Deutsche Bank
Research

Global

Quantitative Strategy
Quantcraft

Date
23 April 2020

The Return of the Machines

Gianpaolo Tomasi
gianpaolo.tomasi@db.com
+44(20)75453460

This is the twenty fifth edition of our Quantcraft series. This periodical outlines new trading and analytical models across different asset classes.

This *Quantcraft* report brings our N-LASR stock selection model back into the spotlight. Five years since the last report, we assess the performance of this Adaboost-inspired alpha aggregation algorithm using a series of new tests, including:

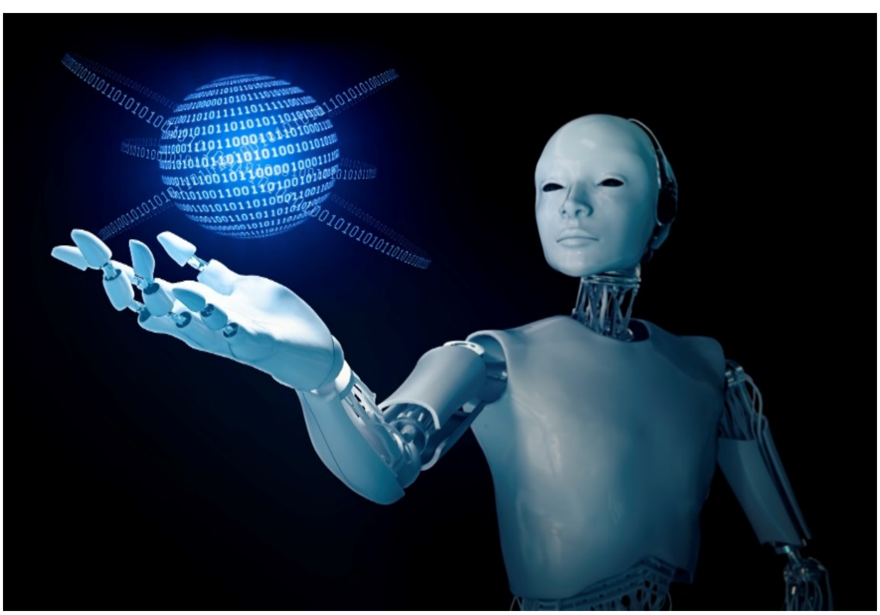
- Performance against other machine learning aggregation methods such as gradient boosting and neural networks;
- Performance decay as a function of execution delay and cost slippage;
- Performance as a function of the number of tradeable assets;
- Performance versus traditional cash equities factor portfolios and industry benchmarks, both prior to and after the last report;

While performance has decayed slightly over the past 5 years, it remains in line with backtest-based projections and attractive whether implemented as a long-short strategy or a long-only portfolio.

The machines are back, and they are here to stay.

Research Team
Vivek Anand
Andreas Farmakas
Jose Gonzalez
Rong Leng
Caio Natividade
Gianpaolo Tomasi
Ganchi Zhang
Blaz Zlicar

Figure 1: Return of the Machines



Source: Deutsche Bank.

Deutsche Bank AG/London

Distributed on: 23/04/2020 16:06:08 GMT

Return of the Machines

1. Introduction

Machine learning models are like chefs. They combine ingredients – the "alphas" – according to a pre-defined set of rules – the "recipe" – and, ideally, make the process transparent enough to avoid being labelled as a "black box".

Machine learning algorithms can be particularly useful when the investor seeks to identify robust relationships in large, and often noisy, datasets. They can be particularly effective in uncovering complex patterns such as non-linear relationships that escape the eyes of standard linear models.[1]

DB's N-LASR (Non-Linear Adaptive Style Rotation Model) is a good case in point. Described in Wang et al. (2012, 2013 and 2014), it is based on *adaptive boosting* to forecast the cross-section of stock returns using a selection of stock features – our "alphas" [2]– as inputs to combine.

For this report, we focused on the 80% most liquid stocks in the MSCI World universe - roughly 1,200 stocks - using 114 features as input and as such there is enough breadth for the model to uncover patterns that will persist out-of-sample. And, by combining such learning firepower with methods that reduce unwanted systematic exposure,[3] we find an attractive proposition to the systematic equity investor.

But, as the reader would rightfully point out, we must address the typical criticisms that supervised learning methods receive when applied to asset predictions. Are there over-fitting issues? Is turnover too high? Is the model simply capturing illiquidity premium? Is it unnecessarily opaque, and is complexity really needed?

This *Quantcraft* is different from our previous coverage of the N-LASR model in the following ways:

- Armed with over 5 years of data since the last report, we can accurately assess performance versus prior expectations. Further, we can also assess performance using different features and universes – all with the same aggregation algorithm.
- Given the significant growth in related literature, we can test new aggregation algorithms against our previous framework.
- We evaluate the efficacy of new ideas that were not considered in our previous reports, such as formulating algorithms in both classification and regression forms, and applying monotonicity constraints – in other words, the model cannot go short a given "alpha". The most promising modifications were made a year ago, and we also show the paper trading performance since.

Section 2 goes through the model construction process, first introducing the stock-specific features to learn from, then describing all learning methods and our test environment. Section 3 shows our general results, while Section 4 goes "under the hood" of N-LASR and Section 5 compares it with traditional equity market neutral and long-only investing whereas Section 7 assesses in-sample versus out-of-sample performance. The Appendix A provides step-by-step details of how the algorithm is built.

[1] The prolific literature on machine learning in finance is encouraging. Applications of neural networks in finance date as early as the late 1990s, when Sill (1999) applied them to corporate bond ratings. More recently, Takeuchi and Lee (2013), Batres-Estrada (2015) and Heaton, Poson and Witte (2016) used tree-based models to predict portfolio returns. Miller et al. (2013, 2015) presented evidence that combining linear and non-linear models can be even more effective than using non-linear models in isolation. Gu, Kelly and Xiu (2018) examine how different learning methods perform in forecasting the cross-section of expected returns and the equity premium, and find that non-linear methods significantly outperform OLS regressions. Rasekhschaffe and Jones (2019) reach similar conclusions using classification instead of regression, and using random forests, neural networks and gradient boosting as non-linear methods. Obviously supervised learning methods are also criticized, especially in the blogosphere. See reddit.com, for instance, for recent reviews.

[2] In this report we used the words "alphas", "features" and "factors" as synonymous.

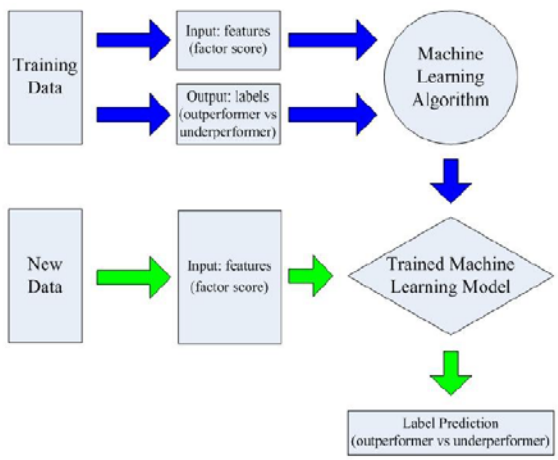
[3] Examples of unwanted exposures include market, volatility industry or sector, and country or region.

2. Methodology

Supervised learning is about collecting features that are expected to be predictive and allocating importance to each based on their ex-ante predictive power.

It starts with how the model is defined – in other words, how it learns. We then feed it with stock-specific features to learn from, as part of the training process. Once the model – the machine – has been calibrated, it is used to make predictions about the future based on new values of those features. The process is illustrated in Figure 2.

Figure 2: Supervised learning for classification summarized



Source: Deutsche Bank

2.1 Our features and methods

As pointed earlier, *features* – or *alphas* – are our input data and hence, what the algorithm is calibrated on. The better their quality, the better our learning and the higher the accuracy of our predictions as the algorithm goes live.

Needless to say, selecting what features to use is arguably the most important step and therefore what this *Quantcraft* starts with.

Equities are privileged in that they benefit from having substantially more asset-specific features than other asset classes. This allows us to feed the "learner" a large number of informative features, and let it identify value accordingly. It is no wonder, therefore, that most of the (promising) research on supervised learning focuses on cash equities.[4]

Figure 3 provides an idea of the features used – a total of 114 computed from Factset data.[5] We divide them across six categories: Technical, Growth, Profitability, Efficiency, Balance Sheet Strength and Value. [6] The alphas are updated on a weekly basis but other rebalance frequencies (daily, monthly) can obviously be used. Importantly, *all fundamental data (i.e. non-price related) is lagged by 3 months before computing the alphas*. This helps reduce the impact of data backfilling.

As for data processing, we applied the following steps:

1. Raw values are cross-sectionally ranked in order to reduce outlier effects. In other words, we map the *raw feature* values of all stocks into percentile rank scores within [0,1].
2. The scores from Step 1 are then de-meaned on a weekly basis inside a sector-region [7] or inside a sector[8] for the non-technical factors only, in order to remove systematic unwanted sector and regional exposures. We used GICS level 1 sectors and 3 geographical regions – North America, Europe and Asia.[9] We do not apply this neutralization step to the technical factors (Momentum, Volatility, Beta and Market Cap [10]), in line with the literature, as the resulting sector-regional biases are often rewarded.
3. In the case of N-LASR, the transformed scores from Step 2 are then cross-sectionally ranked weekly. In the other methods to be described below, they are cross-sectionally z-scored (weekly) instead. These are our *final features*.

We next define our *target variable* – in other words, what we are trying to predict. As per Wang et al. (2014), in order to remove systematic exposures we focus on sector- and region-neutral [11] 4-week forward stock returns. We further divide these returns by the stock-specific 5-year rolling historical volatility[12] for a better cross-sectional comparison. When we frame the problem as regression, we keep the volatility-adjusted, sector-region neutral 4-week forward-looking returns as

[4] Natividade (2013) showed, for instance, much less promising results when testing different supervised learning methods to daily data in FX and other CTA markets.
[5] The literature often uses more features. Gu, Kelly and Xiu (2018), for instance, use 900 signals while Rasekhschaffe and Jones (2019) use 194 features.
[6] There are a few differences with respect to those used in the original model. This is due to infrastructure changes. Note also that Profitability, Efficiency and Balance Sheet Strength are commonly all referred to as 'Quality' alphas.
[7] In the case of MSCI World.
[8] .For regional universes, only sector-neutralisation is applied.
[9] This provided us with $11 \times 3 = 33$ sector-region couples in MSCI World.
[10] Note that our alpha is -Market Cap, the traditional Size factor of the literature (i.e. long small caps, short big caps).
[11] The sector-region neutralization applied to the returns is the same described earlier for the non-technical alphas.
[12] Historical volatility of weekly returns, 5-year lookback window, rolled at every rebalancing date.

they are. When we frame the problem as classification we further rank the above cross-sectionally. The top 30% ranked stocks are assigned a +1 label, and the bottom 30% are assigned a -1 label. The remaining stocks are discarded from the training set.

As per Wang et al. (2014) we used 4 different training models: *long-term*, *short-term*, *seasonal* and *hedge*. They only differ in regards to training window. The long-term and short-term models use 5 years and 1 year of rolling historical data, respectively. The seasonal model uses long-term history [13] for the same calendar month, thereby capturing alpha seasonality. The hedge model is computed, at any point in time, using the worst 50% of the weeks in the previous 10 years, ranked according to the P&L of the aggregate of the other 3 models. The final signal is the equally-weighted average of signals from the 4 models described above.

Figure 3: List of features by category

Profitability 32 features (eg. ROA, ROE)	**Balance Sheet Strength** 28 features (eg Percent accruals, CAPEX to assets)	**Efficiency** 21 features (eg. cash ratio, asset turnover)
Value 17 features (eg. EBIT to EV, dividend yield)	**Growth** 12 features (eg. asset growth, 1Y EPS growth)	**Technical** 4 features (eg. momentu, volatility)

Source: Deutsche Bank

Our third step is the choice of *learners* – in other words, the feature aggregation algorithms. Importantly, and as per Rasekhschaffe and Jones (2019), we include other sophisticated methods for proper comparison. These methods are:

- Random forests (RF): this ensemble method combines predictions from different classification and regression trees (CART), and opts for the most common output from the individual CARTs. Random forests use two steps to generate predictions: (1) compute N equal-sized bootstrapped variations of the training dataset [14] and (2) apply a randomly selected subset of features at each split of the N trees from Step (1). We tested the random-forest method framing the problem both as classification and regression.
- Adaptive boosting and N-LASR: in contrast with Random Forest, AdaBoost focuses on creating a *strong* predictor by combining several *weak* predictors in an iterative fashion. It updates the weights of the observations in the training set at every iteration by underweighting (overweighting) the datapoints where the predictions of the weak learner added at the current iteration are correct (incorrect). N-LASR, as described in the Appendix A, is inspired by Adaboost.
- Gradient boosting (XGB)[15]: this is a generalization of AdaBoost as it can be applied to a number of loss functions. While AdaBoost updates weights at every

[13] Rolling 10-year history.
[14] Using resampling with replacement.
[15] We implemented extreme gradient boosting using the XGB Python library which is much faster and more powerful than the sklearn implementation.

iteration by underweighting (overweighting) datapoints where the ensemble predictions up to that iteration are correct (incorrect), Gradient Boosting fits a weak learner to the *residuals* of the ensemble up to that iteration. It therefore targets the errors in the current ensemble explicitly. We tested gradient boosting framing the problem both as classification and regression.

- Neural networks (NN): while available since the early 1990s, NNs have caught recent attention given their success in computer vision, reinforcement learning and natural language processing problems. NNs are often presented as an interconnected group of units – the neurons – which are organized in layers. During training, the weights connecting consecutive layers are optimized to minimize a pre-defined loss function – typically residual sum of squares for regression and cross-entropy for classification. We tested neural networks framing the problem both as classification and regression.
- Linear regression (NNLS): a simple benchmark, and arguably the simplest supervised learning method to consider. We apply it using a standard non-negative least squares (NNLS) algorithm, thereby forcing non-negativity constraints on the coefficients. In other words, you can't "short" a feature. In this algorithm, the alphas are the explanatory variables of the multivariate regression and the target variable described earlier is the dependent variable.
- Equal weights (EW): the simplest benchmark – it assumes no learning as all alphas are equally-weighted. Predictive power and correlations are not accounted for.

All methods were applied to the aforementioned features and compared against one another. Note that although the alphas are updated weekly, each model is re-calibrated every 4 weeks in order to reduce turnover.[16] Finally, all resulting portfolios rebalance once a week, after all the alphas have been updated.[17]

2.2 Hyper-parameter fitting and portfolio construction

All methods except EW involve parameter training, where the model parameters are optimized as part of the algorithm's own learning process, without human interaction.

That said, the learning guidelines must be separately defined by the researcher. These guidelines typically take the form of another set of parameters – the *hyperparameters* – which are defined before the learning process begins.

We address hyperparameter tuning as follows:

- All relevant hyperparameters are trained over the 1996-2002 period, such that all posterior results (2003-2020) are out-of-sample. These are, respectively, our *validation* and *test* periods.
- For any given method, all 4 sub-portfolios (long-term, short-term, seasonal and hedge) use the same hyperparameters.
- All 4 sub-portfolios used the same hyperparameter configuration.
- All N-LASR hyperparameter levels were kept as per original research reports (see the Appendix for details).
- Specific decisions were made for random forest[18], gradient boosting [19] and neural network [20] approaches. There were no hyperparameters in our linear regressions.
- All other hyperparameters were defined according to default values from sklearn, XGBoost and keras libraries in Python.

Finally, and as alluded to earlier, we also enforced monotonicity constraints in all of our learners. In other words, none of the methods is allowed to *go short* a given alpha. In the case of linear regressions and neural networks, this is implemented through a non-negativity constraint on the coefficients. In gradient boosting and random forests, monotonicity is an explicitly parameter defined for fitting the models. . In the case of N-LASR,

[16] In other words, the parameters (i.e. the betas of NNLS) are computed once every 4 weeks and kept fixed between consecutive model calibrations.

[17] In other words, all positions are kept the same until the next rebalancing date – in one week's time. Note however that all portfolios are marked-to-market daily, and hence performance is evaluated using daily returns.

[18] We kept the number of bootstrap datasets and fraction of features to use at each split to their default value: $N = 10$ and $= \sqrt{k}$, where k is the number of features. We assessed performance as a function of tree depth, testing depths of 2, 3 and 4. We ultimately selected a depth of 3 due to its performance in the validation period.

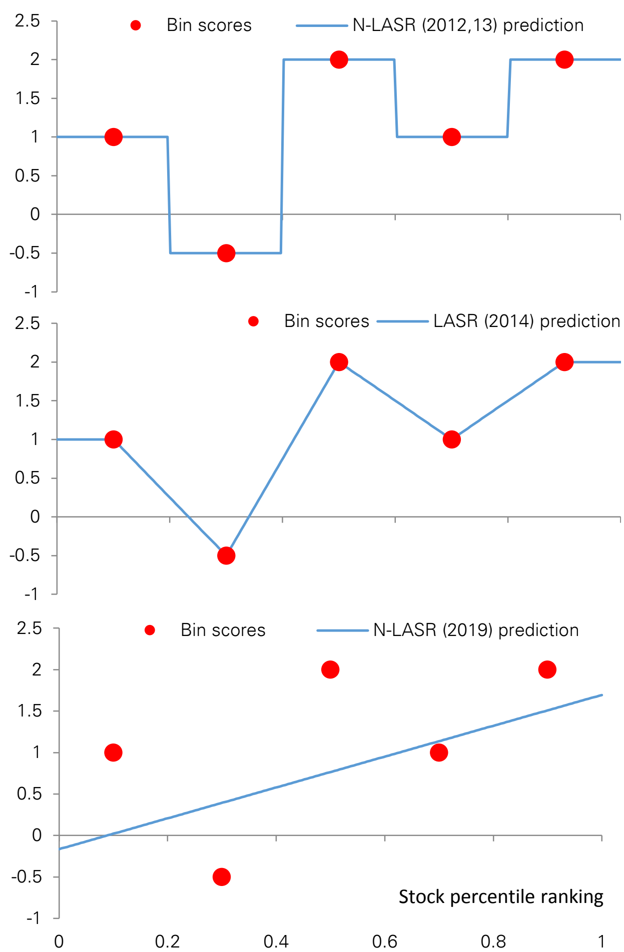
[19] In gradient boosting we focused on 3 hyperparameters:

- The *number of trees* – i.e. estimators – used to create the final prediction. We set that as 30 for consistency with N-LASR.
- The *depth of each tree*, set to 2 so as to enable us to model inter-factor interaction.
- The *learning rate*, which was optimised for during the validation period.

[20] During validation grid-searched the optimal number of hidden layers ($h = 1,2,3$), the number of units in the first layer ($u = 8,16,32$), the activation function (ReLu and hyperbolic tangent), the dropout rate ($d = [0,0.1,0.3,0.5]$) and the number of epochs used during training ($e = 10,20,100$). When $h > 1$ the number of units in the subsequent layer was set to half of the number of units of the previous layer in order to implement the so-called *bottleneck effect.* The parameters selected were $h = 1$, $u = 8$, $d = 0.3$, $e = 20$ and ReLu.

we modified the fitting kernel of the alpha selected on a given iteration during training from piecewise constant as per Wang et al. (2012, 2013), or piecewise linear as per Wang et al. (2014), to a forecast with non-negative slope. This is shown in Figure 4, where the x-axis is the percentile ranking of the alpha selected at the current iteration, the red dots are log-rations described in the Appendix A and the blue line is the forecast associated to that alpha.

Figure 4: N-LASR forecast kernel according to historical versions of the model



Source: Deutsche Bank

With that defined, we now move on to finalize our portfolio construction details for testing:

- We created long-short quintile portfolios[21], where individual positions are signal-weighted. The final positions used in the backtest are then computed as residuals of the regression of the signal on stock-level market betas in order to remove market exposure. Such a step is needed as our learning methods tended to overweight (underweight) low (high) volatility stocks which results in a substantial negative correlation with the market.[22]
- We assumed transaction costs of 5bps (bid-ask spread of 10bps) per dollar traded, and 50bps in annualized borrowing costs for the short positions. We separately assess sensitivity to transaction costs later in this report.
- Positions are executed with a *two-day* lag. In other words, the signal computed after the close of day t is traded market-on-close on day $t + 2$. We address sensitivity to trading delay later in this report.

3. Results

Figures 5-7 contain the main backtest results over the Jan'03 – Jan'20 period – RICs[23], cumulative wealth and performance metrics – all adjusted for transaction costs. As noted earlier, we focus on a *liquid* subset from the whole of MSCI World, so as to ensure that the algorithms are not harnessing illiquidity premia.

It is no wonder that we get encouraging numbers from almost all methods; after all, we are aggregating well established alphas, more than 1 thousand underlying assets and 4 distinct training windows all into one backtest.

But as we delve further into the results, a few other observations come to fore:

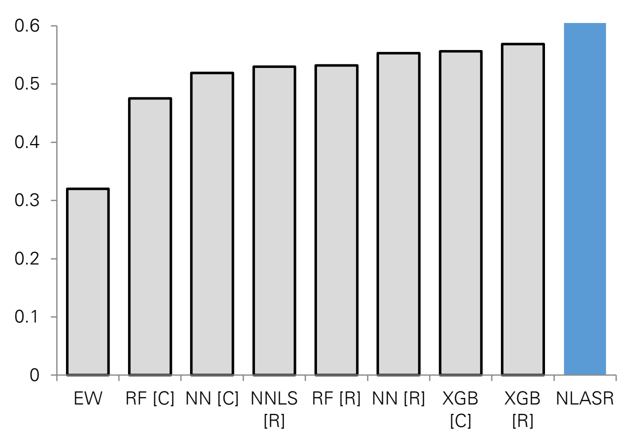
- N-LASR emerges as the outperformer in RIC and backtest metrics alike. It ranks atop in Sharpe ratio, CAGR and drawdown-by-volatility, and also ranks well in turnover terms (3.8% daily, or 19% weekly).
- The naïve, equal weights method significantly underperforms the other approaches. This indicates

[21] As in, long the top 20% ranked stocks and short the bottom 20% ranked stocks.
[22] Our findings are consistent with Rasekhschaffe and Jones (2019) and our prior N-LASR reports. For each method we run weekly cross-sectional regressions of the signal of the top and quintile stocks on the stocks' market betas. The regression residual for each stock is the new signal. Betas were computed using 3 years of weekly returns. Such adjustment led to a significant reduction of the correlation of returns with the market, now within the [-0.15,0.15] range for all methods tested.
[23] Rank Information Coefficient (RIC) is the Spearman Correlation between composite signal and future returns computed using the the entire cross-section.

that some learning can have a powerful impact in cash equity investing. All 114 features are expected to contain *some* predictive power, but weighting the alphas based on their historical performance clearly adds value. The naïve method is the only one that does not apply any dynamic feature discrimination[24] in its selection of weights.

- The least squares method performs well, which challenges the need for further complexity. While such debate extends well beyond the scope of this *Quantcraft*, in order to analyze the issue more in detail we further backtested N-LASR and NNLS in 8 different "regional" universes[25] and the results were consistent with those of MSCI World. In other words, although the two returns streams were comparable in all universes, N-LASR outperformed NNLS in terms of net Sharpe ratio by 5-20% depending on the universe.

Figure 5: Mean percent daily RIC across methods

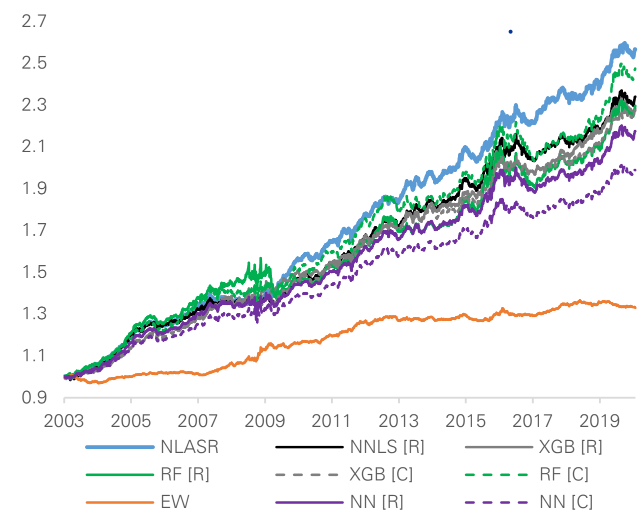


Jan 2003 – Jan 2020 period. Source: Deutsche Bank

At this point, the reader may ask whether the results above are enough to state that these methods are truly different from one another, and if NLASR is a true outperformer. Figure 6 shows that most methods correlate highly which, in fairness, could be expected as all methods use the same input data, training framework and backtest process.

Further, when assessing the statistical significance of the difference between backtests, only the naïve benchmark and neural network methods are, unanimously, statistically different at the 5% significance level.[26] Neural networks are capable of modelling more complex, highly non-linear interactions between alphas and returns, but note that this translated into higher turnover but not better performance.

Figure 6: Backtested wealth curves



Jan 2003 – Jan 2020 period. Transaction costs included. Source: Deutsche Bank

Figure 7: Backtest statistics

	Sharpe ratio	CAGR	Ann vol (%)	Max DD (%)	Daily turnover (%)	Max DD / vol
NLASR	1.64	5.48	3.34	5.95	3.87	1.78
XGB [C]	1.47	4.76	3.23	6.73	5.72	2.08
NNLS [R]	1.38	4.94	3.57	7.33	3.36	2.05
NN [R]	1.37	4.5	3.3	6.46	7.25	1.96
XGB [R]	1.35	4.79	3.55	7.5	6.12	2.11
NN [C]	1.33	3.97	2.98	5.81	6.41	1.95
RF [C]	1.32	5.26	4	8.62	4.28	2.16
RF [R]	1.12	4.83	4.32	12.46	4.55	2.88
EW	0.94	1.62	1.73	3.94	1.11	2.28

Jan 2003 – Jan2020 period. Source: Deutsche Bank

[24] One could debate whether feature discrimination is synonymous to feature momentum, which in itself is a subset of factor momentum in that multiple features may belong to the same investment factor. Stating that supervised learning methods add value solely by capturing factor momentum would be an over-simplification, as we highlight in Section 5. That said, for a recent literature review of advocates of factor momentum in equities, see Gupta and Kelly (2019). For elements of the counter-argument, see DeMiguel, Garlappi and Uppal (2009).

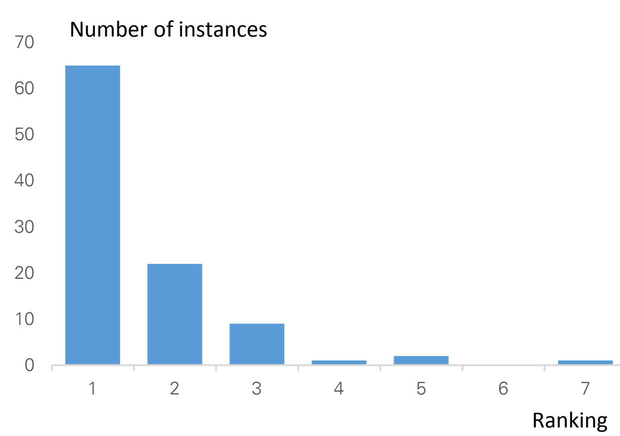
[25] US, Canada, EMEA, Europe-ex-UK, UK, LatAM, Japan and Asia- ex-Japan. These were computed as the intersection between the S&P Broad Market Index BMI and the appropriate countries.

[26] We ran a two-sample paired t-test on every couple of the 9 return streams available, hence a total of 36 tests. The null hypothesis is that, for every pair of return streams, their mean returns are the same. We could not reject this hypothesis at the 5% significance level in tests involving pairwise combinations of XGB, RF, NNLS and N-LASR. The hypothesis was rejected in 4 tests involving NN regressions and 7 involving NN classification. It was unanimously rejected in tests involving the naïve EW benchmark, as would be expected given its much lower relative returns. The test is inspired by Carver (2014).

We also noted earlier that N-LASR outperformed the NNLS method across a wide range of investable universes, but we need a more intense assessment that transcends regions and which includes other methods as well.

With that in mind, we ran an additional exercise involving 100 simulated backtests. We randomly selected 50 alphas from the 114 available and re-ran the same backtest – same data processing, training windows, orthogonalisation and universe of assets – for all methods except neural networks.[27] Figure 8 displays the results of that exercise in more detail, focusing exclusively on where N-LASR ranked versus the rest in each of our iterations. It ranked atop of all methods in 65% of the instances, and top 3 in 96% of the instances, in terms of simulated Sharpe ratio. While not a completely exhaustive exercise, this already suggests the outperformance of N-LASR is unlikely to be a statistical fluke.

Figure 8: Ranking of N-LASR relative to other algorithms in simulated exercise



Source: Deutsche Bank

And as we seek to make sense of the outperformance, an important comment has to be made about hyperparameter tuning. Gradient boosting, random forests and (especially) neural networks are characterized by more hyper-parameters than N-LASR, and their calibration may be challenging. The robustness of performance with respect to hyperparameters helps explain, in our view, why N-LASR produces more attractive results than its peers but it also makes us cognizant that results could change when the dataset is substantially larger (e.g. high-frequency data) as this would make the calibration process more robust.

4. Under the hood of N-LASR

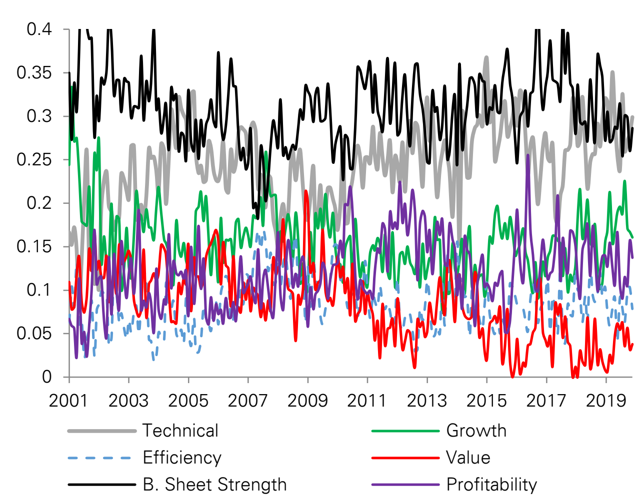
Having reaffirmed our preference for N-LASR, we now zoom into the backtest results so as to better understand the algorithm. We focus on how the weights change over time, how sensitive the results are to breadth, slippage and multiple investment universes, and on factor symmetry – how the long and short legs perform. As before, we focus on Jan'03 – Jan'20 data.

4.1 Historical weight allocation and model performance

As stated earlier, N-LASR uses elements of performance momentum in how it weighs across different alphas. Outperformers in each of the 4 historical training windows are given higher weight going forward, at the detriment of underperformers. As relative performance changes over time, so do the relative weights.

Figure 10 illustrates that argument, by showing the historical weight progression allocated by the aggregate of 4 models to each alpha category. While all categories are represented, the reader will notice a gradual reduction in our allocation to Value to the benefit of Technical features. Other features, most notably linked to Quality and Growth, had their allocation stable through time. Equity Value has been a notable underperformer in recent years, as will be shown in Section 5.

Figure 9: Historical weight distribution of the aggregate model



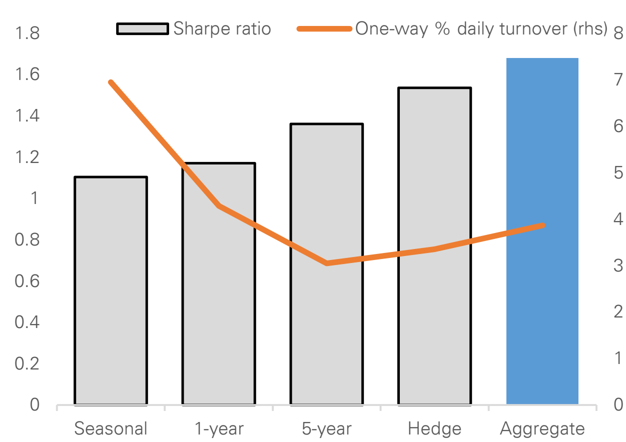
Jan 2003 – Jan 2020 period. Source: Deutsche Bank

[27] Neural networks were excluded due to the higher computational time necessary to estimate their parameters.

Figures 11–13 expand on the above findings to show how each of the four learning models relate to one another and contribute to the total over time across multiple universes. Notably:

- All models correlate highly with one another, suggesting that the diversification effect is moderate.
- The Hedge and 5-year models show the lowest turnover, as expected, as they are trained on longer lookback windows (5 years for both). They also provide the highest risk-adjusted backtest returns, suggesting that models trained on larger datasets perform better.
- The 1-year model, and particularly the seasonal model, are less correlated to the rest. But when it comes to the MSCI World universe, this diversification did not translate into higher returns – only higher turnover instead.
- That said, the power of diversification is made clearer when the 4 models are applied across different regions. As mentioned in Section 3, we further backtested N-LASR in 8 different 'regional' universes using the same alphas and the same implementation details, but higher cost assumptions.[28] Although the performance rankings change according to universe used, the aggregate model ranks atop in all but 3 cases.

Figure 10: Backtest Sharpe ratio and daily turnover across the 4 training models and aggregate (MSCI World universe)



Jan 2003 – Jan 2020 period. Source: Deutsche Bank

Figure 11: Correlation of daily returns across the 4 training models (MSCI World universe)

	Seasonal	1-year	5-year	Hedge	Agg.
Seasonal	1	0.44	0.51	0.50	0.71
1-year	0.44	1	0.67	0.58	0.79
5-year	0.51	0.67	1	0.89	0.92
Hedge	0.50	0.58	0.89	1	0.89
Aggregate	0.71	0.79	0.92	0.89	1

Jan 2003 – Jan 2020 period. Source: Deutsche Bank

Figure 12: Sharpe ratio rankings by training model (multiple regions) and aggregate backtest Sharpe ratio

	Seas.	1-year	5-year	Hedge	Agg.	Sharpe
MSCI World	5	4	3	2	1	1.68
US	3	4	5	2	1	1.45
Canada	2	5	4	1	3	2.00
Japan	4	2	5	3	1	1.59
UK	3	4	5	2	1	2.06
Asia ex-Jp	3	4	5	1	2	2.78
Europe ex-UK	3	4	5	2	1	2.97
EMEA	3	4	5	2	1	3.35
LatAm	5	3	4	1	2	1.18

Sharpe ratios reported net of costs. Jan 2003 – Jan 2020 period. Source: Deutsche Bank

4.2 Sensitivity to costs and trading delay

Aggregation algorithms that use supervised learning are often criticized for three potential shortcomings: high (and unnecessary) turnover, fast signal decay and the need for breadth. We now address each of these.

Figure 14 shows the sensitivity of our backtest Sharpe ratio to transaction cost assumptions. We start with the original guidance - 5bps per dollar traded in bid-mid or mid-ask, and 50bps p.a. in borrowing costs - and worsen both gradually. Figure 15 repeats the exercise but delaying execution from $t+2$, our base case, all the way to $t+20$ business days. Note that the deterioration is linear, as we assume no market impact in this exercise.

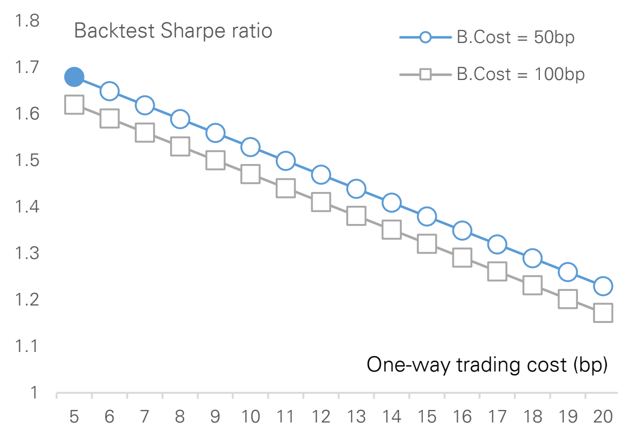
Both figures indicate tame sensitivity to slippage, thereby contesting the standard criticism. In both worse case scenarios, the backtest Sharpe ratio still exceeds 1.0.

In our view, the reader should be unsurprised. As shown earlier, one-way weekly portfolio turnover is moderate (~20%) as most of our features are slow-moving in

[28] One-way trading cost of 10bp and annualized borrowing costs of 100bp.

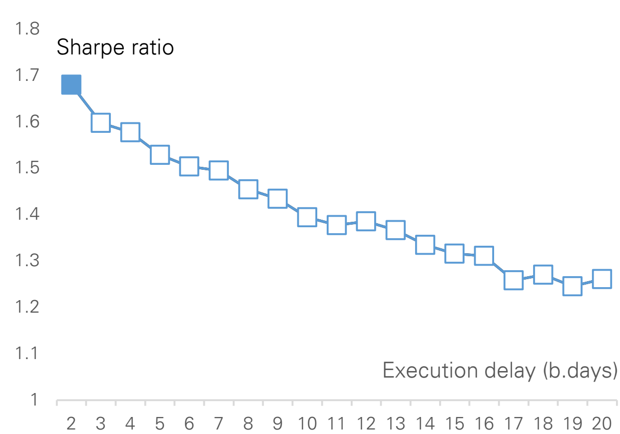
nature and therefore sensitivity to transaction costs and execution delay is minimal.

Figure 13: Sharpe ratio decay according to transaction cost assumptions (execution: t+2, MSCI World universe)



Jan 2003 – Jan 2020 period. Source: Deutsche Bank

Figure 14: Sharpe ratio decay according to execution delay (baseline costs, MSCI World universe)



Jan 2003 – Jan 2020 period. Source: Deutsche Bank

4.3 Sensitivity to breadth

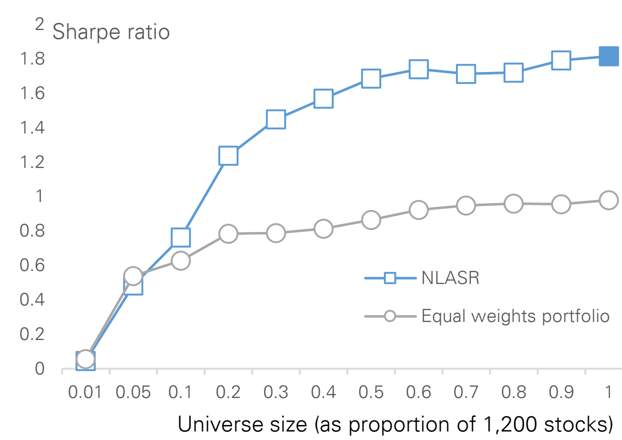
Testing for breadth dependency (i.e. performance as a function of the number of traded instruments) is a more involved exercise, but equally important.

It is broadly accepted that machine-learning methods require large datasets in order to identify reliable patterns that will persist out-of-sample. To assess this issue in the context of our study, we compared N-LASR with our naive (equal-weight) benchmark as a function of the number of traded stocks.

We created smaller sub-universes by randomly sampling a fraction f of stocks in the MSCI World liquid universe where $f \in [0.01, 0.05, 0.1, 0.2, 0.3, \ldots, 1]$, where 1 corresponds to all ~1,200 stocks. We sampled randomly on a semi-annual basis, and repeated the backtest 10 times for each value of f in order to reduce the variance of our results. We then applied the N-LASR algorithm to the new subset of stocks, and compared it with the equal weights used as benchmark.

Figure 16 displays our results, gross of costs. It plots simulated Sharpe ratios for both N-LASR and the naive benchmark according to universe size, as a fraction of the original universe. As per hypothesis, for smaller sample sets the two methods produce very similar results. When dealing with 10% of the original set – 120 assets, the typical number of markets in a traditional CTA portfolio - the difference in results is insignificant. It is only when our pool reaches 20% of the universe (i.e. ~240 stocks) that the difference between approaches becomes notable.

Figure 15: Simulated Sharpe ratios according to universe size



Source: Deutsche Bank

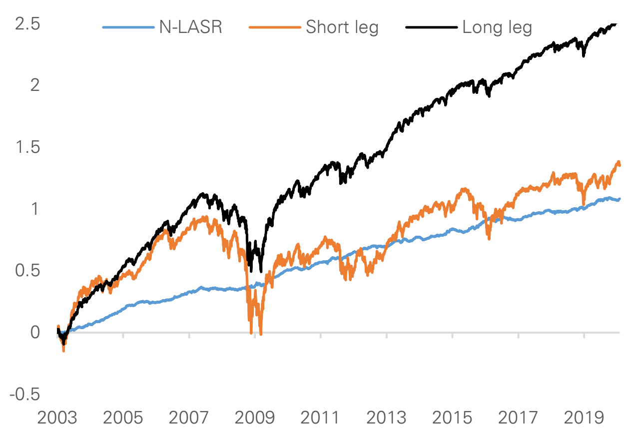
It is no wonder, therefore, that complex supervised learning methods are more popular in investment domains ripe with *breadth* – cash equities, with its high number of assets, being a good example.

4.4 Factor symmetry and focus on the long leg

All our analysis has thus far focused on the net returns of our long-short portfolio. But if we are to look under the hood of N-LASR, it is also useful to understand how the long and short legs differ over time.

Figure 17 shows the arithmetic[29] cumulative returns of both legs, where the position within each leg were computed after the beta neutralization described earlier. As is common with equity factor portfolios the short leg provides a return drag – in other words, overall, the short leg has a positive return. As expected, however, N-LASR's short positions come to fruition during market downturns, as can be seen by the drawdowns in 2008, 2011, 2015 and 2018 – far exceeding those of the long leg. As such, the difference in returns between the two legs has been relatively stable over time. Further, in the backtest the long leg has outperformed the market while the short leg underperformed the market.[30]

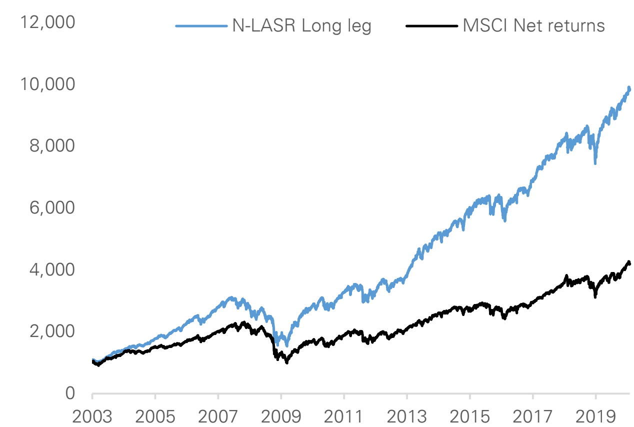
Figure 16: Cumulative returns – N-LASR long and short legs, and their difference



Source: Deutsche Bank

Another natural question is how profitable N-LASR may be as a long-only strategy, especially if compared to the MSCI World Net Return, used as benchmark. Figure 15 addresses this question, by plotting the net wealth of the long leg of N-LASR, corresponding to the top 20% of the stocks in our liquid MSCI World universe[31] versus the benchmark. Backtested performance, after cost[32], looks promising.

Figure 17: N-LASR long-only portfolio versus MSCI World Net Returns Index



Source: Deutsche Bank, Bloomberg.Finance LP

5. Comparison with traditional equity factors

Despite the results shown so far the skeptical reader may still ask: is all this complexity really justified? How does performance compare with a more classical factor investing approach, where the researcher chooses a much smaller set of features for each factor category, allocates the same importance to each feature, and later gives the same amount of risk capital to each category?

This section compares N-LASR with our in-house implementation of traditional equity factor strategies (Momentum, Quality, Value and Low Beta) to assess the value added – if any – by the N-LASR algorithm when compared with more traditional long-short investing. For a fair comparison, we modified the cost assumptions in N-LASR and created one composite PnL stream from the four factor strategies above using inverse volatility weights.[33] The latter adjustment led to an overweight of Quality – the least volatile strategy –, with a mean weight of 38%, and an underweight of Momentum, with mean weight of 14%, while Value and Low Beta had their average weights in between. All strategies use the liquid MSCI World universe.

[29] We chose arithmetic instead of geometric in order to avoid the impact of compounding. This allows us to better evaluate the difference between legs. Please note that in all the other results of this report the compounding effect was always included.

[30] The respective Sharpe ratios for long and short legs, and the market are 1.06, 0.41 and 0.53. Note that these are calculated using arithmetic returns and hence not comparable to the N-LASR Sharpe ratio reported earlier.

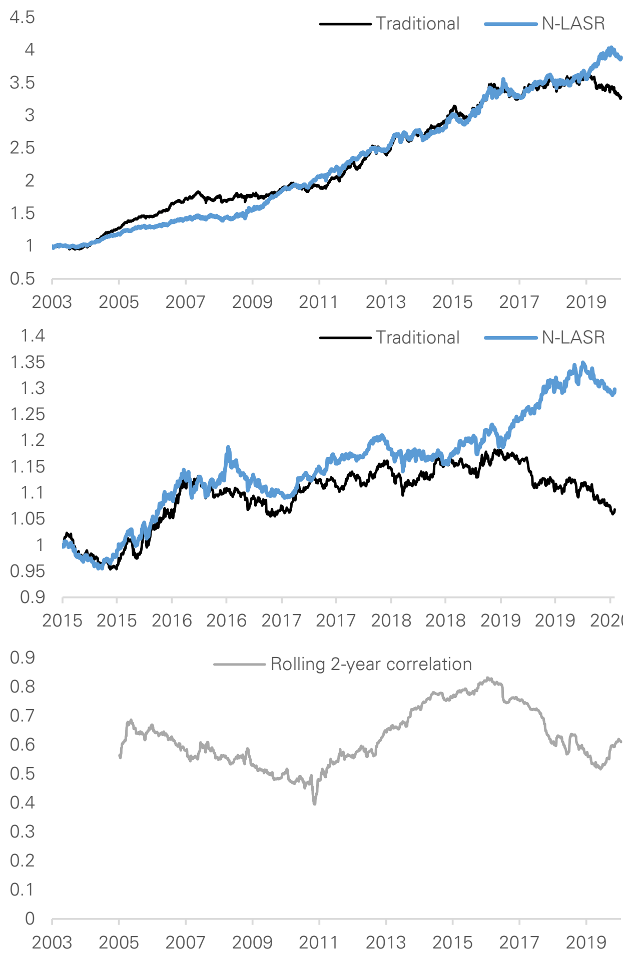
[31] As the orthogonalisation was carried out only to ensure market neutrality for the long-short portfolio, for the long-only portfolio it makes more sense to use the raw composite N-LASR signal (i.e no orthogonalisation) as positions.

[32] We applied taxes on dividends in addition to the 5bps of transaction costs for a fair comparison with the benchmark. As we assumed this to be a funded portfolio, no funding cost was included.

[33] Volatility estimated using daily returns and a rolling 5-year historical window.

Figure 19 displays our backtest results, using the Jan 2003 – Jan 2020 period. Full sample Sharpe ratios are comparable – 1.28 for the factor composite and 1.47 for N-LASR[34] – but the key difference comes in recent years, when N-LASR outperformed.

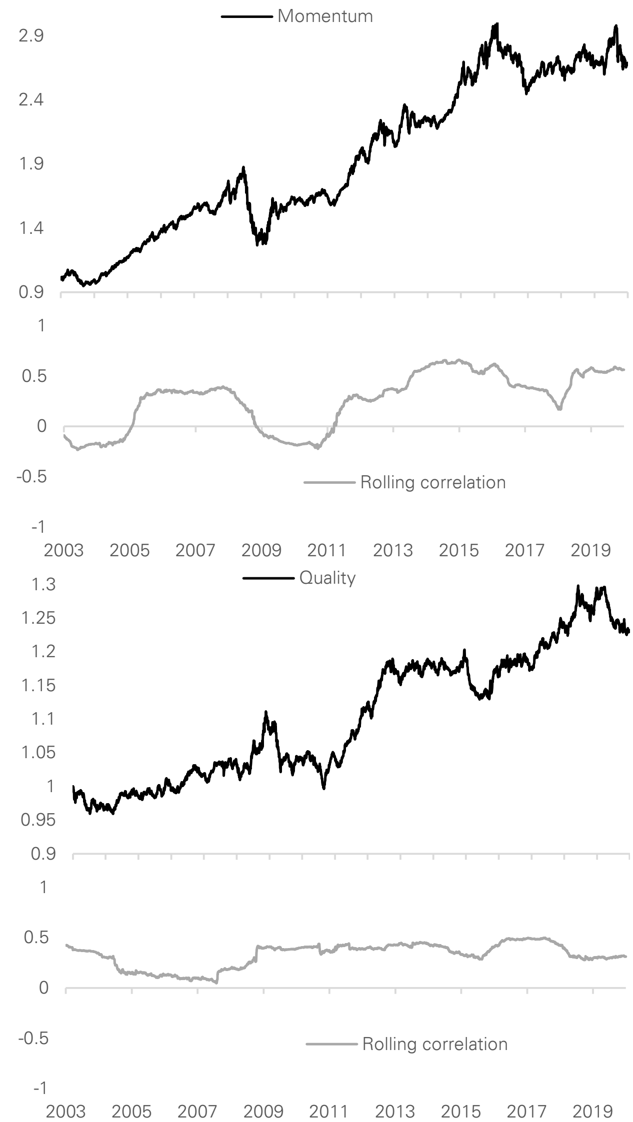
Figure 18: Backtested returns – N-LASR and traditional equity factors



Source: Deutsche Bank

As we dig further into the differences, Figures 20 and 21 show the 2-year rolling return correlations between N-LASR and each of our factor strategies (blue), together with the time series of net wealth of each of them (black).

Figure 19: Traditional factor strategies: Momentum and Quality



Source: Deutsche Bank

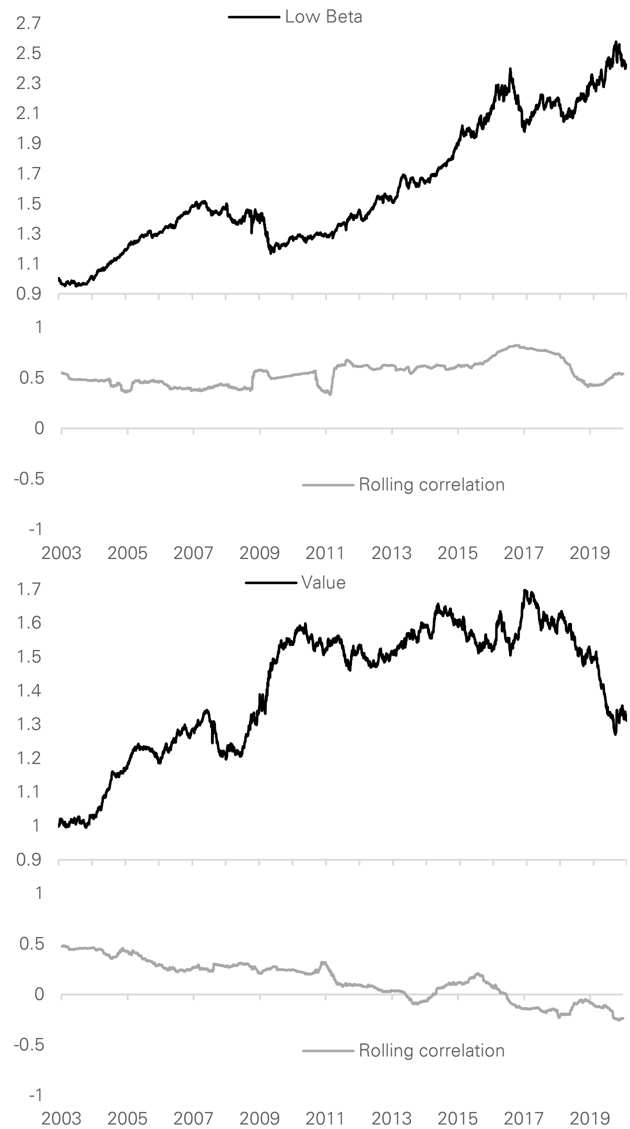
The reader may have noticed an emerging picture: the relationship between N-LASR and the traditional factors is dictated by how the factors are performing. The better a factor has performed, the more that N-LASR loads into the class of features that represents that factor, and vice-versa when a factor starts underperforming.

This explains the time-varying nature of individual factor correlations, most notably with Value. Exposure to metrics such as EBIT/EV, EBITDA/EV and FCV/EV were rewarded right after the "dotcom burst", but less and less so in later years to the point that N-LASR is now

[34] As aforementioned, we changed the cost assumptions used for N-LASR in order to make it exactly comparable with the equity multifactor portfolio and that explains why the SR of 1.47 does not match the full sample SR of 1.64 reported in Figure 7.

negatively related to Value – just as it loads positively to Growth alphas, as per Figure 10.

Figure 20: Traditional factor strategies: Low Beta and Value



Source: Deutsche Bank

Such a dynamic change in exposures is a key feature of the algorithm. It has a natural tendency to load onto alphas which worked well on a specific universe. When tested in the US or Europe universes, for instance, N-LASR loaded more heavily to Momentum; when tested in Japan, the exposure to Momentum was negligible. Needless to say, this relates to how well this investment factor has performed in the respective regions over the years.

By now the reader may be questioning whether N-LASR is just a glorified factor momentum algorithm. We caution against such simplification. Factor momentum algorithms are univariate[35], focus on short formation windows (typically in order of months rather than years) and do not account for other value-add such as diversification – a key feature brought by adaptive boosting.

This takes us to the crux of N-LASR which, to an extent, can be generalized to supervised learning methods in general. *They capture the most important stylized facts in the dataset they're given to learn from*. Performance momentum may be the lesson for cash equities in the past 20 years, just as trend following was the lesson learned when we applied a similar exercise to multi-asset futures contracts.[36]

It also helps us predict when N-LASR is likely to underperform. Sudden crashes in previously outperforming styles will hurt N-LASR; but if these changes become more structural, such that they change the stylized facts of the market, they will also be picked up by the algorithm, which is able to automatically detect them and act accordingly.

6. Performance after December 2014

As noted at the start, it has been over 5 years since our last N-LASR report. It would be appropriate, therefore, to evaluate performance since then. The timing of this *Quantcraft* is also opportune in that markets have just witnessed the so-called "quant bust" of 2020[37], when market neutral strategies underperformed across asset classes.

To be clear, we are not reporting N-LASR "paper trading" performance in this section. There are small differences in the universe of assets, alphas used and source database compared to the last report. More importantly, early last year we implemented non-negativity constraints on the alphas.[38]

Figure 22 reports our performance since the last research report and since the imposition of non-

[35] In other words, they do not take correlations into account. See see Gupta and Kelly (2019) and Anand and Zhang (2020) for examples in cash equities and cross-asset strategies.

[36] See Natividade (2013). Note that the author used 30 years of data at the time, and trend following was quite successful in the 1980s, 1990s and 2000s.

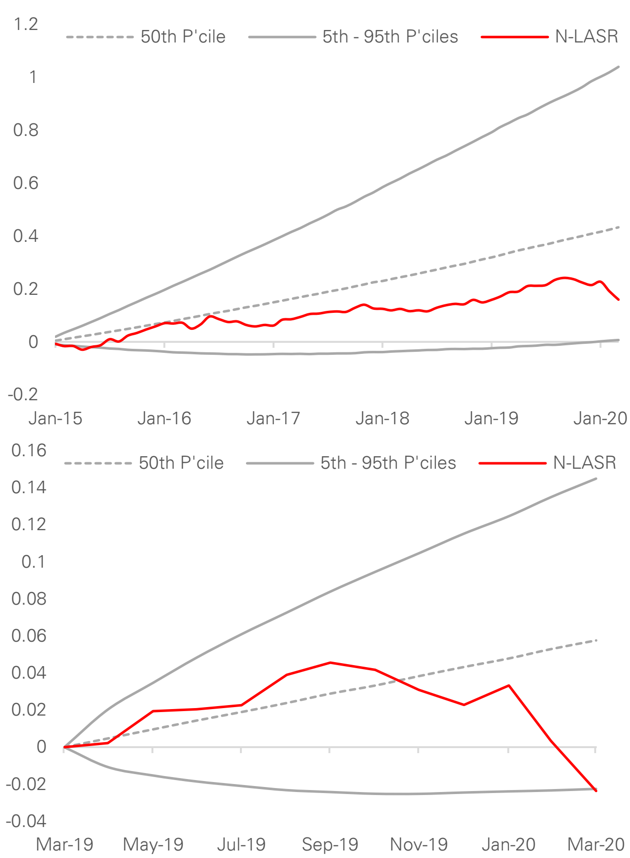
[37] A term coined by Kakushadze (2020).

[38] As per Section 2.2, the algorithm can no longer *go short* any feature.The Appendix describes it in detail. Furthermore, we focus in this paper on a liquid global universe whereas the analysis focussed on regional universes in our previous reports.

negativity constraints (April 2019), up to the end of March 2020.

They are plotted together with expectation bands built using Probability of Divergence (PoD), a concept first introduced in Lopez de Prado (2013) but modified in Anand and Zhang (2020). In summary, we fit an auto-regressive model AR(p)[39] to the N-LASR monthly returns pre-2015 (top chart) and pre- April 2019 (bottom chart), and a Gaussian Mixture Model (GMM) to fit the residuals of the AR calibration. We then use the AR coefficients to generate the deterministic out-of-sample component and add residuals sampled from the GMM to generate 10,000 simulated paths of projected performance post 2015 and post April 2019. Of these, we record the cumulative 5th, 50th and 95th percentiles (black) and plot actual performance (red) against them.[40]

Figure 21: N-LASR performance post previous report and after imposing monotonicity constraints



Source: Deutsche Bank

That the actual performance lies closer to the lower boundary of our simulated projections may seem concerning at first. But as highlighted in Anand and Zhang (2020), only performance outside the "cone" should be investigated for possible cause, whose culprits may vary. In our view, N-LASR currently sits in the bottom half of the cone due to the recent drop in predictive power of traditional features, whose reasons could include increased crowdedness.

Further, its poor performance during the "quant bust" of 2020 is attributed to the sell-off in top factors of recent years – especially Low Beta to which N-LASR has systematic exposure in MSCI World. As highlighted earlier, N-LASR overweights factors that have performed well historically and fast corrections in those lead to model underperformance.[41]

For a more holistic assessment, we compare our results against those of two peer benchmarks for equity market neutral investing: the HFR and Eurekahedge indices.[42] Both are comparable to N-LASR in investment style (equity market neutral) and historical Sharpe ratios.

Figure 23 plots recent returns in all 3 indices as percentiles to their own projection cones, also built using the PoD framework described earlier. We note the following:

- In 2 out of 3 cases, post 2015 performance has generally been in the lower half of our PoD projections. Using the March 31 2020 as our end-point, all 3 are in the bottom half. This supports the prior argument on crowding and subsequent alpha decay. The post Jan'15 Sharpe ratios are 0.76, 0.46 and 0.78 for HFR, Eurekahedge and N-LASR,

[39] The order p of the AR model is selected using Akaike Information Criterion, assessing values of p up to 5

[40] The authors originally use weekly returns. We use monthly in order to make our results comparable with the market benchmarks, whose data is unavailable at a higher frequency. As a result, the N-LASR autoregression coefficents are stronger and hence the cone is wider. If the analysis is done using weekly returns, the results look worse as the cone is narrower. For details on PoD, see Anand and Zhang (2020).

[41] Note that most cash equity factors underperformed during this period.

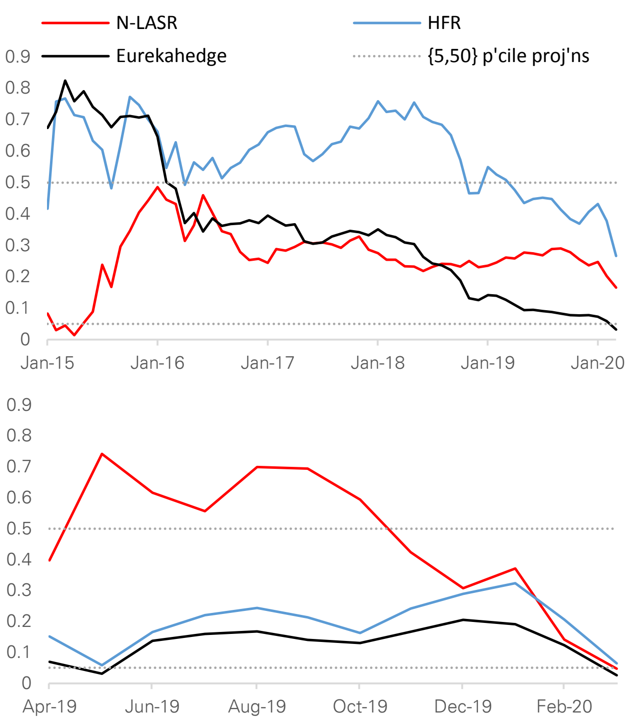
[42] On Bloomberg, these indices are EHFI751 Index (Eurekahedge) and HFRIEMNI Index (HFR).

respectively.[43] Note that the N-LASR numbers are net of transaction costs.

- All 3 indices were hit in the quant bust of 2020. This is no surprise, as most equity factors got hit. All 3 posted negative Sharpe ratios during the Apr'19 – Mar'20 period: -0.69 (HFR), -0.85 (Eurekahedge) and -0.34 (N-LASR).

As such, one can conclude that, while lower, the recent performance of N-LASR is not unusual versus its backtest and comparable benchmarks.

Figure 22: Actual Performance measured as percentiles versus PoD projections*



* Based on data before Jan'15 and Apr'19 for N-LASR and equity market neutral benchmarks
Source: Deutsche Bank

7. Conclusions

This *Quantcraft* report brings our N-LASR model back into the spotlight. Five years since the last report, this Adaboost-inspired alpha aggregation model continues to perform within expectations. The results are also attractive when compared to peer stock selection algorithms, when applied over different universes and when evaluated in funded long-only format.

Some topics are open for debate. The margin of outperformance versus a simpler least squares algorithm, or versus classical factor investing, are examples. It also requires a large universe to be effective, and incurs its worse drawdowns when top historical factors underperform.

That said, N-LASR continues to provide an encouraging source of returns for both return-seeking and diversification-oriented stock investors, whether long-short or long-only.

8. Bibliography

8.1 Quoted references

Anand, V. and Zhang, G. (2020), Keep Calm and Stay Market Neutral, Deutsche Bank Quantcraft, 24 March 2020.

Batres-Estrada, B., (2015) Deep learning for multivariate financial time series. Available at https://www.math.kth.se/matstat/seminarier/reports/M-exjobb15/150612a.pdf

Carver, R. (2014), *Systematic Trading: A unique new method for designing trading and investing systems*, Harriman House, 1st Edition.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009), Optimal Versus Naïve Diversification: How Inefficient is the 1/N Portfolio Strategy?, The Review of Financial Studies, v.22 n5, p.1915 – 1953.

Freund, Y. Schapire, R. (1996) Experiments with a new boosting algorithm. In Machine Learning: Proceedings of the Thirteenth International Conference, pages 148–156

Gu, S., Kelly,B., Xiu,D. (2018) Empirical Asset Pricing via Machine Learning, SSRN Electronic Journal.

[43] Calculated using the volatility of monthly returns. If using the volatility of daily returns, the Sharpe ratio for N-LASR is 0.76. We do not have daily data for the other 2 indices.

Gupta, T. and Kelly, B. (2018), Factor Momentum Everywhere, Working Paper. Available at SSRN: https://ssrn.com/abstract=3300728.

Heaton, J B, Polson, NG., and Witte, JH (2016) Deep Learning in Finance. Available at https://arxiv.org/abs/1602.06561

Kakushadze, Zura (2020), Quant Bust 2020, Working Paper, Available at SSRN: https://ssrn.com/abstract=3570280.

Lopez de Prado, M. (2015), Advances in Quantitative Meta-Strategies, Available at SSRN: https://ssrn.com/abstract=2604812.

Miller, KL. et al. (2013) Size Rotation in the U.S. Equity Market, The Journal of Portfolio Management 39, 116–127.

Miller, KL. et al. (2015) A Risk-Oriented Model for Factor Timing Decisions, The Journal of Portfolio Management 41, 46–58.

Natividade, C. (2013), Machine (Un)learning?, Deutsche Bank Quantcraft, 06 March 2013.

Rasekhschaffe, K. and Jones, R. (2019) Machine Learning for Stock Selection, Financial Analysts Journal vol. 75, no. 3.

Sill, J. (1997), Monotonic Networks. Advances in Neural Information Processing Systems 10 (NIPS 1997).

Takeuchi, L. and Lee, YYA. (2013) Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks, Technical Report, Stanford University.

Wang, S. et al., (2012) The Rise of the Machines, Deutsche Bank

Wang, S. et al, (2013) The rise of the Machines, II, Deutsche Bank Quantitative Strategy

Wang, S. et al, (2014) The rise of the Machines, III, Deutsche Bank Quantitative Strategy

8.2 Other useful references

Alberg, J., and Zachary, L., (2017) Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals, Available at https://arxiv.org/abs/1711.04837

Asness, C. et al. (2000) Predicting Stock Returns Using Industry- Relative Firm Characteristics, SSRN Electronic Journal.

Asness, C., (2016) The Siren Song of Factor Timing aka "Smart Beta Timing" aka 'Style Timing', The Journal of Portfolio Management 42, 1–6.

Bailey, D. and de Prado M. (2012) The Sharpe ratio Efficient Frontier, Journal of Risk, Vol. 15, No.2.

Carhart, M., (1997) On Persistence in Mutual Fund Performance, Journal of Finance 57–82.

Clemen, R., (1989), Combining forecasts: A review and annotated bibliography, International Journal of Forecasting 5, 559–583.

Daniel, K, and Moskowitz, T. (2016) Momentum crashes, Journal of Financial Economics 122, 221–247.

de Prado, ML (2018) *Advances in Financial Machine Learning* (John Wiley & Sons) 1st edition

Fama, E., and French, K., (1992) The cross - section of expected stock returns, The Journal of Finance 47, 427–465.

Fama, E., and French, K., (2017) International tests of a five-factor asset pricing model, Journal of Financial Economics 123, 441–463.

Friedman, J. (2001) Greedy Function Approximation: A Gradient Boosting Machine The Annals of Statistics Vol. 29, No. 5. 1189-1232

Moritz, B., et al. (2016) Tree-Based Conditional Portfolio Sorts: The Relation Between Past and Future Stock Returns. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740751

Nevmyvaka Y., Feng Y., Kearns M. (2006) Reinforcement Learning for Optimized Trade Execution https://www.cis.upenn.edu/~mkearns/papers/rlexec.pdf

Schapire, R., (1990) The Strength of Weak Learnability,. Machine Learning, volume 5, 197–227

Srivastava, N. and Hinton G. (2014), Dropout: A simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research, 15(56):1929–1958.

9. Appendix A

This Appendix provides a step-by-step description of the N-LASR algorithm. While these steps should be self-sufficient, we recommend the interested reader to also review Wang et al. (2012, 2013 and 2014) for a historical progression of the model.

N-LASR is inspired by Adaboost (i.e. adaptive boosting), an algorithm introduced in Freund (1996).

In summary, Adaboost creates a "strong" predictor by optimally comining the features (the "weak learners"). When used for classification, it starts by selecting the weak learner that best minimizes a given error rate, then underweights the datapoints[44] in the training set where the weak learner was correct and overweighs those where it was wrong.

By underweighting the datapoints where the current best weak learner is correct, in the next iteration (where we search for the next best learner) the algorithm encourages the selection of a new weak learner that works well on a different set of datapoints – thereby complementing the previously chosen learner.

As it selects the next best "weak" learner, the algorithm updates the new learner's weight relative to other learners and continues the iteration process until a convergence criterion is met.

In the following sections we describe in more detail how the algorithm works for a given calibration date.

9.1 Preparing the data

One of they parameters of the algorithm is the number of bins K. As in our implementation of N-LASR we employed $K = 5$, in this Appendix we will use the same value although this can be selected arbitrarily as the algorithm will work for any reasonable integer value of K (e.g. $K = 3, K = 10$).

The data is prepared according to the following steps:

1. Rank cross-sectionally all the values of a given alpha at time t using percentile ranking, so that the score for each stock (for that alpha) lies between 0 and 1. In other words, $s_{p_t}^{a} \subset [0,1]$ for every alpha $a \in \{FCF\ to\ Shrhld\ Equity, \dots, Market\ Cap\}$(Figure 2), for every stock p at time t.

2. Volatility-adjust each stock forward return by dividing its 4-week forward return by its 5-year backward (historical) volatility. Compute sector-regional neutral vol-adjusted returns by applying the sector-region neutralization described in Section 2.1. Then rank these pre-processed returns cross-sectionally from 0 to 1 and label them rank-adjusted returns.

3. Also as per Section 2.1, retain for training the stocks whose rank-adjusted returns from Step 2 are below 0.3 or above 0.7. Assign the label $l_{p,t} \in \{-1,1\}$, where the

   former is given to underperformers and the latter to outperformers. Assuming the universe is comprised by 1,200 stocks at a certain point in time, this step will give 360 readings of -1 and 360 readings of +1.

4. Set the weights of all datapoints in the training dataset to $1/N$, where N is the total number of datapoints in the selected lookback window. For instance for the 1-year lookback window, $N = 0.6 \times 1{,}200 \times 52 = 37{,}740$ Here we assumed that the full universe contains 1,200 stocks throughout the lookback period and used 52 as the number of weeks in one year. Hence $w_{i,j} = 1/N$ when iteration $i = 1$, for every datapoint j, t in the training set. These weights will be adjusted later in Step 10.

5. For each alpha a, and for each of $k \in \{1,2,3,4,5\}$ bins, compute the value for each stock inside each bin. For instance, if a certain stock at time t (which we denote as datapoint j) is ranked 0.15 according to ROE, i.e. $s_j^{ROE} = 0.15$, its bin values are $\boldsymbol{\psi}_j^{ROE} = [0.75, 0.25, 0, 0, 0]$.

   These numbers are computed using the inverse of the distance between the rank of the stock and the centers of the 5 bins ($[0.1, 0.3, 0.5, 0.7, 0.9]$), retaining only the 2 closest bins, and then normalizing the bin values so they sum to 1.

9.2 Training

Having formed the bins, the data is now ready for training. The training set is computed by stacking all the data from the lookback window "on top of one another". The size of the lookback window varies according to the 4 models defined in Section 2.1. As such, we remove the time notation t from here onwards.

[44] In our case, a datapoint is stock j at time t; in other words, a stock that was part of the training defined in Step 3 at a specific point in time in the lookback window. For clarity, a given stock may appear in the training window multiple times, nor not at all.

For each iteration $i \in \{1,2,\dots,I\}$, the algorithm is trained according to the following steps:

6. Find the alpha whose *weighted* correlation between ranked scores (Step 1) and rank-adjusted returns (Step 2) is the highest. The weights are identical for all datapoints at iteration 1 (see Step 4) but for later iterations they are different as they have been updated according to Step 10.

7. For each of the 5 bins define 2 additional bins, UP and DOWN: outperformers and underperformers, respectively. For each of the resulting 10 bins and for each datapoint, record the *weighted*[45] sum of the labels assigned in Step 3, thereby computing N_k^{UP} and N_k^{DOWN} for $k \in \{1,2,3,4,5\}$.

   For instance, recall that $s_j^{ROE} = 0.15$ and assume that ROE has been selected as best alpha in Step 6. Assume also that, despite the low score of 0.15, stock j ended up having a high subsequent return and was hence labelled as +1 in Step 3. In that case, set

   $\psi_j^{ROE,UP} = w_j \times \psi_j^{ROE}$ and

   $\psi_j^{ROE,DOWN} = [0,0,0,0,0]$.

   Compute then $\psi_k^{ROE,DOWN}$ and $\psi_k^{ROE,UP}$ for each bin $(k = 1,2,3,4,5)$ by summing the corresponding $\psi_j^{ROE,DOWN}$ and $\psi_j^{ROE,UP}$ over all the datapoints. Note that we end up with 10 numbers, one UP and one DOWN for each of the 5 bins.

8. Define the *bin score*, for each of the 5 bins, as the log of the ratio of weighted sums described in Step 7

   $\theta_k^{ROE} = \log(\psi_k^{ROE,UP} / \psi_k^{ROE,DOWN})$ [46] for each of the bins $(k = 1,2,3,4,5)$.

   We note that it is common, during the first few iterations, for the number of underperformers (the denominator in the equation above) to be high for bin 1 and low for bin 5. Similarly, the number of outperformers (the numerator above) tends to be high for bin 5 and low for bin 1. This is because the top alphas are chosen in the first few iterations.

9. Fit a linear function to the bin scores computed in Step 8. In other words, use the bin centers defined in Step 5 ($[0.1,0.3,0.5,0.7,0.9]$) as independent variable and the bin scores as dependent variable when estimating β and γ:

   $$\begin{bmatrix}\theta_1^{ROE}\\ \theta_2^{ROE}\\ \theta_3^{ROE}\\ \theta_4^{ROE}\\ \theta_5^{ROE}\end{bmatrix} = \begin{bmatrix}1 & 0.1\\ 1 & 0.3\\ 1 & 0.5\\ 1 & 0.7\\ 1 & 0.9\end{bmatrix}\begin{bmatrix}\gamma\\ \beta\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_4\\ \varepsilon_5\end{bmatrix}$$

   This step represents a key innovation from previous N-LASR iterations, as previously shown in Figure 4.[47]

10. If $\beta < 0$, exit the algorithm as at the current iteration we predict a non-monotonic relationship between the current alpha (ROE in this example) and future stock returns.

    If $\beta \geq 0$, compute the return forecasts for each stock. Using the previous example, given that datapoint j has ranking $s_j^{ROE} = 0.15$, the forecast for datapoint j is:

    $\hat{\emptyset}_j^{ROE} = \gamma + \beta \times 0.15$.[48]

11. Now we apply adaptive boosting. Update the weight for datapoint j for the following iteration (i+1) as an function of its return label, alpha prediction and current weight at iteration i:

    $w_{i+1,j} = w_{i\,j}\, e^{-l_j \times \hat{\emptyset}_j^{ROE}}$

    Weights are then normalized so as to sum to 1. We therefore inflate the weight of datapoints whose forecast at current iteration i is incorrect and deflate the weights where the forecast is correct.

12. Return to Step 6 and repeat the exercise until the maximum number of iterations is reached.

9.3 Prediction

For the out-of-sample model predictions we utilize only the alphas selected during training and their respective γ and β estimates. Assume we are at time t and all the current alphas have been computed and ranked cross-sectionally between 0 and 1, as usual. The prediction steps are:

I. For each alpha, compute the forecast for each stock as per Step 10 using the slope β and intercept γ computed in Step 9.

[45] The weights are initially defined in Step 4 and then updated in Step 10.

[46] One must carefully address the corner case when any of the 10 ψ_k is equal to 0, as that would potentially give rise to unwanted results

[47] It is also less sensitive to outlier effects, albeit not immune. We address that through corner case rules.

[48] Note that, as we are fitting the bin scores that can be negative (the logarithm of a ratio), the forecast can be negative.

II. Average all the predictions for each individual stock computed as per Step I. Note that this does not imply the alphas are equally weighted. The slope β computed in Step 9 acts as a weight on the individual alphas.

Appendix 1

Important Disclosures

Additional information available upon request

Prices are current as of the end of the previous trading session unless otherwise indicated and are sourced from local exchanges via Reuters, Bloomberg and other vendors . Other information is sourced from Deutsche Bank, subject companies, and other sources. For disclosures pertaining to recommendations or estimates made on securities other than the primary subject of this research, please see the most recently published company report or visit our global disclosure look-up page on our website at https://research.db.com/Research/Disclosures/CompanySearch. Aside from within this report, important conflict disclosures can also be found at https://research.db.com/Research/Topics/Equities?topicId=RB0002 under the "Disclosures Lookup" and "Legal" tabs. Investors are strongly encouraged to review this information before investing.

Analyst Certification

The views expressed in this report accurately reflect the personal views of the undersigned lead analyst(s). In addition, the undersigned lead analyst(s) has not and will not receive any compensation for providing a specific recommendation or view in this report. Gianpaolo Tomasi

Hypothetical Disclaimer

Backtested, hypothetical or simulated performance results have inherent limitations. Unlike an actual performance record based on trading actual client portfolios, simulated results are achieved by means of the retroactive application of a backtested model itself designed with the benefit of hindsight. Taking into account historical events the backtesting of performance also differs from actual account performance because an actual investment strategy may be adjusted any time, for any reason, including a response to material, economic or market factors. The backtested performance includes hypothetical results that do not reflect the reinvestment of dividends and other earnings or the deduction of advisory fees, brokerage or other commissions, and any other expenses that a client would have paid or actually paid. No representation is made that any trading strategy or account will or is likely to achieve profits or losses similar to those shown. Alternative modeling techniques or assumptions might produce significantly different results and prove to be more appropriate. Past hypothetical backtest results are neither an indicator nor guarantee of future returns. Actual results will vary, perhaps materially, from the analysis.

Additional Information

The information and opinions in this report were prepared by Deutsche Bank AG or one of its affiliates (collectively 'Deutsche Bank'). Though the information herein is believed to be reliable and has been obtained from public sources believed to be reliable, Deutsche Bank makes no representation as to its accuracy or completeness. Hyperlinks to third-party websites in this report are provided for reader convenience only. Deutsche Bank neither endorses the content nor is responsible for the accuracy or security controls of those websites.

If you use the services of Deutsche Bank in connection with a purchase or sale of a security that is discussed in this report, or is included or discussed in another communication (oral or written) from a Deutsche Bank analyst, Deutsche Bank may act as principal for its own account or as agent for another person.

Deutsche Bank may consider this report in deciding to trade as principal. It may also engage in transactions, for its own account or with customers, in a manner inconsistent with the views taken in this research report. Others within Deutsche Bank, including strategists, sales staff and other analysts, may take views that are inconsistent with those taken in this research report. Deutsche Bank issues a variety of research products, including fundamental analysis, equity-linked analysis, quantitative analysis and trade ideas. Recommendations contained in one type of communication may differ from recommendations contained in others, whether as a result of differing time horizons, methodologies, perspectives or otherwise. Deutsche Bank and/or its affiliates may also be holding debt or equity securities of the issuers it writes on. Analysts are paid in part based on the profitability of Deutsche Bank AG and its affiliates, which includes investment banking, trading and principal trading revenues.

Opinions, estimates and projections constitute the current judgment of the author as of the date of this report. They do not necessarily reflect the opinions of Deutsche Bank and are subject to change without notice. Deutsche Bank provides liquidity for buyers and sellers of securities issued by the companies it covers. Deutsche Bank research analysts sometimes have shorter-term trade ideas that may be inconsistent with Deutsche Bank's existing longer-term ratings. Some trade ideas for equities are listed as Catalyst Calls on the Research Website (https://research.db.com/Research/) , and can be found on the general coverage list and also on the covered company's page. A Catalyst Call represents a high-conviction belief by an analyst that a stock will outperform or underperform the market and/or a specified sector over a time frame of no less than two weeks and no more than three months. In addition to Catalyst Calls, analysts may occasionally discuss with our clients, and with Deutsche Bank salespersons and traders, trading strategies or ideas that reference catalysts or events that may have a near- term or medium-term impact on the market price of the securities discussed in this report, which impact may be directionally counter to the analysts' current 12-month view of total return or investment return as described herein. Deutsche Bank has no obligation to update, modify or amend this report or to otherwise notify a recipient thereof if an opinion, forecast or estimate changes or becomes inaccurate. Coverage and the frequency of changes in market conditions and in both general and company-specific economic prospects make it difficult to update research at defined intervals. Updates are at the sole discretion of the coverage analyst or of the Research Department Management, and the majority of reports are published at irregular intervals. This report is provided for informational purposes only and does not take into account the particular investment objectives, financial situations, or needs of individual clients. It is not an offer or a solicitation of an offer to buy or sell any financial instruments or to participate in any particular trading strategy. Target prices are inherently imprecise and a product of the analyst's judgment. The financial instruments discussed in this report may not be suitable for all investors, and investors must make their own informed investment decisions. Prices and availability of financial instruments are subject to change without notice, and investment transactions can lead to losses as a result of price fluctuations and other factors. If a financial instrument is denominated in a currency other than an investor's currency, a change in exchange rates may adversely affect the investment. Past performance is not necessarily indicative of future results. Performance calculations exclude transaction costs, unless otherwise indicated. Unless otherwise indicated, prices are current as of the end of the previous trading session and are sourced from local exchanges via Reuters, Bloomberg and other vendors. Data is also sourced from Deutsche Bank, subject companies, and other parties.

The Deutsche Bank Research Department is independent of other business divisions of the Bank. Details regarding our organizational arrangements and information barriers we have to prevent and avoid conflicts of interest with respect to our research are available on our website (https://research.db.com/Research/) under Disclaimer.

Macroeconomic fluctuations often account for most of the risks associated with exposures to instruments that promise to pay fixed or variable interest rates. For an investor who is long fixed-rate instruments (thus receiving these cash flows), increases in interest rates naturally lift the discount factors applied to the expected cash flows and thus cause a loss. The longer the maturity of a certain cash flow and the higher the move in the discount factor, the higher will be the loss. Upside surprises in inflation, fiscal funding needs, and FX depreciation rates are among the most common adverse macroeconomic shocks to receivers. But counterparty exposure, issuer creditworthiness, client segmentation, regulation (including changes in assets holding limits for different types of investors), changes in tax policies, currency convertibility (which may constrain currency conversion, repatriation of profits and/or liquidation of positions), and settlement issues related to local clearing houses are also important risk factors. The sensitivity of fixed-income instruments to macroeconomic shocks may be mitigated by indexing the contracted cash flows to inflation, to FX depreciation, or to specified interest rates – these are common in emerging markets. The index fixings may – by construction – lag or mis-measure the actual move in the underlying variables they are intended to track. The choice of the proper fixing (or metric) is particularly important in swaps markets, where floating coupon rates (i.e., coupons indexed to a typically short-dated interest rate reference index) are exchanged for fixed coupons. Funding in a currency that differs from the currency in which coupons are denominated carries FX risk. Options on swaps (swaptions) the risks typical to options in addition to the risks related to rates movements.

Derivative transactions involve numerous risks including market, counterparty default and illiquidity risk. The appropriateness of these products for use by investors depends on the investors' own circumstances, including their tax position, their regulatory environment and the nature of their other assets and liabilities; as such, investors should take expert legal and financial advice before entering into any transaction similar to or inspired by the contents of this publication. The risk of loss in futures trading and options, foreign or domestic, can be substantial. As a result of the high degree of leverage obtainable in futures and options trading, losses may be incurred that are greater than the amount of funds initially deposited – up to theoretically unlimited losses. Trading in options involves risk and is not suitable for all investors. Prior to buying or selling an option, investors must review the 'Characteristics and Risks of Standardized Options", at http://www.optionsclearing.com/ about/publications/character-risks.jsp. If you are unable to access the website, please contact your Deutsche Bank representative for a copy of this important document.

Participants in foreign exchange transactions may incur risks arising from several factors, including the following: (i) exchange rates can be volatile and are subject to large fluctuations; (ii) the value of currencies may be affected by numerous market factors, including world and national economic, political and regulatory events, events in equity and debt markets and changes in interest rates; and (iii) currencies may be subject to devaluation or government-imposed exchange controls, which could affect the value of the currency. Investors in securities such as ADRs, whose values are affected by the currency of an underlying security, effectively assume currency risk.

Unless governing law provides otherwise, all transactions should be executed through the Deutsche Bank entity in the investor's home jurisdiction. Aside from within this report, important conflict disclosures can also be found at https://research.db.com/Research/ on each company's research page. Investors are strongly encouraged to review this information before investing.

Deutsche Bank (which includes Deutsche Bank AG, its branches and affiliated companies) is not acting as a financial adviser, consultant or fiduciary to you or any of your agents (collectively, "You" or "Your") with respect to any information provided in this report. Deutsche Bank does not provide investment, legal, tax or accounting advice, Deutsche Bank is not acting as your impartial adviser, and does not express any opinion or recommendation whatsoever as to any strategies, products or any other information presented in the materials. Information contained herein is being provided solely on the basis that the recipient will make an independent assessment of the merits of any investment decision, and it does not constitute a recommendation of, or express an opinion on, any product or service or any trading strategy.

The information presented is general in nature and is not directed to retirement accounts or any specific person or account type, and is therefore provided to You on the express basis that it is not advice, and You may not rely upon it in making Your decision. The information we provide is being directed only to persons we believe to be financially sophisticated, who are capable of evaluating investment risks independently, both in general and with regard to particular transactions and investment strategies, and who understand that Deutsche Bank has financial interests in the offering of its products and services. If this is not the case, or if You are an IRA or other retail investor receiving this directly from us, we ask that you inform us immediately.

In July 2018, Deutsche Bank revised its rating system for short term ideas whereby the branding has been changed to Catalyst Calls ("CC") from SOLAR ideas; the rating categories for Catalyst Calls originated in the Americas region have been made consistent with the categories used by Analysts globally; and the effective time period for CCs has been reduced from a maximum of 180 days to 90 days.

United States: Approved and/or distributed by Deutsche Bank Securities Incorporated, a member of FINRA, NFA and SIPC. Analysts located outside of the United States are employed by non-US affiliates that are not subject to FINRA regulations.

Germany: Approved and/or distributed by Deutsche Bank AG, a joint stock corporation with limited liability incorporated in the Federal Republic of Germany with its principal office in Frankfurt am Main. Deutsche Bank AG is authorized under German Banking Law and is subject to supervision by the European Central Bank and by BaFin, Germany's Federal Financial Supervisory Authority.

United Kingdom: Approved and/or distributed by Deutsche Bank AG acting through its London Branch at Winchester House, 1 Great Winchester Street, London EC2N 2DB. Deutsche Bank AG in the United Kingdom is authorised by the Prudential Regulation Authority and is subject to limited regulation by the Prudential Regulation Authority and Financial Conduct Authority. Details about the extent of our authorisation and regulation are available on request.

Hong Kong SAR: Distributed by Deutsche Bank AG, Hong Kong Branch except for any research content relating to futures contracts within the meaning of the Hong Kong Securities and Futures Ordinance Cap. 571. Research reports on such futures contracts are not intended for access by persons who are located, incorporated, constituted or resident in Hong Kong. The author(s) of a research report may not be licensed to carry on regulated activities in Hong Kong and, if not licensed, do not hold themselves out as being able to do so. The provisions set out above in the 'Additional Information' section shall apply to the fullest extent permissible by local laws and regulations, including without limitation the Code of Conduct for Persons Licensed or Registered with the Securities and Futures Commission. This report is intended for distribution only to 'professional investors' as defined in Part 1 of Schedule of the SFO. This document must not be acted or relied on by persons who are not professional investors. Any investment or investment activity to which this document relates is only available to professional investors and will be engaged only with professional investors.

India: Prepared by Deutsche Equities India Private Limited (DEIPL) having CIN: U65990MH2002PTC137431 and registered office at 14th Floor, The Capital, C-70, G Block, Bandra Kurla Complex Mumbai (India) 400051. Tel: + 91 22 7180 4444. It is registered by the Securities and Exchange Board of India (SEBI) as a Stock broker bearing registration no.: INZ000252437; Merchant Banker bearing SEBI Registration no.: INM000010833 and Research Analyst bearing SEBI Registration no.: INH000001741. DEIPL may have received administrative warnings from the SEBI for breaches of Indian regulations. Deutsche Bank and/or its affiliate(s) may have debt holdings or positions in the subject company. With regard to information on associates, please refer to the “Shareholdings” section in the Annual Report at: https://www.db.com/ir/en/annual- reports.htm.

Japan: Approved and/or distributed by Deutsche Securities Inc.(DSI). Registration number - Registered as a financial instruments dealer by the Head of the Kanto Local Finance Bureau (Kinsho) No. 117. Member of associations: JSDA, Type II Financial Instruments Firms Association and The Financial Futures Association of Japan. Commissions and risks involved in stock transactions - for stock transactions, we charge stock commissions and consumption tax by multiplying the transaction amount by the commission rate agreed with each customer. Stock transactions can lead to losses as a result of share price fluctuations and other factors. Transactions in foreign stocks can lead to additional losses stemming from foreign exchange fluctuations. We may also charge commissions and fees for certain categories of investment advice, products and services. Recommended investment strategies, products and services carry the risk of losses to principal and other losses as a result of changes in market and/or economic trends, and/or fluctuations in market value. Before deciding on the purchase of financial products and/or services, customers should carefully read the relevant disclosures, prospectuses and other documentation. 'Moody's', 'Standard Poor's', and 'Fitch' mentioned in this report are not registered credit rating agencies in Japan unless Japan or 'Nippon' is specifically designated in the name of the entity. Reports on Japanese listed companies not written by analysts of DSI are written by Deutsche Bank Group's analysts with the coverage companies specified by DSI. Some of the foreign securities stated on this report are not disclosed according to the Financial Instruments and Exchange Law of Japan. Target prices set by Deutsche Bank's equity analysts are based on a 12-month forecast period.

Korea: Distributed by Deutsche Securities Korea Co.

South Africa: Deutsche Bank AG Johannesburg is incorporated in the Federal Republic of Germany (Branch Register Number in South Africa: 1998/003298/10).

Singapore: This report is issued by Deutsche Bank AG, Singapore Branch (One Raffles Quay #18-00 South Tower Singapore 048583, 65 6423 8001), which may be contacted in respect of any matters arising from, or in connection with, this report. Where this report is issued or promulgated by Deutsche Bank in Singapore to a person who is not an accredited investor, expert investor or institutional investor (as defined in the applicable Singapore laws and regulations), they accept legal responsibility to such person for its contents.

Taiwan: Information on securities/investments that trade in Taiwan is for your reference only. Readers should independently evaluate investment risks and are solely responsible for their investment decisions. Deutsche Bank research may not be distributed to the Taiwan public media or quoted or used by the Taiwan public media without written consent. Information on securities/instruments that do not trade in Taiwan is for informational purposes only and is not to be construed as a recommendation to trade in such securities/instruments. Deutsche Securities Asia Limited, Taipei Branch may not execute transactions for clients in these securities/instruments.

Qatar: Deutsche Bank AG in the Qatar Financial Centre (registered no. 00032) is regulated by the Qatar Financial Centre Regulatory Authority. Deutsche Bank AG - QFC Branch may undertake only the financial services activities that fall within the scope of its existing QFCRA license. Its principal place of business in the QFC: Qatar Financial Centre, Tower, West Bay, Level 5, PO Box 14928, Doha, Qatar. This information has been distributed by Deutsche Bank AG. Related financial products or services are only available only to Business Customers, as defined by the Qatar Financial Centre Regulatory Authority.

Russia: The information, interpretation and opinions submitted herein are not in the context of, and do not constitute, any appraisal or evaluation activity requiring a license in the Russian Federation.

Kingdom of Saudi Arabia: Deutsche Securities Saudi Arabia LLC Company (registered no. 07073-37) is regulated by the Capital Market Authority. Deutsche Securities Saudi Arabia may undertake only the financial services activities that fall within the scope of its existing CMA license. Its principal place of business in Saudi Arabia: King Fahad Road, Al Olaya District, P.O. Box 301809, Faisaliah Tower - 17th Floor, 11372 Riyadh, Saudi Arabia.

United Arab Emirates: Deutsche Bank AG in the Dubai International Financial Centre (registered no. 00045) is regulated by the Dubai Financial Services Authority. Deutsche Bank AG - DIFC Branch may only undertake the financial services activities that fall within the scope of its existing DFSA license. Principal place of business in the DIFC: Dubai International Financial Centre, The Gate Village, Building 5, PO Box 504902, Dubai, U.A.E. This information has been distributed by Deutsche Bank AG. Related financial products or services are available only to Professional Clients, as defined by the Dubai Financial Services Authority.

Australia and New Zealand: This research is intended only for 'wholesale clients' within the meaning of the Australian Corporations Act and New Zealand Financial Advisors Act, respectively. Please refer to Australia-specific research disclosures and related information at https://australia.db.com/australia/content/research-information.html Where research refers to any particular financial product recipients of the research should consider any product disclosure statement, prospectus or other applicable disclosure document before making any decision about whether to acquire the product. In preparing this report, the primary analyst or an individual who assisted in the preparation of this report has likely been in contact with the company that is the subject of this research for confirmation/clarification of data, facts, statements, permission to use company-sourced material in the report, and/or site-visit attendance. Without prior approval from Research Management, analysts may not accept from current or potential Banking clients the costs of travel, accommodations, or other expenses incurred by analysts attending site visits, conferences, social events, and the like. Similarly, without prior approval from Research Management and Anti-Bribery and Corruption ("ABC") team, analysts may not accept perks or other items of value for their personal use from issuers they cover.

Backtested, hypothetical or simulated performance results have inherent limitations. Unlike an actual performance record based on trading actual client portfolios, simulated results are achieved by means of the retroactive application of a backtested model itself designed with the benefit of hindsight. Taking into account historical events the backtesting of performance also differs from actual account performance because an actual investment strategy may be adjusted any time, for any reason, including a response to material, economic or market factors. The backtested performance includes hypothetical results that do not reflect the reinvestment of dividends and other earnings or the deduction of advisory fees, brokerage or other commissions, and any other expenses that a client would have paid or actually paid. No representation is made that any trading strategy or account will or is likely to achieve profits or losses similar to those shown. Alternative modeling techniques or assumptions might produce significantly different results and prove to be more appropriate. Past hypothetical backtest results are neither an indicator nor guarantee of future returns. Actual results will vary, perhaps materially, from the analysis.

Additional information relative to securities, other financial products or issuers discussed in this report is available upon request. This report may not be reproduced, distributed or published without Deutsche Bank's prior written consent.

David Folkerts-Landau
Group Chief Economist and Global Head of Research

Pam Finelli
Global Chief Operating Officer
Research

Anthony Klarman
Global Head of
Debt Research

Michael Spencer
Head of APAC Research

Steve Pollard
Head of Americas Research
Global Head of Company
Research

Gerry Gallagher
Head of European
Company Research

Andreas Neubauer
Head of Germany Research

Peter Milliken
Head of APAC
Company Research

Jim Reid
Global Head of
Thematic Research

Francis Yared
Global Head of Rates Research

George Saravelos
Global Head of FX Research

Peter Hooper
Global Head of
Economic Research

International Production Locations

Deutsche Bank AG
Deutsche Bank Place
Level 16
Corner of Hunter & Phillip Streets
Sydney, NSW 2000
Australia
Tel: (61) 2 8258 1234

Deutsche Bank AG
Equity Research
Mainzer Landstrasse 11-17
60329 Frankfurt am Main
Germany
Tel: (49) 69 910 00

Deutsche Bank AG
Filiale Hongkong
International Commerce Centre,
1 Austin Road West,Kowloon,
Hong Kong
Tel: (852) 2203 8888

Deutsche Securities Inc.
2-11-1 Nagatacho
Sanno Park Tower
Chiyoda-ku, Tokyo 100-6171
Japan
Tel: (81) 3 5156 6000

Deutsche Bank AG London
1 Great Winchester Street
London EC2N 2EQ
United Kingdom
Tel: (44) 20 7545 8000

Deutsche Bank Securities Inc.
60 Wall Street
New York, NY 10005
United States of America
Tel: (1) 212 250 2500