Completed 21 Nov 2017 03:17 PM EST
Disseminated 22 Nov 2017 12:15 AM EST

J.P.Morgan

North America Fixed Income Strategy
22 November 2017

Do androids dream of electric bonds?

Machine learning in interest rate markets

- We explore to what extent machine learning can enhance or automate investor decision making in fixed income markets …
- … by developing daily, automated trading decisions for liquid, benchmark rates products held for one-week to one-month, feeding market data to a variety of machine learning techniques
- After optimizing, we find that a range of strategies for trading 10-year Treasuries outperform both random and all-long returns at high statistical significance and a headline Sharpe above 1.2 ...
- ... but struggle to outperform for other instruments—including swap spreads and volatility.
- We find substantial evidence that ML can play a role informing investor decision making, particularly to help time optimal execution

US Fixed Income Strategy

Munier Salem AC
(1-212) 270-0371
munier.salem@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Jay Barry
(1-212) 834-4951
john.f.barry@jpmorgan.com

Jason Hunter
(1-212) 270-0034
jason.x.hunter@jpmorgan.com

Devdeep Sarkar
(1-212) 834-2096
devdeep.sarkar@jpmorgan.com

Phoebe A White
(1-212) 834-3092
phoebe.a.white@jpmorgan.com

Alix Tepper
(1-212) 622-9461
alix.tepper@jpmorgan.com

Luke Y Chang
(1-212) 834-7568
luke.chang@jpmorgan.com

J.P. Morgan Securities LLC

See page 14 for analyst certification and important disclosures.

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Machine Learning in Interest Rate Markets

Machine learning is not new to financial markets. The pricing models of mortgage backed securities have for decades relied on accurately predicting borrower refinancing behavior via data-driven, empirical models. Credit agencies form much of their bond ratings outlook by analyzing historical relationships between a company's financials and their likelihood of default. Insurance companies do likewise with actuarial models. More recently, hedge funds have developed "statistical arbitrage" strategies, typically relying on ML and data mining techniques to transform market indicators into a mean reverting signal that can inform rapid, short-term buy/sell decisions. All these endeavors involve appealing to a large bank of historical data to infer predictive information.

Yet something feels very new and exciting about machine learning these days. Cars use it to drive themselves. Phones use it to recognize their owners. Streaming services use it to build eerily well curated playlists of music you've never heard before. Against this backdrop, the question naturally arises: **can machine learning enhance—or even automate—investment decision making?**

We set out to explore this question in the context of U.S. interest rate markets, developing daily, automated trading decisions for benchmark rates products held for one-week to one-month. Rather than explicitly program an algorithm with hard-coded rules (look for positive carry, be careful around Fed meetings, mind the auction cyclicals) we instead feed raw market data into a variety of machine learning techniques, to see if they can independently discover high performing trading strategies. The results suggest certain approaches do a good—i.e., statistically significant—job of timing and sizing trades. In this way, we find substantial evidence that ML can play a role informing investor decision making, particularly to optimize execution.

The task at hand

Our problem domain—the set of 'raw' ex-ante features we use to inform our trade decisions—consists of daily close levels across USD interest rates and beyond. Within fixed income these marks pertain to Treasuries, swaps, swaptions, OIS, TIPS and international interest rates. Beyond rates products we also include levels from the corporate bond, agency mortgage, FX, commodities and equities markets, along with domestic and global economics data. Finally, we include binary 'date flags' for FOMC meetings, payrolls releases and month-ends. **Exhibit 1** summarizes this broad feature set in a bit more detail. For USD rates products we include levels for all benchmark tenors (and expiries/tails), and for a majority of variables we consider 1-week, 1-month and 3-month changes in levels along with 3-month, 6-month, 1-year and 2-year trailing z-scores. Altogether this consists of roughly 1,250 raw input features.

This exhaustive list of factors reflects our "agnostic" preliminary approach to the problem of machine learning in financial markets. A strong case could be made for using our rates expertise to curate a shorter, more focused list of drivers for each asset, reducing the number of features and taking care to build signals devoid of strong collinearities (e.g. using level, slope and curvature of the Treasuries curve, rather than all the benchmark yields themselves; or isolating the residual of a regression of VIX versus 1Mx10Y swaption vol, rather than including both outright). Here we instead throw in the kitchen sink, relying on PCA and normalization (described in the next section) to both reduce dimensionality and remove co-dependencies— perhaps the most extreme departure from our normal approach to rates strategy. As we discuss in the next section, this generic approach comes at the cost of a lack of transparency.

Exhibit 1: We employ a comprehensive set of input features within and beyond USD interest rate markets

Treasuries	yield, yield error, carry, specialness, matched-maturity swap spreads, OIS swap spreads, market depth, repo rates, term premium
Swaps	Realized volatility, carry
Swaptions	Implied vol surface, skew
OIS	Rates
TIPS	Breakevens
MBS	Mortgage basis, hedge adjusted carry, convexity, option-adjusted duration, OAS, positioning
Cross-Asset	HG and HY bond indices, European rates, equity index levels and volatility, FX indices and carry, commodity indices
Econ	Global and regional and domestic economic indices and surprise indices, various date flags
Dates	Flags for FOMC meetings, payrolls and month-ends

Source: J.P. Morgan

Our models were trained and tested over two separate timespans: a post-crisis period from mid-2008 to 2016 and a longer 'millennial' period from 2000-2016. The first time-period comes with the advantage of a richer

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

feature set: many of our input factors have only been reliably and frequently tracked since the financial crisis. The second time period covers an obviously broader set of market conditions, including the complete arc of the global financial crisis, albeit with a sparser set of factor inputs. We elected to "quarantine" all 2017 data, which was not used for training or testing purposes under any circumstances, until we had arrived at our final, tuned and productionized machine learning strategies. We describe the rationale for this in greater detail in the next section.

The strategies we developed were tasked with executing trades on a daily basis. The trade structure (e.g., a 1Mx10Y ATMF swaption straddle) and a horizon of either one week or one month were stipulated in advance. Thus the only decision to make was that of positioning: whether to buy or sell, and in what size. In some cases the trade size was uniform for each day, and the predictor simply needed to elect to buy or sell. In other cases the algorithm was allowed to arrive at an optimal trade size. A slightly more advanced approach would allow the predictor to also select the proper structure on a daily basis, and/or arrive at the decision of when to stop out of the trade on its own. We looked primarily at three different trade constructs: 10-year Treasuries, 10-year matched maturity swap spreads, and 1Mx10Y ATMF swaption straddles. Appendix A briefly describes how we simulate the P/L from each.

Thus our problem boils down to predicting the optimal daily trade positioning, given our set of ex-ante input factors. Given this limited time frame and the slow cadence of our data (daily closes), this gives us a limited sample size of roughly 1500 testing days for the post-crisis set and roughly 3500 testing days for the millennial period. Worse still the sample points will exhibit a strong degree of autocorrelation with adjacent days (forward and backward over the length of the trade horizon), even if the underlying features and asset performance exhibit *no* autocorrelation, simply because these trades are composed of many of the same daily returns. For this reason the number of plausibly independent observations is much lower.

Finally, **our metric for success is risk-adjusted returns, typically captured with Sharpe ratio for linear instruments and non-parametric Sharpe for volatility**. Returns across all financial products exhibit a high degree of kurtosis—outsize moves whose elevated frequency of occurrence is starkly inconsistent with that of a normal distribution. Because of this, sample means and standard deviations are noisy and heavily biased by extreme moves. Said another way, **how you were positioned on one or two days where the market moved aggressively can exert an outsize influence on these performance metrics**. For this reason we carefully outline in the next section how we developed a proper performance benchmark for our ML techniques. Beyond this issue of sample noise, there are simple strategies that are known to perform consistently well over our timeframes of interest (e.g., buying bonds; selling vol). **Thus we require our ML techniques to not only clear the threshold of shot noise, but also, stacked on top of that, the threshold of simpler strategies.**

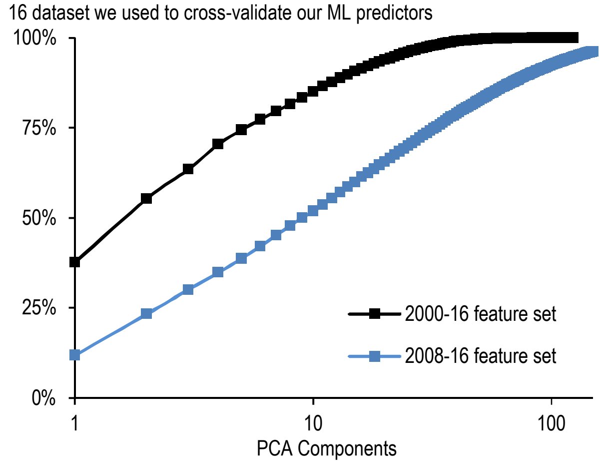
Beyond global risk-adjusted returns, we also explore other features of the strategies' behavior (volatility, average returns, and the fraction of days you buy or sell). From this we find **certain ML techniques can converge to multiple, qualitatively distinct solutions of comparable performance**, depending on the choice of the technique's so-called 'hyperparameters'.

A framework for applying machine learning to fixed income trades

In this section we outline our standardized implementation approach that, given a particular ML technique (e.g., random forest), takes us from raw input features and a PnL series to a set of optimal daily trade instructions. Along the way we highlight some common machine learning pitfalls that we have taken some care to avoid.

Exhibit 2: PCA allows us to reduce the dimensionality of our input feature set

Cumulative explained variance of PCA components for the 2008-16 and 2000-16 dataset we used to cross-validate our ML predictors



Source: J.P. Morgan

To start, we sharply reduce the dimensionality of our feature set via principal component analysis (PCA), as

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

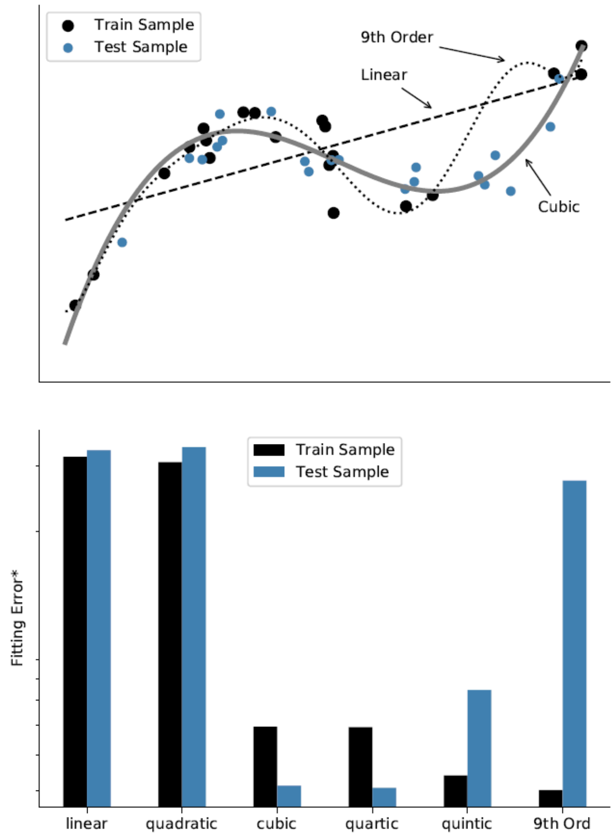
well as de-trending and normalizing each factor. We apply the transformation to all our continuous variables, excluding binary "date flags" from the process. This results in a set of orthogonal (and thus uncorrelated) features, the first few of which—the 'principal components'—explain a majority of the variance in the original raw inputs. Exactly how many components we feed to each ML technique is itself a "hyperparameter" of the technique. As illustrated in **Exhibit 2**, 10 components capture roughly half the variation in the richer 2008-16 dataset, while capturing roughly 80% of variation on the longer, but sparser 2000-16 set. Approaching 90% of the variation requires ~100 components for the 08-16 set, our primary sample for this piece. While including more components captures a greater fraction of the raw input data's variation it can reduce many techniques' ability to converge to a sufficiently well-generalized predictor.

Given our reduced feature set, we next turn to the task of fitting our ML technique of interest. For any prediction scheme, from ordinary least squares (OLS) regression to artificial neural networks (ANN), there are an array of so-called hyperparameters the investigator must choose before setting an algorithm off to find the optimal solution. Properly tuning these parameters can greatly enhance the technique's success at prediction, while an improper choice, disguised by impressive goodness-of-fit metrics "in-sample" can lead to abysmal results out-of-sample. To illustrate this point, we turn to the simple example of OLS regression with a 1-dimensional feature set, as illustrated in **Exhibit 3**. The sample data clearly exhibits non-linear behavior, and is perhaps well explained by a polynomial, at which point we're left to consider how high a degree of polynomial is appropriate in fitting the points. The polynomial degree is one of OLS's hyperparameters—a value that can be chosen prior to applying the standard least-squares minimizing algorithm, and then iterated upon (tuned) until we're satisfied with the results. Too low an order (e.g., linear) and we miss important trends. But too high an order (e.g., 9th-order polynomial), and the regression begins to "memorize" its training space, oscillating franticly between points in a way that does not generalize well to out-of-sample data.

How then to choose the optimal answer? Cross-validation provides a generic approach to arriving at an optimal set of hyperparameters. The data is partitioned into two sets dubbed "training" and "testing" samples. The predictor is then fitted on the training sample and then its performance is evaluated on the testing set. In Exhibit 3 we find in the case of OLS that while ramping up the polynomial order invariably improves goodness of fit within the training zone, performance degrades dramatically out of sample for high-order fits.

Exhibit 3: Cross-validation can help select optimal choices among an ML technique's hyperparameters; for example, in the case of ordinary least squares regression, polynomial order

Top: A hypothetical 1-dimensional dataset split into a "train" and "test"samples; ordinary least squares regression is applied of varying polynomial order, fitted on the train sample **Bottom:** Fitting error* both in- and out-of-sample for the same data set and regression predictors



*Fitting error defined as log of the sum of the squares of the residuals (SSR) between the predicted y-values and the data points, both within the training sample each predictor was fitted on (Train Sample) and outside this sample (Test Sample).

Source: J.P. Morgan

For many problem domains, sample data are partitioned into training and test sets via random selection. For time series data such as financial returns, this approach fails due to the issue of autocorrelation. To the extent markets tomorrow look and perform similar to markets today, randomly placing adjacent days in training and test samples will cause information to bleed between the two sets. Worse still, even if markets exhibit *no* intrinsic autocorrelation, our P/L series itself necessarily exhibits *strong* autocorrelation since the full holding period returns on adjacent days are built from many of the same daily returns. For this reason, our testing set always consists of a contiguous block of days that occurs

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

entirely *after* an expanding or rolling set of training days. We also remove the first *n* days from the testing set, where *n* is our trade's holding period, since the final points within the training space are built from daily returns on those days.

Cross validation is not a silver bullet. For starters, by endlessly stepping through hyperparameter space in search of optimal test-set performance, we are in effect fitting those parameters with the test set. Hyperparameters that perform well in both the training and test set may go on to generalize poorly to new data that never participated in the cross-validation exercise. Second, as we'll see in the next section, many of these methods turned loose on high dimensional data with only a modest sample size (as is certainly the case in this piece) have enough degrees of freedom to produce a very wide range of performance levels. In this case, a blind approach to cross-validation, simply cherry-picking a solution with a high Sharpe ratio from an essentially randomly distributed set of fitted results, will lead to a strategy that veers into massive losses on novel data points. This is equivalent to assessing the performance of a million lousy dart players, taking notice of one who hits ten bullseye in a row by sheer, dumb luck, and then mistakenly branding him an 'expert player.' This issue is especially acute for financial strategies, given the aforementioned high kurtosis of daily returns.

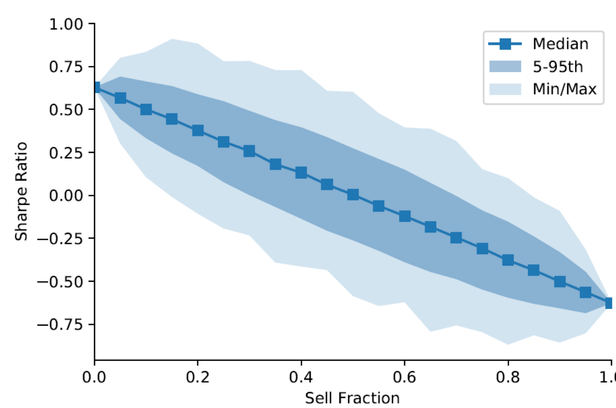
We address these concerns by adopting two additional safeguards. The first is to "quarantine" a final, contiguous set of days that occurs after the testing set, which we set aside from use throughout the cross-validation and hyperparameter selection process. In our case this was the entirety of 2017 thus far, a 10-month period. Only in the final days of our analysis, having selected the optimal strategy for each ML technique, did we set them loose on this quarantined set.

The second safeguard we employed was to adopt a proper baseline significant Sharpe threshold with which to compare the success of each fitted predictor: a threshold Sharpe above which the favorable signal is highly unlikely to have arisen by random chance. As discussed in the previous section, this threshold must clear not only the froth of random, fat-tailed daily returns, but also the performance of simple strategies like buying bonds or selling volatility every day. To generate this value, we take the daily sizing decisions of our predictor and randomly shuffle those decisions (permute them) and calculate the new Sharpe. We then repeat this exercise many thousands of times (essentially a Monte Carlo [MC] approach), recording the 95th and 99th percentile Sharpe of these randomly permuted returns along with the maximal Sharpe from all the trials.

Exhibit 4 shows the results of this exercise for buying or selling 10-year Treasuries with a uniform trade size and holding for a one-week period every day from 2010 to the end of 2016. The median, 95th and maximal randomly discovered Sharpe ratios from 3,000 random trials are shown as a function of sell fraction, i.e., the percentage of days you were short instead of long. **The median Sharpe threshold drops from roughly 0.6 for an all-long strategy, to 0.0 for a 50/50 split**, to -0.6 for all-short over this period. The maximal Sharpe from 3,000 random trials peaks near a sell fraction of 20%, just shy of 0.95, meaning if you were required to go short 1 out of 5 trades, you have a roughly 1:3,000 chance of achieving those risk-adjusted returns by dumb luck. For all our predictors, it is the "residual Sharpe" in excess of these percentiles of randomly permuted decisions that we consider when declaring a run an improvement over sheer dumb luck. Within this framework, it is much more impressive if a machine learning strategy with a 50% sell fraction produces a Sharpe of 0.75 over this timeframe than if an 80%-long strategy does likewise.

Exhibit 4: Doing much better than an all-long strategy by random chance requires a lot of luck, especially when forced to go short some fraction of the time

Sharpe ratio percentile thresholds* for daily trades of 10-year Treasuries held for one week from 2010-16; unitless



*Percentiles come from 3000 trials randomly permuting daily buy/sell decisions at a fixed percentage of selling days (sell fraction) and then calculating Sharpe ratio for each trial.

Source: J.P. Morgan

In the context of cross-validation, where we can run thousands of different hyperparameter choices across the techniques, the clearance burden is quite high: we look for results that outperform the highest Sharpe found in the MC simulation by a comfortable margin. But the interpretation is quite different after cross-validation, when we run a couple of hand-selected 'best' predictors

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

and test their mettle on 'quarantined' 2017 data. There we are essentially running a hypothesis test on each single trial, and the solutions need only clear the 95% threshold (they can of course do much better than this, but a more stringent requirement can elevate the risk of so-called 'Type-II errors'—e.g., false negatives). This procedure is a generalized approach to the classic t-score hypothesis test.

Beyond these two safeguards we also took some care to eliminate solutions that exhibited red flags. In particular we disqualified solutions with exorbitantly high in-sample accuracies (approaching 100%, e.g. memorizing the sample set), and "fluke" solutions—isolated out-performers that did not exist within a cluster of nearby hyperparemeters that performed similarly. We were most confident when a great solution was not especially sensitive to the exact choice of hyperparameter or when dialing up/down a parameter showed a clear, intuitive trend in terms of in-/out-of-sample accuracy, and out-of-sample volatility and Sharpe.

Finally, we turn to the issue of what exactly our ML techniques were tasked with fitting and predicting. Supervised machine learning can be broken down into two broad tasks: classification and regression. Either task is readily applicable to the problem at hand. Our predictors could either attempt to predict the magnitude and direction of returns (regression) or simply whether the asset will rally, sell-off or move sideways (classification). We limited our investigations to the latter approach, exploring classification schemes. For the majority of our work we trained the techniques to classify just two outcomes: rally or selloff, and thus whether to buy or sell. We also looked into multinomial classification, with three outcomes (rally, selloff or roughly unch'd) and four outcomes (rally and selloff broken into two different size classes).

When buying or selling in equal magnitude chunks each day, a buy/sell/hold approach incrementally outperformed simple buy/sell or transacting in multiple discrete sizes. **However, we found *substantially improved* success from a simple binary classifier (buy or sell) when we sized the trade based on the classifier's level of conviction.** All the methods we employed could produce not only a prediction of whether or not the market would rally or selloff but also the probability with which it was certain of the correct category. We then sized the trade according to the Kelly Criterion, a simple two-outcome bet-sizing strategy. Assuming expected gains and losses are of roughly equal magnitude (more or less accurate for a simple duration trade) the optimal size to maximize expected returns is $S = 2*P - 1$ where P is the likelihood you made the right choice (S will be between 0 and 1, since P is necessarily above 50%).

Results from cross-validation

We adopted an iterative train/test approach to selecting optimal ML predictors to produce our daily trading signals, exploring a broad range of "hyperparameters" for each technique. Before launching into the results of this exercise, we enumerate and briefly describe each method employed. As explained in the previous section, we limited this preliminary work to the task of classification (will the market rally or selloff?), and present these techniques in that context.

Classical Techniques

k-Nearest Neighbors (KNN): To classify a new data point, the predictor looks at the training samples closest to the new point in feature space (nearest neighbors). The mode of those samples forms the prediction class (e.g. find the k dates when the market most closely resembled its current conditions, and see whether buying or selling worked well on the majority of those days).

Decision Trees: Here the predictor essentially builds a decision flow chart. Starting at the root node of the tree, the predictor separates the data by a threshold value of one feature, the feature and threshold chosen to produce two branches of data that are as cleanly separated as possible. The process continues down through the branches of the tree, until a decision is reached as to which class the data belongs to.

Support Vector Machines (SVM): SVMs attempt to find a slice through the feature space that best separates disparate outcomes. For a binominal predictor in two dimensions, this can be visualized as a line drawn to separate two prediction classes (e.g., buys and sells) as cleanly as possible, picking in particular a line with the largest gap between itself and the data points; in many dimensions the line becomes a 'hyperplane'. This approach is dubbed "linear" SVM. Of course, many datasets are not cleanly separable by a line or plane, so various "kernel" functions can be employed to transform the original feature set into a higher dimensional space where the plane succeeds. We employed a few non-linear kernels, among them polynomial functions, radial basis functions (RBF) and sigmoidal functions, with varying levels of success.

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Ensemble Techniques

Random Forest (RF): Random forests create an ensemble of decision trees. Each tree can be built from a randomized sub-sample of the data, and the decisions at each node can be made from a randomly selected subset of input features. This lack of determinism makes for a more diverse, healthy ecosystem of predictors from which to draw the modal outcome (majority rules).

'Deep Learning' Techniques

Artificial Neural Networks (ANN): ANNs use a set of linked 'neurons' to transform feature inputs into an output class prediction. The input feature values are passed as signals to an initial layer of neurons, and each constituent neuron passes its input signal through an activation function, the output of which is fired off to the next layer of neurons. The strength of the connection between each pair of neurons—i.e. the weight of each linkage—is adjustable, and just like a living creature's brain, the network learns by strengthening connections that were active during a correct trial classification, and weakening those active during a mistaken prediction.

Initial cross–validation results

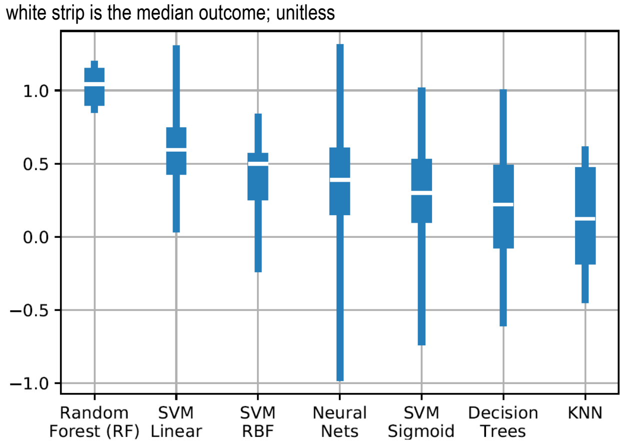
We performed the train/test cross-validation procedure outlined in the previous section on all of the above algorithms, beginning with daily trades of 10-year on-the-run Treasury notes, held for one week. We trained each predictor using the broad set of input factors available for the 2008-16 period, holding off on using all 2017 data which we placed in "quarantine." Towards the end of this section, we broaden the discussion to other asset classes, hold periods and training epochs.

Across techniques and regardless of parameter choices, we found a substantial improvement in outcomes when we sized trades based on the classifier's level of conviction. For this we use the Kelly Criterion, ($S = 2*P-1$, where P represents your perceived likelihood that the prediction is correct). Going forward we present only results using this sizing approach.

For each technique, we marched over a broad "grid" of hyperparameters, fitting the predictors on a training sub sample and then testing its performance on a later, distinct test sample. We then computed summary statistics for that particular run, such as training and test sample hit rates, daily buy/sell fractions and ultimately Sharpe ratios. From the start, we eliminated solutions with excessively high training hit rates (e.g. predictors that simply memorize). All surviving predictors are summarized in **Exhibit 5,** which shows the distribution of test-sample Sharpe ratios broken out by ML technique. From this chart we found several ML techniques, for a subset of hyperparameter choices, produced Sharpe ratios meaningfully higher than uniform buying (see Exhibit 4 and discussion therein).

Exhibit 5: Trading 10-year Treasuries, we found several ML techniques could produce predictors with out-of-sample Sharpe ratios meaningfully above an all-long strategy

Distribution of Sharpe ratio broken out by ML technique; each predictor was trained to take daily positions in 10-year Treasuries* over the test sample†; the thin lines show the min/max range, thick lines the inner-quartile range, and white strip is the median outcome; unitless



* Positions were taken daily throughout the test period (see next note), holding the then-on-the-run 10-year note. Trades were sized based on the predictor's level of conviction, following the Kelly Criterion assuming a symmetric payout distribution, e.g. $S = 2*P - 1$, where P is between 50% and 100%.

† Predictors were trained on data beginning in mid-2008 and tested out-of-sample beginning in early 2010. The first 5 days were removed from the testing period, and Sharpe ratios and sell fractions were then computed on the remaining out-of-sample period of roughly 1.5 years. The training window was then expanded four times, until all dates up until 12/30/2016 were tested. Throughout this predictor-selection and evaluation process, data from 2017 was held in "quarantine" and not under consideration.

Source: J.P. Morgan

It's tempting to simply select a few high performers (with Sharpes in excess of 1.2) and declare success. This is, after all, out-of-sample performance. However, given our sparse data set, the outsize importance of a few timeframes when the market moves substantially, and the relative opaqueness of many of these methods, cherry-picking these predictors and calling it a day strikes us as highly irresponsible.

Instead we looked for structure and consistency in these results. As **Exhibit 6** illustrates, the various ML predictors arrived upon a very broad range of trading strategies, at least measured by the fraction of days they went short (sell fraction). While random forests showed a high degree of structure and consistency (two tight clumps, both in excess of unity Sharpe), ANN predictors were all over the map. The best ANN predictors appear

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

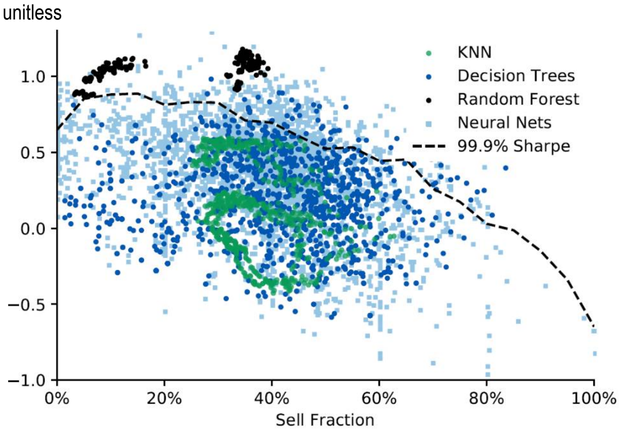
Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

as isolated outperformers, and we had a lot of trouble identifying why exactly they out-performed. Neighboring choices of hyperparameters produced drastically different performance and behavior. Worse still, the performance was highly sensitive to a random seed, used to initiate the optimization scheme that solves for the optimal linkage weights, suggesting the algorithm often did not converge to the globally optimal network. These issues are all likely symptoms of our sparse dataset; the complexity of markets; and ANNs' known issues when applied towards prediction tasks, as opposed to object recognition, where it has enjoyed much success.

Exhibit 6: Looking at risk-adjusted returns across a range of experiments, classical techniques were sensitive to choices of hyperparameters, RF outperformed, and ANNs were erratic

Sharpe ratio versus fraction of days short for various ML predictors tested on 10-year Treasuries* held daily for 5 days over our post-crisis sample space†; unitless



* Positions were taken daily throughout the test period (see next note), holding the then-on-the-run 10-year note. Trades were sized based on the predictor's level of conviction, following the Kelly Criterion assuming a symmetric payout distribution, e.g. S = 2*P – 1, where P is between 50% and 100%.

† Predictors were trained on data beginning in mid-2008 and tested out-of-sample beginning in early 2010. The first 5 days were removed from the testing period, and Sharpe ratios and sell fractions were then computed on the remaining out-of-sample period of roughly 1.5 years. The training window was then expanded four times, until all dates up until 12/30/2016 were tested. Throughout this predictor-selection and evaluation process, data from 2017 was held in "quarantine" and not under consideration.

Source: J.P. Morgan

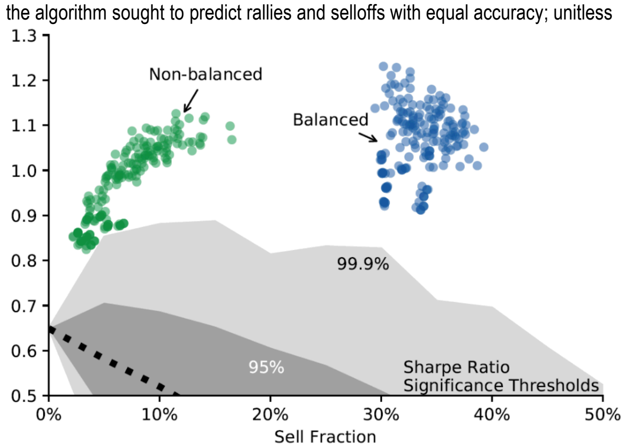
While random forest succeeded almost regardless of the choice of hyperparameters, and ANNs performed sporadically and in a seemingly irrelevant fashion to their selection, the suite of classical methods we employed sat somewhere in between. Decision trees, KNN and SVM techniques produced a broad spectrum of Sharpe ratios and sell fractions (Exhibit 6), but clear trends emerged when the data was broken out by choice of hyperparameters. In general, low-depth decision trees, also pruned for nodes with too few samples points, generalized more accurately. KNN schemes with a broad sample of nearby neighbors (*k~80* as opposed to, say, 10) performed better out of sample. Yet while these properly tuned classical techniques showed some promise, the clear 'winner' was random forest.

A closer look at random forest

By essentially every metric we probed, the most promising and consistent results came via the random forest (RF) technique, the only "ensemble" method we explored in this preliminary work. As the name suggests, a random forest is built from an ensemble (forest?) of decision trees. Each tree is trained on a randomly selected subset of the input data, and for each prediction, the random forest reports the modal answer among all the trees—essentially majority rules. In either analogy or contrast to real-world legislative bodies (depending on your point of view), random forests benefit from many dissenting voices. Unfortunately, even after sub-sampling the data, trees can often look annoyingly similar, and so the algorithm also randomly selects a subset of input features the tree can use to form decisions at each branch. Through all these efforts, the resulting ensemble of trees is typically less susceptible to overfitting.

Exhibit 7: Random forest predictors, trading 10-year Treasuries out-of-sample from 2011-16 consistently produced Sharpe ratios above 1.0, well above the dumb-luck threshold

Sharpe ratio versus fraction of days short for random forest predictors tested on 10-year Treasuries* held daily for 5 days over our post-crisis sample space†; the 'balanced' (blue) and 'non-balanced' (green) clumps denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy; unitless



* Positions were taken daily throughout the test period (see next note), holding the then-on-the-run 10-year note. Trades were sized based on the random forest's level of conviction, following the Kelly Criterion assuming a symmetric payout distribution, e.g. S = 2*P – 1, where P is between 50% and 100%.

† Our random forest predictors were trained on data beginning in mid-2008 and tested out-of-sample beginning in early 2010. The first 5 days were removed from the testing period, and Sharpe ratios and sell fractions were then computed on the remaining out-of-sample period of roughly 1.5 years. The training window was then expanded four times, until all dates up until 12/30/2016 were tested. Throughout this predictor-selection and evaluation process, data from 2017 was held in "quarantine" and not under consideration.

Source: J.P. Morgan

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

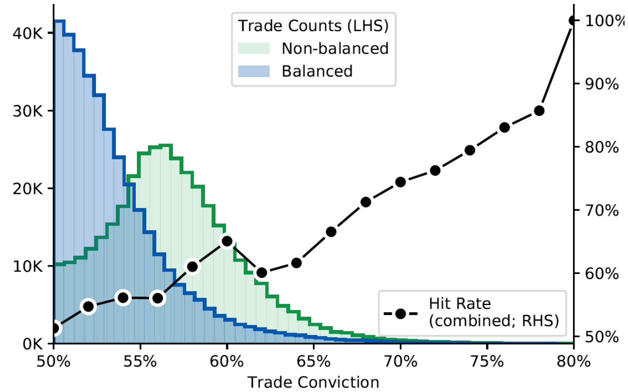
Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Since RF was by far the most successful technique we employed, we take some time to present its performance and general behavior on the cross-validated sample data, again for 10-year Treasuries, traded daily and held for one-week periods, from 2008-16. After eliminating parameter choices that showed obvious signs of over-fitting, we were left with predictors consistently producing Sharpe Ratios in excess of unity (**Exhibit 7**). The solutions separate into two general clumps by the fraction of days they go short based on the setting of one parameter that dictates how important it is for correct predictions to be "balanced." In the context of bonds, where buying every day can generate a decent enough hit rate, this parameter effectively asks the algorithm to spot opportunities to short with as much accuracy as opportunities to buy. We note both solutions outperform all long by a comfortable margin, despite markedly different aggregate behavior. While for the sake of uniformity and simplicity none of the results presented here include transaction costs, for the random forest results in Exhibit 7, standard bid/ask on 10-year Treasuries degrade all Sharpe ratios (including the benchmark all-long and randomly permuted results) by a bit less than 0.1 (~10%).

Exhibit 8: The highest performers were optimized to predict selloffs with more accuracy and were more 'timid'—having less confidence in their predictions and often trading in smaller size

Distribution of days on which the RF predictors* had X% confidence ('conviction') in their decision to go long or short (LHS, count) broken out by weighting strategy†, also shown is realized hit-rate vs conviction (RHS; %)



* RF cross-validated on 10-year Treasury performance (daily trades, 1-week holding) from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs.

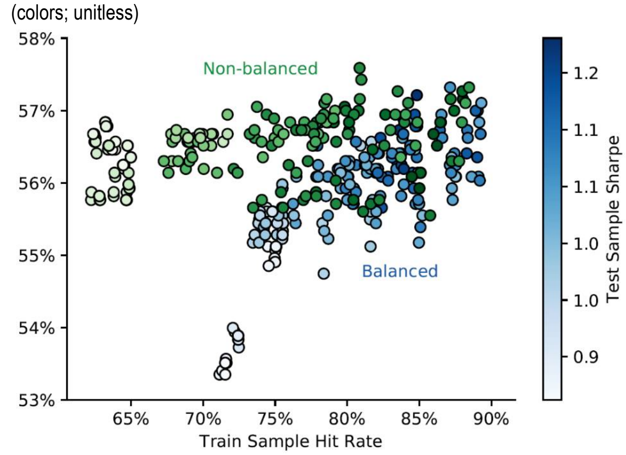
Source: J.P. Morgan

As mentioned above, across techniques we found sizing trades by the predictor's level of conviction substantially boosted Sharpe ratios, despite the trades having the same "hit rate" in calling buys or sells. For RF we found the balanced and non-balanced clumps had disparate levels of conviction. The balanced predictors behaved in a markedly more 'timid' fashion, with most days having a conviction ~50% (very low conviction), than the non-balanced ones where the modal conviction was ~56% (**Exhibit 8**). These two distinct personalities did not preclude either from performing well in aggregate. **What's more, all the RF predictors appeared to be decently "self-aware" with realized test-sample hit rates trending much higher on days when they had high conviction (where they traded in large size). This suggests one of the most compelling applications of ML is in timing execution.**

We also found that the out-of-sample hit rate of our RF predictors was not especially well correlated with Sharpe ratio (**Exhibit 9**). For non-balanced predictors, which had very low sell fractions, test sample hit rate stood firm at roughly 57% for all predictors, regardless of training-sample performance. However, as training hit-rate continued to climb, so did risk-adjusted returns, again thanks to a better understanding of how confident the predictor was in the call it was making. This behavior was also true of the balanced schemes, though to a lesser extent.

Exhibit 9: Among optimized predictors, Sharpe Ratio continued to climb as in-sample "hit-rate" (calling rallies and selloffs) improved, despite out-of-sample hit-rate saturating

Test (out-of-)sample hit-rate vs train (in-)sample hit rate*, broken out by weighting strategy† (LHS; %), with shading connoting test sample Sharpe ratio (colors; unitless)



* RF cross-validated on 10-year Treasury performance (daily trades, 1-week holding) from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs

Source: J.P. Morgan

An important question to ask is *when* these ML techniques out-performed simply buying bonds. **Exhibit 10** shows that outperformance came in bursts, particularly when ML schemes managed to avoid large draw-downs such as following the taper tantrum of 2013

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

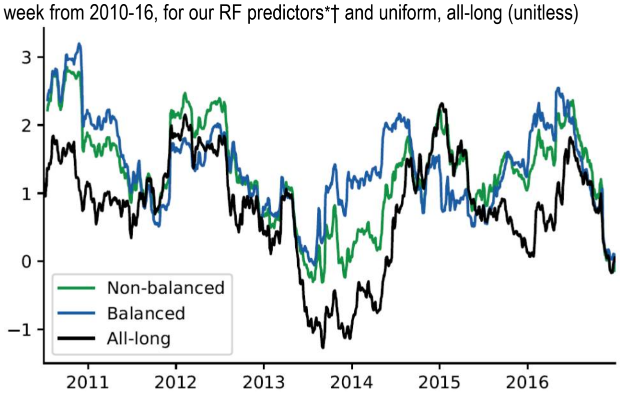
Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

and the widening in rates in late 2015. Within the 2010-16 testing timeframe, and across all our candidate RF predictors, performance rarely fell below the simple all-long strategy. This stands in contrast to the behavior of the randomly permuted decisions that also struck upon a high global Sharpe, where an enormous outperformance in one regime was often tempered by a large underperformance a year or two later.

Exhibit 10: RF prediction strategies consistently outperformed a simple all-long strategy through the Taper Tantrum and late 2015

Rolling 1-year Sharpe ratio for daily trades of 10-year Treasuries held for one week from 2010-16, for our RF predictors*† and uniform, all-long (unitless)



* RF cross-validated on 10-year Treasury performance (daily trades, 1-week holding) from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs

Source: J.P. Morgan

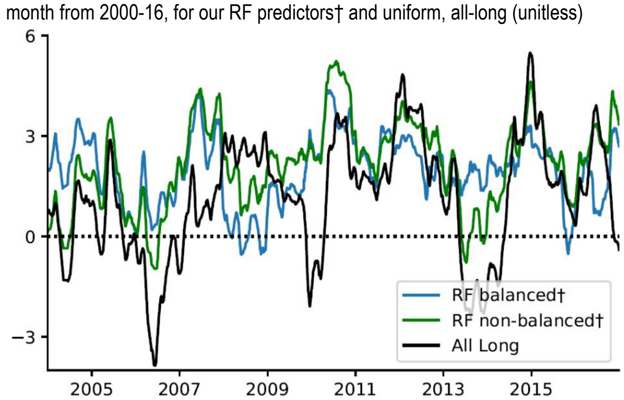
The results we've presented for 1-week hold, 10-year Treasury trades using our broad, 2008-16 feature set generalize well to both monthly-hold strategies and predictors cross-validated over the longer but lower-dimensional, 2000-16 feature set. For both weekly- and monthly-hold bond trades tested from 2003-16, RF predictors managed to avoid all the major drawdowns associated with daily buy strategies (**Exhibit 11**), though they do dip into the red, on a rolling 1-year basis, on occasion.

Emboldened by our success with duration trades, we next explored how well machine learning techniques perform on spreads and volatility. For spreads we again looked to the 10-year sector, trading matched-maturity swap spreads to on-the-run Treasuries. And for volatility we traded 1Mx10Y swaption straddles (no delta hedging) held for both one week and to maturity. Once more we found RF predictors provided the best performance, with behavior again bifurcated across balanced and non-balanced class weight schemes.

However, whereas with bonds the bulk of RF predictors produce Sharpes well in excess of both our 95th and 99.9th percentile significance thresholds, the story is more mixed for spreads and vol (**Exhibit 12**). In the case of spreads, there is some ambiguity around the proper 'benchmark' strategy to compare against. Over our standard cross-validation period of the post-crisis era, narrowers have been the winning spread trade on average. However, the "default" positive carry position is to buy spreads (wideners). So which to choose as a baseline? For this exercise we chose *selling* spreads, though for our final test on the quarantined 2017 data, we'll compare to *buying* spreads. This inconsistency is perhaps unfair to our ML predictors, and one interesting question is whether they'll beat both strategies over the relevant timeframes. With all that in mind, Exhibit 12's results are somewhat lackluster: the entirety of RF candidate predictors mostly fall above 95% of our randomly permuted signal Sharpe values. But they lie comfortably within the maximum Sharpe found from the Monte Carlo exercise (roughly speaking, the 99.9% threshold). This is not the clean outperformance we saw with duration.

Exhibit 11: Shifting our cross-validation to the 2000-16 feature set, we found RF strategies manage to avoid major drawdowns associated with all-long Treasury strategies

Rolling 1-year Sharpe ratio for daily trades of 10-year Treasuries held for one month from 2000-16, for our RF predictors† and uniform, all-long (unitless)



* RF cross-validated on 10-year Treasury performance (daily trades, 1-month holding) from 2000-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs.

Source: J.P. Morgan

For volatility, disappointment begins to really set in. Here we measure performance not by Sharpe, but rather a non-parametric ratio better suited to the non-linear payoff's skewed, long-tailed returns. Trading swaption straddles, our non-balanced predictors simply converged to the popular systematically short-gamma strategy, while the balanced predictors *underperformed* systematic shorts by a wide gap. While the vol strategists among us can breathe easy for the moment being, we note that this underperformance may have much to do with our

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

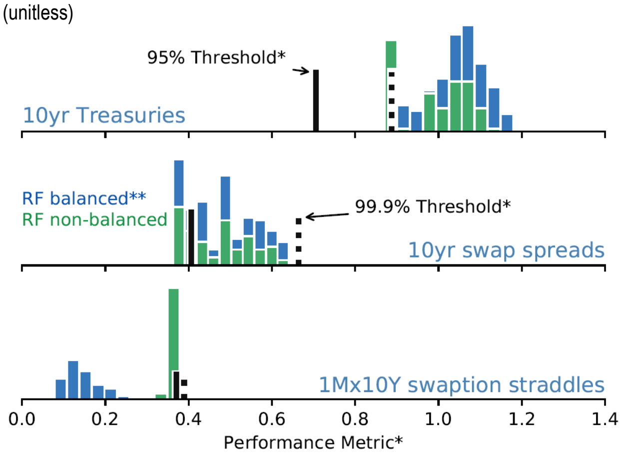
Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

categorization/sizing scheme, which was designed with a symmetric P/L distribution in mind.

Exhibit 12: Applied to spreads, our RF predictors had far more trouble out-performing daily purchases of narrowers; for volatility, our best performers converged to simply selling vol

Distribution of performance metrics* for daily trades of Treasuries, swaps and swaption straddles, held for one week, from 2008-16, for our RF predictors† (unitless)



* Performance metrics denote Sharpe ratio for Treasuries and swaps, and a non-parametric ratio for volatility, built from the average of 1) Non-parametric Sharpe: the average of median versus inter-quartile range; 2) Sterling ratio: median returns versus median losses; and 3) drawdown ratio: returns versus 5th percentile as expected returns versus downside risk. All metrics applied on out-of-sample test set data from 2010-16.

** RF cross-validated from 2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details. Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs.

Source: J.P. Morgan

Performance on Quarantined 2017 Data

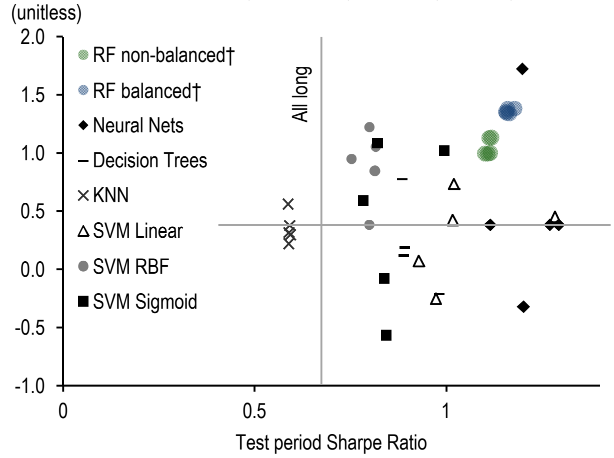
The results presented thus far came from an iterative, experimental fit-then-test approach. While the success of the most robust predictors proved insensitive to the details of implementation, we nonetheless fully admit that during the cross validation process parameters were tuned, design decisions were tweaked, and techniques and strategies were proposed and then abandoned, until out-of-sample performance delivered a respectable risk-adjusted return. Along the way, high level information from our test set has inevitably seeped into the fitting and selection process.

Thus from the start we elected to "quarantine" data from 2017, removing it from all analysis until we had selected the top candidate predictors from each ML technique. This final sample serves to validate the predictors. We have a testable hypothesis: that the best predictors, particularly the random forest predictors, will continue to out-perform a daily buying strategy. And we have a framework for judging the statistical significance of the outperformance (outlined in the Framework Implementation Section).

Beyond this hypothesis-testing framework, testing performance on quarantined data is particularly important for financial time series, which are notoriously non-stationary. That said, 2017 is not exactly a brand new era of financial markets and performance therein says very little about how these predictors would fare through a large, exogenous shock to rates markets.

Exhibit 13: RF predictors handily—and consistently—outperformed all-long Treasury strategies on our quarantined 2017 dataset, consistent with their behavior in cross-validation

Sharpe ratio on 'quarantined' 2017 data versus on test-set data for various ML predictors trained* to trade 10-year Treasury notes daily for 5-day hold periods (unitless)



* All predictors cross-validated on 10-year Treasury performance (daily trades, 1-week holding) from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details. For each predictor we pre-selected the 5 top candidates from each technique before setting it loose on the 'quarantined' 2017 data. Absolutely no information from 2017 was used while training and vetting these predictors.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs.

Source: J.P. Morgan

Exhibit 13 summarizes this exercise, again for daily trades of 10-year Treasuries (1-week hold), showing Sharpe ratios for the top 5 predictors selected in the cross-validation process. Consistent with those results, RF once again stands out as the most consistent out-performer, with balanced RF predictors producing a Sharpe of 1.35 over 2017, compared to 0.38 for all long. The technique outperformed all classical techniques, where the top 5 predictors proved very inconsistent; among them, only SVM RBF consistently performed at-or-above all long. The single best predictor on the quarantine set was a lone neural network that produced a Sharpe in excess of 1.72. However, the four other candidate ANNs did not reproduce this result: three essentially matched all-long, while a fourth produced a negative Sharpe—among the worst across all techniques.

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

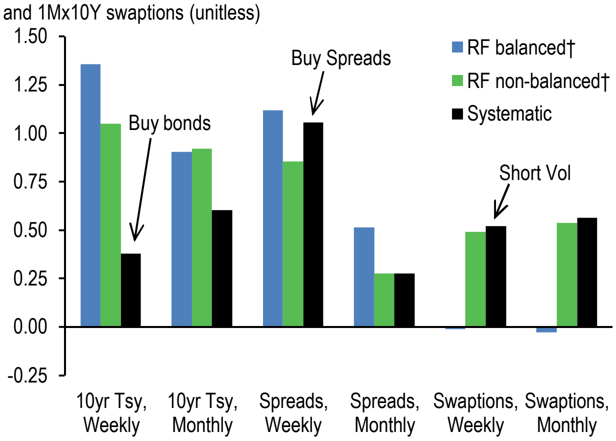
Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

That the highest performing classical and ANN predictors from cross-validation often failed to deliver on the quarantined data is not too surprising, from our perspective. Many of these methods are known to suffer from overfitting, particularly on sparse data such as our daily trades. Neural networks, for all the accolades heaped upon them in the realm of object recognition, are notoriously difficult to properly train on complex problems, and it's not obvious they offer an edge in the realm of time series data.

Exhibit 14: Our predictors fared best on duration over the 'quarantined' 2017 period, performing comparably to simple, systematic strategies on swap spreads and swaption straddles

Performance* of RF predictors** on 'quarantined' 2017 data compared to systematic strategies for 10-year Treasuries, matched-maturity swap spreads, and 1Mx10Y swaptions (unitless)



* For Treasuries and swap spreads, "performance" refers to annualized Sharpe Ratio, for 1Mx10Y swaptions, we instead use the average of non-parametric Sharpe, Sterling and drawdown ratios. Nonparametric Sharpe ratio: median versus inter-quartile range; Sterling ratio: median returns versus median losses; and drawdown ratio: returns versus 5th percentile as expected returns versus downside risk.

† Balanced and non-balanced denote whether or not the algorithm sought to predict rallies and selloffs with equal accuracy—balanced predictors cared more about spotting selloffs

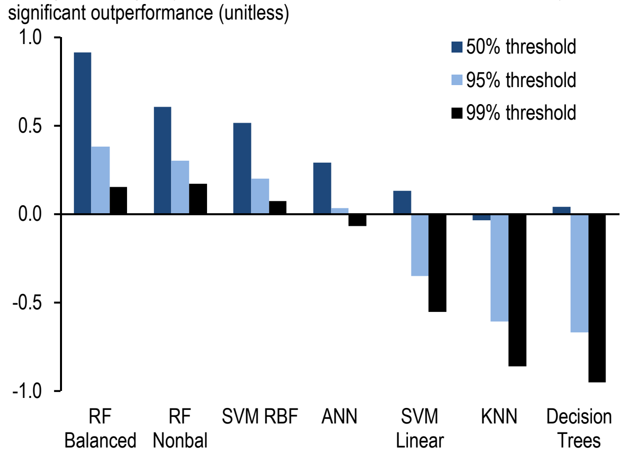
** All predictors cross-validated on daily trades from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details. For each predictor we pre-selected the 5 top candidates from each technique before setting it loose on the 'quarantined' 2017 data. Absolutely no information from 2017 was used while training and vetting these predictors.

Source: J.P. Morgan

We found during cross-validation that even our best performing ML techniques were less successful outperforming in the realm of spreads, and could not offer a substantial improvement over the risk-adjusted returns of systematic short-vol strategies. This trend persisted into the quarantine period (**Exhibit 14**) where over both weekly and monthly timescales RF predictors failed to deliver a meaningful pickup over the aforementioned simple strategies, though they at least managed to pull even. For spreads, it is worth giving credit to random forest predictors for consistency: they incrementally outperformed the success of *narrowers* in cross-validation and then those same strategies went on to roughly match the success of *wideners* in quarantine. For vol, it's likewise far too early to throw in the towel, particularly as many of the details of our implementation (in particular how we size trades based on probabilities) is oriented towards a symmetric distribution of risk, which is more appropriate for linear products.

Exhibit 15: Random forest and SVM RBF strategies outperformed all-long by a statistically significant margin on the 2017 data

Residual Sharpe* of various ML predictors** on 'quarantined' 2017 data for strategies trading 10-year Treasuries; positive values denote statistically significant outperformance (unitless)



*Residual Sharpe: for each predictor, we take its daily trade decisions (trade size and direction) and randomly permute (shuffle) them across all days in the quarantine period, re-computing Sharpe for this randomized set. We repeat the exercise thousands of times, taking the 50th, 95th and 99th percentile outcomes. These Sharpes are then subtracted from the predictor's Sharpe. If this residual Sharpe is positive, we deem it statistically significant to that percentile level. In the context of our quarantine set, this is somewhat analogous to a one-sample hypothesis test.

† All predictors cross-validated on daily trades from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 7 for more details. For each predictor we pre-selected the 5 top candidates from each technique before setting it loose on the 'quarantined' 2017 data. No information from 2017 was used while training and vetting these predictors.

Source: J.P. Morgan

A final question to ask regarding these quarantine results is: how significant is the out-performance? For this we again turn to a Monte Carlo approach, taking the trade sizes and directions for each ML predictor in 2017 and randomly permuting them to compute a Sharpe significance threshold. The performance of each predictor against this threshold, which we call its 'residual Sharpe' is, in this context, loosely analogous to a t-statistic in a one-sample hypothesis test. **Exhibit 15** shows this across ML techniques, averaged across the five candidate predictors. From this we find RF once more comfortably outperforms, at both the 95th and 99th percentile level, as does SVM RBF. The remainder of methods failed to meet this benchmark.

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Conclusions

- We explore to what extent machine learning can enhance or automate investor decision making in fixed income markets by developing daily, automated trading decisions for liquid, benchmark rates products held for one-week to one-month, feeding market data to a variety of machine learning techniques.
- Specifically we look the 'classical techniques': k-nearest neighbors (KNN), decision trees, and support vector machines (SVM); the 'ensemble' technique random forest; and artificial neural networks (ANN)
- Among the various techniques, random forest (RF) showed the most promise, consistently producing Sharpe ratios in excess of unity while trading 10-year Treasuries, both in cross-validation and on 'quarantined' 2017 data. This performance came for both 1-week and 1-month hold periods, using a broad set of factors available from 2008 and a more limited set available since 2000.
- Sizing daily trades by the ML predictor's level of conviction, via the Kelly Criterion, substantially enhanced Sharpe Ratios, across timeframes, asset classes, hold periods and ML techniques. Further, the predictor's perceived conviction in its decision correlated well with its realized hit rate. This suggests a promising application of ML in fixed income is in aiding investors in optimal execution.
- RF predictors clumped into two broad categories, one with higher daily conviction and a lower sell-fraction, and another that went short duration 30-40% of the time, and was a bit more timid in its conviction and sizing. While disparate in aggregate behavior, both sets of strategies performed well in cross-validation and on quarantined data.
- Trading 10-year matched-maturity swap spreads, RF predictors incrementally outperformed narrowers on the cross-validation dataset, and then the same predictors went on to roughly match wideners on 2017 data. While this consistency is promising, our confidence in the technique for spreads is somewhat tepid.
- Trading 1Mx10Y ATMF swaption straddles, RF predictors failed to outperform the popular systematically short-gamma strategy.

Appendix A: Simulating P/L

Below we briefly outline details of our daily trades. All P/Ls are computed using daily close marks. In all cases, we ignore transaction costs.[1]

For **10-year Treasuries** we purchase the current issue, financing the bond in overnight repo including specials. If a new security is auctioned during the trade period of either one-week or one-month, we do *not* roll to the new bond (thus at the end of the trade, we may be holding the first old bond). Trade performance is given by the total return, consisting of the change in the bond's *dirty* price, less total repo cost.

For **10-year matched maturity swap spreads**, we again buy the most recently auctioned 10-year Treasury and simultaneously pay fixed on a PVBP-weighted notional of an on-market swap with maturity date matched to the Treasury. We hold both instruments for the trade horizon (do not roll into the new OTR), and compute total return including carry from both the bond and swap.

For **1Mx10Y ATMF swaption straddles** we hold the straddle, without delta hedging, for either one week or one month. In the former case, we compute the change in premium. For the latter we compare initial premium to terminal payoff.

[1] We exclude transaction costs for the sake of uniformity and simplicity. However, from one random forest cross-validation exercise run on 10-year Treasuries assuming typical bid/ask, we found these costs degrade the Sharpe ratio by less than 0.1 (~10%).

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Disclosures

Analyst Certification: The research analyst(s) denoted by an "AC" on the cover of this report certifies (or, where multiple research analysts are primarily responsible for this report, the research analyst denoted by an "AC" on the cover or within the document individually certifies, with respect to each security or issuer that the research analyst covers in this research) that: (1) all of the views expressed in this report accurately reflect his or her personal views about any and all of the subject securities or issuers; and (2) no part of any of the research analyst's compensation was, is, or will be directly or indirectly related to the specific recommendations or views expressed by the research analyst(s) in this report. For all Korea-based research analysts listed on the front cover, they also certify, as per KOFIA requirements, that their analysis was made in good faith and that the views reflect their own opinion, without undue influence or intervention.

Company-Specific Disclosures: Important disclosures, including price charts and credit opinion history tables, are available for compendium reports and all J.P. Morgan–covered companies by visiting https://www.jpmm.com/research/disclosures, calling 1-800-477-0406, or e-mailing research.disclosure.inquiries@jpmorgan.com with your request. J.P. Morgan's Strategy, Technical, and Quantitative Research teams may screen companies not covered by J.P. Morgan. For important disclosures for these companies, please call 1-800-477-0406 or e-mail research.disclosure.inquiries@jpmorgan.com.

Analysts' Compensation: The research analysts responsible for the preparation of this report receive compensation based upon various factors, including the quality and accuracy of research, client feedback, competitive factors, and overall firm revenues.

Other Disclosures

J.P. Morgan ("JPM") is the global brand name for J.P. Morgan Securities LLC ("JPMS") and its affiliates worldwide. J.P. Morgan Cazenove is a marketing name for the U.K. investment banking businesses and EMEA cash equities and equity research businesses of JPMorgan Chase & Co. and its subsidiaries.

Options related research: If the information contained herein regards options related research, such information is available only to persons who have received the proper option risk disclosure documents. For a copy of the Option Clearing Corporation's Characteristics and Risks of Standardized Options, please contact your J.P. Morgan Representative or visit the OCC's website at https://www.theocc.com/components/docs/riskstoc.pdf

Legal Entities Disclosures
U.S.: JPMS is a member of NYSE, FINRA, SIPC and the NFA. JPMorgan Chase Bank, N.A. is a member of FDIC. **U.K**.: JPMorgan Chase N.A., London Branch, is authorised by the Prudential Regulation Authority and is subject to regulation by the Financial Conduct Authority and to limited regulation by the Prudential Regulation Authority. Details about the extent of our regulation by the Prudential Regulation Authority are available from J.P. Morgan on request. J.P. Morgan Securities plc (JPMS plc) is a member of the London Stock Exchange and is authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority. Registered in England & Wales No. 2711006. Registered Office 25 Bank Street, London, E14 5JP. **South Africa**: J.P. Morgan Equities South Africa Proprietary Limited is a member of the Johannesburg Securities Exchange and is regulated by the Financial Services Board. **Hong Kong**: J.P. Morgan Securities (Asia Pacific) Limited (CE number AAJ321) is regulated by the Hong Kong Monetary Authority and the Securities and Futures Commission in Hong Kong and/or J.P. Morgan Broking (Hong Kong) Limited (CE number AAB027) is regulated by the Securities and Futures Commission in Hong Kong. **Korea**: This material is issued and distributed in Korea by or through J.P. Morgan Securities (Far East) Limited, Seoul Branch, which is a member of the Korea Exchange(KRX) and is regulated by the Financial Services Commission (FSC) and the Financial Supervisory Service (FSS). **Australia**: J.P. Morgan Australia Limited (JPMAL) (ABN 52 002 888 011/AFS Licence No: 238188) is regulated by ASIC and J.P. Morgan Securities Australia Limited (JPMSAL) (ABN 61 003 245 234/AFS Licence No: 238066) is regulated by ASIC and is a Market, Clearing and Settlement Participant of ASX Limited and CHI-X. **Taiwan**: J.P.Morgan Securities (Taiwan) Limited is a participant of the Taiwan Stock Exchange (company-type) and regulated by the Taiwan Securities and Futures Bureau. **India:** J.P. Morgan India Private Limited (Corporate Identity Number - U67120MH1992FTC068724), having its registered office at J.P. Morgan Tower, Off. C.S.T. Road, Kalina, Santacruz - East, Mumbai – 400098, is registered with Securities and Exchange Board of India (SEBI) as a 'Research Analyst' having registration number INH000001873. J.P. Morgan India Private Limited is also registered with SEBI as a member of the National Stock Exchange of India Limited (SEBI Registration Number - INB 230675231/INF 230675231/INE 230675231), the Bombay Stock Exchange Limited (SEBI Registration Number - INB 010675237/INF 010675237) and as a Merchant Banker (SEBI Registration Number - MB/INM000002970). Telephone: 91-22-6157 3000, Facsimile: 91-22-6157 3990 and Website: www.jpmipl.com. For non local research reports, this material is not distributed in India by J.P. Morgan India Private Limited. **Thailand**: This material is issued and distributed in Thailand by JPMorgan Securities (Thailand) Ltd., which is a member of the Stock Exchange of Thailand and is regulated by the Ministry of Finance and the Securities and Exchange Commission and its registered address is 3rd Floor, 20 North Sathorn Road, Silom, Bangrak, Bangkok 10500. **Indonesia**: PT J.P. Morgan Securities Indonesia is a member of the Indonesia Stock Exchange and is regulated by the OJK a.k.a. BAPEPAM LK. **Philippines**: J.P. Morgan Securities Philippines Inc. is a Trading Participant of the Philippine Stock Exchange and a member of the Securities Clearing Corporation of the Philippines and the Securities Investor Protection Fund. It is regulated by the Securities and Exchange Commission. **Brazil**: Banco J.P. Morgan S.A. is regulated by the Comissao de Valores Mobiliarios (CVM) and by the Central Bank of Brazil. **Mexico**: J.P. Morgan Casa de Bolsa, S.A. de C.V., J.P. Morgan Grupo Financiero is a member of the Mexican Stock Exchange and authorized to act as a broker dealer by the National Banking and Securities Exchange Commission. **Singapore:** This material is issued and distributed in Singapore by or through J.P. Morgan Securities Singapore Private Limited (JPMSS) [MCI (P) 202/03/2017 and Co. Reg. No.: 199405335R], which is a member of the Singapore Exchange Securities Trading Limited and/or JPMorgan Chase Bank, N.A., Singapore branch (JPMCB Singapore) [MCI (P) 059/09/2017], both of which are regulated by the Monetary Authority of Singapore. This material is issued and distributed in Singapore only to accredited investors, expert investors and institutional investors, as defined in Section 4A of the Securities and Futures Act, Cap. 289 (SFA). This material is not intended to be issued or distributed to any retail investors or any other investors that do not fall into the classes of "accredited investors," "expert investors" or "institutional investors," as defined under Section 4A of the SFA. Recipients of this document are to contact JPMSS or JPMCB Singapore in respect of any matters arising from, or in connection with, the document. **Japan**: JPMorgan Securities Japan Co., Ltd. and JPMorgan Chase Bank, N.A., Tokyo Branch are regulated by the Financial Services Agency in Japan. **Malaysia**: This material is issued and distributed in Malaysia by JPMorgan Securities (Malaysia) Sdn

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Bhd (18146-X) which is a Participating Organization of Bursa Malaysia Berhad and a holder of Capital Markets Services License issued by the Securities Commission in Malaysia. **Pakistan**: J. P. Morgan Pakistan Broking (Pvt.) Ltd is a member of the Karachi Stock Exchange and regulated by the Securities and Exchange Commission of Pakistan. **Saudi Arabia**: J.P. Morgan Saudi Arabia Ltd. is authorized by the Capital Market Authority of the Kingdom of Saudi Arabia (CMA) to carry out dealing as an agent, arranging, advising and custody, with respect to securities business under licence number 35-07079 and its registered address is at 8th Floor, Al-Faisaliyah Tower, King Fahad Road, P.O. Box 51907, Riyadh 11553, Kingdom of Saudi Arabia. **Dubai**: JPMorgan Chase Bank, N.A., Dubai Branch is regulated by the Dubai Financial Services Authority (DFSA) and its registered address is Dubai International Financial Centre - Building 3, Level 7, PO Box 506551, Dubai, UAE.

Country and Region Specific Disclosures
U.K. and European Economic Area (EEA): Unless specified to the contrary, issued and approved for distribution in the U.K. and the EEA by JPMS plc. Investment research issued by JPMS plc has been prepared in accordance with JPMS plc's policies for managing conflicts of interest arising as a result of publication and distribution of investment research. Many European regulators require a firm to establish, implement and maintain such a policy. Further information about J.P. Morgan's conflict of interest policy and a description of the effective internal organisations and administrative arrangements set up for the prevention and avoidance of conflicts of interest is set out at the following link https://www.jpmorgan.com/jpmpdf/1320742677360.pdf. This report has been issued in the U.K. only to persons of a kind described in Article 19 (5), 38, 47 and 49 of the Financial Services and Markets Act 2000 (Financial Promotion) Order 2005 (all such persons being referred to as "relevant persons"). This document must not be acted on or relied on by persons who are not relevant persons. Any investment or investment activity to which this document relates is only available to relevant persons and will be engaged in only with relevant persons. In other EEA countries, the report has been issued to persons regarded as professional investors (or equivalent) in their home jurisdiction. **Australia:** This material is issued and distributed by JPMSAL in Australia to "wholesale clients" only. This material does not take into account the specific investment objectives, financial situation or particular needs of the recipient. The recipient of this material must not distribute it to any third party or outside Australia without the prior written consent of JPMSAL. For the purposes of this paragraph the term "wholesale client" has the meaning given in section 761G of the Corporations Act 2001. **Germany:** This material is distributed in Germany by J.P. Morgan Securities plc, Frankfurt Branch which is regulated by the Bundesanstalt für Finanzdienstleistungsaufsicht. **Hong Kong:** The 1% ownership disclosure as of the previous month end satisfies the requirements under Paragraph 16.5(a) of the Hong Kong Code of Conduct for Persons Licensed by or Registered with the Securities and Futures Commission. (For research published within the first ten days of the month, the disclosure may be based on the month end data from two months prior.) J.P. Morgan Broking (Hong Kong) Limited is the liquidity provider/market maker for derivative warrants, callable bull bear contracts and stock options listed on the Stock Exchange of Hong Kong Limited. An updated list can be found on HKEx website: http://www.hkex.com.hk. **Korea:** This report may have been edited or contributed to from time to time by affiliates of J.P. Morgan Securities (Far East) Limited, Seoul Branch. **Singapore:** As at the date of this report, JPMSS is a designated market maker for certain structured warrants listed on the Singapore Exchange where the underlying securities may be the securities discussed in this report. Arising from its role as designated market maker for such structured warrants, JPMSS may conduct hedging activities in respect of such underlying securities and hold or have an interest in such underlying securities as a result. The updated list of structured warrants for which JPMSS acts as designated market maker may be found on the website of the Singapore Exchange Limited: http://www.sgx.com. In addition, JPMSS and/or its affiliates may also have an interest or holding in any of the securities discussed in this report – please see the Important Disclosures section above. For securities where the holding is 1% or greater, the holding may be found in the Important Disclosures section above. For all other securities mentioned in this report, JPMSS and/or its affiliates may have a holding of less than 1% in such securities and may trade them in ways different from those discussed in this report. Employees of JPMSS and/or its affiliates not involved in the preparation of this report may have investments in the securities (or derivatives of such securities) mentioned in this report and may trade them in ways different from those discussed in this report. **Taiwan**: This material is issued and distributed in Taiwan by J.P. Morgan Securities (Taiwan) Limited. According to Paragraph 2, Article 7-1 of Operational Regulations Governing Securities Firms Recommending Trades in Securities to Customers (as amended or supplemented) and/or other applicable laws or regulations, please note that the recipient of this material is not permitted to engage in any activities in connection with the material which may give rise to conflicts of interests, unless otherwise disclosed in the "Important Disclosures" in this material. **India:** For private circulation only, not for sale. **Pakistan:** For private circulation only, not for sale. **New Zealand:** This material is issued and distributed by JPMSAL in New Zealand only to persons whose principal business is the investment of money or who, in the course of and for the purposes of their business, habitually invest money. JPMSAL does not issue or distribute this material to members of "the public" as determined in accordance with section 3 of the Securities Act 1978. The recipient of this material must not distribute it to any third party or outside New Zealand without the prior written consent of JPMSAL. **Canada:** The information contained herein is not, and under no circumstances is to be construed as, a prospectus, an advertisement, a public offering, an offer to sell securities described herein, or solicitation of an offer to buy securities described herein, in Canada or any province or territory thereof. Any offer or sale of the securities described herein in Canada will be made only under an exemption from the requirements to file a prospectus with the relevant Canadian securities regulators and only by a dealer properly registered under applicable securities laws or, alternatively, pursuant to an exemption from the dealer registration requirement in the relevant province or territory of Canada in which such offer or sale is made. The information contained herein is under no circumstances to be construed as investment advice in any province or territory of Canada and is not tailored to the needs of the recipient. To the extent that the information contained herein references securities of an issuer incorporated, formed or created under the laws of Canada or a province or territory of Canada, any trades in such securities must be conducted through a dealer registered in Canada. No securities commission or similar regulatory authority in Canada has reviewed or in any way passed judgment upon these materials, the information contained herein or the merits of the securities described herein, and any representation to the contrary is an offence. **Dubai:** This report has been issued to persons regarded as professional clients as defined under the DFSA rules. **Brazil**: Ombudsman J.P. Morgan: 0800-7700847 / ouvidoria.jp.morgan@jpmorgan.com.

General: Additional information is available upon request. Information has been obtained from sources believed to be reliable but JPMorgan Chase & Co. or its affiliates and/or subsidiaries (collectively J.P. Morgan) do not warrant its completeness or accuracy except with respect to any disclosures relative to JPMS and/or its affiliates and the analyst's involvement with the issuer that is the subject of the research. All pricing is indicative as of the close of market for the securities discussed, unless otherwise stated. Opinions and estimates constitute our judgment as of the date of this material and are subject to change without notice. Past performance is not indicative of future results. This material is not intended as an offer or solicitation for the purchase or sale of any financial instrument. The opinions and recommendations herein do not take into account individual client circumstances, objectives, or needs and are not intended as recommendations of particular securities, financial instruments or strategies to particular clients. The recipient of this report must make its own independent decisions regarding any securities or financial instruments mentioned herein. JPMS distributes in the U.S. research published by non-U.S. affiliates and accepts responsibility for its contents. Periodic updates may be provided on companies/industries based on company specific developments or announcements, market conditions or any other publicly available information. Clients should contact analysts and execute transactions through a J.P. Morgan subsidiary or affiliate in their home jurisdiction unless governing law permits otherwise.

Munier Salem
(1-212) 270-0371
munier.salem@jpmorgan.com

Zhan Zhao
(1-212) 834-7218
zhan.zhao@jpmorgan.com

Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

"Other Disclosures" last revised November 11, 2017.