WOLFE RESEARCH

LUO'S QES RESEARCH

Quantitative Research
Economics
Portfolio Strategy

MULTI-DIMENSIONAL ALPHA

February 24, 2017

STYLE ROTATION, MACHINE LEARNING, AND THE QUANTUM LEAP

QES Handbook of Active Investing, Part III

- **Advanced Topics in Systematic Equity Modeling.** This paper forms the third part of our *QES Handbook of Active Investing* series. In Part I (*The Big and the Small Sides of Big Data*), we discussed data contents and data science. In Part II (*Signal Research and Multifactor Models*), we reviewed the methodologies of factor backtesting and multifactor modeling techniques. In this research, we address some of the more advanced topics. In particular, we focus on how to incorporate style rotation/factor timing and machine learning in equity models, and finally introduce the first of our global stock-selection models – the LEAP (L-Economic Alpha Processing).
- **Factor Timing, Style Rotation, and Macro Overlay.** The average return of most conventional alpha factors has declined significantly in recent years, while volatility has increased and furthermore, downside risk has spiked. Factor returns are no longer stationary, subject to periodic structural breaks, and sensitive to risk-on and risk-off regimes. On the positive side, it does provide opportunities for factor timing. We show how global macro research techniques can be used in systematic style rotation.
- **Machine Learning.** Machine learning is not about random forest, deep learning, or any fancy algorithm. Rather, it marks a fundamental shift in investment philosophy. A precise function form is rarely assumed, a thorough search of pattern recognition algorithms is highly encouraged, and parameter tuning is part of the model building. More importantly, there are a range of techniques (e.g., resampling, bagging, and boosting) in machine learning to ensure the backtesting results are as truthful as possible.
- **The Quantum LEAP.** Lastly, we are excited to introduce our first global stock-selection model – LEAP (L-Economic Alpha Processing) model. The LEAP model takes advantage of all the unique features in our research infrastructure (from Big Data contents, signal research, style rotation to machine learning). The model is based on our experience in the past 16 years and has exceptional performance in all regions. We also offer data feeds based on the model. Please contact us for details.

Source: gettyimages.com

Yin Luo, CFA, CPA
YLuo@wolferesearch.com

Sheng Wang
SWang@wolferesearch.com

Javed Jussa
JJussa@wolferesearch.com

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

Table of Contents

A LETTER TO OUR READERS

Welcome to Part III of the QES Handbook series

This is Part III of our initial launch paper and the third in the QES Handbook series. In Part I – *The Big and the Small Sides of Big Data*, we provided an overview of the 30 year history of active investing. We discussed the Big Data evolution, the vast different contents that are available to investors, and data modeling. In the second paper (see Luo, et al [2017b]), we elaborated on how to translate from data into knowledge, i.e., how to conduct single factor backtesting and more importantly, how to build multifactor models.

Conventional stock selection models rely on a number of unrealistic assumptions, e.g., factor returns are linear in nature and stable over time. In this paper, we introduce a few advanced topics that are seldom covered in academic or practitioners' research – factor timing, factor selection, machine learning, and the next generation of alpha models.

As shown in Luo, et al [2017a], the average return of most conventional alpha factors has declined significantly in recent years, while volatility has increased and furthermore, downside risk has spiked. Factor returns are no longer stationary, subject to periodic structural breaks, and sensitive to risk-on and risk-off regimes. On the positive side, it does provide opportunities for factor timing. Style rotation has always been a controversial topic within the quantitative investing community. The opponents argue that the breadth of style rotation is limited (as we have far fewer factors than stocks), timing is difficult (if not impossible), time series data history is short, and factor timing increases portfolio turnover. However, since the performance of style rotation models tends to be uncorrelated to the excess return from stock selection strategies, even a modest skill can still provide great diversification benefit. Discretionary factor selection and style timing is still the dominant approach in the industry. The biggest problem with such an *ad hoc* process is that it is not reproducible, which violates the first principle of systematic investing. We shall focus our research on systematic style rotation.

The vast majority of finance research relies on linear models. Long/short hedged portfolio backtesting, mean-variance optimization, and linear multifactor models are easy to implement, traceable, and intuitive. As shown in Luo, et al [2017a, 2017b], the linearity assumption is increasingly being challenged. Many investors are still suspicious about machine learning in finance. Moreover, machine learning is often confused for data mining or data snooping. We would argue that the current practice of investment research is more sensitive to data mining than a properly designed machine learning model. In practice, analysts almost always search through a large number of factors, models and tune the parameters until they find what they want to see. While the data backtesting is out-of-sample, the model searching is essentially in-sample. Machine learning is completely different – a precise function form is rarely assumed, a thorough search of pattern recognition algorithms is highly encouraged, and parameter tuning is part of the model building. There are a wide range of techniques in machine learning to mitigate overfitting, e.g., dividing the data into trading, testing, and validation samples, resampling, bagging, and boosting. Although most people perceive machine learning as a mountain of algorithms from random forest or deep learning, it is not about any specific technique. Rather, machine learning marks a fundamental shift of research philosophy.

Now, after everything that we have shown, the question is "so what?". Can we really do better than the average? Can we beat the flip of a coin? In this paper, we introduce one of our flagship global stock-selection models – the LEAP (L-Economic Alpha Processing). The LEAP model takes advantage of our cutting edge technology infrastructure, parallel computing, deep factor library and access to Big Data, style rotation and macro overlay, and machine learning. More importantly, the LEAP model also emphasizes full transparency and strives to be reasonably intuitive. A full-blown machine learning based model will be introduced in the next few months in a separate paper.

The LEAP model reflects our 16-year experience in quantitative modeling and a decade of experience in applying machine learning in active investing. The LEAP model has a Sharpe ratio of 4.0x in the US and 4.7x in Europe – considerably higher than a typical multifactor model. The downside risk of the LEAP model is materially lower than our benchmark strategy. Furthermore, it demonstrates decent performance in the US and Japan (the two most difficult markets for alpha generation), in March-May 2009 (the largest drawdown witnessed by many quant funds), and in 2016 (the most recent painful experience). Lastly, for global equity portfolio managers, the LEAP models are uncorrelated across different regions, offering far greater diversification benefit than traditional quantitative or fundamental active strategies.

The LEAP model covers over 11,000 stocks globally, in nine regions (US, Canada, Europe, UK, AxJ, Japan, Australia and New Zealand, LATAM, and emerging EMEA), all on a daily basis. We plan to offer a data feed of the LEAP model alpha in the near term. Please contact us for details.

Any feedback and suggestion are more than welcome!

Regards,

Yin, Javed, Sheng, Gaurav, Kartik, and Luo's QES team

FACTOR TIMING, STYLE ROTATION AND MACRO OVERLAY

In this section, we provide a brief overview of our style rotation model. A more in-depth discussion on our global macro research (including style rotation) will be covered in a series of forthcoming research papers.

In Academia, style timing, along with market timing, has always been a controversial topic. There were a few rather dated papers on style rotation prior to 2008 (e.g., Bernstein [1995], Sorensen and Lazzara [1995], Kao and Shumaker [1999], Levis and Liodakis [1999], and Asness et al. [2000]). Similarly, in the practitioners' world, factor selection was mostly based on manager's discretion and factor weighting was primarily static in the same pre-2008 period.

The 2007 quant crisis and the subsequent 2008 global financial crisis have triggered a round of strong interest on factor timing or style rotation in academic research. For example, Limthanakom and Collver [2010] document style momentum and find macroeconomic variables have predictive power of future style returns. Ardia, Boudt, and Wauters [2016] argue for the economic benefits of timing style factors. There remains considerable suspicion on the possibility of style timing. For example, Corbett [2016] finds that US equity mutual fund managers who frequently change styles (market, size, value and momentum) underperform their peers.

THE BASICS

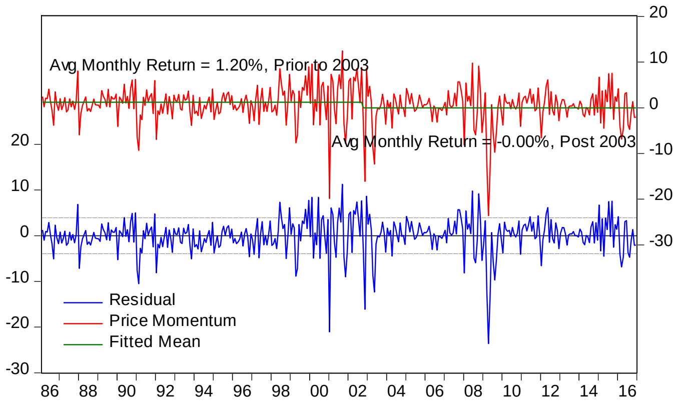
As we elaborate with great details in Luo, et al [2017a, 2017b], there is strong evidence suggesting that:

- Factor returns are time varying;
- The average return across factors has been declining; and
- Factor returns have periodic structural breaks and large outliers

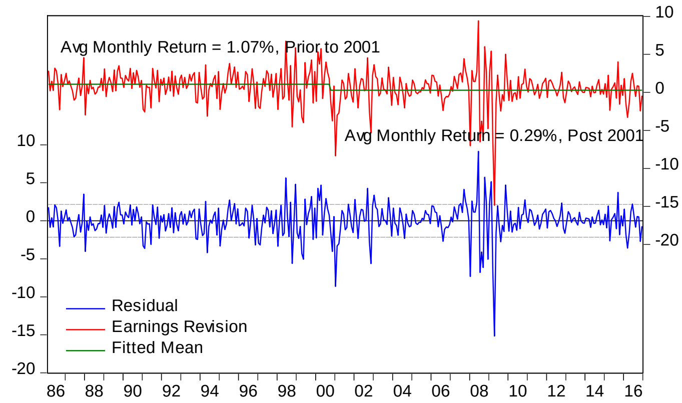
Figure 1 (A) and (B) show the breakpoint regression (see Bai and Perron [1998]) on the time series return of value (earnings yield) and price momentum (12M-1M total return) in the US. It is evident that both time series have at least one breakpoint. The average monthly returns of both factors were around 1% in the 1980s/1990s, but are approaching 0% in recent years.

Figure 1 The Structural Changes in Factor Performance

A) Price Momentum



B) Earnings Revision



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

WHY DOES IT MATTER?

To measure the upside from factor timing, we show a simple simulation. We start from the same eight factors in the benchmark model (see Luo, et al [2017b]):

- 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 earnings 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
- Quality – Profitability: Return on equity – we like firms with high ROEs
- Quality – Leverage: Debt/Equity ratio – we prefer companies with low financial leverage
- Quality – Earnings Quality: Sloan's accruals – we buy companies with low accruals

To demonstrate the potential upside from style rotation, we create a perfect foresight model, assuming that we have 100% accuracy in predicting the next month's factor return. Essentially, at each month end, we conduct a Grinold & Kahn optimization (essentially a mean-variance optimization on the factor space):

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

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

There are two input parameters for the Grinold & Kahn optimization – the vector of expected factor IC ($\widetilde{IC}$) and the covariance matrix (Σ_{IC}).

For the perfect foresight model, we set the vector of predicted factor IC ($\widetilde{IC}$) as the actual next month's factor IC. This essentially assumes that we are 100% correct in predicting future factor performance. The risk model, i.e., the factor covariance matrix is estimated using a 10-year rolling window sample covariance matrix. Therefore, we assume that we have perfect skill in predicting factor return, but no special skill in forecasting risk.

For this optimization, we do not add any minimum weight constraint. Therefore, we can use the close-form formula $\widehat{\omega_t} = \Sigma_{IC}^{-1} IC_{t+1}$ to derive factor weights. Lastly, we use this perfect foresight model as our stock-selection model and track the model performance over time.

Compared to our benchmark model BM[1] (see Figure 2 A), the perfect foresight model's performance (as measured by risk-adjusted IC) is almost eight times higher (see Figure 2 B). If we could predict factor performance with 100% precision, our stock selection model never had a single down month in the past 23 years.

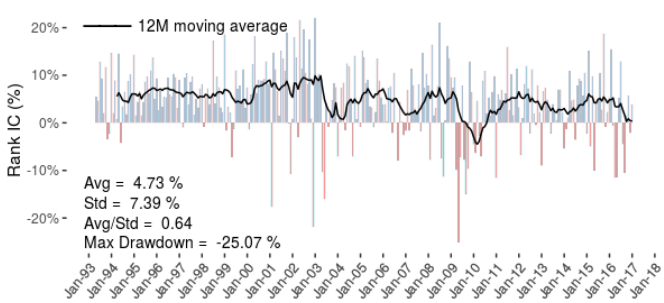
Globally, we observe the same impact. The perfect foresight model boosts the risk-adjusted IC by 300%-700% (see Figure 2 C) and Sharpe ratio by 300%-900% (see Figure 2 D) in the nine regions.

Factor weights in the perfect foresight model change dramatically from month to month, with very high turnover (see Figure 2 E). The average signal autocorrelation drops by over 80% from the benchmark model (see Figure 2 F). This once again highlights the point that, if we do have great predictive power, the portfolio is most likely to have high turnover. Therefore, high turnover itself is neither a friend nor foe. It critically depends on our skill, transaction costs, size of the portfolio and a few other parameters.

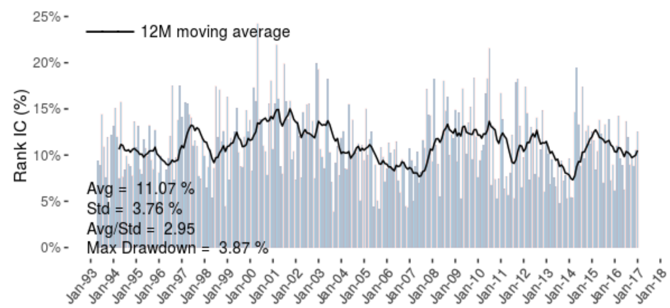
[1] The benchmark model (BM) equally weighs the eight factors. Detailed can be found in Luo, et al [2017b].

Figure 2 The Upside Potential from Style Rotation

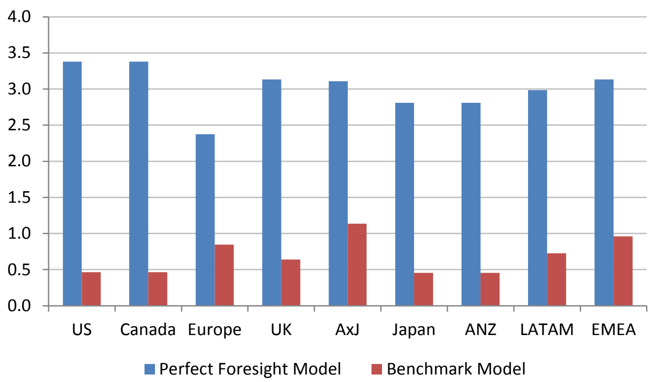
A) Benchmark Model without Style Rotation, US



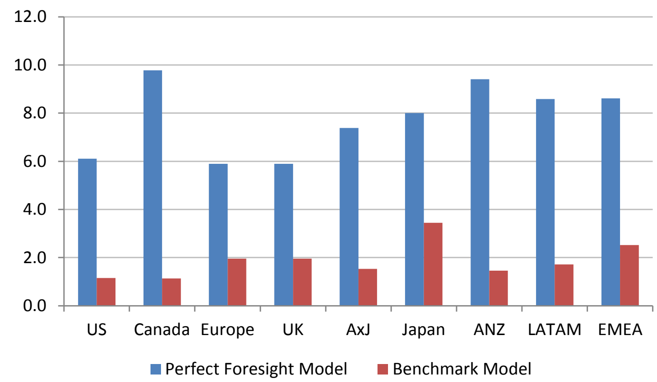
B) Perfect Foresight Model, US



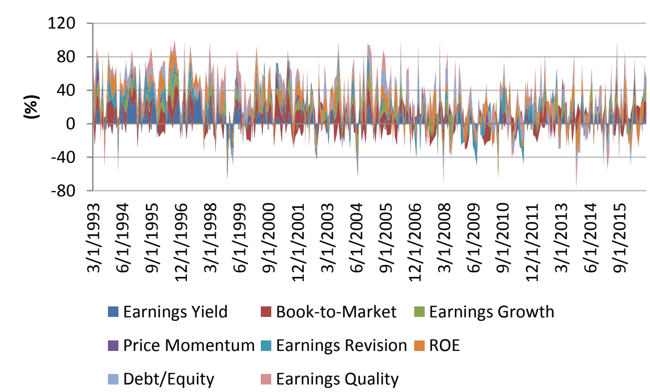
C) Global Comparison, Risk-Adjusted IC



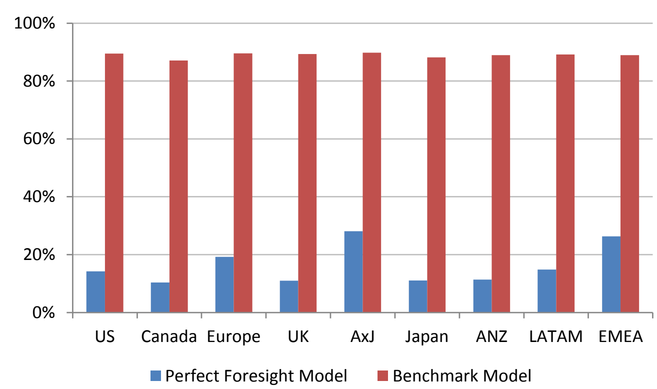
D) Global Comparison, Sharpe Ratio



E) Factor Weights in the Perfect Foresight Model, US



F) Factor Autocorrelation, Global



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

MODELING TECHNIQUES

Broadly speaking, there are two modeling techniques for style rotation – cross sectional and time series. The most powerful approach to model style factor returns are generally different from the ones used for stock selection. We have a far smaller number of factors than stocks. A cross-sectional ranking is likely to be difficult. Therefore, time series regression techniques are likely to be more effective.

For demonstration purpose, in this section, we choose 15 common style factors to show how style rotation can be applied in practice:

- Earnings Yield, Trailing
- Dividend Yield
- Book-to-Market
- EBITDA/EV
- Expected 5Y EPS Growth
- Expected FY1/FY0 EPS Growth
- Price Momentum (12M-1M)
- Price Reversal (1M)
- Earnings Revision (FY1 EPS, 3M)
- ROE
- Debt/Equity Ratio
- Earnings Quality (Sloan's Accruals)
- Beta
- Amihud Illiquidity
- Size (Log Market Capitalization)

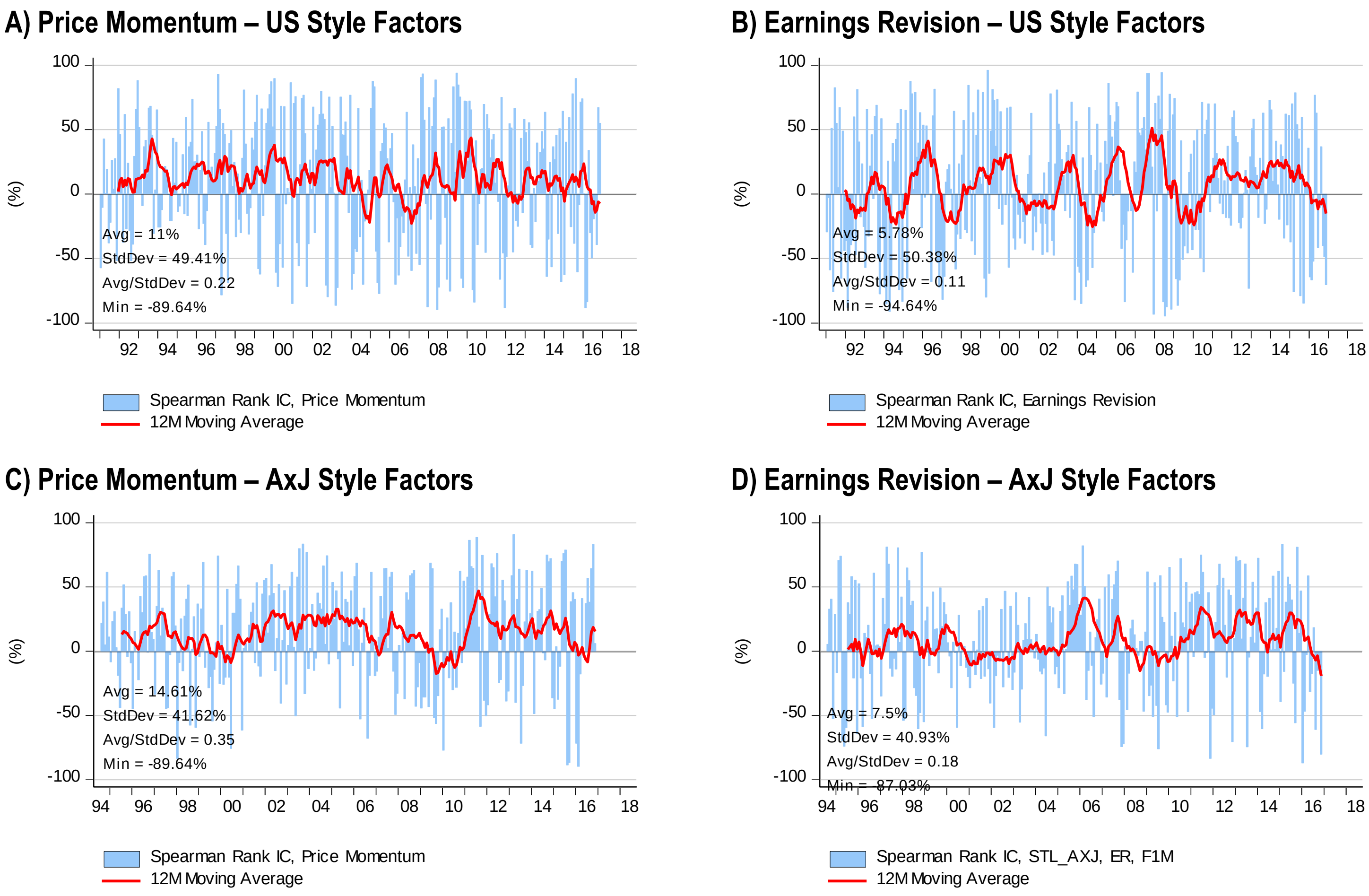
For each of the above 15 factors, we construct a long/short quintile portfolio, neutralized for country and sector effect. Stocks are equally weighted in both long and short sides. We track the 15 factor portfolios, for each of the nine regions (US, Canada, Europe, UK, AxJ, Japan, ANZ, LATAM, and EMEA).

Cross-Sectional Approach

Using the cross sectional approach, the global macro modeling process mirrors the bottom-up stock selection. We perform our analysis cross sectionally, for all 15 style factors at the same point in time. The predictors used in cross sectional analysis are typically asset specific, e.g., the valuation and price momentum of each style factor. For example, Zaremba and Szyzka [2016] document significant factor momentum in emerging markets, using the Polish equity market as an example.

Figure 3 (A) to (D) show the performance of price momentum and earning revisions in equity factor rotation, from a cross sectional context. The price momentum factor is defined as the cumulative return of each style portfolio over the past 12 months, similar to how price momentum is defined at the stock level. Similarly, earnings revision is defined the difference of median earnings revision at the long quintile and short quintile portfolio. Both factors appear to have reasonable predictive power of next month's style portfolio returns, in both US and AxJ. However, we need to keep in mind that the breadth of our investment universe is rather limited – we only have 15 style portfolios.

Figure 3 The Predictive Power of Momentum and Earnings Revision in Style Rotation



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

Time Series Approach

The time series approach models each asset classes (e.g., style factors) individually. The predictors used in time series models can be applied to most asset classes, e.g., GDP growth, inflation, yield spread, and other economic and capital market variables. We form our return forecast for each asset classes independently[2]. Then we compare the return predictions of each asset class jointly to build our portfolio. Using a simple regression model as an example, the typical time series model can be specified as:

[2] Time series analysis offers a large selection of modeling techniques. We are certainly not limited to single equation models. All style portfolios can be modeled jointly, via SUR (Seemingly Uncorrelated Regression), VAR (Vector Autoregressive), VECM (Vector Error Correction Model), etc. Detailed discussion will be covered in a forthcoming research paper.

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

Where,

$f_{i,t}$ is the return of style portfolio i at time t;

$\alpha_{i,t}$ and $\beta_{i,k,t}$ are the coefficients to be estimated by the regression, for factor portfolio i at time t;

$E_{k,t-1}$ is the kth macro variable at time $t-1$;

K is the number of predictors; and

$\varepsilon_{i,t}$ is the regression residual of asset i at time t.

In a simple example that we want to estimate the relationship between value premium (e.g., the return of the earnings yield factor) on one macro variable, say, the lagged value of industrial production (MoM percentage change). When we estimate the model for momentum factor return, we still use the same predictor – the MoM change in industrial production. The intercept ($\alpha_{i,t}$) and slope ($\beta_{i,k,t}$) are different for value versus momentum factors.

The functional form does not have to be linear and the estimation technique is not limited to ordinary least squares. A general setup of the model is:

$$f_{i,t} = \varphi\left(\boldsymbol{\theta}_{i,t}\right)\boldsymbol{E}_{t-1} + \varepsilon_{i,t}$$

Where,

$\varphi(.)$ Is the functional form;

$\boldsymbol{\theta}_{i,t}$ is a vector of model parameters to be estimated empirically, for of asset i at time t; and

$\boldsymbol{E}_{t-1}$ is the $(K \times 1)$ vector of macro variables at time $t-1$.

The time series model is estimated by each asset class, using either a rolling window or an expanding window. Once the model parameters are estimated, the prediction of return for the next period, i.e., $t+1$ is therefore:

$$\hat{f}_{i,t+1} = \hat{\alpha}_{i,t} + \sum_{k=1}^{K} \hat{\beta}_{i,k,t} E_{k,t}$$

Where,

$\hat{f}_{i,t+1}$ is the predicted return for asset i at time t; and

$\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,k,t}$ are the estimated model coefficients.

Similarly, a more general form of prediction is:

$$\hat{f}_{i,t+1} = \varphi\left(\widehat{\boldsymbol{\theta}}_{i,t}\right)\boldsymbol{E}_{t}$$

Where, $\widehat{\boldsymbol{\theta}}_{i,t}$ is the vector of estimated model parameters.

Global Macro Database

Similar to the global equity database discussed in Luo, et al [2017a], we also have a parallel system for our global macro data. A thorough discuss of our global macro database is beyond the scope of this research and will be covered in a forthcoming research. There are a large number of potential macro variables to predict style factor returns. In this section, we briefly discuss a few examples.

Policy Uncertainty

Political uncertainty dominated the world in 2016, with Britain's shocking exit from the European Union, the surprising result of the US election, and the multiple geopolitical events around the world. The world in 2017 is set to surpass 2016, with multiple proposed profound changes by the new US administration, German and French elections, and the potential trade conflicts regarding to NAFTA, TPP, US/China, and the list goes on.

In a series of papers, three economic professors (see Baker, Bloom, and Davis [2015]) construct a suite of policy-related economic uncertainty indices for the major economies, e.g., US, Canada, Europe, UK, China, Japan, etc. The policy uncertainty indices are mostly constructed using key word searches[3] from each country/region's leading newspapers. As shown in Figure 4 (A) and (B), the policy uncertainty indices have reached all-time highs in almost all major countries. The policy uncertainty indices all have a positive serial correlation (i.e., high uncertainty tends to be followed by high risk in the near future). In the longer term, they show mean-reversal patterns[4].

For demonstration purpose, we further fit a Markov Regime Switching (MRS) model on the US policy uncertainty index. Figure 4 (C) shows the regime classification, where the highlighted areas indicate "high uncertainty". Each high uncertainty regime lasts for about 11 months on average (while low uncertainty regimes are about 18 months). It is evident that the risk-on/risk-off regime switching occurs more frequently around and after recessions and major geopolitical events.

The returns of our 15 style portfolios are significantly different in high versus low uncertainty regimes. Initially, it might be surprising to note that the defensive styles (e.g., dividend yield, low beta, and price momentum) actually perform substantially worse, while cyclical factors (e.g., small cap, book-to-market) generate superior returns in a high uncertainty environment (see Figure 4 D). Investors are probably either aware of the mean-reversal nature of policy uncertainty, or overly complacent at turning points. When uncertainty reaches an extremely high level, managers expect risk to come down and embrace risk-on styles.

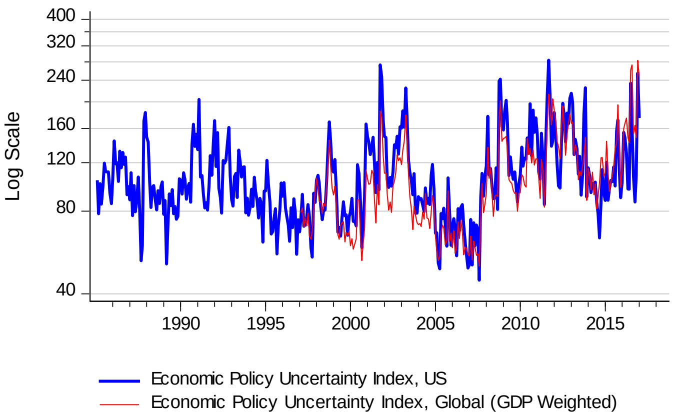
The distribution of policy uncertainty index is highly skewed to the right, meaning that we are more likely to see extremely high uncertainty than what implied by a normal distribution (see Figure 4 E). Lastly, As shown in Figure 4 (F), in ultra-high uncertainty periods (defined as above the two standard deviation band), investors become even more contrarian and start to chase cyclical styles such as size (small cap) and book-to-market and penalize low beta and price momentum. Mean reversal or StatArb styles tend to perform well in a risk-on/risk-off environment.

[3] For example, the key words can be "uncertain", "uncertainty", "economic", "economy" and other policy-relevant terms. For US, the authors also incorporate number of tax code provisions scheduled to expire over the next 10 years and the estimate dispersions from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters.

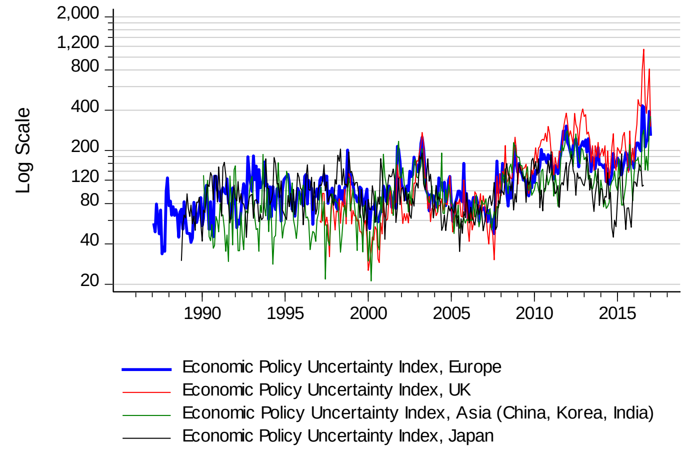
[4] An augmented Dickey-Fuller test strongly rejects the Null hypothesis of a unitroot process, suggesting the policy uncertainty indices as mean-reversal.

Figure 4 Political Uncertainty and Factor Performance

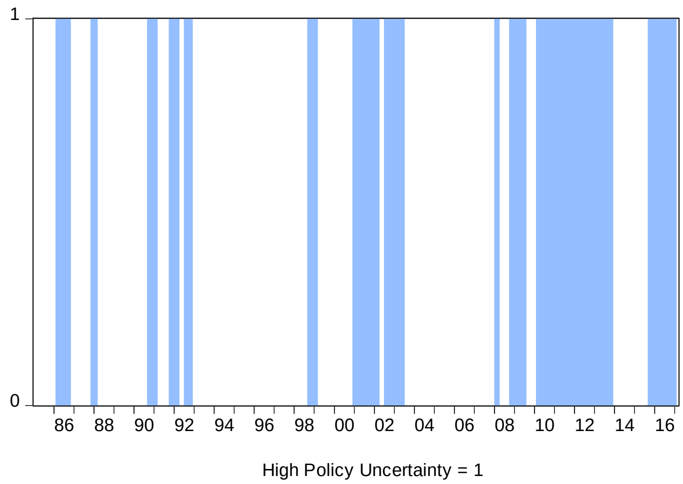
A) Policy Uncertainty Index in the US



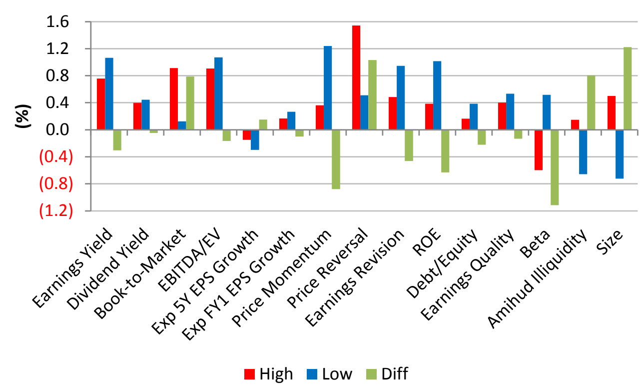
B) Policy Uncertainty Index around the World



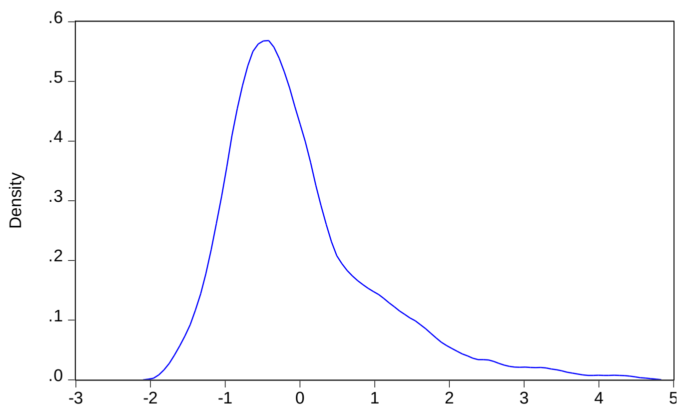
C) Policy Uncertainty Regime in the US



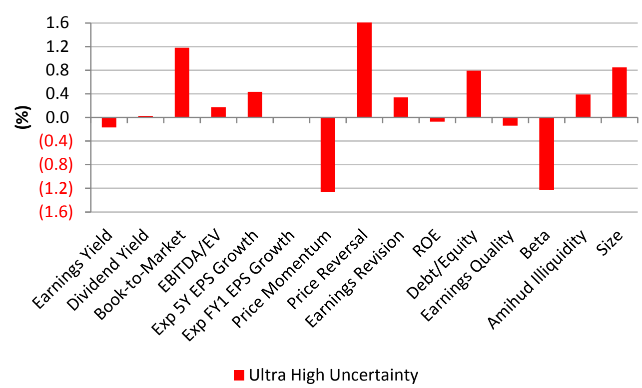
D) Factor Performance under Policy Uncertainty Regimes



E) The Distribution of Policy Uncertainty Index



F) Factor Performance in Ultra High Uncertainty Periods



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

Nowcasting and Economic Cycle

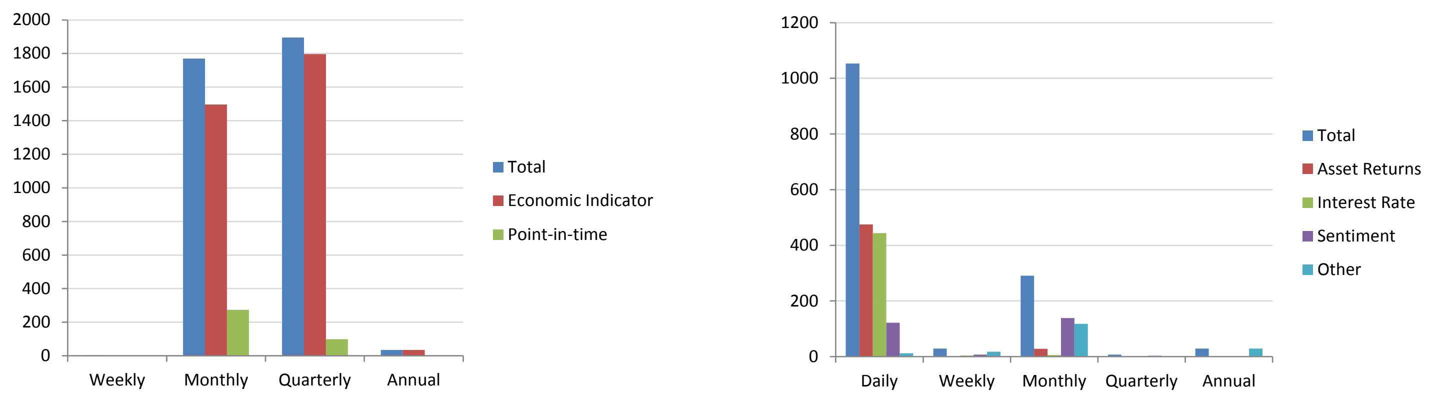
One of the most actively researched fields in macroeconomics in recent years is Nowcasting. Nowcasting is a combination of now and forecasting. It has been used for a long time in meteorology and recently ported into economics. It is about predicting the present, the recent past and the near future. The classic example is GDP.

US quarterly GDP typically comes in three flavors – advance (release toward the end of the month after the quarter end), revised (the second month after quarter end), and final (the third month after quarter end). Therefore, before we can even forecast the next quarter's GDP, as of the quarter end, we barely just have the data for the previous quarter's GDP and do not even know the current quarter yet.

Beyond quarterly GDP, most economic data series are of monthly frequency. There are also weekly or even daily data, especially financial market data. As shown in Figure 5 (A), monthly and quarterly frequencies account for the vast majority of economic variables, which is very different from capital market variables (mostly in daily frequency, see Figure 5 B). How to consistently model data series of different frequencies has been one of the biggest challenges in macroeconomic research.

Figure 5 Frequency of Economic and Capital Market Data

A) Economic Data **B) Capital Market Data**



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

The basic principle of Nowcasting is to get a better and accurate estimate of the state of the economy, as increasingly more information becomes available. Market participants monitor many economic data series, form expectations, and revise the assessment whenever realizations diverge significantly from prior views.

We will introduce our macroeconomic modeling and forecasting in a forthcoming research. Here, let's show an example of our economic growth Nowcasting index[5], which summarizes all key economic indicators released every day into one indicator.

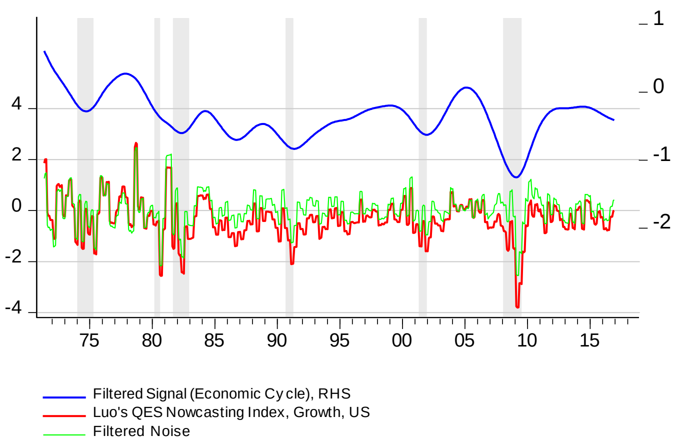
To extract the underlying signal (i.e., economic cycle) from noise, we apply the Hodrick-Prescott filter (see Hodrick and Prescott [1997]). Figure 6 (A) and (B) show our Nowcasting economic growth index

[5] We have four Nowcasting indices (growth, anticipated growth, inflation, and employment) for about 40 countries and regions.

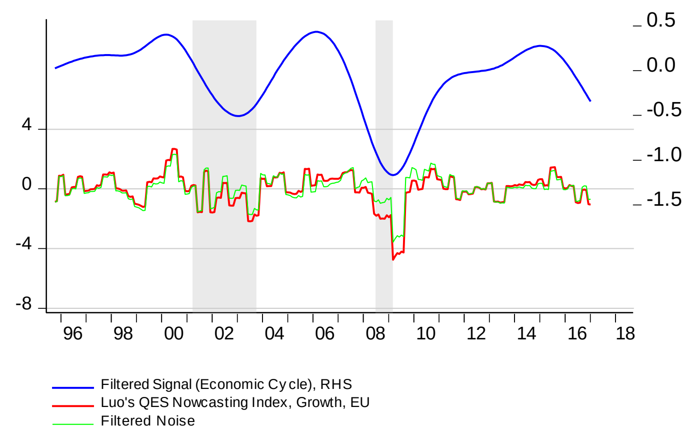
for the US and Europe[6]. US economic growth is at the peak level. In Europe, economic environment remains challenging.

Figure 6 Economic Nowcasting Index

A) US Growth



B) European Growth



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

To understand the predictive ability of our Nowcasting index, we conduct two simple linear regressions:

$$f_{i,t} = \varphi_{i,0} + \varphi_{i,1}E_t + \epsilon_{i,t} \quad [1]$$

And

$$f_{i,t+1} = \varphi_{i,0} + \varphi_{i,1}E_t + \epsilon_{i,t} \quad [2]$$

Where,

$f_{i,t}$ is the return of style factor i at time t, and

E_t is our nowcasting economic growth index at time t

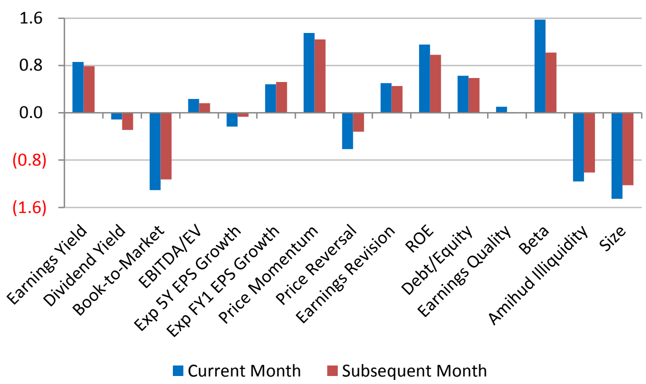
The first regression [1] reveals the contemporaneous relationship between economic growth and factor performance, while the second equation [2] states whether the current economic situation can predict the next month's factor return.

As shown in Figure 7 (A), our US Nowcasting index is highly correlated to current month's and equally predictive to next month's factor returns. Similar to what we observe for policy uncertainty, the relationship between economic growth and factor return is also contrarian in nature – when economic growth is strong, defensive factors (e.g., low beta, price momentum) tend to perform well, while book-to-market, dividend yield, and price reversal styles are more likely to survive in economic downturns. The pattern is similar in Europe, albeit weaker (see Figure 7 B). Lastly, we notice that the coefficients for current month's and next month's factor returns are almost identical.

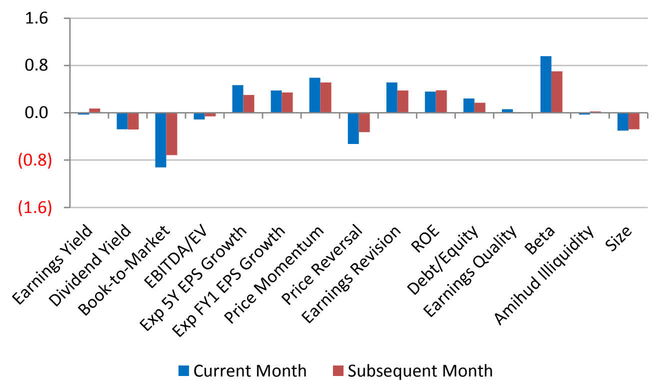
[6] The shaded areas indicate past recessions.

Figure 7 The Explanatory and Predictive Power of Economic Growth on Factor Performance

A) US



B) Europe



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

Capital Market Variables

With the Trump administration's proposed fiscal stimulus (infrastructure spending, tax cuts, etc.) at the time when the US economy is running at full employment, it is likely to trigger inflation and more hawkish Federal Reserve interest rate hikes. The consensus is that the US rates are likely to rise in the coming months (see Figure 8 A). We can't directly use the bond yield in our models, because it shows a staggering downward trend in the past 30 years[7]. We take a simple transformation by subtracting its own 12-month moving average:

$$NormalizedYield_t = NominalYield_t - \frac{1}{12}\sum_{\tau=1}^{12} NorminalYield_{t-\tau+1}$$

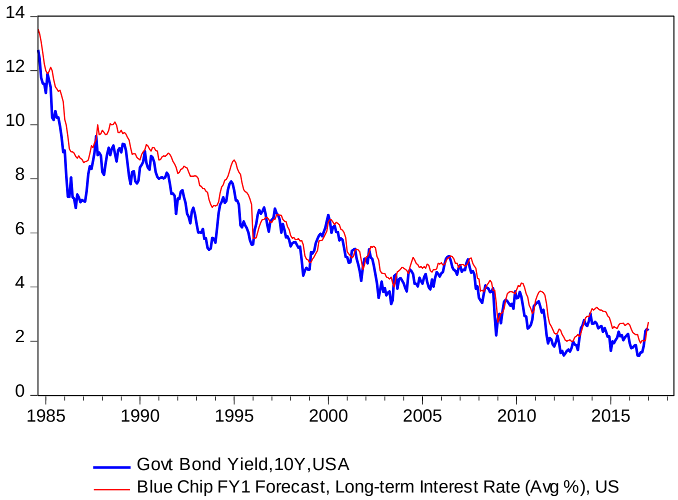
The normalized yield shows far more attractive time series properties (see Figure 8 B for the US and C for Europe/UK/Japan).

We then carry out the same set of two regressions (using current month's and next month's factor returns, respectively) against the normalized long-term bond yield. As shown in Figure 8 D, rising interest rates are detrimental to the concurrent performance of dividend yield, price reversal, and ROE factors. However, the implication for forward factor returns can be quite different. In fact, if today's interest rate is high, we are better off by investing in price momentum and low beta styles for the future.

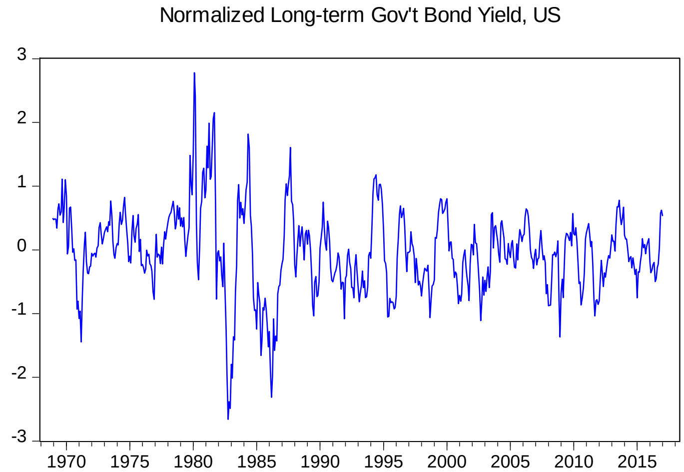
[7] In econometrics jargon, the US long-term interest rate is a trending time series.

Figure 8 US 10-Year Treasury Bond Yield

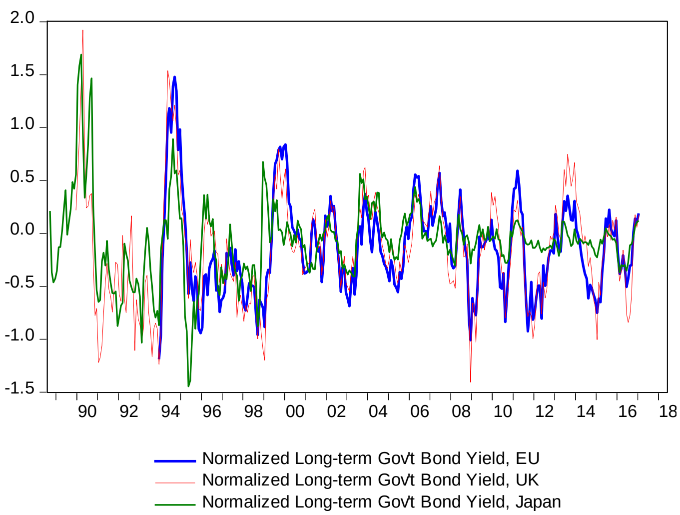
A) Current and Expected Bond Yield, US



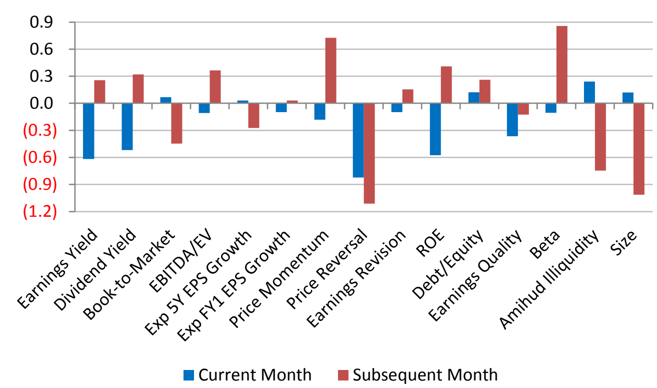
B) Normalized 10Y Treasury Bond Yield, US



C) Normalized Long-Term Gov't Bond Yield, Global



D) Bond Yield and Factor Returns



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

Seasonality

Seasonality in asset returns has long been documented. The three best known examples are:

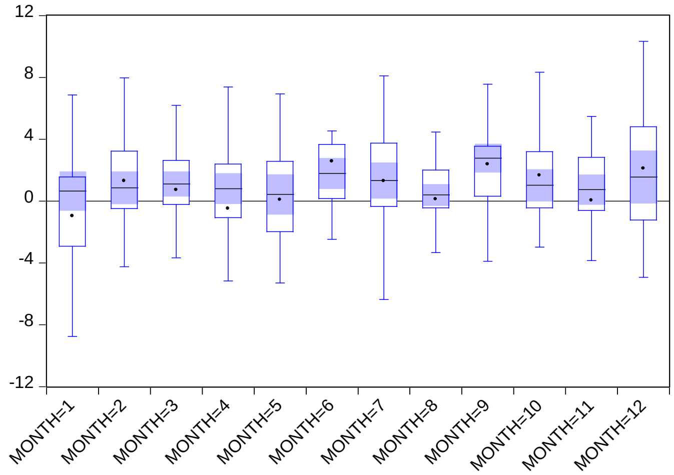
- **January effect** (see Rozeff and Kinney [1976]) states that small cap and high beta stocks outperform large cap and low beta stocks in January.
- **Sell in May and go away** (also known as the Halloween effect) suggests that stock market tends to suffer from weaker returns in the months from May to September (see Bouman and Jacobsen [2002]).
- **December tax loss selling and window dressing** anomaly argues that investors want to sell poorly performing stocks (e.g., low momentum) in December for tax reasons or have the

incentive to hold high quality/winning stocks for window dressing. Either argument suggests that momentum and quality styles should perform well in December.

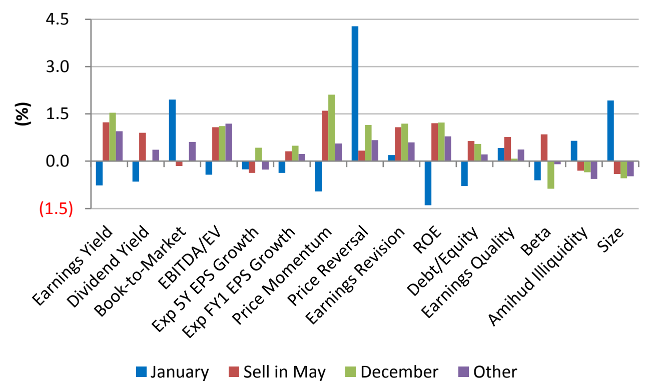
Figure 9 (A) shows the distribution of momentum factor return by calendar month. It is evident that momentum return is much higher in December, during the summer, and much weaker in January. Examining the three seasons (i.e., January, Sell in May, and December), January effect appears to be strong for most factors, followed by the Sell in May effect (see Figure 9 B).

Figure 9 The Seasonality of Factor Returns

A) Price Momentum, by Calendar Month



B) Factor Return in January, Summer, and December



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

This section briefly deliberates our thesis on factor timing and style rotation. We will revisit this topic when we present our global stock selection model – the LEAP model. More in-depth coverage of global macro research will be addressed in a series of forthcoming papers.

Machine Learning in Investment Management

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. Figure 10 (A) shows the interest in machine learning since 2004, using Google Trends as a proxy.

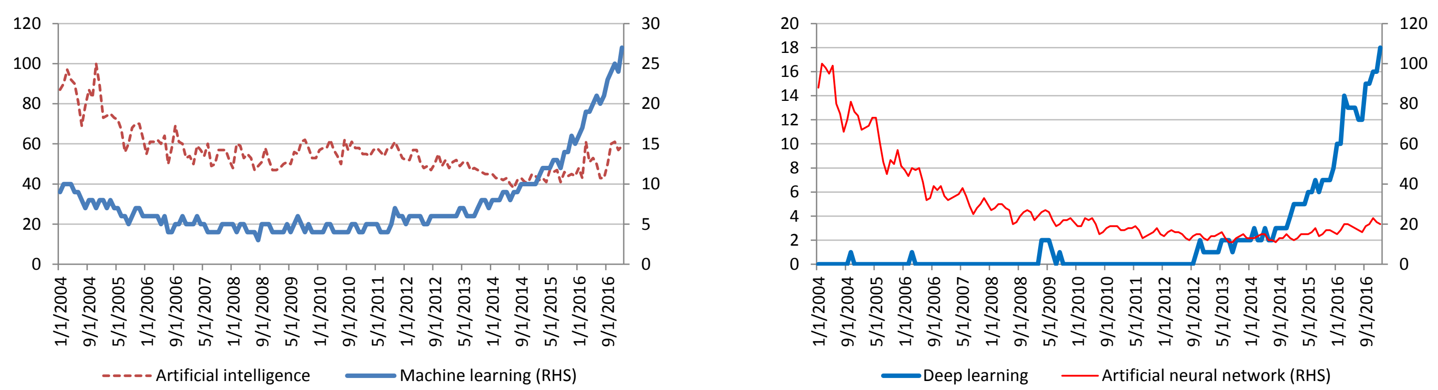
Despite the tremendous success of deploying artificial intelligence in many fields from drug development, physics, computer games to driverless cars, machine learning was traditionally perceived as data mining and received limited interest in mainstream finance research and institutional investing. However, the tide has changed rapidly in the past few years. We are suddenly seeing enormous interest in machine learning techniques from many of our clients globally. We have always been strong believers and proponents that machine learning is a critical tool for investment research and it may well represent the future of investing.

Machine learning is also known as data mining, pattern recognition, predictive modeling, etc. This is a rapidly evolving field, where new algorithms are constantly developing and old ones go out of favor. For example, Figure 10 (B) shows that deep learning[8] is generating great interest, while artificial neural network has been losing momentum.

Most people refer to machine learning as the various algorithms such as random forest, classification trees, and deep learning. However, machine learning is much more than the tools and techniques for uncovering patterns within data. Machine learning is a new philosophy and represents the entire process of forming a hypothesis, collecting data, fitting, selecting, and validating models.

Figure 10 Google Trends Interest in Big Data and Machine Learning

A) Machine Learning and Artificial Intelligence **B) Deep Learning versus Neural Network**



Sources: Google Trends, Wolfe Research Luo's QES

This section is meant to be a gentle introduction of machine learning in quantitative equity investment. We plan to publish extensively in the space of machine learning in investing in the coming months. For the purpose of this discussion, we avoid theories and mathematical details. Interested readers please refer to Hastie, et al [2008].

[8] The famous AlphaGo machine developed by Google and won against the international Go champion Lee Sedol, is powered by the deep learning algorithm. See https://deepmind.com/research/alphago/ for details.

A Fundamental Shift of Investment Philosophy

Traditional finance and economic research generally follows a structured process (see Figure 11 “Finance Theory” panel):

- **Establish a Hypothesis.** The first step in research is typically to establish a set of hypotheses. For example, the researcher may suggest that, due to the agency problem, company executives may engage in activities to benefit themselves personally rather than maximizing shareholder value. Company management may resort to aggressive accounting practice to boost immediate profit. However, such activities can’t last forever. Eventually, it will trigger the suspicion from the market or regulators. As a result, such companies tend to underperform the market in the long run. Most of the time, the functional form of the relationship is also pre-specified by the analyst.
- **Data Collection.** Once researchers build their hypothesis, they collect data to verify their theory.
- **Model Construction.** Once we have the data in place, we can use various statistical tools to analyze the data and see if the data supports our hypothesis. In addition to our traditional regression and econometric models, the toolbox includes all machine learning algorithms. However, in practice, the ordinary least squares (OLS) regression and other linear models dominate empirical finance and economics research, while machine learning is often criticized as overfitting.
- **Validation.** The most important criterion for most analysts is the so-called economic intuition, i.e., whether the proposed model makes intuitive sense. However, as we discussed in Luo, et al [2017b], coming up with a great story for a given pattern is probably not very difficult. The problem is whether such a pattern truly exists. More importantly, whether an investment strategy will continue to perform has absolutely nothing to do with how well we tell the story.
- **Implementation.** If our theory is appealing, data is sufficient, and our models support our hypothesis, we are more likely to implement our model in practice.

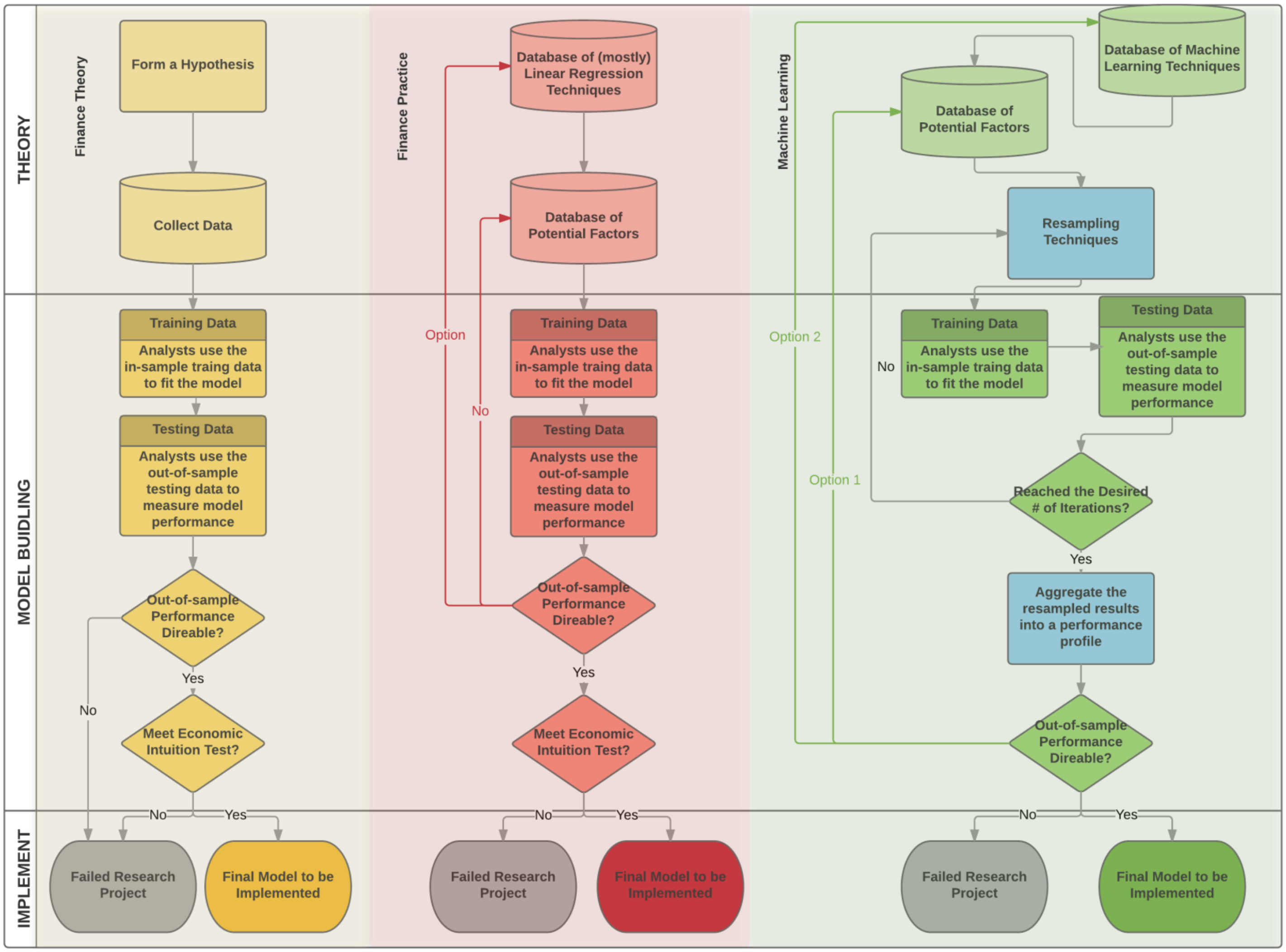
In reality, few researchers follow the above steps strictly. Rather, most analysts conduct an implicit loop, i.e., when our data and models do not support our theory, we try a different suite of models, until we “find” the desirable pattern (see Figure 11 “Finance Practice” panel). The so-called data mining bias refers to the case that we repeatedly model the same set of data multiple times, until we find an attractive pattern. Researchers tend to only publish their positive findings, but rarely (if ever) tell the readers how many times they have “abused” the data. We often hear the comment that “all backtestings work” refers to precisely that.

The machine learning philosophy, on the surface, seems to be quite similar. However, it has a few fundamental differences from classic finance research (see Figure 11 “Machine Learning” panel):

- **Hypothesis.** Classic finance requires us to define our hypothesis in a clear manner. Using the same agency issue example, researchers have to specify how to measure accounting manipulation. Most of the time, analysts also specify the functional form (most likely to be a linear relationship or at least can be proxied by a linear model). On the other hand, machine learning has far fewer restrictions. Researchers typically include a large number of potential predictive variables and let the model identify patterns. Domain knowledge is critical for machine learning and analysts normally include as many potential factors in the model as possible.

- **Data Collection.** In the machine learning literature, this step is where it marries with Big Data. Due to the speed and scalability of machine learning algorithms, researchers always want more data. Some data may not be obviously related to our hypothesis, but could still be useful.
- **Model Construction.** The key difference between classic finance and machine learning is that machine learning allows for a larger number of searches to find patterns. Parameter tuning is highly encouraged, rather than being forbidden. However, it is crucial to ensure that the search is honest. In practice, researchers typically split their data into three sub-sets: training, testing, and validation. In classic finance, researchers pretend they only process the data once, while in reality, most people just use manual searches to find the best outcome and publish the results.
- **Validation.** Economic intuition is critical for classic finance and economic research, but it is normally not emphasized in machine learning. The results of machine learning models may or may not be intuitive. Sometimes, it is even difficult to present a transparent illustration for models such as support vector machine and deep learning. In machine learning, analysts do not count on intuition alone to make their decisions. Rather, machine learning has its own specific rules for model validation.
- **Implementation.** Traditional modeling, on the surface, seems to be less likely to suffer from data snooping bias. However, in reality, precisely because it does not have explicit algorithms to control for repeated modeling of the same dataset, it is actually more likely to be overfitted. Machine learning algorithms, on the other hand, explicitly search a large range of possible patterns to identify the optimal ones. It primarily relies on the robustness of the algorithms and more importantly, the explicit methods of validation.

Figure 11 A Fundamental Shift in Investment Philosophy



Sources: Wolfe Research Luo's QES

A Brief Introduction of Machine Learning Algorithms

There are a wide range of machine learning algorithms and the list is constantly growing, as new techniques are being added to the toolbox.

As shown in Figure 12, at the top layer, machine learning can be divided into **supervised** and **unsupervised** learning. In simple terms, supervised learning studies the relationship between a set of input predictors (aka factors in investment management) and output (i.e., the variable that we try to predict, aka independent variable). On the other hand, unsupervised learning only concerns the structure among the set of input variables.

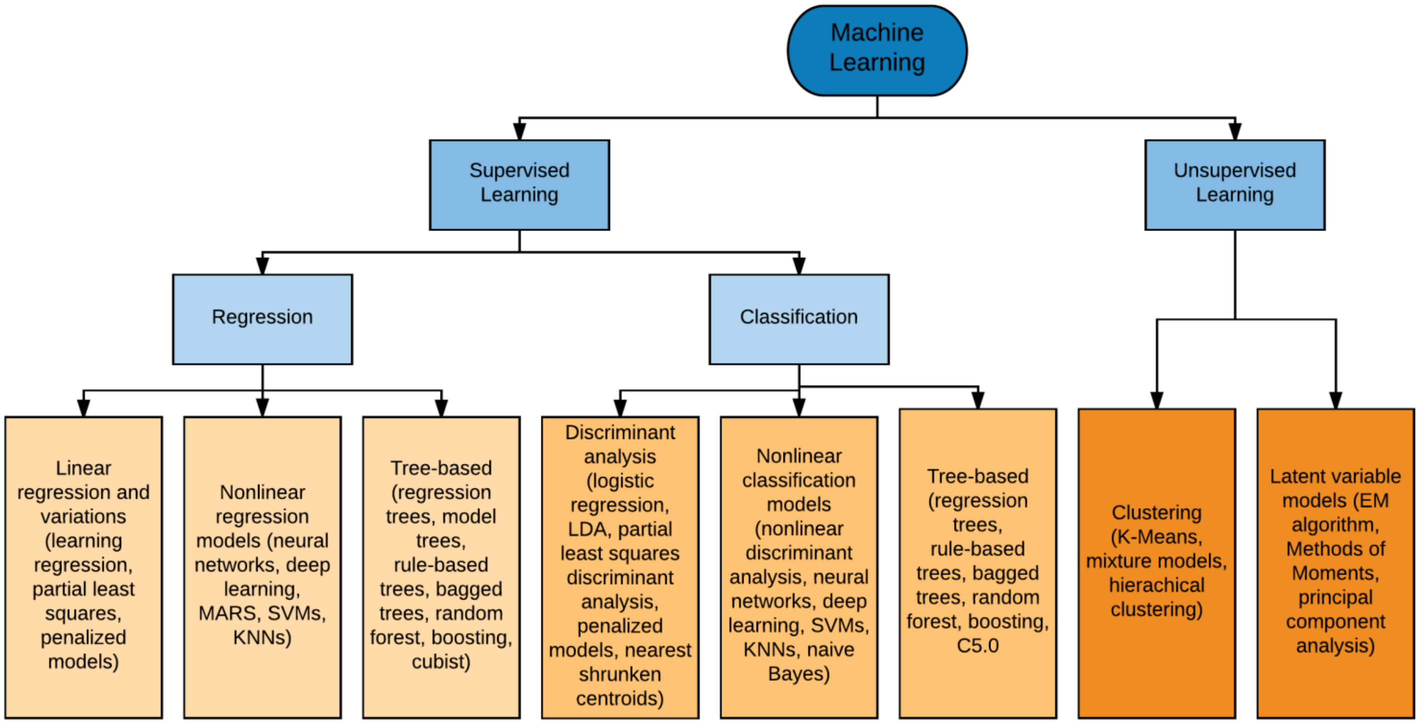
Supervised learning can be further divided into regression and classification problems.

- **Regression.** In a regression setting, the output variable is normally continuous, such as stock return.

- **Classification.** In a classification problem, the output variable is categorical, such as whether a company will be taken over or not.

To many people's surprise, linear regression also belongs to the broad machine learning toolbox. In this paper, we briefly discuss a few algorithms underlying our first global stock selection model. A more in-depth coverage of machine learning will be addressed in a series of forthcoming research papers.

Figure 12 A List of Machine Learning Algorithms



Sources: Wolfe Research Luo's QES

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 for factor k at time t; K is the number of factors; $f_{i,k,t-1}$ is the score of factor k for of stock i at time $t-1$; and $\varepsilon_{i,t}$ is the regression residual.

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., neural networks (NNs), 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, in our previous research, we built models to identify potential takeover targets, dividend hikes/cuts, accounting fraud, etc. In those applications, the output variable is categorical, e.g., yes and no.

In some occasions, the natural output might be continuous, e.g., stock returns. However, there are good reasons to translate the problem from continuous to discrete. Instead of forecasting stock returns, we can 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. Models may over-fit 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. Discretionary managers naturally prefer models with higher hit rates; therefore a classification setting may fit their needs better.

Classification and Regression Trees and Ensemble Models

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[9] 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 regression, the prediction is typically the mean value for each region. 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:

- The predictor to split on and the value of the split,

[9] 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 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.

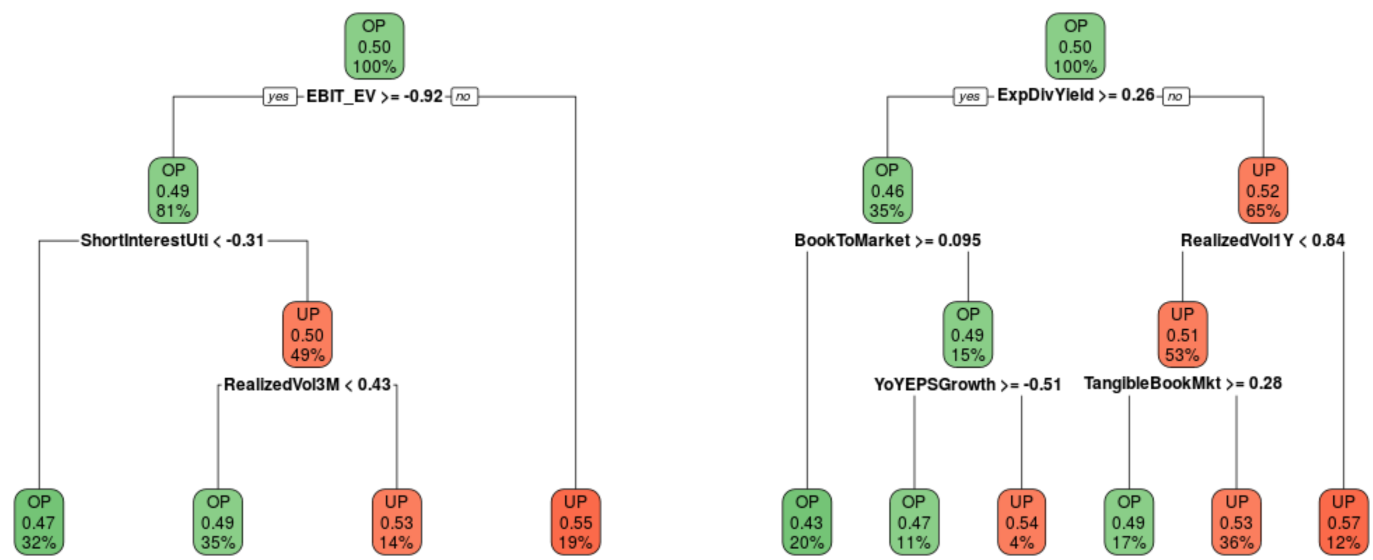
Bagged, random forest (see Breiman [2001]), boosting (See Valiant [1984], Kearns and Valiant [1989]), AdaBoost (see Freund and Schapire [1996]), Stochastic Gradient Boosting (see Friedman, et al [2010]), and other ensemble models using classification trees are likewise extended to classification settings. In most cases, tree-based algorithms were initially developed for classification problems and later extended to regression context.

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 all of our ~400 factors in our factor library to predict this Outperformer/Underperformer classification. Figure 13 (A) shows a fitted CART model for the US market as of January 2017, using the past 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 those cheaper stocks (based on EBIT/EV), we further divide them, conditional on short interest (utilization) – stocks that are cheap and have low short interest are put into the very left hand branch and classified as Outperformers. The middle branch is further split on realized volatility – cheap stocks with high short interest and high (low) volatility are classified as Underperformers (Outperformers).

The CART model for US stocks is rather small – with three factors and four terminal nodes. On the other hand, the model for Japan is more complex, with five factors and six terminal nodes (see Figure 13 B). The CART model for Japan is very valuation centric – dividend yield, book-to-market, and tangible book-to-market account for three of the five factors.

Figure 13 CART Models for US and Japan

A) US **B) Japan**



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.

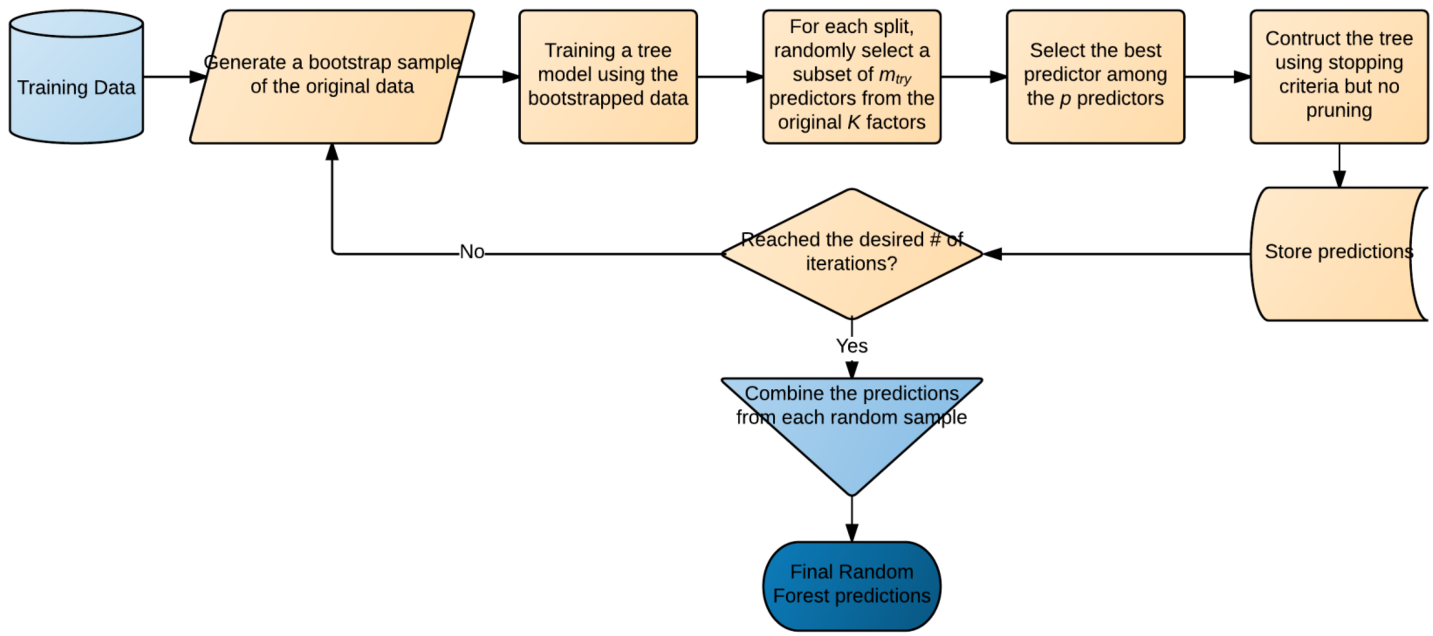
Random Forest (RF)

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 14 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[10] with a non-pruned tree. Then it repeats the same procedure many times[11]. Finally, we take the average from each CART model's prediction to derive our final forecast.

[11] 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.

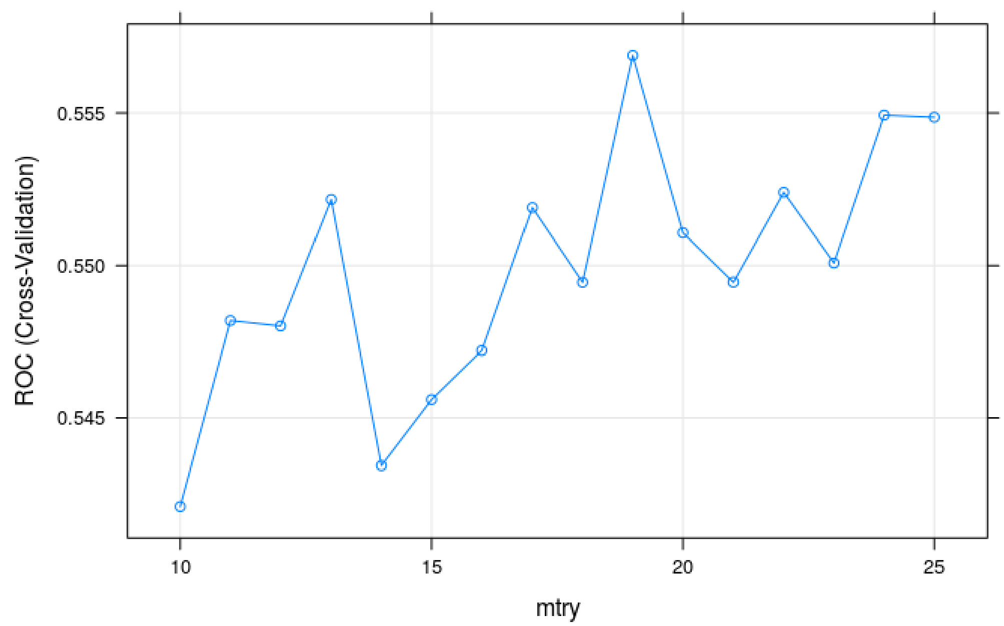
Figure 14 Random Forest Flowchart



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. As a demonstration, we fit a random forest model, using the same setup as in the CART model, for each of the nine regions, using all of our factors. Figure 15 shows the sensitivity of model performance (on the y-axis, hit rate) to the m_{try} parameter in Canada. As we vary our m_{try} parameter from 10 to 25, the model performance stays almost the same.

Figure 15 The Sensitivity of Model Performance on the Tuning Parameter



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

In our previous research, we have successfully applied the random forest algorithm in a number of different applications, from news sentiment, stock return forecasting, to predicting takeover targets. 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[12]

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.

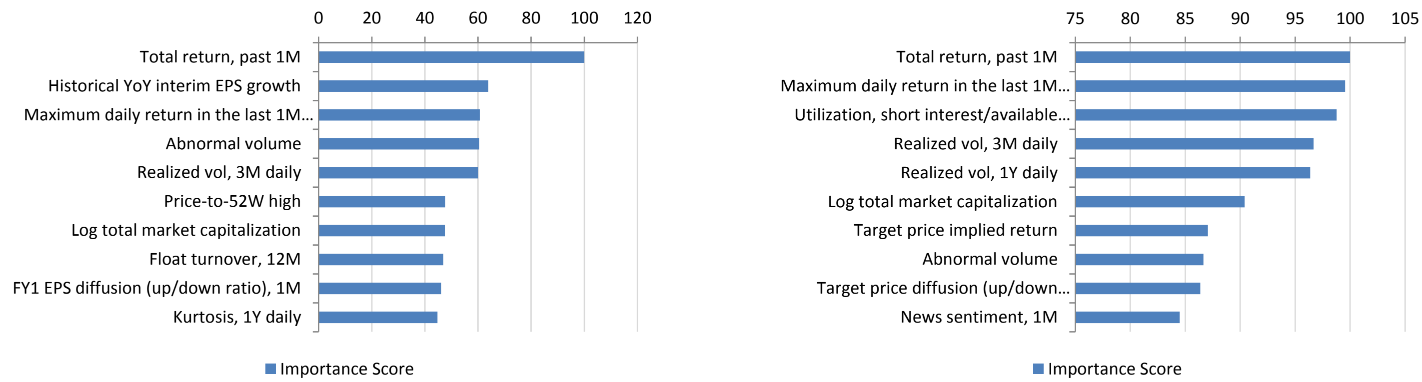
Figure 16 (A) shows the top 10 factors selected by the random forest algorithm, as of December 31, 1996, using 10-year trailing monthly data. Because we require all factors to have complete 10-year data, we only have about 100 factors at the end of 1996. It is somewhat surprising to see that the year-over-year 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 2016, half of the top 10 factors remain the same as in 1996, which highlights the stability of the random forest model. More

[12] 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.

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 16 The Top 10 Factors Selected by the Random Forest Algorithm

A) Data as of 12/31/1996, US | **B) Data as of 12/31/2016**



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

Adapting Machine Learning Algorithms to Investing

It all seems to be natural to extend the array of machine learning algorithms in finance. However, before we get too excited, there are a few hurdles that we need to address first. We can't directly and naïvely port machine learning algorithms to investing.

Cross Sectional versus Time Series Analysis

Traditional machine learning algorithms were mostly developed for cross sectional analysis. For example, machine learning has been successfully implemented in medical research to classify patients into ER rooms, in pharmaceutical and biotech research to find new drugs, and in insurance fraud detection. In all these areas, we can sample data from different periods and we can expect models constructed using one period of data to be equally predictive for another period of testing data.

For finance, however, time is an extremely important dimension. In some occasions, time dimension is the dominant one. For example, when we try to predict GDP growth using other macroeconomic variables, we know many economic variables are serially correlated, which, if used properly, should improve our prediction accuracy. However, typically machine learning algorithm can't be easily adapted to time series prediction. More importantly, in time series prediction, our sample size is normally much smaller than cross sectional analysis, where thousands and even millions of observations are normal.

In stock selection modeling, we deal with both cross sectional and time series data; therefore, machine learning algorithms are more likely to be useful. However, we still need to worry about autocorrelation for many predictors. For example, our earnings yield data is likely to be serially correlated from month to month, because companies report their financials only once a quarter (or semi-annual, tri-annual, or even annually). Fortunately, the dependent variables (e.g., stock returns that we want to predict) are generally not correlated over time.

In stock selection models, samples are not independent. We pool together essentially the same stocks over time. Although there are new companies added and old ones being removed, from one period to the next, the vast majority of stocks remain the same. One way to conduct more objective cross validation is to tune the parameters using data from one region and test the performance on a different region.

Dimensionality

The other challenge in investing is that we potentially have a large number of factors/predictors. In our example, we have around 400 factors in our factor library and that number is growing over time. In an ideal setup, we would hope that our sample size (N) is much greater than the number of predictors (P), i.e., $N \gg P$. Some algorithms (e.g., CART) are more effective than the others (e.g., multivariate regression) in the case of $N \approx P$.

Furthermore, factors are likely to be correlated, which means some factors have redundant information. Both issues are not unique to finance, but they need to be properly addressed in our models.

In machine learning literature, the techniques used to reduce the dimensionality of factors are called signal extraction or feature extraction. Principal component analysis (PCA) is a commonly used approach. The problem with PCA is that it does not consider the response in summarizing information. A potential better alternative is the PLA (Partial Least Squared) regression technique.

We discussed factor selection in Luo, et al [2017b] and will address the topic in more details in the next section.

Imbalanced Sample

Lastly, the sample can also be highly imbalanced. One area that we have done extensive research in the past is to predict takeover targets. At any given point in time, only a small number of companies are in the process of being acquired. Therefore, a model that predicts no takeover would have a very high hit rate, but obviously, that naïve model is unlikely to be useful to investors. The typical approach treating highly unbalanced data is to use matched samples. For example, we could randomly sample the same number of non-takeover companies with similar characteristics (e.g., sector, size, and valuation) as our target companies, and use that sample to train our model.

Interpretability

The greatest challenge of implementing machine learning in investing, however, rests on its interpretability. Few portfolio managers are comfortable investing in a model that they do not even understand themselves. Furthermore, managers often need to communicate what they invest and how they invest to asset owners. Although not all machine learning algorithms are difficult to interpret, the general perception is that they behave more like “black-box” models. In this paper and our forthcoming publications, we attempt to challenge this traditional perception. The variable importance score from the random forest model is a great example of showing some structure from a complex algorithm.

Will Machine Eventually Replace Human Analysts?

One of the most common questions and probably the most controversial one is whether machine will one day replace human analysts. In our opinion, the answer is yes and no.

Many of the repetitive tasks such as data entries, plotting a chart, summarizing data, and potentially even interviewing/analyzing company management and writing reports can be and will probably be replaced by computers.

On the other hand, analysts with skills of both finance and machine learning will be in huge demand. As computing power becomes stronger and stronger, cheaper and cheaper, and more and more machine learning algorithms are becoming readily available, implementing a machine learning model to find investment opportunities seems to become easier over time. However, as Rodriguez [2011] points out, the credibility of model building has weakened especially as the window to data access and analysis tools has widened.

We do not believe that subject knowledge of finance, domain expertise on specialized datasets and machine learning are mutually exclusive. Quite to the contrary, we think they are complementary. This view is also shared by some researchers (e.g., Ayres [2007]).

DATA MINING, OVERFITTING, AND DATA SNOOPING

With the massive improvement in computing speed and the readily available machine learning algorithms, the greatest risk of applying machine learning to investment is overfitting. Modern machine learning algorithms can be highly flexible and adaptive to identify complex relationships. However, they can easily identify an artificial pattern that does not really exist and of course, will not have much true out-of-sample predictive power. This bias is often referred as data mining, overfitting, or data snooping, i.e., we abuse our data repeatedly until we find a pattern that we want to see. It has been extensively studied in the forecasting field (see Clark [2004]) and finance (see Bailey, et al [2014] and Luo, et al [2017b]).

The real out-of-sample test is to evaluate our strategy using live trading data. However, when we conduct investment research, we rarely have that luxury. Therefore, in this paper, we propose a few alternative approaches. Specifically, we describe how we can measure, with statistical confidence, whether our backtesting is likely to work out-of-sample.

Model Selection and Tuning

As listed in Figure 12, there are a large number of classification and regression techniques we can use. Moreover, for each of these algorithms, there are many parameters that we can tweak. For example, even with linear regressions, we have the choice of including/excluding variables, adding higher order and interaction terms, and fine-tuning other dimensions (e.g., weighted least square regression, quantile regression). How to best measure the honest out-of-sample performance is the first and probably the most important step to ensure that we do not overfit our model.

In-sample versus Out-of-sample

Using the training data, we can build our model, analysts may attempt to try a wide range of algorithms. For each algorithm, researchers may further fine-tune and pick the best set of parameters. The performance of such model on the same set of in-sample data is called in-sample performance or apparent performance, which is most likely to be overly optimistic. In a simple example, if we have the same number of predictors as our sample size, even a linear regression model would achieve an R^2 of 100%. However, that model is unlikely to have much predictive power over a new set of data. Although the in-sample bias is well documented and understood, we still see a surprisingly large number of studies showing only in-sample performance. In Luo, et al [2017b], we show that in-sample factor selection models inflate performance by over 100% in many occasions.

The traditional approach is to split our data into training and testing. We construct our model using the training data set and then measure the model's performance in the testing data (i.e., out-of-sample data). In investment research, the most common way is to train the model using historical data and measure the performance using the subsequent periods of data. It can be done either as a one-time excise or a rolling backtest. In most academic research, the researchers build the model using one set of historical data and report the performance in the subsequent period of data. Alternatively, analysts may repeat the process, at each given point in time, fit one model using data that was available as of the given point in time, then test the model's performance in the subsequent period. The same process will be repeated using either an expanding or rolling window.

The problem is that analysts have the tendency of building multiple models and repeat the same process. In the end, they will only report the one model with the best "out-of-sample" performance. The reality is that it is out-of-sample in terms of data, but is in-sample in terms of model selection.

Resampling, Cross validation, and Bootstrap

When the sample size is small, splitting the data into training, testing, and validation may not be practical. More importantly, a single testing data set may not make sufficiently accurate judgment on the model performance (see Hawkins, Basak, and Mills [2003]). In this case, an effective technique called cross validation can be employed.

In resampling, we randomly split our data into training and testing sets. We build the model using the training set and then test the performance in the testing set. This process is repeated multiple times and the results are aggregated and summarized.

The most commonly used resampling technique is the k-Fold Cross Validation. The sample data is randomly partitioned into k sets of roughly equal size. A model is fit using all the data except the first subset (called the first fold). The model performance is measured using the first hold-out sample. The process is repeated for each of the other partitions. The k resampled estimates of performance are summarized, usually with the mean and standard error. The choice of k is usually five or 10, but there is no theory to suggest the "optimal" k. k is bounded between one (i.e., the traditional approach without resampling) and N (i.e., the size of the sample). In the case of $k = N$, it is called Leave-One-Out Cross Validation (LOOCV). As k grows larger, the bias (i.e., the difference between the estimated and true values of performance) gets smaller, but the computation needs increase. Even with today's computing power, for large scale simulation, setting k too big can still impose a heavy computational burden. As Molinaro [2005] pointed out, the LOOCV and 10-fold cross validation yield similar results. Therefore, in practical terms, we can normally set $k = 10$.

Based on cross validation, we may choose simply the one with best performance. Alternatively, the "one-standard error" rule is also often used. The principal is that we prefer a simpler and more parsimonious model, i.e., we choose the simplest model whose performance is within a single standard error of the one with the best accuracy, as originally suggested by Breiman et al [1984].

The Variance-Bias Trade-off

The model selection process, in essence, is about the trade-off of variance versus bias. We can decompose the MSE (Mean Squared Error) as follows, if we assume that the data are statistically independent and the residuals have a theoretical mean of zero and a constant variance of σ^2:

$$E(MSE) = \sigma^2 + (Model\ Bias)^2 + Model\ Variance$$

The first part, σ^2 is the irreducible noise, which can't be reduced by modeling.

The second element – the bias of the model reflects how close the functional form of the model can get to the true relationship, i.e., the difference between the estimated value and the true value. Generally speaking, more complex, adaptive, and flexible models tend to have lower bias. However, if a model over-fits the data, it will not be able to predict a different set of data (i.e., high variance).

The last component – the model variance – reflects the sensitivity of our estimated model to small perturbations in data. Variance can also be interpreted as the inconsistency in predictions from one another, over different training sets, not whether they are accurate or not. Variance is inversely related to robustness. Loosely defined, simple models tend not to over-fit and have low variance, but if a model is too simple, it may not be flexible enough to capture the true relationship (i.e., high bias).

Using an extreme example, if we predict the future return of a stock using its own sample mean, the model would have very low variance (i.e., it is insensitive to the training data), but it suffers from very high bias (as it will not predict future return very well).

The task of model building is to find an optimal balance between bias and variance (or flexibility and robustness). In another word, it is about the trade-off between prediction accuracy and model robustness.

L-ECONOMIC ALPHA PROCESSING (LEAP) MODEL

In Luo, et al [2017a], we study how Big Data and data science can be applied in active investing, while in a follow-up paper (see Luo, et al [2017b]), we transform data into knowledge, discuss signal research, and present common multifactor models. In this paper, in the "Machine Learning" section, we elaborate how sophisticated pattern recognition techniques can be used in finance. Lastly, in the "Style Rotation" section, we argue that macro data and information should be used in factor timing.

In this section, we bring together everything that we have examined so far, and introduce the first of our proprietary global stock selection models – the LEAP or L-Economic Alpha Processing model. The LEAP model takes advantage of the latest machine learning techniques, macroeconomic research, but more importantly, also emphasizes interpretability, transparency and relies primarily on the more familiar linear modeling techniques. A more complex nonlinear machine learning model will be released in the near future.

MODEL BACKGROUND

First of all, let's define some basic parameters for the LEAP model. It is a global stock selection model that aims to predict near- to mid-term alpha (one to 12 months), for about 11,000 stocks, currently representing 46 countries in the MSCI ACWI. The inputs for the LEAP model comprise ~400 factors in our factor library. Currently, most of the factors are still on the more traditional side in six style categories: value, growth, price momentum & reversal, sentiment, quality, and alternative. However, as we are building up our Big Data infrastructure, we will add more and more unconventional signals. The LEAP model structure is purely systematic, from factor selection, weighting, style prediction, alpha forest, to portfolio construction. Therefore, as new factors are added to the factor library, the LEAP algorithm automatically assess whether any of them should join the model.

Global Universe

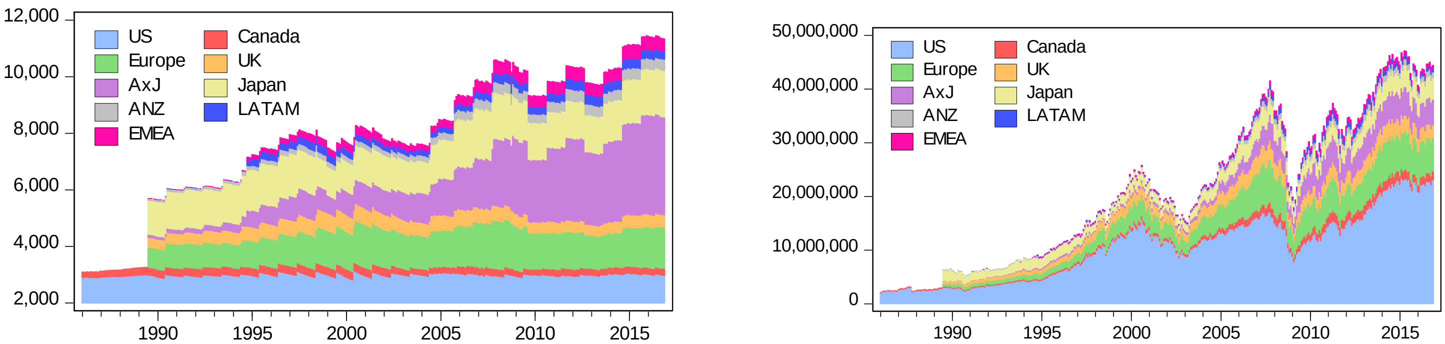
Similar to the generic multifactor model suite, we also have nine regional LEAP models – one for each region:

- 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 17 (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 17 B).

Figure 17 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

Data Pre-Processing

All factors in our factor library are normalized using our proprietary ranking-inverse normal transformation, at each given point in time (see Luo, et al [2017a] for details). We also adjust for major risk factors, e.g., country, sector, size, and beta.

Model Estimation and Forecasting Frequency

The LEAP model is re-fitted for each of the nine regions monthly, i.e., once a model structure is determined, it stays the same during the month, with the same factors and weights. However, factor scores (exposures) change every day, as new information arrives. Therefore, the model is updated daily. We offer clients with data feeds via SFTP and other delivery options. Please contact us for details.

Forecasting Horizon

The LEAP model is calibrated to predict one-month-ahead return. Although the forecasting horizon may appear to be short, as we will present later, the model's information decay is reasonably slow and lasts for beyond a year. Therefore, the LEAP model is equally useful to long- and short-term investors.

Factor Selection

For the LEAP model, we start from our factor library, which includes around 400 factors, from all categories discussed in Luo, et al [2017b]. The list of factors constantly grows over time.

As discussed in Luo, et al [2017b], dimension reduction is a formidable challenge for multifactor models. How do we choose among these many factors to design a parsimonious model. In this paper, we further elaborate on the two factor selection techniques discussed before – the wrapper and filter methods. The LEAP model uses a combination of both approaches to eliminate redundant variables.

NONLINEAR MACHINE LEARNING COMPONENT

To capture the most salient nonlinear features in our data, we use two powerful and fairly transparent machine learning algorithms – RF (Random Forest) and CART (Classification and Regression Trees).

Feature Selection via Random Forest

The RF algorithm automatically conducts feature selection and tends to be fairly robust to dimensionality (i.e., large number of factors) and multicollinearity.

There are two major downsides for the RF algorithm. First of all, because it comprises hundreds and thousands of single tree models, it is difficult to present and not very easy to understand. Second, it can also be very slow. In our experiment, it takes about two hours on a 25-core Linux grid to fit a single RF model. Therefore, even with parallel computing, to fit the model each month for 30 years is likely to take a long time.

To overcome the two issues, but also take advantage of the feature selection ability, we use the RF algorithm as a factor (feature) selection tool. At the end of each year, we fit the RF model for each of the nine regions to select the most predictive factors (in either a linear or more likely to be non-linear fashion), on a suite of Linux servers designed for parallel computing[13].

The factors selected by the RF algorithm can still be correlated. Therefore, we perform another filtering process by removing highly correlated factors. The final list, around 50-80 factors, goes into the CART model below.

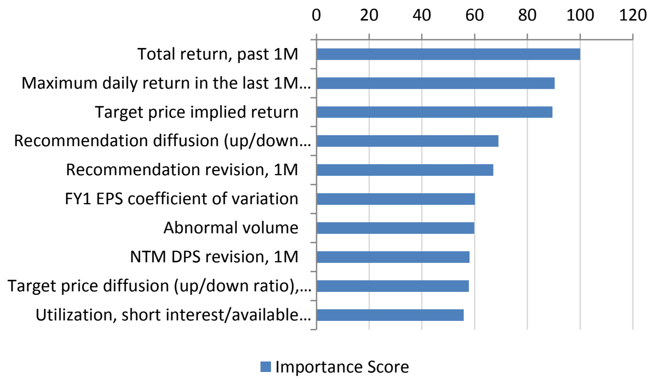
Figure 18 (A) to (F) shows the top 10 factors selected by the FR algorithm for Canada, Europe, UK, AxJ, Japan, and ANZ, respectively, as of December 31, 2016. There are a few interesting patterns to note:

- The most predictive factors appear to be very similar across regions;
- Short-term reversal, our lottery factor, target price-implied return, abnormal volume, historical year-over-year earnings growth, realized volatility, and recommendation diffusion ratio are top ranked in most regions;
- Short-term reversal, in particular, shows up top of the list, even in those regions that we do not observe strong reversal patterns, e.g., ANZ, which suggests that the patterns identified by the random forest algorithm is likely to be very different from traditional linear backtesting.

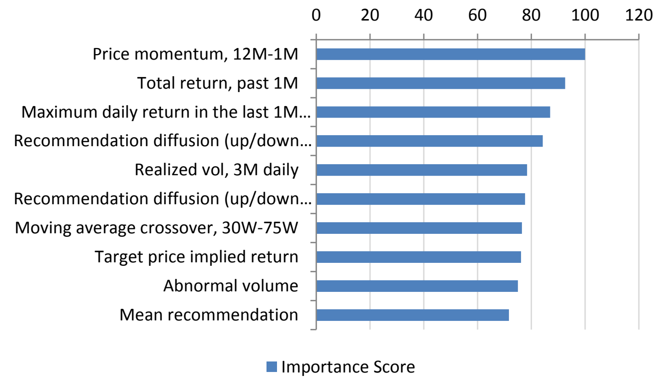
[13] This process generally takes less than three hours.

Figure 18 Top 10 Factors Selected by the Random Forest Algorithm, as of December 31, 2016

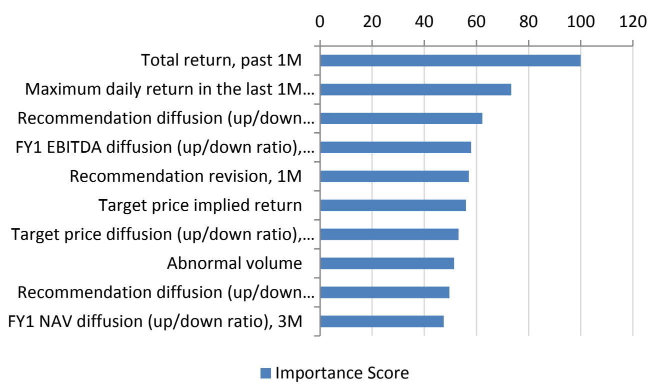
A) Canada



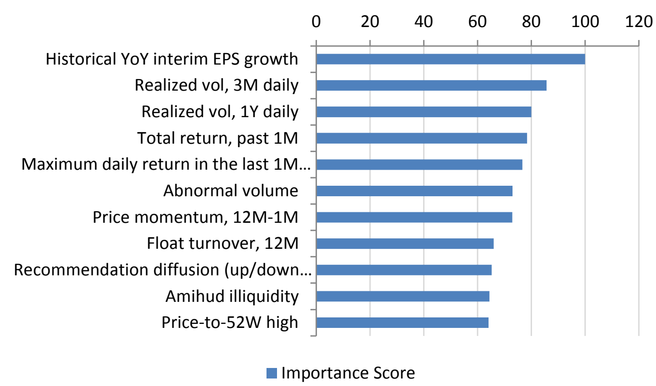
B) Europe



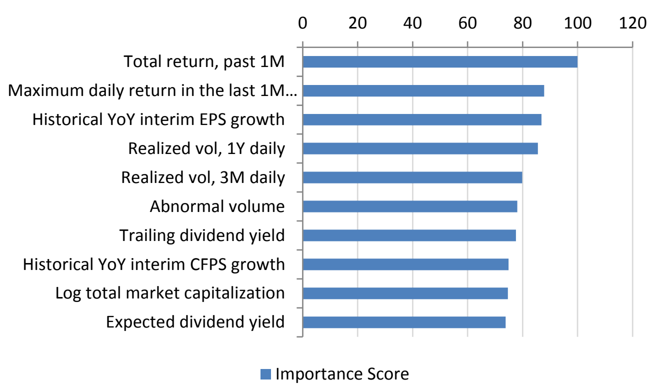
C) UK



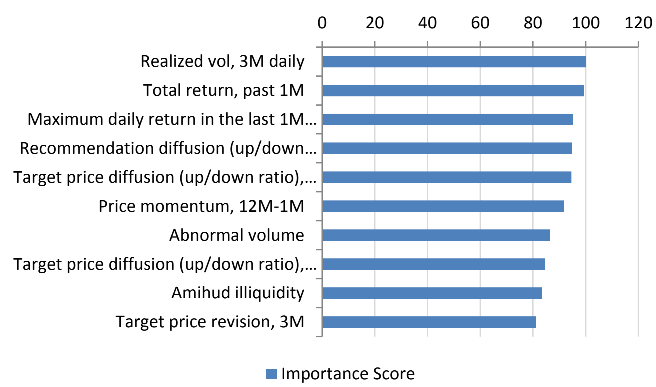
D) AxJ



E) Japan



F) ANZ



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

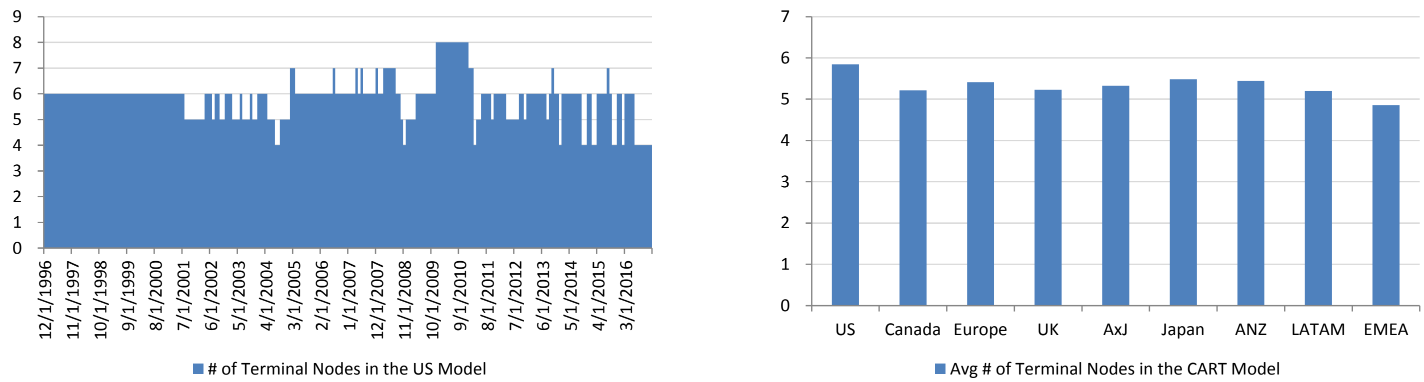
CART

The CART model tends to be fairly simple, transparent, and fast. However, the model is sensitive to dimensionality (i.e., large number of predictors) and correlated factors. The RF and related factor pre-processing step above help us reduce dimensionality and remove correlated factors. The other big downside of CART, i.e., performance tends to be not very strong, will be addressed in the linear model section below. Because RF and CART are closely related, the RF factor selection process is a hybrid approach combining the wrapper and filter methods.

In the CART component, we intentionally keep the CART model simple and parsimonious. As shown in Figure 19 (A), there are only four to eight terminal nodes in the CART component in the US model throughout the history. Across the nine regions, the CART model contains around four to six terminal nodes on average (see Figure 19 B).

Figure 19 # of Terminal Nodes in the CART Model

A) US **B) Average # of Terminal Nodes, Global**



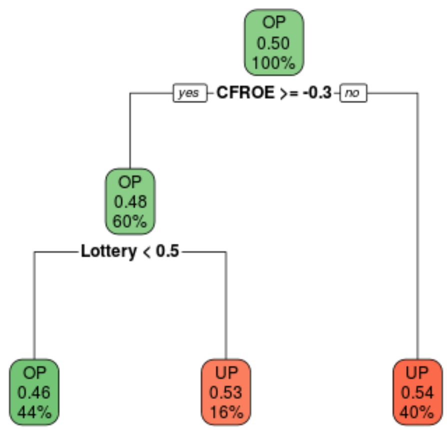
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Figure 20 (A) to (D) show the current CART models for Canada, Europe, AxJ, and LATAM, respectively. The nonlinear model is indeed quite intuitive. The factors used in the model and the direction of the splits are mostly as expected.

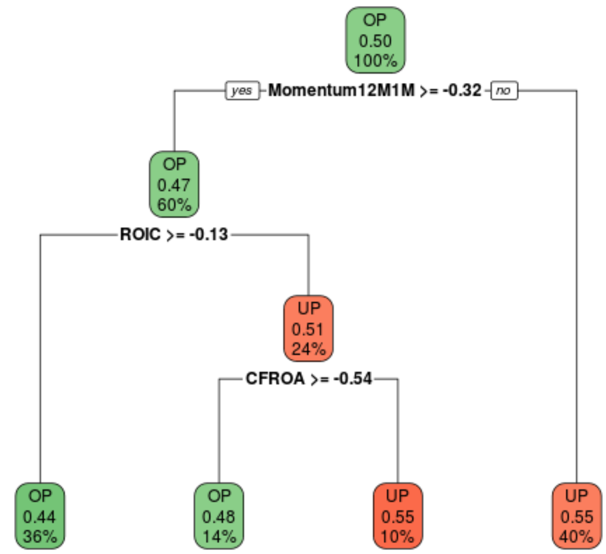
We then create a dummy variable for each of the terminal nodes. The argument is that stocks falling into each of the terminal node may have different payoff patterns, which will be incorporated into a more traditional linear multifactor model.

Figure 20 Current CART Model Structure

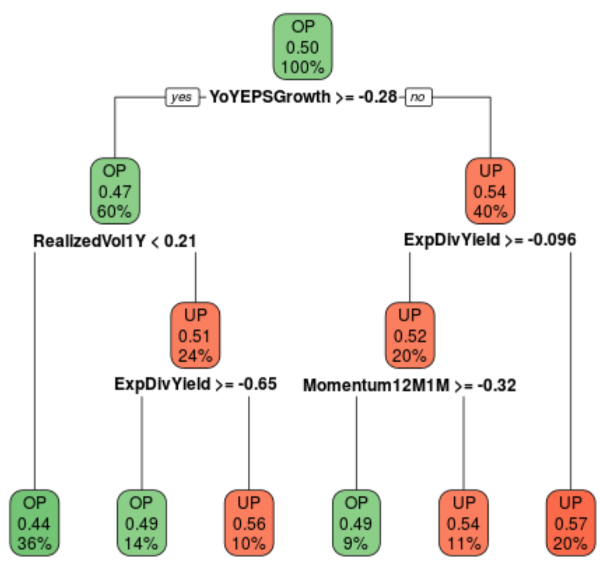
A) Canada



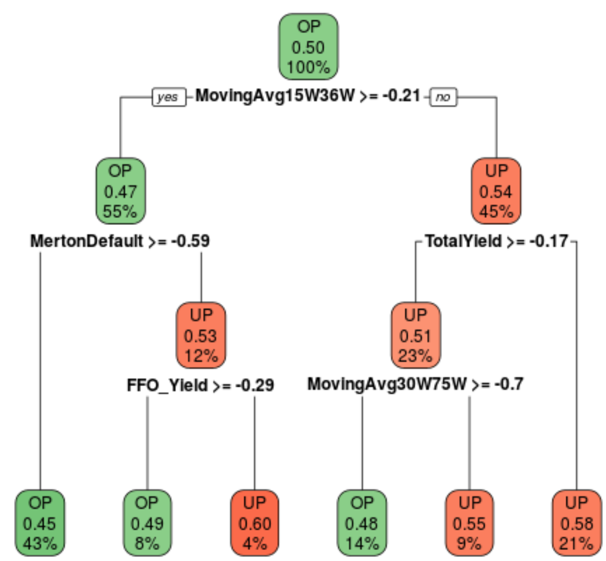
B) Europe



C) AxJ



D) LATAM



Note: For CART model graphs, **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

After the RF factor selection process, the CART model picks factors at each point in time. Figure 21 (A) to (F) show the top 10 most frequently chosen factors by the RF-CART model. It appears that factor turnover is more modest in Europe and Japan than the other regions – the top two factors in these regions show up in the model almost 40% of the time.

The best nonlinear factors appear to be quite different from our perceived best signals. For example, in the US market, consensus recommendation revision, EBIT/EV, and one-month reversal are the top three factors.

In some occasions, the factors in the RF-CART model seem to be counter-intuitive. For example, in Canada (see Figure 21 B), the second most frequently chosen factor in the one-month return – normally labeled as one-month reversal or mean-reversal, i.e., stocks with the biggest rallies last month tend to underperform in the subsequent month. However, we also know that reversal is not a particularly strong signal in Canada, as the Canadian market is more dominated by price momentum.

Compared to the mean-reversal effect in the US (see Figure 22 B), the reversal signal performance is much weaker in Canada (see Figure 22 A). However, there is something much deeper behind the sluggish performance of reversal in Canada. As shown in Figure 22 C, the patter is much of a U-shape curve, in that stocks with the highest and lowest returns in recent days are both penalized in Canada. On the other hand, the pattern in the US is a much neater linear monotonic straight line (see Figure 22 D).

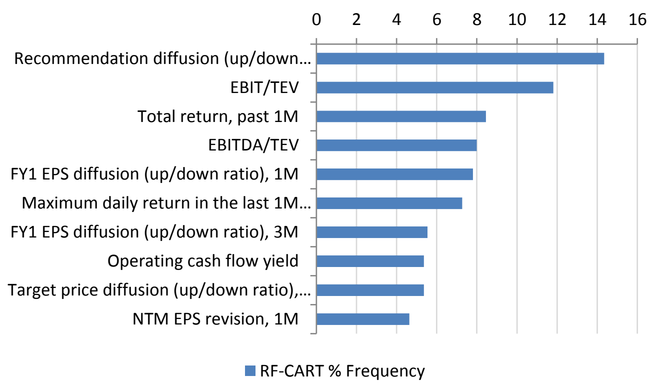
Furthermore, as shown in Figure 22 E, the RF-CART model is able to identify that, after controlling for price momentum, one-month reversal becomes a useful classifier in Canada. This is a hidden relationship that is otherwise impossible to retrieve from linear models.

Lastly, Figure 22 F shows how the one-month reversal signal enters the RF-CART model in the US. It clearly demonstrates a very different picture from how it predicts returns in Canada. In the US market, price reversal is a strong classifier and serves as the first split.

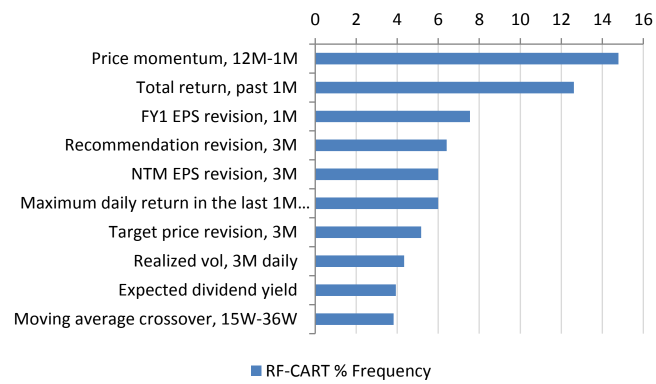
- On the left branch, underperforming stocks in the past with attractive valuations are expected to do well in the future.
- On the right branch, stocks with the strong return in the last month are further classified on earnings revision – in that, if past stellar returns are attributable to earnings revision, they are labeled as Outperformers; otherwise, they are predicted to underperform.

Figure 21 Top 10 Most Frequent Factors Selected by the RF-CART Model

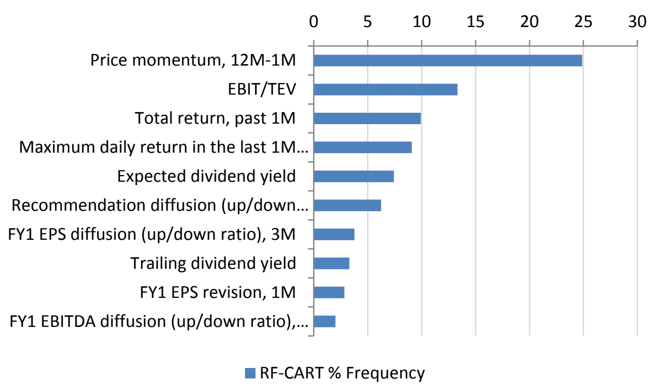
A) US



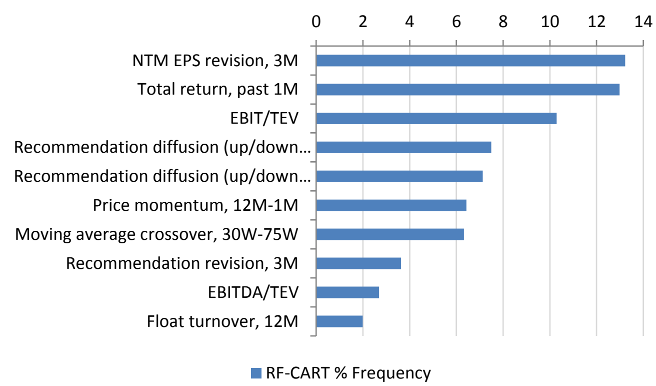
B) Canada



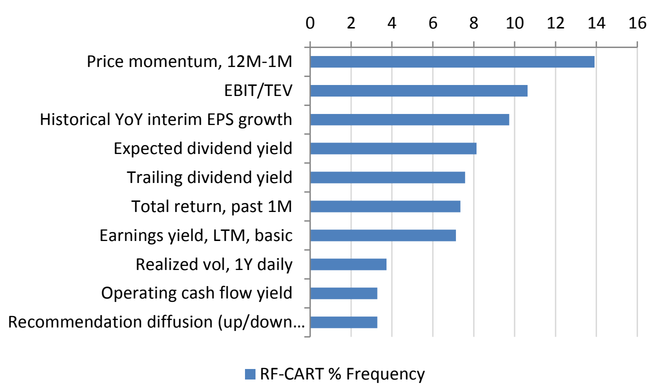
C) Europe



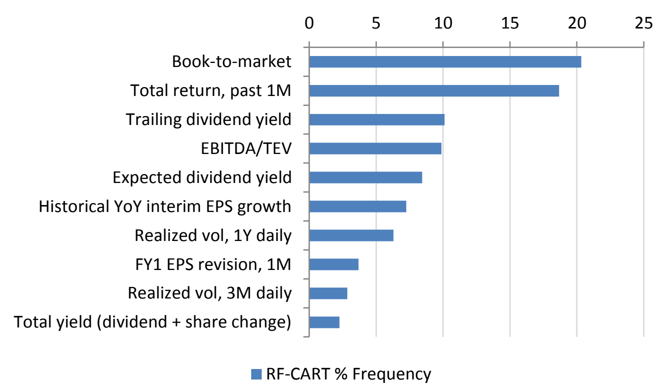
D) UK



E) AxJ



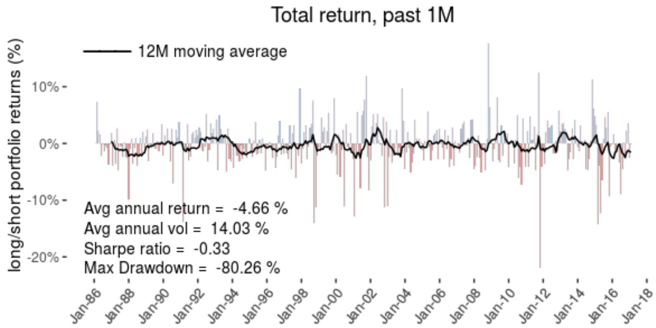
F) Japan



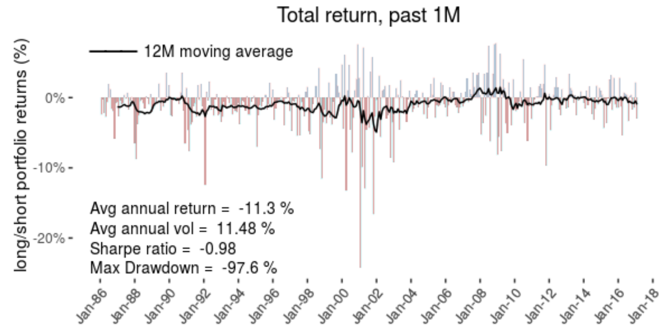
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Figure 22 Mean-Reversal in Canada

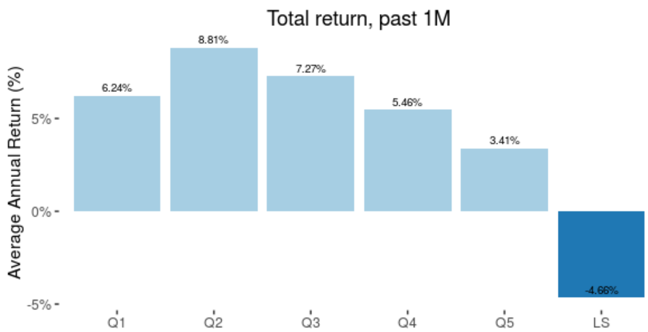
A) One-Month Reversal in Canada



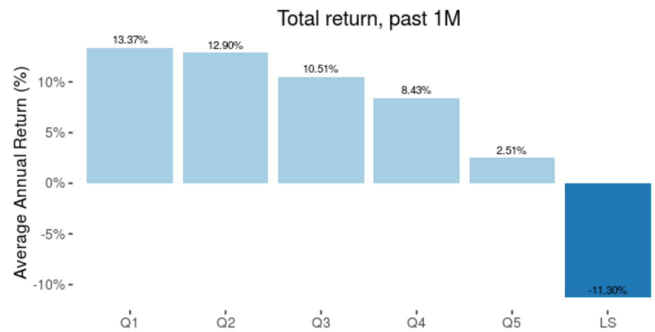
B) One-Month Reversal in the US



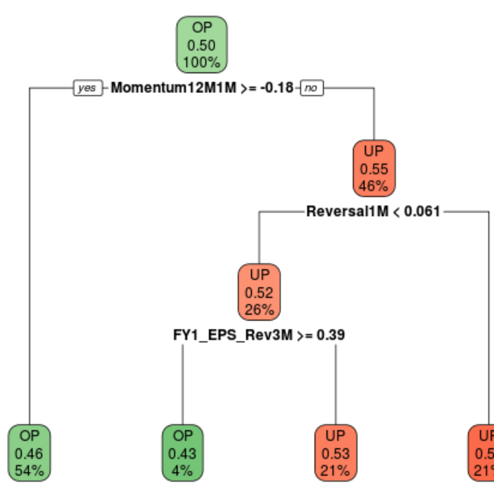
C) Quintile Portfolios, One-Month Reversal, Canada



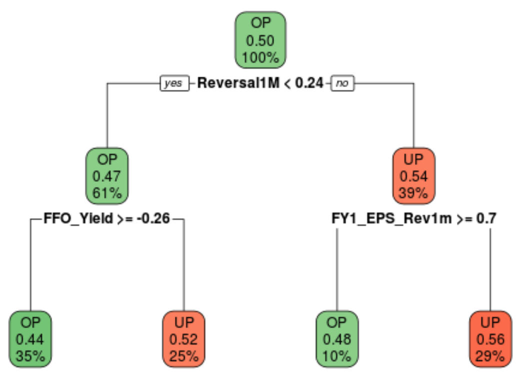
D) Quintile Portfolios, One-Month Reversal, US



E) RF-CART in Canada, 2/28/1998



F) RF-CART in the US, 8/31/2004



Note: For CART model graphs, **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

Linear Panel Data Econometric Component

The core of the LEAP model is still a traditional linear multifactor model. However, we add a few innovative features to deal with the most challenging part of linear regression models.

Linear regression models suffer from a number of weaknesses:

- **Dimensionality.** We have ~ 400 factors in our factor library. If we include all the factors in a linear regression, we need to estimate too many coefficients and the model tends to be highly unstable. In some markets, such as Canada, we have more factors than the number of stocks. In that case, an ordinary least squares regression can't even be estimated.
- **Multicollinearity.** Many of the ~400 factors are highly correlated. Linear models become highly unstable when the predictors are highly correlated (see Luo, et al [2017b] for demonstration of the impact of multicollinearity on model performance).
- Linear regression models do not automatically conduct **feature selection**.
- Linear regression models can't automatically handle **nonlinear patterns**. Although analysts can manually add higher order (e.g., quadratic or cubic) and interaction terms, it will further intensify the dimensionality and feature selection problem.

The RF-CART component introduced above captures the most important nonlinear patterns in our data. To deal with dimensionality and multicollinearity, we use both wrapper and filter feature selection methods in the linear regression part.

Modified Fama-MacBeth Regression

Each month, we perform the following cross-sectional Fama-MacBeth regression, on the previous 120 months of data:

$$r_{i,t} = \sum_{k=1}^{K_t} \beta_{k,t} f_{i,k,t-1} + \sum_{m=1}^{M_t} \gamma_{m,t} D_{i,m,t-1} + \epsilon_{i,t}$$

Where,

$r_{i,t}$ is the normalized and neutralized return for stock i in period t;

$\beta_{k,t}$ is the estimated coefficient (i.e., orthogonal return) for factor k in period t;

$f_{i,k,t-1}$ is the normalized and neutralized score of stock i, for factor k, in period $t-1$;

$\gamma_{m,t}$ is the estimated coefficient for the mth terminal node from our CART model in period t;

$D_{i,m,t-1}$ is the dummy variable for the mth terminal node from the CART model in period $t-1$;

$\epsilon_{i,t}$ is the regression residual (i.e., the random noise that can't be explained by our factors), for stock i in period t;

$i = 1, \dots, N_t$, where N_t is the number of stocks in our universe in period t;

$k = 1, \dots, K_t$, where K_t is the number of factors in period t; and

$m = 1, \dots, M_t$, where M_t is the number of terminal nodes from the CART model in period t.

However, as we pointed out in the earlier sections, linear regression is particularly sensitive to dimensionality (i.e., too many factors) and multicollinearity problems. Linear regression does not perform variable selection (feature selection). Therefore, we introduce two factor selection algorithms.

Filtering

The first feature selection algorithm is based on the filtering philosophy. At each month end, we conduct a univariate backtesting for each factor in our factor library. We then select the top 20% of factors from each one of the seven style categories (value, growth, momentum/reversal, sentiment, quality, alternative and Big Data), based on the performance in the previous 120 months. Performance is measured by risk-adjusted IC (i.e., $@mean(IC_t)/@stdev(IC_t)$). Factors that do not have 120 months of data are not included. Since we have around 400 factors, we end up with roughly 80 factors per month.

To deal with multicollinearity, we compute all pairwise correlation, based on the time series of factor IC's, i.e., performance correlation. Then we automatically remove those factors that are highly correlated with other factors.

Wrapper

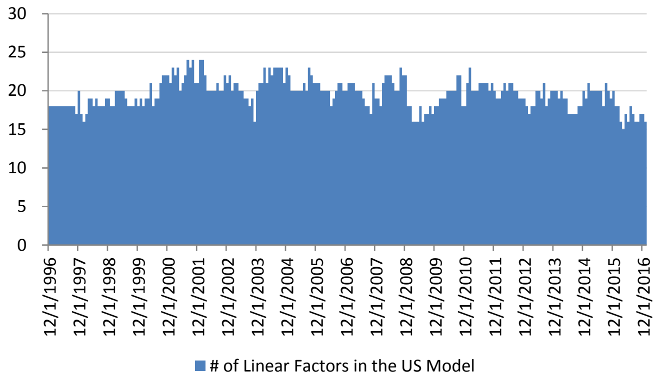
After the filtering stage, we conduct the second factor selection processing using a wrapper method – a backward selection procedure. At each month, we fit our cross sectional regression model with all factors. Then at each step, we remove the one with the highest p-value, until it reaches the desired threshold. We also require that we have at least one factor from each of the seven style buckets to maintain a fair representation. This is also computationally extensive. Suppose that we start with 80 factors, to eliminate one variable, we need to fit the cross-sectional regression 120 times. To reduce the number of factors from 80 to 20 would require $(80-20)\times 120 = 7200$ cross-sectional regressions. Even though linear regression can be processed fairly fast with matrix manipulation (e.g., about 0.01 second), for 30 years of monthly backtesting, it still takes about seven to eight hours with a quad-core computer. Conducting the backtesting on a 25-core Linux grid significantly reduces the computing time and we can finish the linear model backtesting within two hours for the US market.

In the final fitted linear model, we can retrieve the time series coefficient (orthogonal return) of each factor, $\beta_{k,t}$. Therefore, we essentially assume factor returns are time-varying. We make a further assumption that these factor returns are conditional on macroeconomic regimes and therefore, can be predicted based on macro variables discussed in the "Style Rotation" section. We will elaborate this point in the next section.

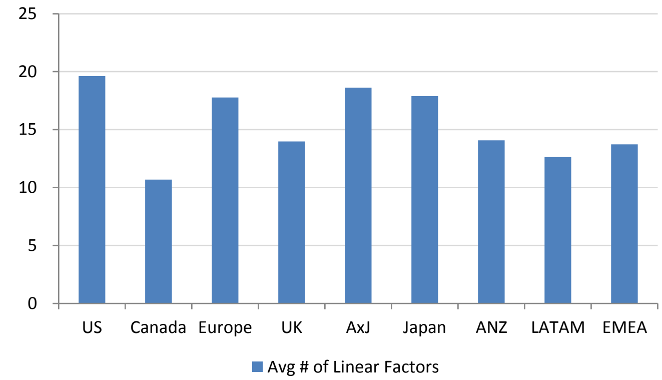
Eventually, the linear model selects around 15-25 factors in the US (see Figure 23 A). Globally, the linear component also appears to be fairly parsimonious, with around 10-20 factors on average (see Figure 23 B).

Figure 23 # of Linear Factors in the Model

A) US



B) Average # of Linear Factors, Global



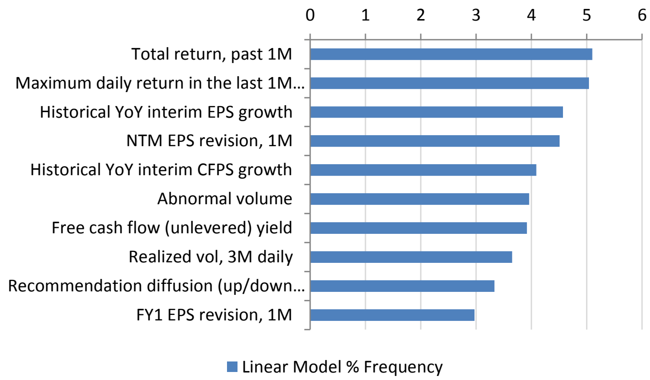
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

The linear model is also fairly intuitive. For example, Figure 24 (A) to (F) show the most frequently used linear factors in the US, Canada, Europe, UK, AxJ, and Japan. Compared to the most common RF-CART factors (see Figure 21), it is apparent that the frequencies for linear factors are much lower.

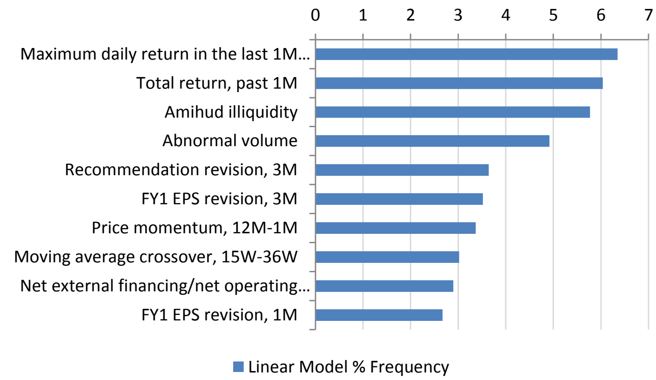
Some factors show up in most regions, e.g., one-month reversal, our lottery factor, recommendation diffusion/revision ratios, abnormal volume, free cash flow yield (or EBIT/EV), and to a lesser extent, price momentum.

Figure 24 Top 10 Most Frequent Linear Factors

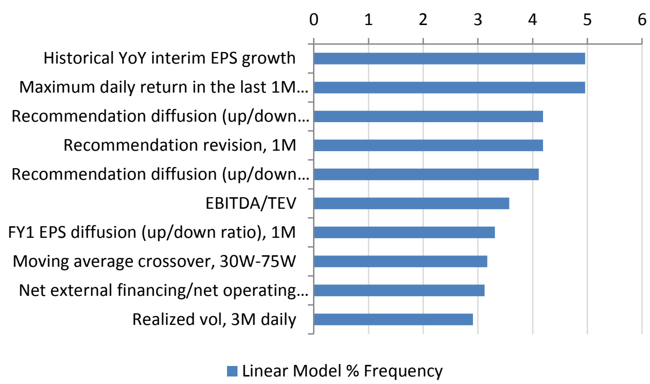
A) US



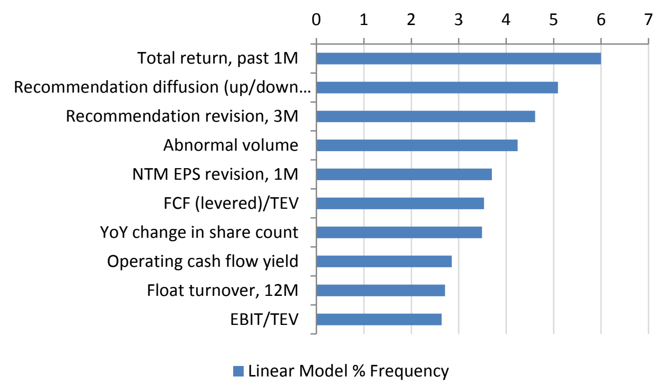
B) Canada



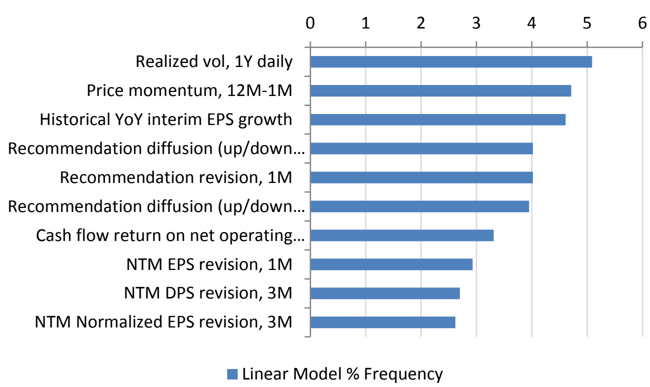
C) Europe



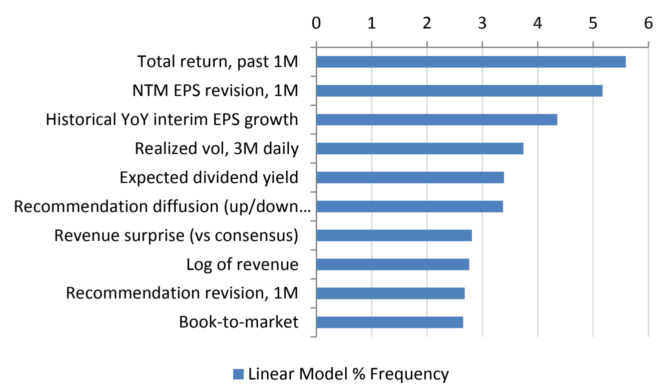
D) UK



E) AxJ



F) Japan



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

STYLE ROTATION

Dynamic factor modeling and in particular, style rotation has gained popularity since 2007 quant meltdown and 2008 global financial crisis. Investors gradually realize that factor returns are unstable, subject to structural breaks, and even worse, decay over time (see Luo, et al [2017a, 2017b], McLean and Pontiff [2012]). Factor rotation, or style rotation, potentially adds a new and uncorrelated source of alpha. There is significant suspicion among investors on the feasibility of style rotation. The typical arguments against style rotation include:

- Style rotation has low **breadth**. Instead of thousands of stocks for multifactor models, we only have a handful of factors to rotate from.
- Timing factor returns is extremely **difficult** or maybe not even possible. This argument follows similar logic as market timing. Because the Sharpe ratio of style rotation tends to be much lower than stock selection, mixing them also drags down performance.
- Style rotation significantly increases portfolio **turnover**.

While all above three points have merits, it does not mean style rotation is not useful. Despite of the low breadth and modest performance, style rotation tends to be uncorrelated to stock selection. The diversification benefit is particularly useful in an environment when the average return from traditional factors is low and unstable. Similarly, although style rotation injects additional turnover to the portfolio, the extra boost in performance may still outweigh transaction costs.

The linear factor selection and estimation process discussed in the previous sections are likely to choose those factors with relatively stable payoff patterns over time. Therefore, the unconditional returns from these factors are expected to be positive and significant. The style rotation element in the LEAP model, therefore, is expected to be only incremental.

We explore a relatively simple model for our style forecast. Mathematically, we fit the following time series equation each month:

$$\beta_{k,t} = \mu_k + \sum_{h=1}^{H} \varphi_{k,h} E_{h,t-1} + \sum_{l=1}^{L} \psi_{k,l} \beta_{k,t-l} + \epsilon_{k,t}$$

$$\gamma_{m,t} = \theta_m + \sum_{h=1}^{H} \nu_{m,h} E_{h,t-1} + \sum_{l=1}^{L} \chi_{m,l} \gamma_{m,t-l} + \varsigma_{m,t}$$

Where,

$\beta_{k,t}$ is the factor return (estimated from the previous section) for factor k in period t;

μ_k is the intercept term, i.e., the unconditional factor return, expected to be positive;

$E_{h,t-1}$ is the hth macro variable at period $t-1$, i.e., we use previous period's macro variable to predict the subsequent period's factor return;

$\varphi_{k,h}$ is the estimated coefficient for the kth factor on the hth macro variable;

$\psi_{k,l}$ is the estimated coefficient for the kth factor on its own lth lagged value, i.e., we assume some time series dependence;

$\epsilon_{k,t}$ is the regression residual for the factor return equation;

$\gamma_{m,t}$ is the CART node return (estimated from the previous section) for node m in period t;

θ_m is the intercept term for node m, i.e., the unconditional return, expected to be positive for nodes classified as Outperformers and negative for nodes classified as Underperformers;

$E_{h,t-1}$ is the hth macro variable at period $t-1$, i.e., we use previous period's macro variable to predict the subsequent period's factor return;

$\upsilon_{m,h}$ is the estimated coefficient for the mth terminal node on the hth macro variable;

$\chi_{m,l}$ is the estimated coefficient for the mth terminal node on its own lth lagged value, i.e., we assume some time series dependence;

$\varsigma_{k,t}$ is the regression residual for the node return equation;

$k = 1, \ldots, K_t$, where K_t is the number of factors in period t; and

$h = 1, \ldots H$, where H is the number of macro variables; and

$l = 1, \ldots L$, where L is the number of time lag.

Once the above equation is estimated, using the previous 120 months of data, we can use it to forecast the following period's factor returns:

$$\hat{\beta}_{k,t+1} = \hat{\mu}_k + \sum_{h=1}^{H} \hat{\varphi}_{k,h} E_{h,t} + \sum_{l=1}^{L} \hat{\psi}_{k,l} \beta_{k,t-l+1}$$

$$\hat{\gamma}_{k,t+1} = \hat{\theta}_k + \sum_{h=1}^{H} \hat{\upsilon}_{k,h} E_{h,t} + \sum_{l=1}^{L} \hat{\chi}_{k,l} \beta_{k,t-l+1}$$

Where, $\hat{\mu}_k$, $\hat{\varphi}_{k,h}$, $\hat{\psi}_{k,l}$, $\hat{\theta}_k$, $\hat{\upsilon}_{k,h}$, $\hat{\chi}_{k,l}$ and are estimated using rolling 120 months of data.

The choice of what macro variables to include in the style rotation model is the next question we want to address. We decide to choose the variables based on a few criteria:

- **Transparency and intuition.** Because the LEAP model strives to be transparent and intuitive, we choose macro variables for the same reason.
- Representing the major facets of the **underlying economy**. We choose a balance of macroeconomic variables (e.g., business cycle), policy and economic uncertainty, market sentiment (e.g., VRP, credit spread), central bank and the yield curve (e.g., yield curve), international trade (as reflected in currency), and commodities (proxyed by crude oil), and seasonality (e.g., January, Sell-in-May, or December).
- **Global versus local**. For each of the nine markets, we choose local indicators for most economic variables. As the global capital markets are fully integrated, we occasionally also use US, global, or super-regional variables (e.g., US VRP[14]).

[14] We have done extensive research on our VRP indicator. VRP or Variance Risk Premium represents the difference between the market implied variance and realized variance. The VRP is a contrarian indicator.

- **Parsimoniousness**. At each given point in time, we use a rolling 120 months of data to estimate factor returns; therefore, for each factor, we only have 120 months of return data. Therefore, it is essential to have a small set of variables.

In the end, we choose five to nine variables for each region (see Figure 25). More details about how these variables are selected will be covered in a forthcoming research on our global macro strategy.

Figure 25 Variables in the Style Rotation Model

Region	VRP	Political Uncertainty	Economic Uncertainty	Interest Rate	Equity Market	Credit Spread	Business Cycle	Currency	Oil & Commodities	Seasonality
US	US VRP	US	US	US	US	US Corporate	Philly Fed ADS			√
Canada	US VRP	Canada	Canada	US				CAD/USD		
LATAM	US VRP	US	US		US			Trade Wgted USD		
Europe	US VRP	Europe	US		Europe	Europe High Yield	Exp Europe GDP	EUR/USD		√
UK		UK	UK	UK	Global	US High Yield		GBP/USD		√
Emerging EMEA	US VRP	Europe	US		Global			Trade Wgted USD		
AxJ		US	AxJ	US	AxJ	US High Yield	Exp AxJ GPD		Oil Shock	
Japan			Japan			US High Yield	Japan Nowcasting			√
ANZ	US VRP	Australia	ANZ	Australia	Australia		Exp ANZ GDP	AUD/USD	Baltic	

Sources: Wolfe Research Luo's QES

Factor Return Shrinkage

Because of the dimensionality issue (120 months of time series data, but a large number of macro predictors), the predicted return for each factor is subject to large estimation error. Therefore, we may want to try to further shrink the prediction towards long-term (120-month) and short-term (12-month) averages[15]. The shrinkage estimators also reduce factor turnover and therefore portfolio turnover.

$$\tilde{\beta}_{k,t+1} = \frac{1}{3}\left(\hat{\beta}_{k,t+1} + \frac{1}{120}\sum_{\tau=1}^{120}\beta_{k,t-\tau+1} + \frac{1}{12}\sum_{\tau=1}^{12}\beta_{k,t-\tau+1}\right)$$

$$\tilde{\gamma}_{k,t+1} = \frac{1}{3}\left(\hat{\gamma}_{k,t+1} + \frac{1}{120}\sum_{\tau=1}^{120}\gamma_{k,t-\tau+1} + \frac{1}{12}\sum_{\tau=1}^{12}\gamma_{k,t-\tau+1}\right)$$

Stock Return Prediction

To summarize, the predicted return from LEAP model comes from the following three models:

$$\hat{r}_{i,t+1} = \sum_{k=1}^{K_t}\tilde{\beta}_{k,t+1}f_{i,k,t} + \sum_{m=1}^{M_t}\tilde{\gamma}_{m,t+1}D_{i,m,t}$$

- **RF-CART.** First of all, each stock, based on its factor scores, shall fall into one RF-CART terminal node, $D_{i,m,t}$. This is the nonlinear element.
- **Linear Factors and Returns.** Second, from the final linear model, we should know the exact factors ($f_{i,k,t}$) being selected. From the shrinkage-based style rotation model, we should know

[15] The long-term (120-month) average is also the sample mean.

the coefficient/return to be applied to each factor ($\tilde{\beta}_{k,t+1}$) and for each RF-CART terminal node ($\tilde{\gamma}_{m,t+1}$).

- **LEAP Model.** Lastly, we know the factor scores and RF-CART terminal node for each stock, we should be able to compute the implied return prediction for each stock.

Model Contribution

To summarize, let's compare our LEAP model with the Benchmark Model (BM)[16].

- The BM is a traditional linear multifactor model, by equally weighting eight common factors (trailing earnings yield, book-to-market, FY1 expected earnings growth, price momentum, earnings revision, ROE, debt/equity ratio, and earnings quality). The BM represents the most conventional structures, e.g., common factors, static and naïve weighting scheme, linear functional form.
- The LEAP model employs a number of sophisticated modeling techniques:
  - A wide range of ~400 factors, including an increasing number of unconventional Big Data signals;
  - Nonlinear classification machine learning algorithms – random forest and CART;
  - Innovative factor selection techniques from both wrapper and filter categories;
  - Robust time series cross sectional panel data econometric model
  - Incorporating our strength in global macro data and style rotation

To demonstrate the incremental benefit from each of the main modeling techniques, we choose two of the most challenging markets for alpha generation – the US and Japan to show the incremental alpha from each of the three key elements:

- A large factor library, linear factor selection algorithm, and panel data econometrics,
- RF-CART nonlinear factor selection and modeling, and
- Style rotation

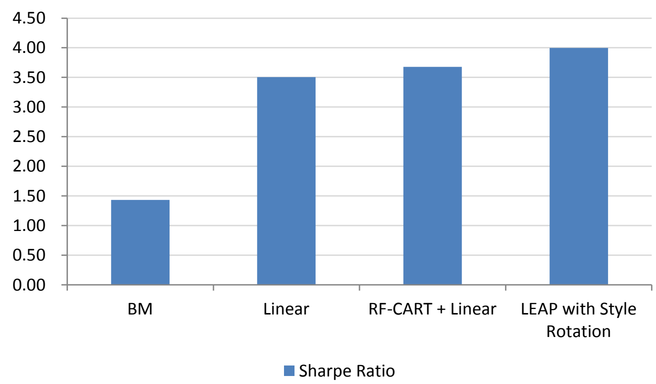
As shown in Figure 26 (A) and (B), clearly the first element (i.e., factor library, linear factor selection, and panel data econometrics) adds the most significant alpha, by increasing the Sharpe ratio by 145% and 123% in the US and Japan, respectively. The RF-CART nonlinear component further lifts Sharpe ratio by 5% and 6% in the two regions. Finally, the style rotation model boosts the performance by another 9% and 14%, respectively.

If we measure performance by risk-adjusted IC, we see a similar improvement from the BM to the most sophisticated LEAP model. More importantly, the maximum drawdown in the US (and Japan) is cut down from -33% (and -21%) to -4% (and -7%).

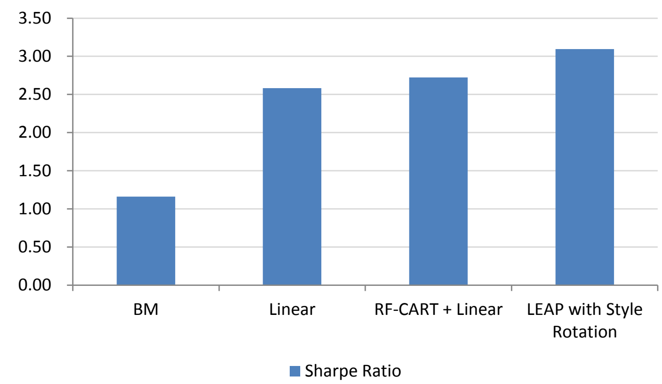
[16] The BM is defined in Luo, et al [2017b].

Figure 26 The Incremental Benefits from the LEAP Model

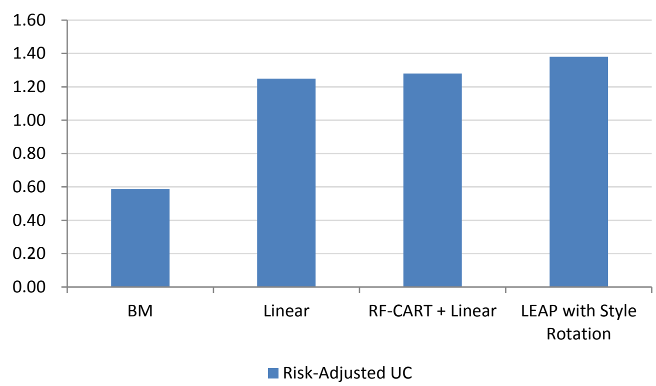
A) Sharpe Ratio, US



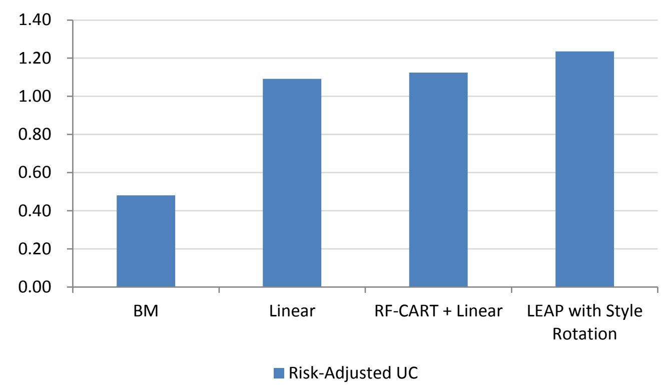
B) Sharpe Ratio, Japan



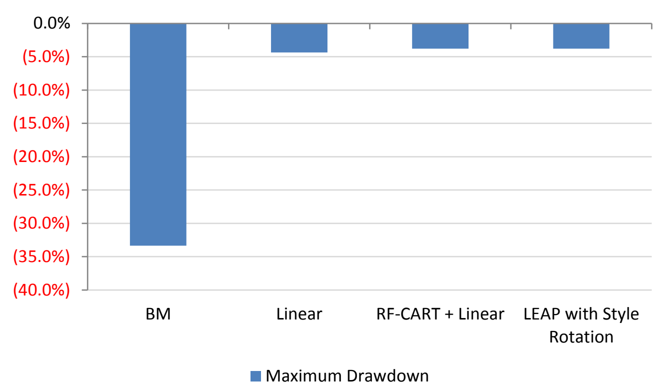
C) Risk-Adjusted IC, US



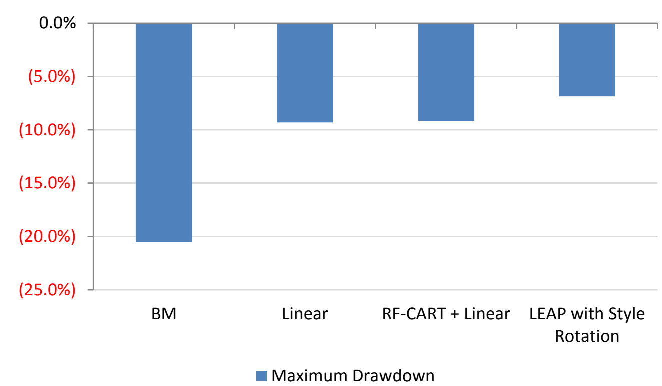
D) Risk-Adjusted IC, Japan



E) Maximum Drawdown, US



F) Maximum Drawdown, Japan



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

MODEL'S OUT-OF-SAMPLE PERFORMANCE

In this section, we show the LEAP model's predictive power. As discussed in Luo, et al [2017b], we measure a model's predictive power using primarily:

- Long/short quantile portfolio returns
- Spearman rank IC

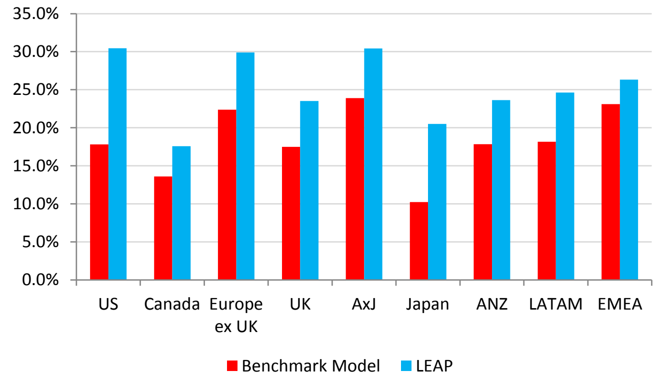
A more realistic portfolio simulation, incorporate typical institutional constraints (e.g., turnover, leverage, risk control, etc.) and transaction costs will be presented in Part IV of this research series.

Performance across Regions

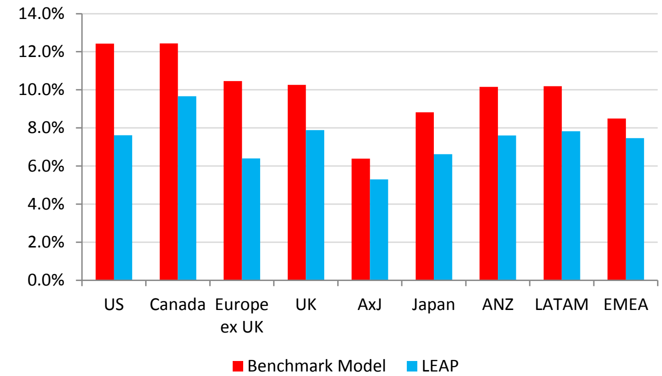
The LEAP model delivers superior returns (see Figure 27 A) and lower risk (see Figure 27 B) than the Benchmark Model (BM) in all nine regions. After adjusting for risk, the LEAP model almost triples the Sharpe ratio in the US (see Figure 27 C). The lift in performance in the US, Japan, and Europe is particularly noticeable (see Figure 27 C and D). On a monthly basis, the LEAP model has generated positive alpha almost 90% of the time in all regions (see Figure 27 E). Furthermore, the LEAP model is able to reduce the downside risk by almost 90% in the US and to a large extent in other regions (see Figure 27 F).

Figure 27 LEAP Model Performance across Regions

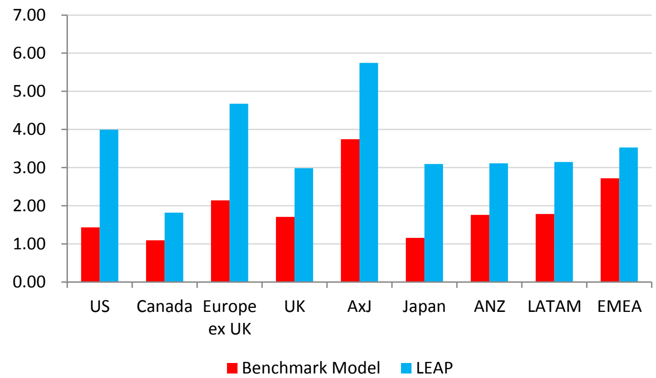
A) Long/Short Quintile Portfolio Annual Return



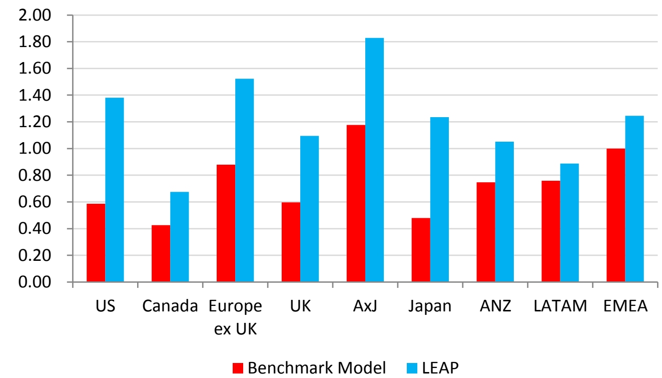
B) Long/Short Quintile Portfolio Annual Volatility



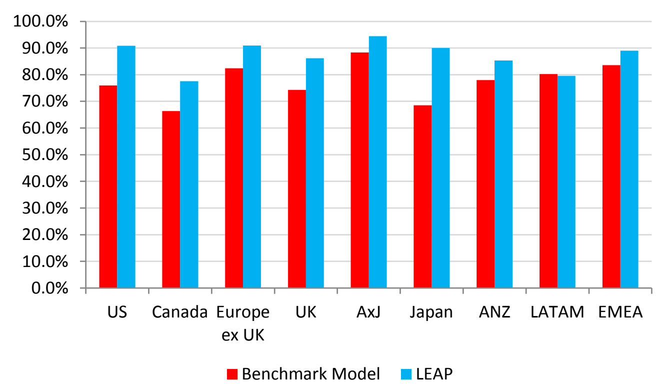
C) Sharpe Ratio



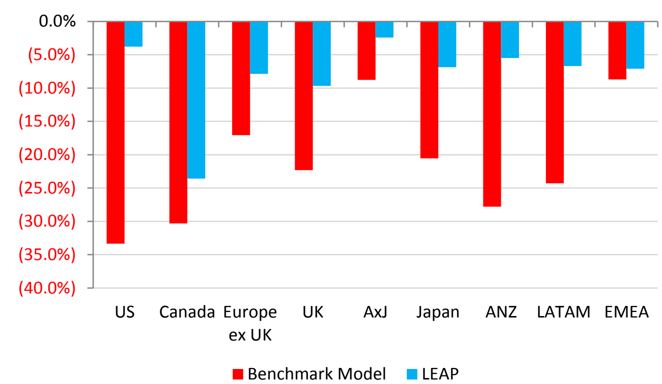
D) Risk-Adjusted Rank IC



E) Hit Rate



D) Maximum Drawdown



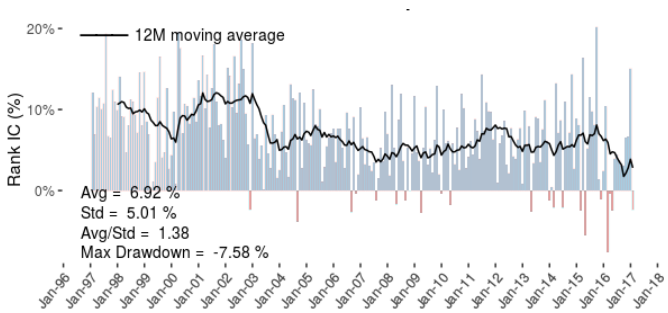
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Performance over Time

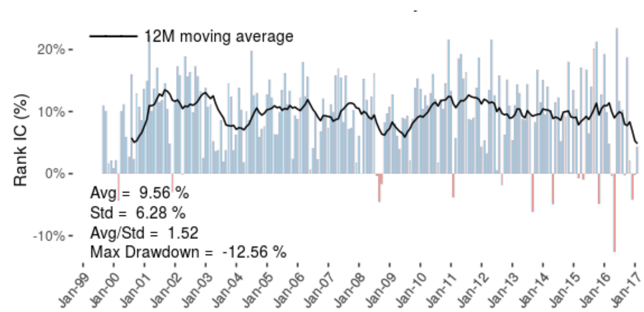
Figure 28 (A) to (D) show the LEAP model's rank IC over time, for the US, Europe, AxJ, and Japan. The model's performance has stayed relatively stable over the past 20 years.

Figure 28 LEAP Model – Rank IC

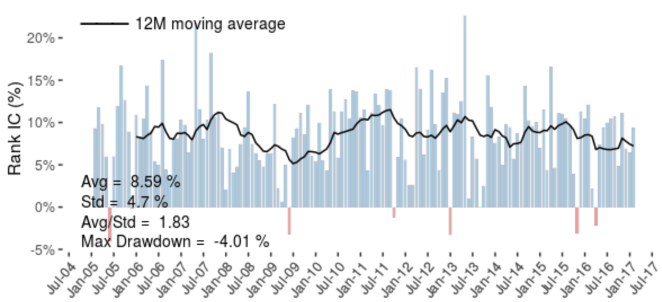
A) US



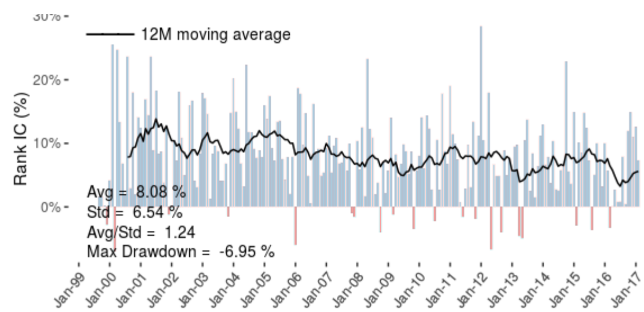
B) Europe



C) AxJ



D) Japan



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

A Case Study of Two Challenging Periods

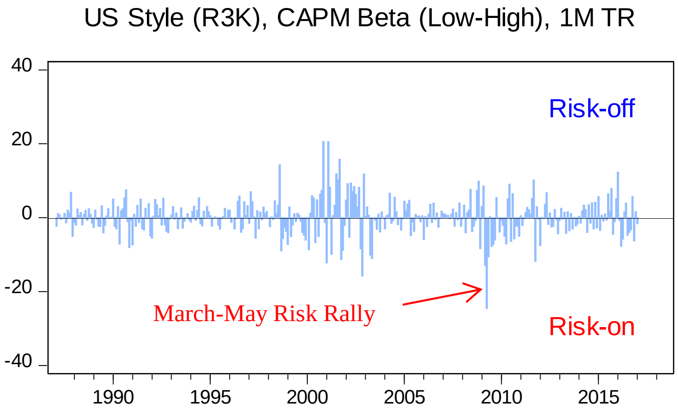
As seen in the previous section, the LEAP model has outperformed the benchmark model considerably in all nine regions. The model's performance has also been fairly consistent over time. In this section, we want to zoom into two specific periods that many traditional quantitative investors struggled –March-May 2009 and then 2016.

March-May 2009

As detailed in Luo, et al [2017a], we witnessed a drastic risk rally in March-May 2009. While the US economy reached the bottom of the recession, the market sentiment improved sharply, which triggered a strong appetite for risky assets (see Figure 29 A). In less than three months, price momentum and value factors were down by -45% and -30%, respectively (see Figure 29 B).

Figure 29 March-May Risk Rally

A) The Return of the Low Beta Factor



B) March-May 2009 Risk Rally

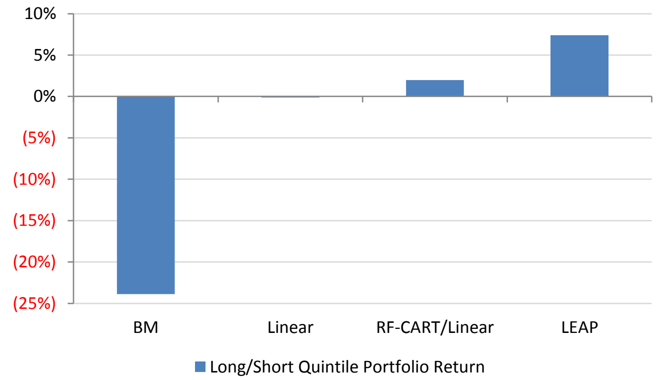
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

As shown in Figure 30 (A), in the three-month period, the BM suffered a loss of -24%. Interestingly, the linear econometric component completely eliminated the drawdown. The RF-CART nonlinear element further lifted the return to a positive territory. Lastly, the style rotation model boosted the performance to an upswing of 7.4% in three month.

- **Linear Factor Selection Model.** As shown in Figure 30 (B), the linear factor selection algorithm picked a set of 16 factors on March 31, 2009 to predict stock returns for April 2009. The list of the 16 factors is almost entirely different from the ones in the BM. In particular, the LEAP model chose short-term reversal over mid-term price momentum.
- The **RF-CART model** provided some additional diversification benefit. It is interesting to note that mid-term price momentum did enter the model, but in a nonlinear fashion (see Figure 30 C).
- Lastly, the **style rotation model** delivered superior return during the three months, which further highlights the diversification benefit from macro timing. As shown in Figure 30 D, the style rotation model was correct in predicting a negative return from value (free cash flow yield) and quality (net external financing/net operating assets), and assigning significant weight to reversal.

Figure 30 The Anatomy of March-May Risk Rally

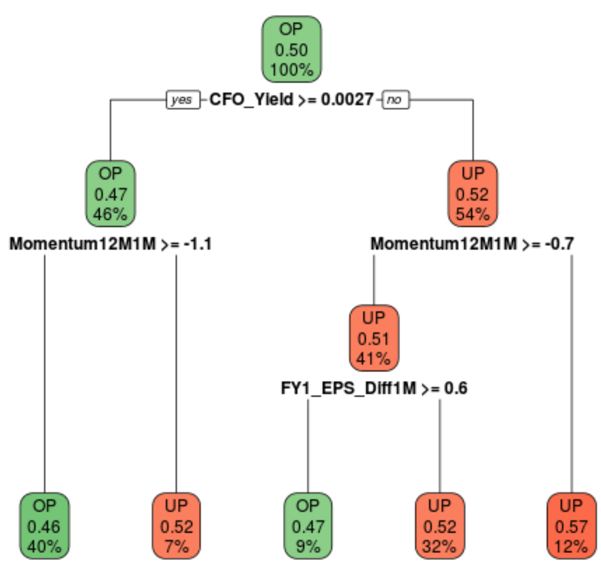
A) Model Performance, Cumulative Return, March-May 2009



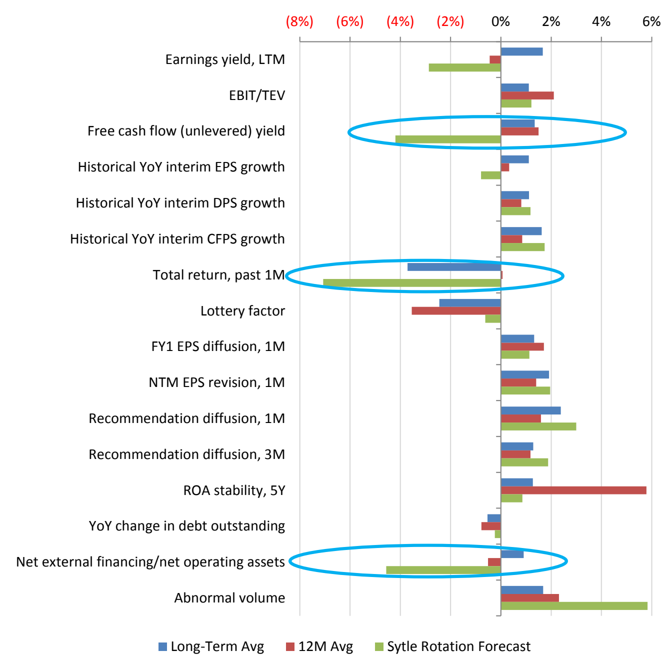
B) Linear Factors Selected, 3/31/2009

Style Category	BM	LEAP
Value	Earnings yield, LTM Book-to-market	Earnings yield, LTM EBIT/TEV Free cash flow (unlevered) yield
Growth	FY1 Exp EPS growth	Historical YoY interim EPS growth Historical YoY interim DPS growth Historical YoY interim CFPS growth
Momentum/Reversal	Price momentum, 12M-1M	Total return, past 1M Lottery factor
Sentiment	Earnings revision, 3M	FY1 EPS diffusion, 1M NTM EPS revision, 1M Recommendation diffusion, 1M Recommendation diffusion, 3M
Quality	ROE Debt/Equity Earnings quality, accruals	ROA stability, 5Y YoY change in debt outstanding Net external financing/net operating assets
Alternative		Abnormal volume

C) RF-CART, 3/31/2009



D) Style Rotation Prediction, 3/31/2009



Note: For CART model graphs, **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.

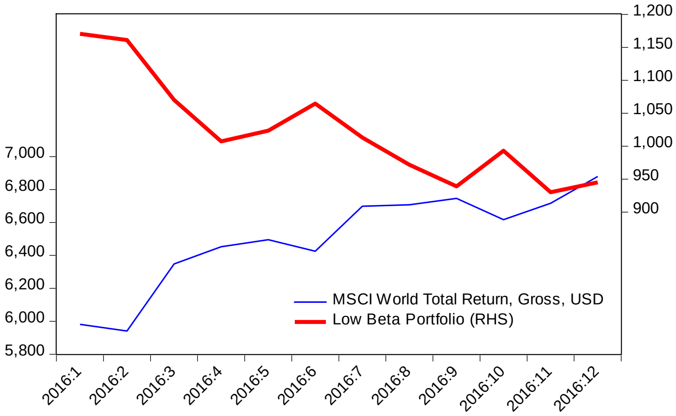
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES

Zoom into 2016

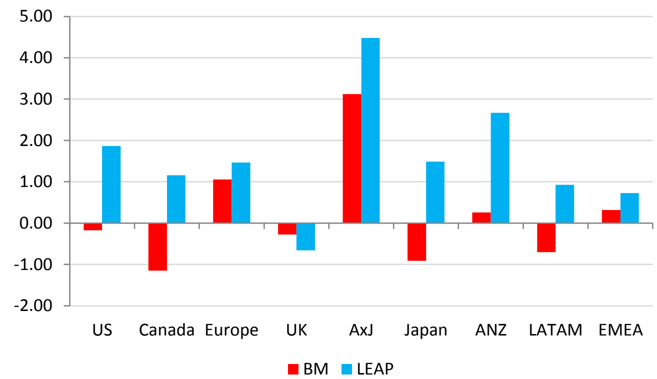
Similar to the March-May 2009 experience, the year of 2016 was also mostly dominated by a risk-on sentiment. As shown in Figure 31 (A), despite of political shocks from Brexit and the surprising election in the US, the global market gathered steam in 2016. As investors' sentiment improved and the "animal spirits" built up, the return of the low risk factor plunged – investors embraced risky assets for the most part of 2016. For managers with a positive tilt towards price momentum, high quality, and defensive stocks, 2016 was another difficult year. The LEAP model again proves to be extremely resilient. As shown in Figure 31 B, the LEAP model outperformed the BM in eight of the nine regions. In the US, Canada, Japan, and LATAM, in particular, the BM suffered from hefty losses, while the LEAP model delivered decent returns.

Figure 31 A Deep Dive into 2016

A) The Return of the Low Beta Factor



B) Model Performance Comparison, Sharpe Ratio



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

Model Turnover and Decay

No model is perfect, which also applies to the LEAP model. The downside of the model lies in its high turnover. As shown in Figure 32 (A), the average serial correlation of the LEAP model alpha in the US is around 45% – in line with typical earning revision type of factors, but much lower than most fundamental factors. Globally, the LEAP model also has lower signal autocorrelation (therefore, higher turnover) than the benchmark model (see Figure 32 B).

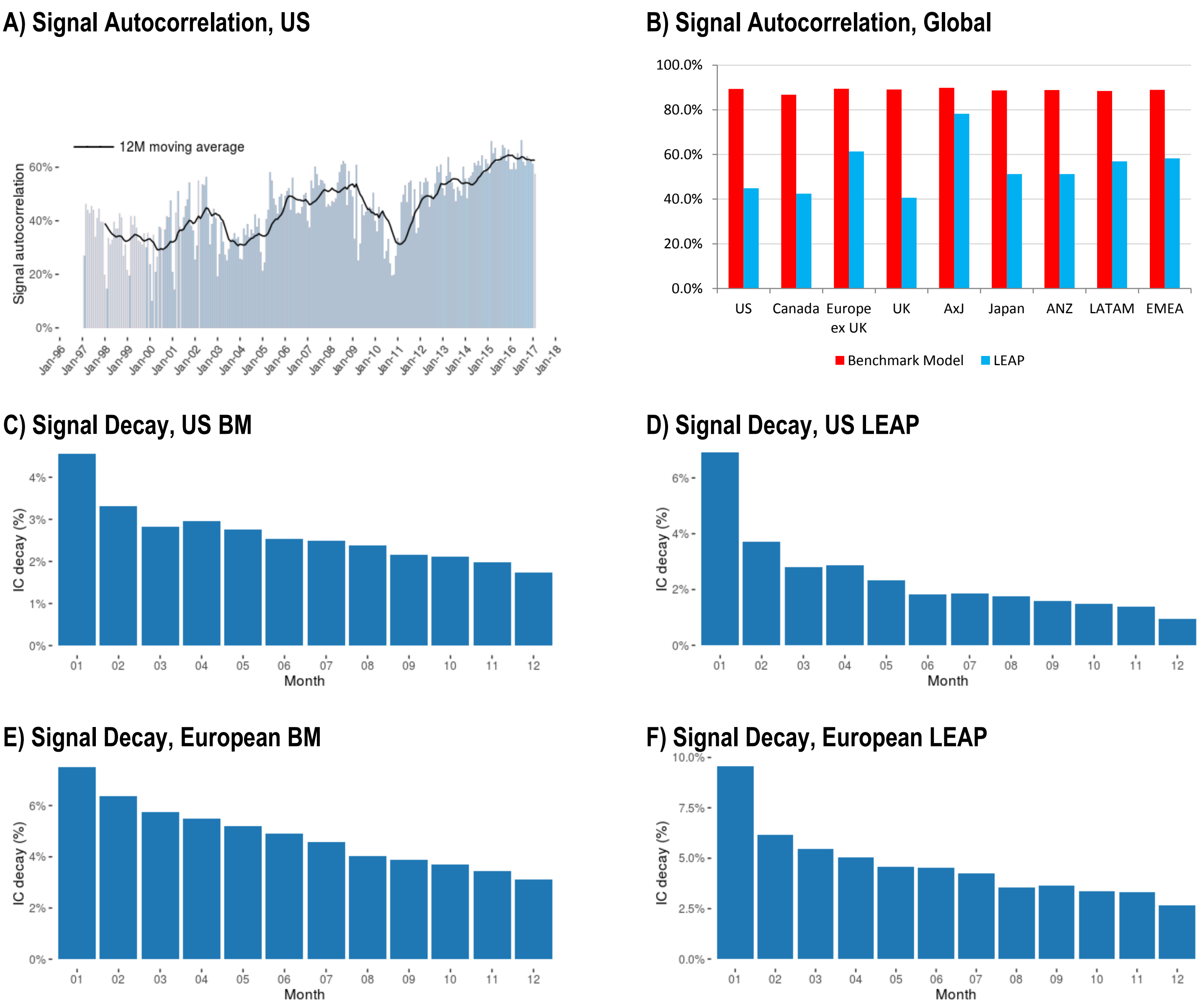
Despite of the relatively high turnover, the LEAP model's forecasting ability sustains for more than a year (see Figure 32 D and F). The LEAP model has a stronger predictive power over the BM for a forecasting horizon up to six months and then is on par with the BM for longer term projection.

In our previous research, we find that high turnover by itself is not necessarily detrimental (see Luo, et al [2017b]). In liquid markets such as the US, Europe, and Japan, a model's predictive power is normally more important than controlling for turnover. On the other hand, in emerging markets, transaction costs tend to be high and it is critical to incorporate turnover in portfolio implementation. Fortunately, the LEAP model's turnover appears to be much more modest in emerging markets such as AxJ, LATAM, and EMEA.

In our previous research, we also find that even for a long-term buy-and-hold manager, a strong model with high turnover can still be used for timing purpose. For example, a manager may decide to invest (or divest) a given name, but is not sure when to enter the transaction. The LEAP model can be particularly used in those occasions.

We will elaborate how to take into account of turnover and transaction costs in portfolio construction in Part IV of this research series.

Figure 32 LEAP Model – Turnover and Decay



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

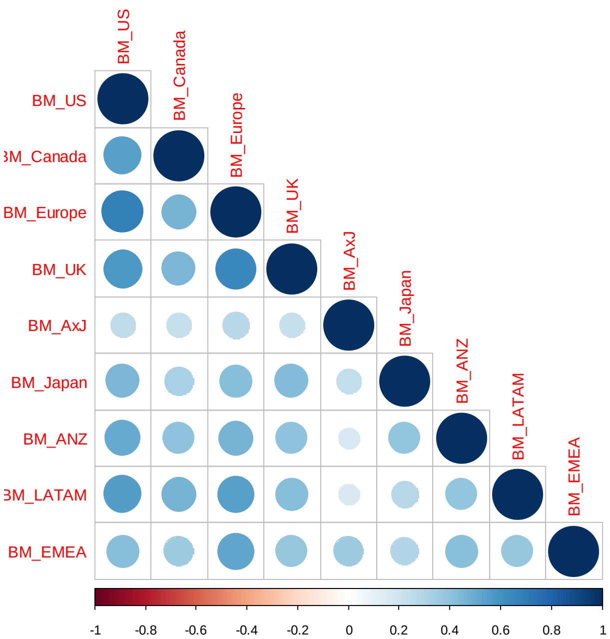
Diversification Benefit

The Benchmark Model (BM), being a more generic strategy, is somewhat correlated across regions. As shown in Figure 33 (A), the average pairwise correlation among the BM strategies is about 41%. On the other hand, as a purer alpha signal, the LEAP model is substantially less correlated across regions (see Figure 33 B). Indeed, the average pairwise correlation among the nine regions is only

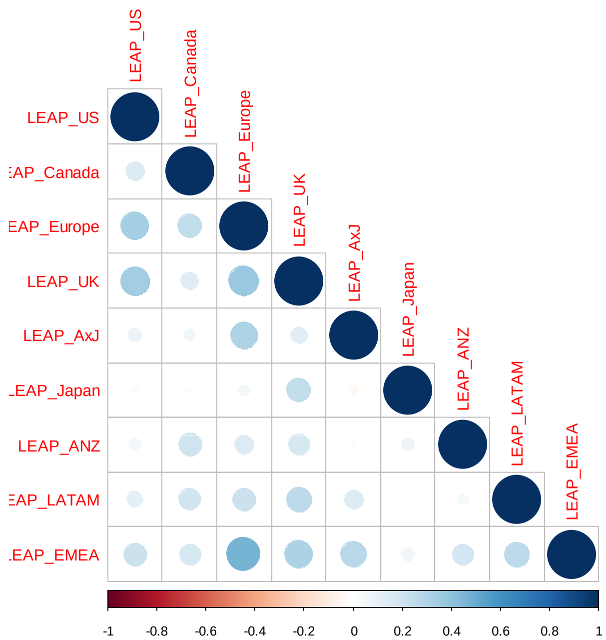
17%, less than half of the BM correlation. Investing in the LEAP model globally provides a much higher diversification benefit for active managers. A cluster analysis on the BM and LEAP model across regions also reveal very different pictures (see Figure 33 C and D).

Figure 33 LEAP Model Diversification Benefit

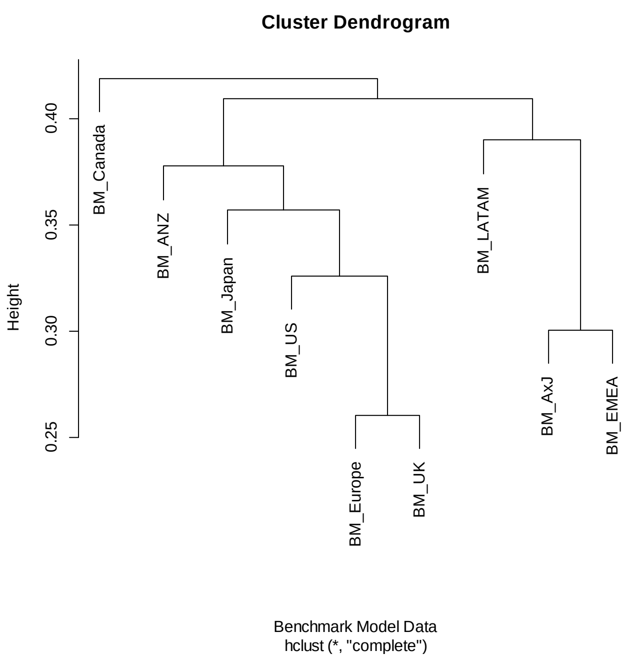
A) BM Correlation



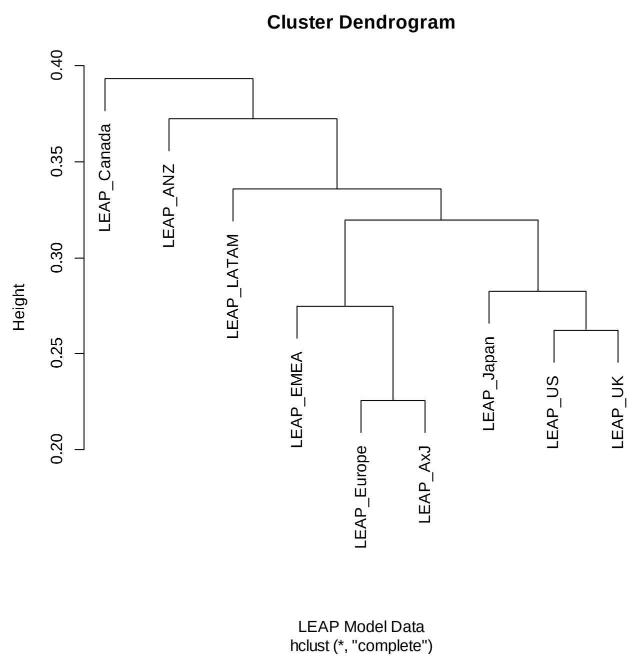
B) LEAP Correlation



C) BM Cluster Analysis



D) LEAP Cluster Analysis



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

Performance Summary

Figure 34 (A) and (B) summarize the key statistics for the BM and LEAP, respectively.

Figure 34 Performance Summary

A) LEAP

LEAP	Avg # of Stocks	Start Date	End Date	Long/Short Quintile (%)	Annualized Vol (%)	Sharpe Ratio	Rank IC (%)	Risk-Adjusted IC	Hit Rate	Max Drawdown (%)	Serial Correlation
US	2,984	1/31/1997	1/31/2017	30.5%	7.6%	4.00	6.9%	1.38	90.9%	(3.8%)	44.9%
Canada	256	1/31/1997	1/31/2017	17.6%	9.7%	1.82	5.4%	0.67	77.6%	(23.6%)	42.4%
Europe ex UK	1,369	8/31/1999	1/31/2017	29.9%	6.4%	4.67	9.6%	1.52	91.0%	(7.9%)	61.3%
UK	485	8/31/1999	1/31/2017	23.5%	7.9%	2.98	7.1%	1.09	86.2%	(9.7%)	40.6%
AxJ	1,719	1/31/2005	1/31/2017	30.4%	5.3%	5.74	8.6%	1.83	94.5%	(2.4%)	78.2%
Japan	1,386	8/31/1999	1/31/2017	20.5%	6.6%	3.10	8.1%	1.24	90.0%	(6.9%)	51.2%
ANZ	348	8/31/2004	1/31/2017	23.6%	7.6%	3.11	6.7%	1.05	85.3%	(5.5%)	51.2%
LATAM	297	11/30/2004	1/31/2017	24.6%	7.8%	3.15	6.6%	0.89	79.6%	(6.7%)	56.9%
EMEA	425	12/31/2004	1/31/2017	26.3%	7.5%	3.53	8.7%	1.25	89.0%	(7.1%)	58.3%

B) Benchmark Model

BM	Avg # of Stocks	Start Date	End Date	Long/Short Return	Annualized Vol	Sharpe Ratio	Rank IC (%)	Risk-Adjusted IC	Hit Rate	Max Drawdown (%)	Serial Correlation (%)
US	2,984	1/31/1997	1/31/2017	17.8%	12.4%	1.43	4.6%	0.59	75.9%	(33.4%)	89.4%
Canada	256	1/31/1997	1/31/2017	13.6%	12.4%	1.09	4.1%	0.43	66.4%	(30.3%)	86.8%
Europe ex UK	1,368	8/31/1999	1/31/2017	22.4%	10.5%	2.14	7.5%	0.88	82.4%	(17.1%)	89.4%
UK	485	8/31/1999	1/31/2017	17.5%	10.3%	1.71	5.3%	0.60	74.3%	(22.3%)	89.1%
AxJ	1,680	12/31/2004	1/31/2017	23.9%	6.4%	3.74	6.9%	1.18	88.4%	(8.8%)	89.9%
Japan	1,386	8/31/1999	1/31/2017	10.2%	8.8%	1.16	4.3%	0.48	68.6%	(20.5%)	88.7%
ANZ	348	8/31/2004	1/31/2017	17.8%	10.2%	1.76	6.1%	0.75	78.0%	(27.8%)	88.8%
LATAM	297	11/30/2004	1/31/2017	18.2%	10.2%	1.78	6.5%	0.76	80.3%	(24.3%)	88.4%
EMEA	425	12/31/2004	1/31/2017	23.1%	8.5%	2.72	7.3%	1.00	83.6%	(8.7%)	88.9%

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

ACTIVE VERSUS PASSIVE

In this paper and Luo, et al [2017b], we define a few multifactor models, varying from alternative beta, traditional active, and our innovative LEAP model:

- **BM (Benchmark Model).** This is a proxy for typical quant funds. We equally weight eight common stock-selection factors.
- **G&K (Grinold & Kahn).** The G&K model is similar to the BM. However, rather than equal weighting, we apply a mean-variance optimization on the factor space to derive the weight for each factor dynamically over time.
- **GMV (Global Minimum Variance).** The GMV model is purely risk based, by minimizing the expected volatility of the combined multifactor model. The GMV model overweighs factors that are less volatile and uncorrelated to other factors.
- **ABP (Alternative Beta Portfolio).** The ABP combines eight factor portfolios using a risk parity (equal risk contribution) algorithm. It is also purely risk based.

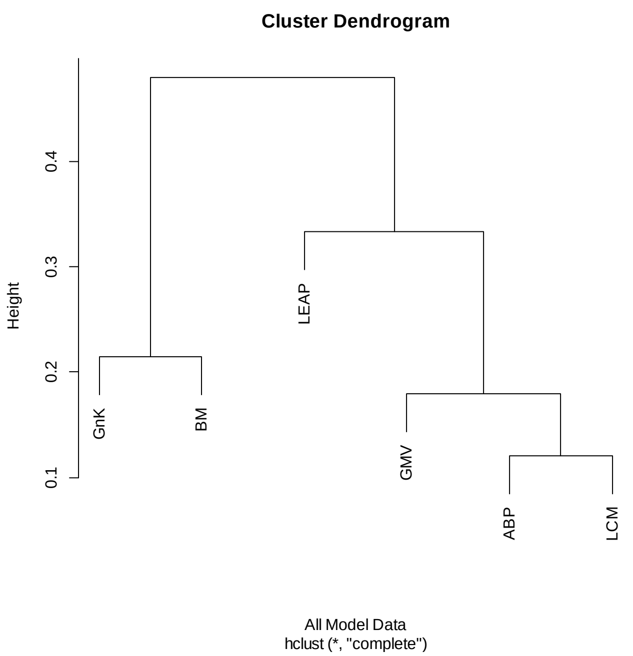
- **LCM (Least Crowded Model).** The LCM applies our proprietary minimum tail dependence algorithm to derive factor weights, by overweighing factors that are least correlated to other factors at the tail.
- **LEAP (L-Economic Alpha Processing).** The LEAP model is based on sophisticated machine learning and econometric techniques, with a macro style rotation overlay, as a proxy for a pure alpha model.

A cluster analysis reveals that the LEAP model is very different from all the other portfolios (see Figure 35 A). The two alternative beta portfolios – ABP and LCM first join each other, followed by the GMV – all three strategies are risk based allocations. The two naïve alpha models – BM and G&K form another distinctive cluster.

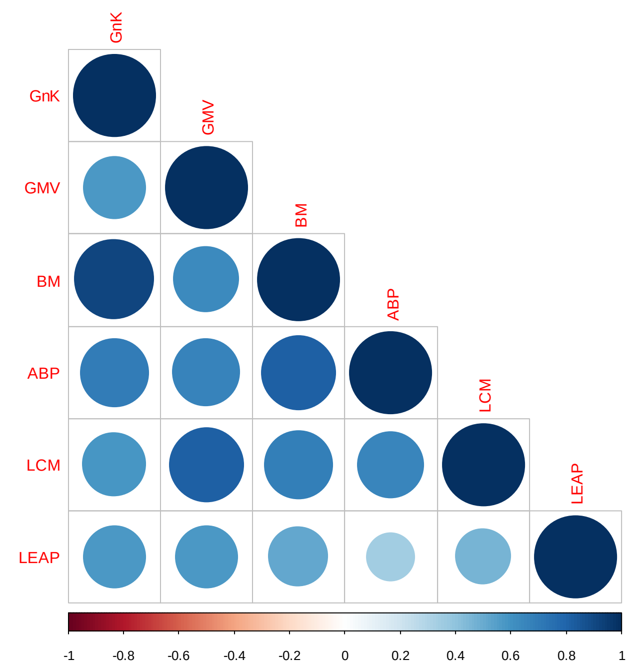
Similarly, a correlation plot (see Figure 35 B) confirms our view that the LEAP model is uncorrelated to other strategies.

Figure 35 A Comparison of All Common Models

A) A Cluster Analysis



B) Correlation Plot



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

Forthcoming Research

As of now, we have not incorporated any of the real-life institutional constraints yet. For example, we assume that we can go long and short any stock for an unlimited quantity. In reality, we may face many restrictions on foreign ownership, compliance restricted list, and more importantly, we may not be able to locate the necessary borrow for our shorting requirements. We have not taken into account of transaction costs. Although trading is becoming cheaper every day, with commissions approaching to zero, the biggest part of execution cost is the hidden element of market impact and bid-ask spread. We assume that we can leverage, short, and face no turnover limit, while in practice most institutional funds may have many portfolio constraints.

On the positive side, we have not yet incorporated risk in portfolio construction. Balancing risk and return is one crucial element of portfolio implementation. We still need to conduct performance attribution and understand the sources of risk and return for our portfolio. We potentially may want to engage in active hedging to eliminate undesirable risk exposures. All of these may help us further boost risk-adjusted performance.

In the next few weeks, we will publish the last part of our launch series, precisely on all of the above portfolio implementation issues – *Part IV: Risk, Portfolio Construction, Trade Execution, and Performance Attribution*.

In addition to the four-part introduction of Big Data and Machine learning in global equity investing, we are also working on a number of issues:

- From Nowcasting to Forecasting – Economics and Portfolio Strategy in the New Age
- Industry-Specific Models in Global Banking and Insurance Industries
- Accounting Quality, Fraud Detection, and Corporate Governance
- Factors based on Alternative Data Sources
- Machine Learning in Global Stock Selection

BIBLIOGRAPHY

Ardia, D., Boudt, K., and Wauters, M. [2016]. "The Economic Benefits of Market Timing the Style Allocation of Characteristic-Based Portfolios", *North American Journal of Economics and Finance*, Vol. 37, 38-62

Ayres, I. [2007]. *Super Crunchers: Why Thing-by-Numbers Is The New Way to Be Smart*, Bantam

Asness, C., Friedman, J.A., Krial, R.J., and Liew, J.M. [2000]. "Style Timing: Value versus Growth", *Journal of Portfolio Management*, 26, 50-60

Bai, J., and Perron, P. [1998]. "Estimating and Testing Linear Models with Multiple Structural Changes", *Econometrica*, 66, 47-78

Bailey, D.H., Borwein, J.M., De Prado, M.L., and Zhu, Q.J. [2014]. "Pseudo-Mathmatics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance", SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2308659

Baker, S.R., Bloom, N., and Davis, S.J. [2015]. "Measuring Economic Policy Uncertainty", NBER Working Paper

Bernstein, R. [1995]. *Style Investing: Unique Insight into Equity Management*, New York: John Wiley & Sons

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

Clark, T. [2004]. "Can Out-of-Sample Forecast Comparisons Help Prevent Overfitting?", *Journal of Forecasting*, 23(2)

Corbett, A. J. [2016]. "Are Style Rotating Funds Successful at Style Timing? Evidence from the US Equity Mutual Fund Market", SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2736672

Friedman, J., Hastie, T., and Tibshirani, R. [2010]. "Regularization Paths for Generalized Linear Models via Coordinate Descent", *Journal of Statistical Software*, 33(1), 1-22

Freund, Y., and Schapire, R. [1996]. "Experiments with a New Boosting Algorithm", *Machine Learning: Proceedings of the Thirteenth International Conference*

Hastie, T., Tibshirani, R., and Friedman, J. [2008]. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2nd Edition

Hawkins, D., Basak, S., and Mills, D. [2003]. "Assessing Model Fit by Cross-Validation", *Journal of Chemical Information and Computer Sciences*, 43(2)

Kao, D.L., and Shumaker, R.D. [1999]. "Equity Style Timing", *Financial Analysts Journal*, 55, 37-48

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*

Levis, M., and Liodakis, M. [1999]. “The Profitability of Style Rotation Strategies in the United Kingdom”, *Journal of Portfolio Management*, Fall 1999, Vol 26, No. 1, 73-86

Limthanakom, N., and Collver, C.D. [2010]. “Style Momentum”, SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1351006

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

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

McLean, R.D., and Pontiff, J. [2012]. “Does Academic Research Destroy Stock Return Predictability?”, Working Paper, https://www.rhsmith.umd.edu/files/Documents/Departments/Finance/DavidMclean.pdf

Molinaro, A. [2005]. “Prediction Error Estimation: A Comparison of Resampling Methods”, *Bioinformatics*, 21(15)

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

Rodriguez, M. [2011]. “The Failure of Predictive Modeling and Why We Follow the Herd”, *Technical Report*, Concepcion, Martinez & Bellido.

Sorensen, E.H., and Lazzara, C.J. [1995]. “Equity Style Management: The Case of Growth and Value”, in *Equity Style Management: Evaluating and Selecting Investment Styles* (Klein, R.A, and Lederman, J. eds), Chicago: Irwin

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

Zaremba, A., and Szyszka, A. [2015]. “Is There Momentum in Equity Anomalies? Evidence from an Emerging Market”, SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2778045

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.