score from MALTA-StatArb for stock i at week t, and

λ is our smoothing coefficient.

The higher the smoothing coefficient λ, the more weight we put to delayed signal; therefore, the lower the turnover of our model. If $\lambda = 0$, we use our original model without any smoothing, while if $\lambda = 100\%$, we only use the previous week's signal.

Figure 66 shows the signal autocorrelation and turnover for two sample smoothing functions – a modestly-smoothed signal ($\lambda = 40\%$) and a highly-smoothed model ($\lambda = 80\%$). As we increase our smooth parameter from 40% to 80%, signal autocorrelation jumps considerably by almost 30% (see Figure 66A), while turnover falls sharply by nearly 50% (see Figure 66B).

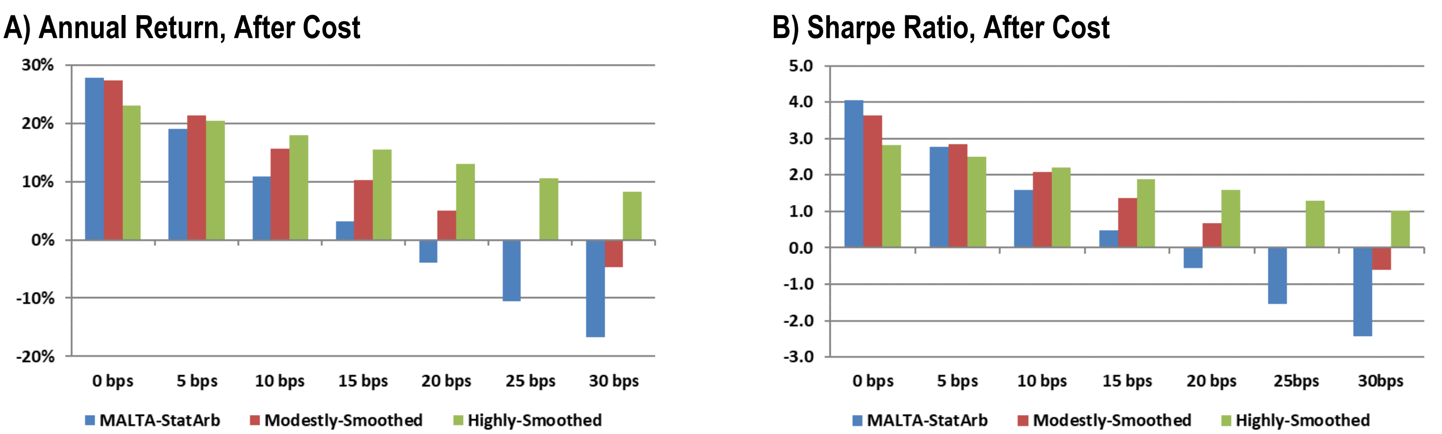
Figure 66 Reduced Turnover and Increase Autocorrelation



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Figure 67 shows the after-cost performance for the same two smoothing functions in Europe. If there is no cost to trade, the original MALTA-StatArb beats the modestly-smoothed version, which in turn, outperforms the highly-smoothly model. The breakeven cost is about 5bps, when the three versions are on par. As we dial up our trading cost assumption, the highly-smoothed model dominates. At around 17bps cost, the original model produces no excess return, while the modestly-smooth model still survives as long as cost is below 25bps. Lastly, even at a high cost of 30bps, the highly-smoothed model still delivers a Sharpe ratio of 1.0x.

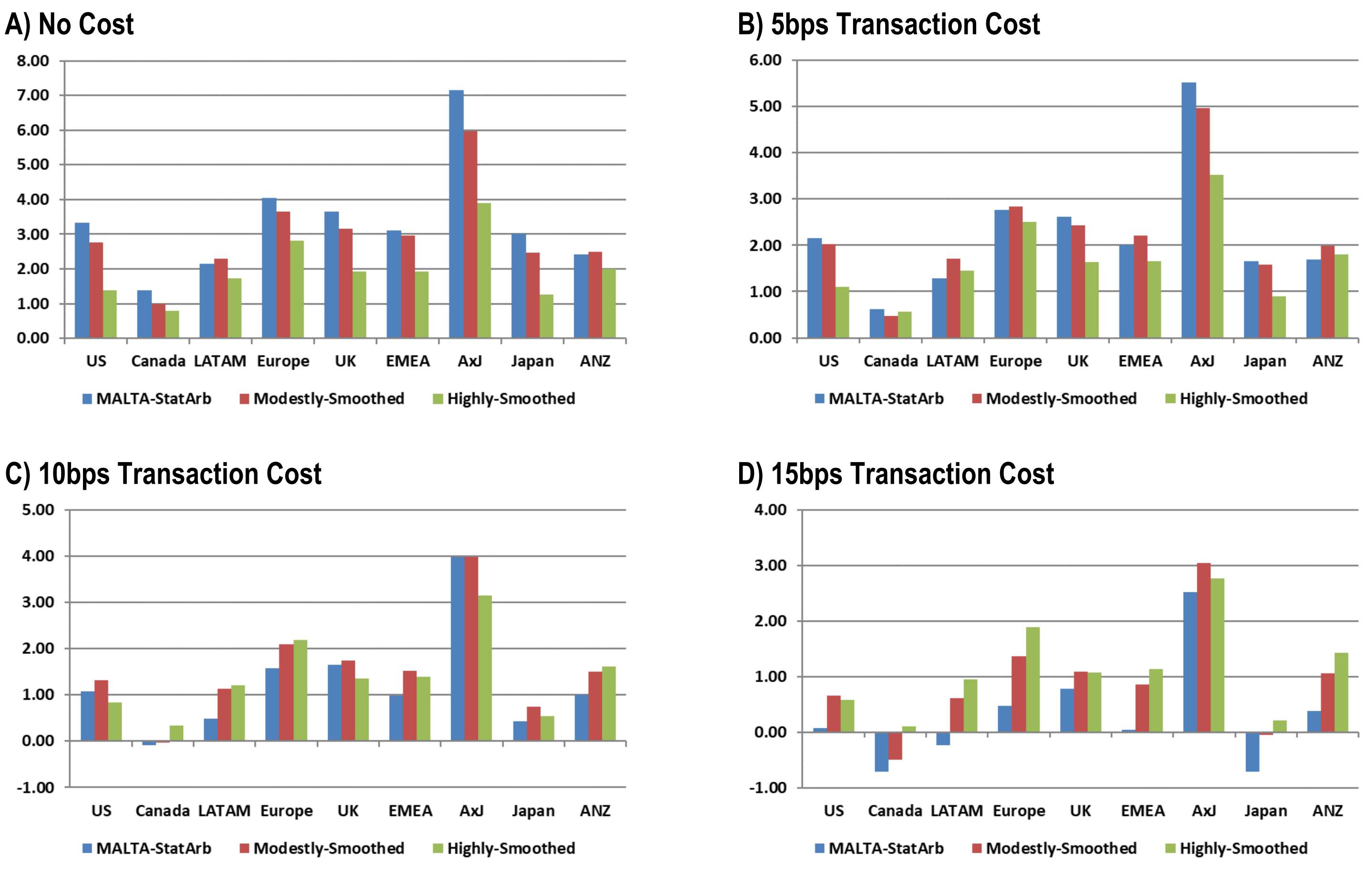
Figure 67 After-Cost Performance Comparison, Smoothed Signal in Europe



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Globally, when transaction cost is low (e.g., below 5bps), the original MALTA-StatArb model is the preferred choice in most regions (see Figure 68A and B). At around 10bps cost, the two smoothed models deliver similar performance (see Figure 68C). When it is expensive to trade (e.g., at or above 15bps cost), the highly-smoothed model dominates in most regions.

Figure 68 After-Cost Performance Comparison, Smoothed Signal, Global Evidence



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

ALIGNING MODEL TRAINING WITH TRADE EXECUTION

In our previous research on portfolio construction, we found that there are a number cases that our alpha predictions can be mis-aligned with our risk models:

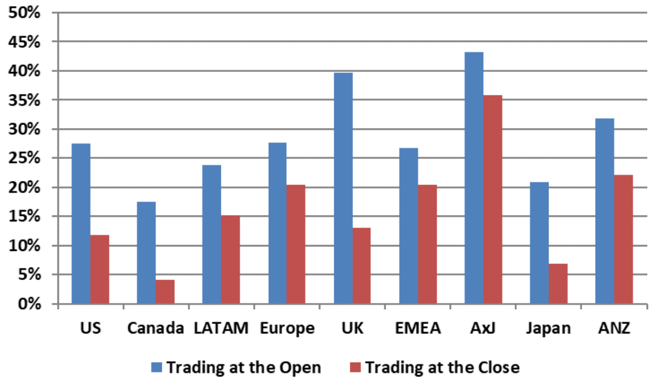
- **Factor mis-alignment.** Alpha models are often developed in-house, while risk models are mostly from commercial risk vendors. Therefore, the factors used in the alpha and risk models may have some overlap, but most likely are different. The optimizer may emphasize the difference between the alpha and risk models and therefore, produce a sub-optimal portfolio.
- **Horizon and decay mis-alignment.** Similar to factor mis-alignment, alpha model could have a much longer (or shorter) investment horizon than (therefore, different signal decay from) the risk model, which generates insufficient trades (or unnecessary turnover).

In this section, we address the third mis-alignment issue – the alignment between model development and trade execution. As of now, we have backtested and assumed that we always trade at the next day's *open* price. To make it clear of how this specific version is calibrated, we rename the model MALTA-StatArbOpen. However, opening volume can be erratic and is generally lower than at the close. As shown in Figure 69, moving our execution from the open to the close, performance plunges.

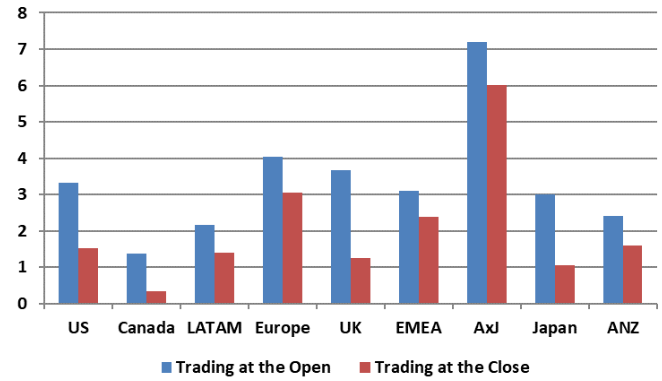
Because we calibrate our MALTA-StatArbOpen model based on the next day's open price to the next Friday's close, the model is optimized to capture the pattern accordingly. When we execute at the first day's close, we lose significant predictive power, due to information decay.

Figure 69 Performance Comparison, Open versus Close Trade Execution, MALTA-StatArb Model

A) CAGR



B) Sharpe Ratio



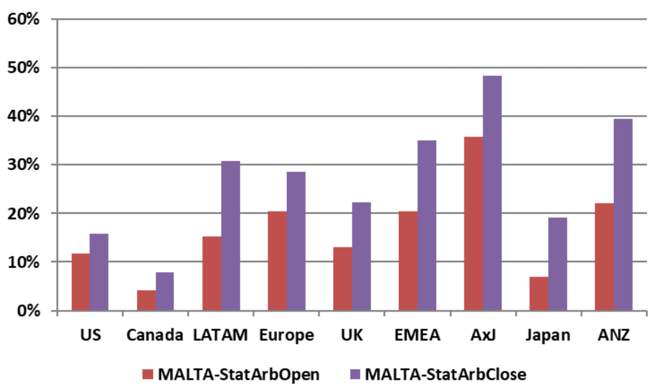
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

To better align with our execution strategy, we re-train a new model – MALTA-StatArbClose to predict the return computed from the next day's close until Friday's close. Otherwise, the new model is exactly the same as MALTA-StatArbOpen.

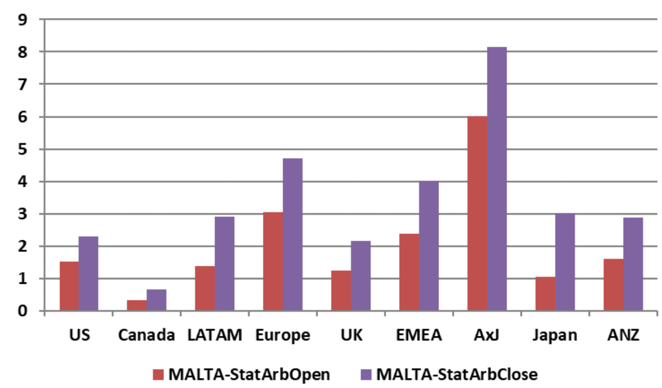
As expected, the new MALTA-StatArbClose model performs much better in every region, when we trade at the next day's closing price (see Figure 70), due to the alignment in model training and trade execution. Sharpe ratio more than doubles in most regions. In Japan, Sharpe ratio almost triples to 3.0x.

Figure 70 Performance Comparison, Trading at the Next Day's Close

A) CAGR



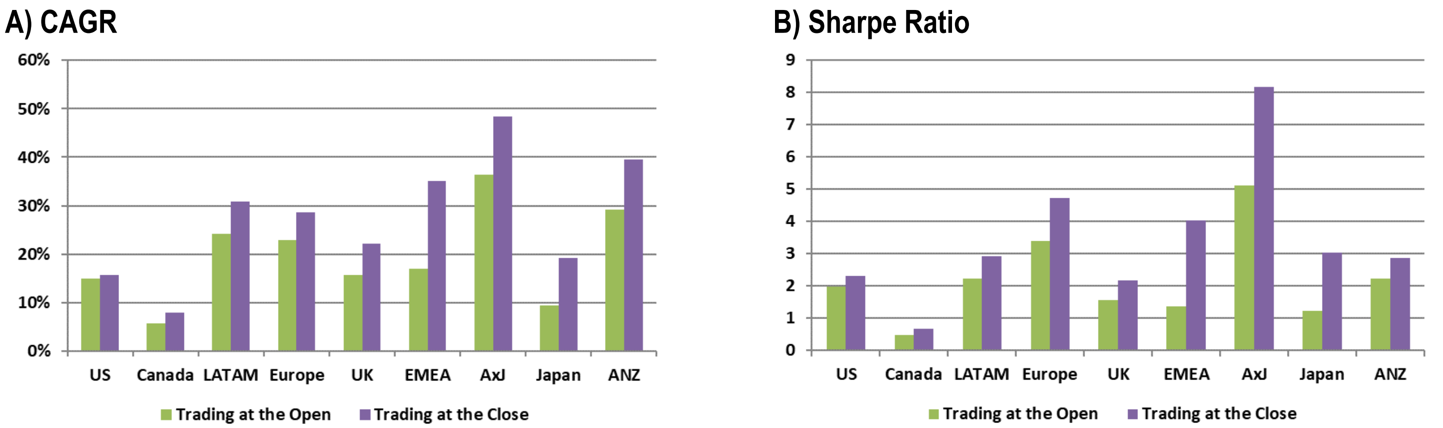
B) Sharpe Ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

To further illustrate the model development and trade execution alignment issue, as shown in Figure 71, the MALTA-StatArbClose model performance actually falls – in some cases, significantly – if we execute at the earlier opening price.

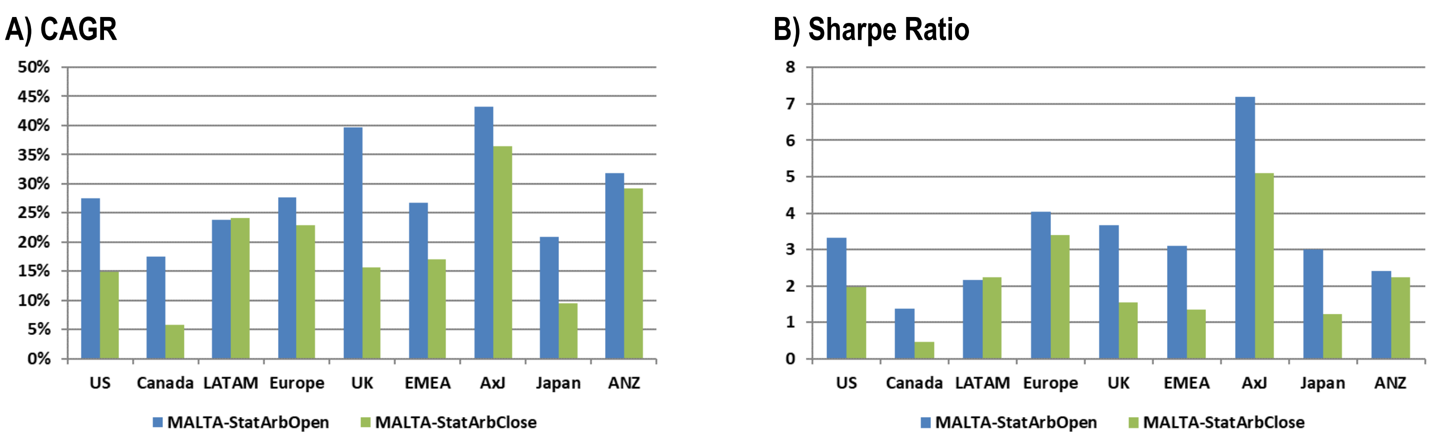
Figure 71 Performance Comparison, Open versus Close Trade Execution, MALTA-StatArbClose Model



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Similarly, as shown in Figure 72, the MALTA-StatArbClose model is not necessarily better than our original MALTA-StatArbOpen model. If we trade at the next day's open, the MALTA-StatArbClose actually trails behind the base case model (which is calibrated for an opening price execution) in all nine regions.

Figure 72 Performance Comparison, MALTA-StatArbOpen vs MALTA-StatArbClose, Trading at the Next Day's Open



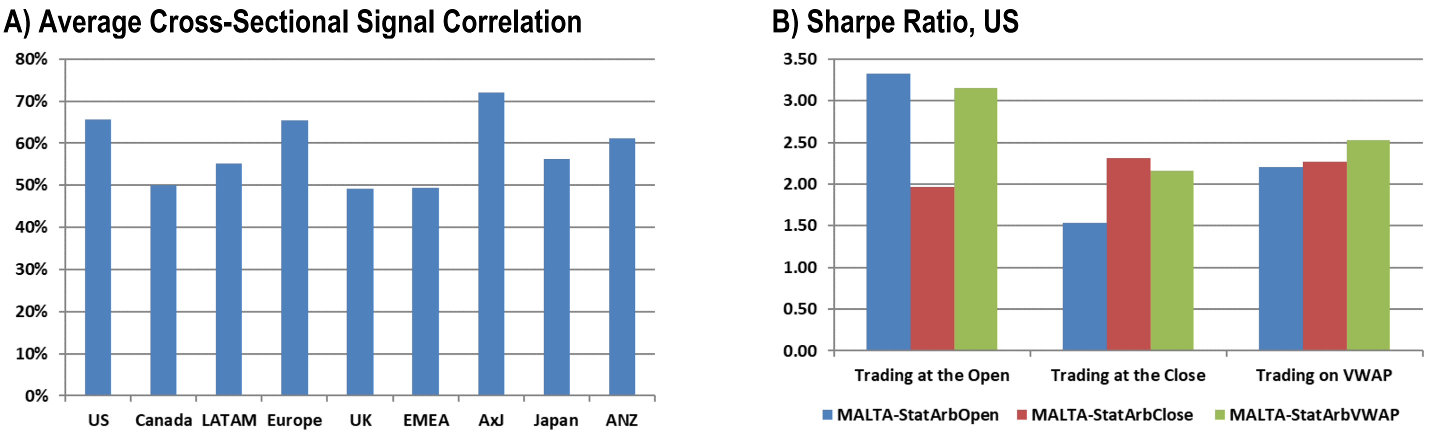
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Optimal Trade Execution

Lastly, since MALTA-StatArb and MALTA-StatArbClose models are only modestly correlated (see Figure 73A), the combined MALTA-StatArbVWAP model may be able to capture further diversification benefit. More importantly, instead of trading at the open or close, we could take advantage of a more passive VWAP (Volume Weighted Average price) algorithm, which has much larger liquidity than the open/close. As shown in Figure 73(B), the combined model delivers the best performance with a VWAP execution, while the proper aligned model development/trade execution still prevails in the other two cases (i.e., MALTA-StatArbOpen is optimized for open and MALTA-StatArbClose is best suited for close). Therefore, we would recommend the following intraday execution for our statistical arbitrage model:

- Execute a small portion (e.g., 10%) at the opening price, using the MALTA-StatArbOpen model
- Submit the majority of the trades (e.g., 70%) to a passive VWAP algorithm, based on the MALTA-StatArbVWAP model
- Complete the rest of the orders (e.g., the remaining 20%) at the closing price, based on the MALTA-StatArbClose model

Figure 73 Intraday Execution



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

Performance Boost with Daily Rebalance

As discussed in *Multi-Dimensional Alpha: Risk, Portfolio Construction, and Performance Attribution* (see Luo, et al [2017d]), keeping our total annual turnover the same, more frequent rebalance (e.g., monthly rather than quarterly) almost always improves performance. The performance boost is due to two reasons. First, the predictive power of almost all proprietary information decays over time; therefore, more frequent rebalances allow us to take advantage of more timely information. Secondly, for the same annual turnover, more frequent rebalances mean small trades at each rebalance, which lowers our market impact cost.

As a suite of statistical arbitrage models, the MALTA-StatArb can be implemented with intraday trading, daily rebalance, or weekly execution. Although the model is re-estimated once a week after Friday's market close, the model prediction is made once a day, with three outputs:

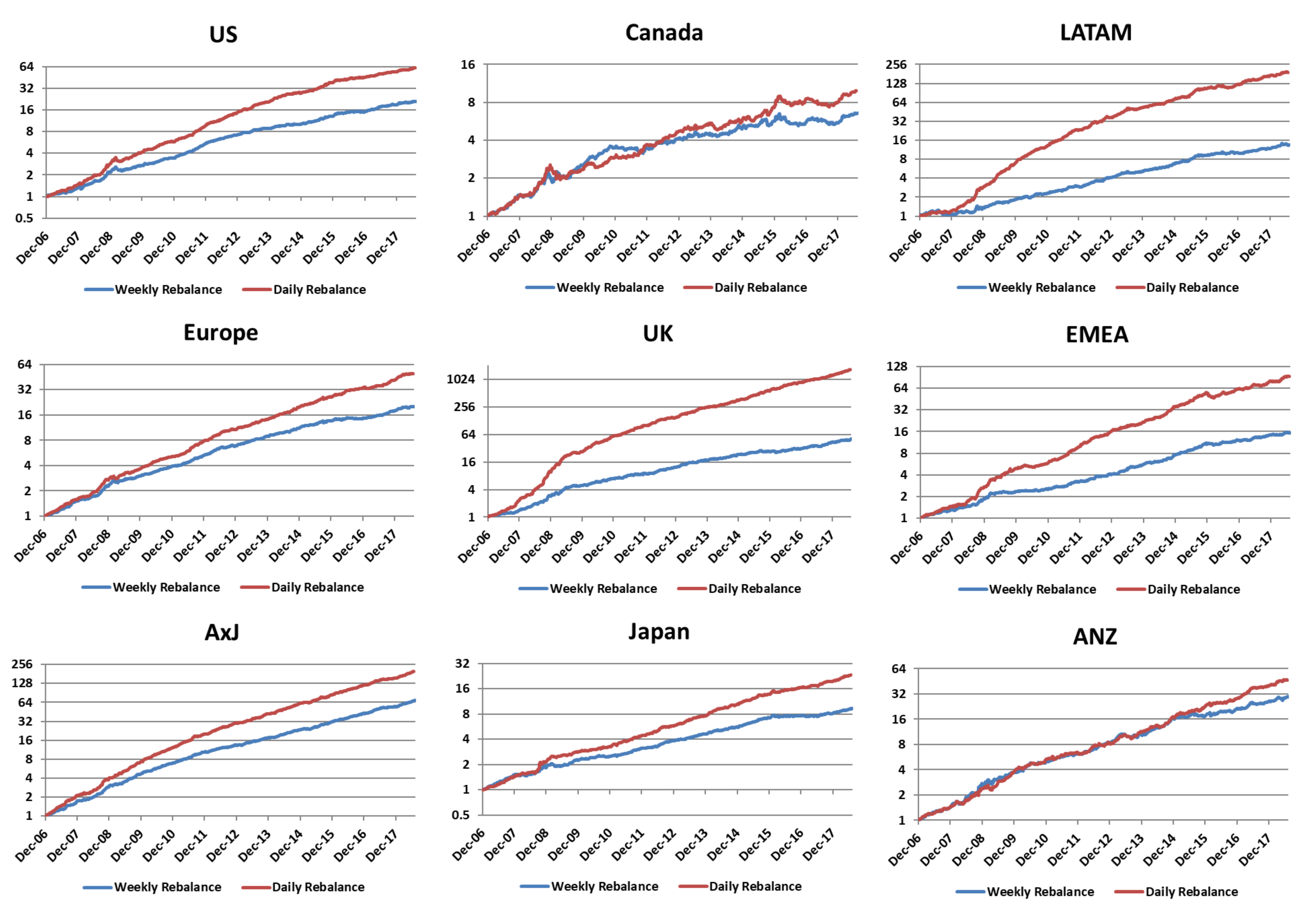
- MALTA-StatArbOpen – calibrated and designed for opening price trade execution
- MALTA-StatArbVWAP – combined model and optimized for intraday passive VWAP algorithm
- MALTA-StatArbClose – calibrated and ideally for closing price participation

As a demonstration, in this section, we compare a daily rebalance with a weekly rebalance to show the incremental value added from more frequent transactions, for each of the nine regions:

- Daily Rebalance. We use the MALTA-StatArbOpen model to trade at the opening price every business day, without any turnover constraint, buying the top 10% stocks (equally weighted) and shorting the bottom decile stocks (equally weighted).
- Weekly Rebalance. We use the MALTA-StatArbOpen model trade at the open, on the first business day of each week, without any turnover constraint, using the same long/short decile portfolio.

As shown in Figure 74, daily rebalanced portfolio consistently and considerably outperforms a lower frequency weekly implementation, in all regions of the world.

Figure 74 Cumulative Performance, Decile Portfolios



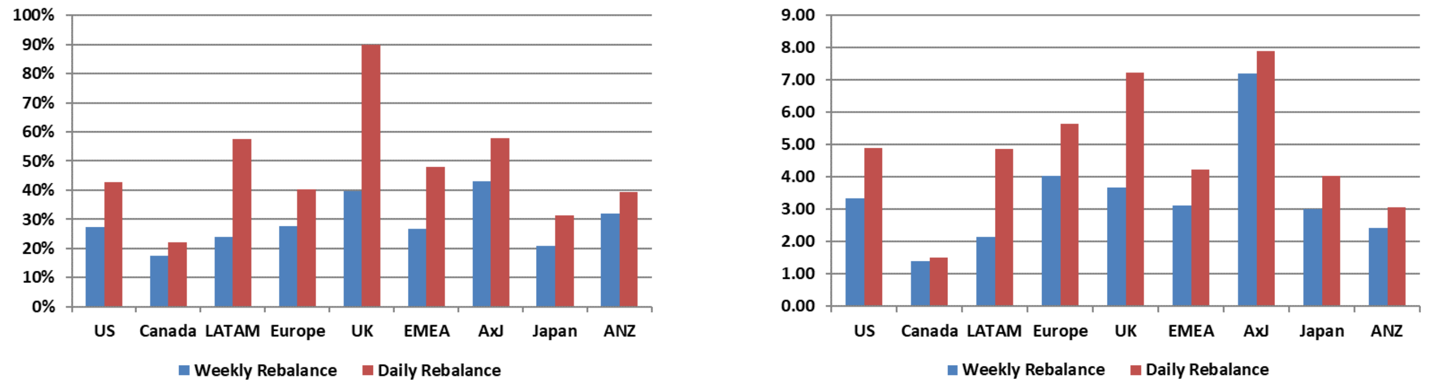
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

As shown in Figure 75, the daily rebalanced portfolio enjoys a much higher Shape ratio than the weekly traded strategy, especially in the UK, US, Europe, and LATAM.

Figure 75 Performance Boost with More Frequent Rebalance

A) Annualized Return

B) Sharpe Ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, IHS Markit, Wolfe Research Luo's QES

BIBLIOGRAPHY

Breiman, L. [1996]. “Bagging Predictors”, Machine Learning, 24(2), 123-140

Breiman, L. [2001]. “Random Forests”, *Machine Learning*, 45, 5-32

Breiman, L., Friedman, J., Olshen, R., and Stone, C. [1984]. *Classification and Regression Trees*, Chapman and Hall, New York

Bouman, S., and Jacobsen, B. [2002]. “The Halloween Indicator, Sell in May and Go Away: Another Puzzle”, *American Economic Review*, 92, 1618−1635

Carhart, M.M. [1997]. “On Persistence in Mutual Fund Performance”, Journal of Finance, 52, 57-82

Fama. E.F., and MacBeth, J. [1973]. “Risk, Return, and Equilibrium: Empirical Tests”, *Journal of Political Economy*, May-June 1973.

Fama, E., and French, K.R. [1993]. “Common Risk Factors in the Returns on Stocks and Bonds”, *Journal of Financial Economics* 33(1): 3-56.

Fama, E., and French, K.R. [1996]. “Multifactor Explanations of Asset Pricing Anomalies”, *Journal of Finance*, Vol. 51, 55-84.

Freund, Y. [1995]. “Boosting a Weak Learning Algorithm by Majority”, *Information and Computation*, 121, 256-285.

Friedman, J.H. [1999]. *Stochastic Gradient Boosting*, available at here

Grinold, R., and Kahn, R. [1999]. *Active Portfolio Management, A Quantitative Approach for Producing Superior Returns and Controlling Risk*, McGraw-Hill, 2nd Edition.

Jussa, J., Luo, Y., and Wang, S. [2017a]. “Quant CSI”, Wolfe Research Luo’s QES, March 13, 2017, available at here

Jussa, J., Wang, S., Rohal, G., Alvarez, M., and Luo, Y. [2017b]. “Unlocking Hidden Value”, Wolfe Research Luo’s QES, October 19, 2017, available at here

Jussa, J., Rohal, G., Luo, Y., Wang, S., and Alvarez, M. [2018a]. “Seeking Alpha from Insider Transaction”, Wolfe Research Luo’s QES, February 20, 2018, available at here

Jussa, J., Alvarez, M., Wang, S., Luo, Y., and Rohal, G. [2018b] “Stocks Most Exposed to Rising Interest Rates”, Wolfe Research Luo’s QES, March 20, 2018, available at here

Kearns, M., and Valiant, L. [1989]. “Cryptographic Limitations on Learning Boolean Formulae and Finite Automata”, in *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*

Luo, Y., Jussa, J., and Wang, S. [2017a]. “The Big and The Small Sides of Big Data”, Wolfe Research Luo’s QES, February 8, 2017, available at here

Luo, Y., Jussa, J., and Wang, S. [2017b]. “Signal Research and Multifactor Models”, Wolfe Research Luo’s QES, February 16, 2017, available at here

Luo, Y., Jussa, J., and Wang, S. [2017c]. “Style Rotation, Machine Learning, and The Quantum LEAP”, Wolfe Research Luo’s QES, February 24, 2017, available at here

Luo, Y., Jussa, J., and Wang, S. [2017d]. "Risk, Portfolio Construction, and Performance Attribution", Wolfe Research Luo's QES, May 9, 2017, available at here

Luo, Y., Alvarez, M., Jussa, J., Wang, S., and Rohal, G. [2017e]. "From Nowcasting to Forecasting", Wolfe Research Luo's QES, September 7, 2017, available at here

Luo, Y., Alvarez, M., Jussa, J., Wang, S., and Rohal, G. [2017e]. "Systematic Global Macro Investing – Launching our Global Macro Forecasts", Wolfe Research Luo's QES, September 7, 2017, available at here

Luo, Y., Alvarez, M., Jussa, J., Wang, S., and Rohal, G. [2018a]. "Why the Market Volatility has been so Low?", Wolfe Research Luo's QES, January 30, 2018, available here

Luo, Y., Rohal, G., Wang, S., Jussa, J., Alvarez, M., and Zhong, J. [2018b]. "Banking on the Banks – Welcome to BALI", Wolfe Research Luo's QES, September 4, 2018, available at here

Mantegna, R.N. [1999]. "Hierarchical Structure in Financial Markets", *European Physical Journal*, vol 11, 193-197

Rohal, G., Luo, Y., Jussa, J., and Wang, S. [2017a]. "Text Mining Unstructured Corporate Filing Data", Wolfe Research Luo's QES, April 20, 2017, available at here

Rohal, G., Luo, Y., Alvarez, M., Jussa, J., and Wang, S. [2017b]. "New Insights in Short Interest – A Global Perspective", Wolfe Research Luo's QES, August 15, 2017, available at here

Rohal, G., Luo, Y., Alvarez, M., Jussa, J., and Wang, S. [2018a]. "Tone at the Top? Quantifying Management Presentation", Wolfe Research Luo's QES, January 23, 2018, available at here

Rohal, G., Luo, Y., Jussa, J., Alvarez, M., Wang, S., and Zhong, J. [2018b]. "Pay for Performance", Wolfe Research Luo's QES, May 22, 2018, available at here

Rozeff, M. S., and Kinney, W. R. [1976]. "Capital Market Seasonality: The Case of Stock Returns". *Journal of Financial Economics*, October 1, 1976

Quinlan, R. [1993]. *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers

Schapire, R.E. [1990]. "The Strength of Weak Learnability", *Machine Learning*, 45, 197-227.

Schapire, R. E., Freund, Y., Bartlett, P., and Lee, W. S. [1998]. "Boosting the margin: A new explanation for the effectiveness of voting methods". *Annals of Statistics*, 26(5), 1651-1686.

Valiant, L. [1984]. "A Theory of the Learnable", Communications of the ACM, 27, 1134-1142

Wang, S., Luo, Y., Jussa, J., Alvarez, M., and Rohal, G. [2017a]. "Machine Learning Takeovers", Wolfe Research Luo's QES, September 12, 2017, available at here

Wang, S., Luo, Y., Alvarez, M., Jussa, J., and Rohal, G. [2017b]. "The Silk Road to China", Wolfe Research Luo's QES, December 1, 2017, available at here

Wang, S., Luo, Y., Jussa, J., Alvarez, M., and Rohal, G. [2018a] "Systematic Alpha from Risk Arbitrage (SARA)", Wolfe Research Luo's QES, May 15, 2018, available here

Wang, S., Luo, Y., Jussa, J., Alvarez, M., and Rohal, G. [2018b] "Global Systematic Risk Arbitrage – SARA Global", Wolfe Research Luo's QES, August 28, 2018, available here

Wang, S., Ai, H., Yamashita, T., & Lao, S. [2010]. “Combined top-down/bottom-up human articulated pose estimation using Adaboost learning”. In *Pattern Recognition* (ICPR), 2010 20th International Conference on (pp. 3670-3673). IEEE.

Wu, B., Ai, H., Huang, C., & Lao, S. [2004]. “Fast rotation invariant multi-view face detection based on real adaboost”. In Automatic Face and Gesture Recognition, 2004. *Proceedings. Sixth IEEE International Conference* on (pp. 79-84). IEEE.

DISCLOSURE SECTION

Analyst Certification:
The analyst of Wolfe Research primarily responsible for this research report whose name appears first on the front page of this research report hereby certifies that (i) the recommendations and opinions expressed in this research report accurately reflect the research analysts' personal views about the subject securities or issuers and (ii) no part of the research analysts' compensation was, is or will be directly or indirectly related to the specific recommendations or views contained in this report.

Other Disclosures:
Wolfe Research, LLC does not assign ratings of Buy, Hold or Sell to the stocks it covers. Outperform, Peer Perform and Underperform are not the respective equivalents of Buy, Hold and Sell but represent relative weightings as defined above. To satisfy regulatory requirements, Outperform has been designated to correspond with Buy, Peer Perform has been designated to correspond with Hold and Underperform has been designated to correspond with Sell.

Wolfe Research Securities and Wolfe Research, LLC have adopted the use of Wolfe Research as brand names. Wolfe Research Securities, a member of FINRA (www.finra.org) is the broker-dealer affiliate of Wolfe Research, LLC and is responsible for the contents of this material. Any analysts publishing these reports are dually employed by Wolfe Research, LLC and Wolfe Research Securities.

The content of this report is to be used solely for informational purposes and should not be regarded as an offer, or a solicitation of an offer, to buy or sell a security, financial instrument or service discussed herein. Opinions in this communication constitute the current judgment of the author as of the date and time of this report and are subject to change without notice. Information herein is believed to be reliable but Wolfe Research and its affiliates, including but not limited to Wolfe Research Securities, makes no representation that it is complete or accurate. The information provided in this communication is not designed to replace a recipient's own decision-making processes for assessing a proposed transaction or investment involving a financial instrument discussed herein. Recipients are encouraged to seek financial advice from their financial advisor regarding the appropriateness of investing in a security or financial instrument referred to in this report and should understand that statements regarding the future performance of the financial instruments or the securities referenced herein may not be realized. Past performance is not indicative of future results. This report is not intended for distribution to, or use by, any person or entity in any location where such distribution or use would be contrary to applicable law, or which would subject Wolfe Research, LLC or any affiliate to any registration requirement within such location. For additional important disclosures, please see www.wolferesearch.com/disclosures.

The views expressed in Wolfe Research, LLC research reports with regards to sectors and/or specific companies may from time to time be inconsistent with the views implied by inclusion of those sectors and companies in other Wolfe Research, LLC analysts' research reports and modeling screens. Wolfe Research communicates with clients across a variety of mediums of the clients' choosing including emails, voice blasts and electronic publication to our proprietary website.

This report is limited for the sole use of clients of Wolfe Research. Authorized users have received an encryption decoder which legislates and monitors the access to Wolfe Research, LLC content. Any distribution of the content produced by Wolfe Research, LLC will violate the understanding of the terms of our relationship.