TradeRunner

Ensemble learning-driven systematic trading in interest rate markets

- We revisit and expand our analysis of systematic trading signals in U.S. interest rate markets generated via machine learning techniques, this time focusing on ensemble methods
- Gradient boosting machines (GBM) clearly outperform other classical techniques with some key differences relative to random forest (RF) classifiers
- GBM tends to offer both higher returns and higher variance, resulting in a modest improvement in overall Sharpe ratio versus RF
- We find GBM-driven trading signals produce comparable risk-adjusted returns across the curve in cross-validation with a very high level of statistical significance
- In quarantine, though all tenors outperform passive longs, only 5- and 30-year maturities do so at 90+% confidence …
- … and those sets of signals clearly succeed in both enhancing returns when yields decline and minimizing losses in a bear market

US Fixed Income Strategy

Munier Salem AC
(1-212) 270-0317
munier.salem@jpmorgan.com

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

Jimmy (Guanjie) Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

J.P. Morgan Securities LLC

See page 4 for analyst certification and important disclosures.

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

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

TradeRunner: Ensemble learning-driven systematic trading in interest rate markets

- **We revisit and expand our analysis of systematic trading signals in U.S. interest rate markets generated via machine learning techniques, this time focusing on ensemble methods**
- **Gradient boosting (GBM) of decision trees clearly outperforms other classical techniques with some key differences relative to random forest (RF) classifiers**
- **GBM tends to offer both higher returns and higher variance, resulting in a modest improvement in overall Sharpe ratio versus RF**
- **We find GBM-driven trading signals produce comparable risk-adjusted returns across the curve in cross-validation with a very high level of statistical significance**
- **In quarantine, though all tenors outperform passive longs, only 5- and 30-year maturities do so at 90+% confidence …**
- **… and those sets of signals clearly succeed in both enhancing returns when yields decline and minimizing losses in a bear market**

Last year we introduced a simple framework for executing daily, automated trading in U.S. interest rate products using machine learning techniques trained on historical market performance. We fed learners a broad range of market observable data and pricing information (yields, carry, positioning data, performance in equities, across DM markets, etc.) and then labeled each trading day as "rally" or "selloff" based on the subsequent 5-day performance of the asset in question. The classifiers were trained and tested out-of-sample ("cross-validated") using daily closing levels from 2008-16 and then set loose on a "quarantined" set of data from January 2017 onwards (see *Do androids dream of electric bonds?* M. Salem et al., 11/22/17, hereafter referred to as *Androids*). **We experimented with a broad set of off-the-shelf machine learning techniques and found so-called "ensemble" methods, in particular random forest (RF) classifiers, had the most success trading various products**. In particular, we could consistently produce risk-adjusted returns well in excess of a "passive" uniform buying benchmark while trading 10-year U.S. Treasury notes. The algorithm was moderately successful trading swap spreads but had trouble with swaption straddles (volatility) where it failed to outperform systematic selling strategies.

In the present work, we revisit this ML-based approach to trading interest rates, focusing exclusively on U.S. Treasuries, but broadening to four benchmark points along the curve (2s, 5s, 10s and 30s). We also introduce an additional ensemble learning approach to trading known as "gradient boosting," which we find, by a few compelling metrics, is incrementally more successful than RF. We conclude by updating quarantine performance of these classifiers, which we've been tracking, in real time, since publishing last November.

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

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

Gradient boosting: another successful approach to ML trading

In our previous work, we considered the performance of classical machine learning techniques, such as k-nearest neighbors (KNN), support vector machines (SVM), and decision trees versus passive longs in of 10-year notes. Highly structured/regularized techniques failed to spot any relative value, producing consistently low outperformance (e.g., SVM linear, low-depth trees), whereas techniques allowed to capture detailed interactions and non-linearities had a habit of memorizing in-sample behavior and sporadically producing outperformance in cross-validation that failed to generalize to data held in quarantine during the model engineering phase (e.g., decision trees, KNN).

In that same piece, we found the most success with random forest classifiers—the sole example of an 'ensemble' learning technique employed therein. **Ensemble techniques aggregate many "weak" learners together to achieve enhanced prediction performance**. In the case of random forest, those weak learners are decision trees—excellent at capturing nuanced interactions and nonlinearities, but notorious for overfitting and memorizing in-sample behavior. The "forest" comes from producing many decision trees formed from random sub-samples of the original dataset, where each branch of each tree is then afforded a randomly chosen subset of candidate input features to choose among when partitioning the dataset. **This results in a diverse array of 'weak' learners that then come together to vote on the ensemble's final prediction**.

Another popular ensemble method is gradient boosting, which again aggregates a number of weak learners in the hopes of a resultant aggregate learner with enhanced predictive power. Many common implementations again employ decision trees as the underlying weak learners. The process begins by training a single decision tree (typically a one-branch "stump" that simply predicts the global modal outcome) and computes the errors/residuals of that initial tree. A subsequent tree is then fit to the residuals and used to enhance the prediction of the first tree. The process is then repeated many times until some threshold accuracy is reached. There's a good deal of flexibility and formalism around what 'residuals' and 'accuracy' mean exactly and how to achieve an optimal outcome, and, like random forests, random sub-sampling of samples and features typically enhances the final ensemble learner.

In this work, we present results using gradient boosting machines (GBM) to implement our ML trading strategy and compare its performance and behavior to random forest (RF). In the prior piece, we found RF strategies produced Sharpe ratios in cross-validation on the order of unity, outperforming the all-long benchmark by a statistically significant margin (as judged by a Monte-Carlo shuffle of buy/sell signals at fixed sell fraction). While tuning model "hyperparameter" choices (such as number of trees, depth of each tree, sorting criteria, etc.) incrementally improved performance, the behavior in terms of risk-adjusted returns and percentage of days when the algorithm recommended selling rather than buying ("sell fraction") was tightly clumped into two small clusters determined by just one hyperparameter, controlling "class-weighting." When class-weighting was turned off, the training scheme sought to teach the learner to optimize global performance, across rallies and selloffs. When class-weighting was set to "balanced," the learner was over-penalized for missing selloffs, which occur less frequently. Perhaps intuitively, this bifurcated behavior into a "non-balanced" clump with lower sell fraction (roughly 15-20%) and another "balanced" clump with higher sell fraction (40-45%).

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

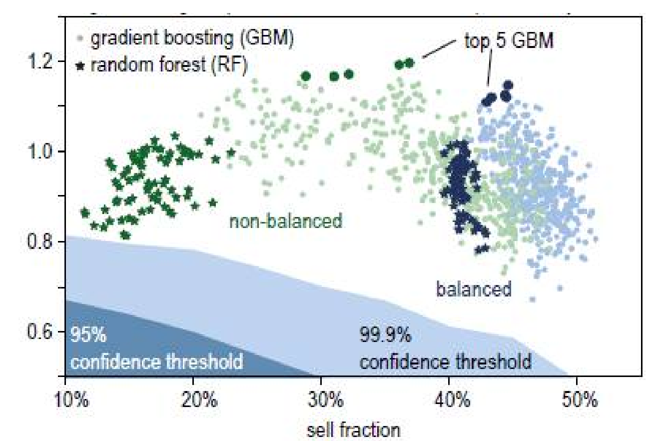
Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

Exhibit 1 shows an updated version of these performance metrics for RF, for a cross-validation period spanning from 2009-16, along with a comparison to the new GBM approach. **In the present section, we focus exclusively on 10-year maturities, but later broaden our analysis to consider other points across the curve**. Similar to RF, GBM classifiers managed to consistently outperform all-long regardless of hyper-parameter choices. Further, the best GBM classifiers incrementally outperformed the top RF classifiers, achieving a 10% boost in Sharpe ratio to roughly 1.1-1.2. **In mild contrast to RF, the sell fraction behavior of GBM is much more varied, with non-balanced GBM classifiers producing a sell fraction anywhere from 20-48% and the balanced scheme selling with a frequency of 40-52%**. While this range of behavior seems broad compared to RF, it is still a concise, well-behaved portion of parameter space that stands in stark contrast to other non-ensemble classifiers we've explored (e.g., decision trees, simple neural nets; see *Androids*). **Among the top-performing GBMs, as measured by Sharpe ratio in cross-validation, sell fraction behavior is more tightly clustered, with both non-balanced and balanced schemes yielding notably higher sell fractions than their RF counterparts**. This is an important change in behavior that we tend to view as an improvement, for reasons explained shortly.

Exhibit 1: We find gradient boosting machines (GBM) perform incrementally better than random forest classifiers (RF) trading duration, with GBM spanning a broader range of buy/sell behavior

Sharpe ratio versus fraction of days short for machine learning (ML) 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



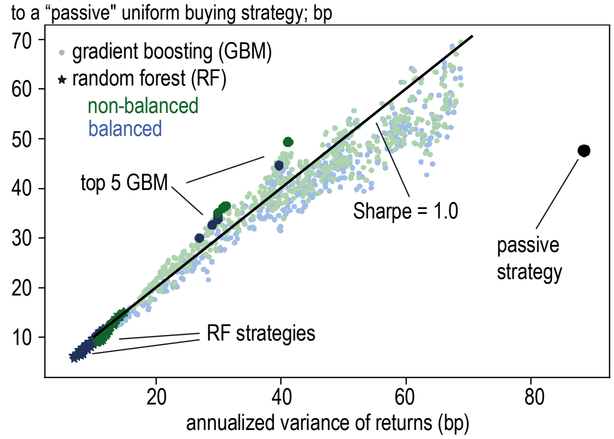
Notes: Positions were taken daily throughout the test period, holding the then-on-the-run Treasury note/bond. Trades were sized based on the gradient boosting machine (GBM)'s level of conviction, following the Kelly Criterion assuming a symmetric payout distribution, e.g. S = 2*P – 1, where P is between 50% and 100%.

Our classifiers' predictors were trained on data beginning in mid-2008 and tested out-of-sample beginning in 2009. 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 2017-18 was held in "quarantine" and not under consideration.

Source: J.P. Morgan

Exhibit 2: Many GBM instances produced dramatically higher returns in cross-validation, though this came at the expense of equally higher variance

Average annualized return versus annualized variance for ML classifiers, compared to a "passive" uniform buying strategy; bp



Source: J.P. Morgan

While these two ensemble approaches to ML-driven trading appear quite similar in terms of risk-adjusted returns and frequency of shorting, stark differences emerge on an outright returns basis, as shown in Exhibit 2. While RF classifiers span a comparatively modest range of 10-20bp in annualized return,

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

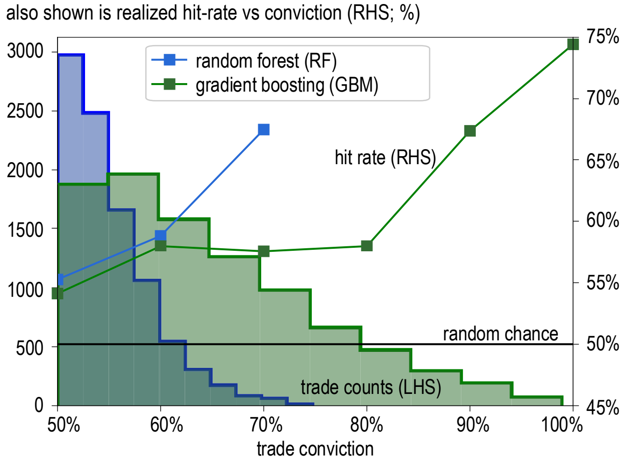
Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

GBM classifiers span a much broader range, with certain hyperparameter choices yielding 50bp+ annualized returns. This outsized headline performance comes with a much higher variance in returns, which rise in lock-step, leading to a risk-adjusted (e.g., Sharpe) frontier that never breaches 1.2. **Our top performing GBM classifiers along this frontier likewise span a comparatively broad range of return/variance behavior compared to RF**.

We can show in a fairly straightforward fashion that this divergence in outright returns behavior boils down to the question of a classifier's "confidence." **Both GBM and RF learners can report how confident they are in any given prediction in the form of a probability from 50% to 100%—the former meaning the classifier considers the outcome a complete toss-up, the latter meaning it considers its prediction a sure thing**. Implementation-wise, this number comes from the observed "purity" of samples within the constituent decision trees: if the particular branch of each tree being used to call the market on that day has a healthy mixture of both buys and sells, the classifier will report low confidence. If the branch in question (of each tree) is purely composed of either buys or sells with few counter examples, it will report high confidence. **We use this confidence metric to size our trades, utilizing the Kelly Criterion, which we've found quite effective at meaningfully boosting performance (see *Androids*)**.

Exhibit 3: GBM learners traded with higher confidence than RF, while enjoying a similar realized hit-rate in cross-validation …

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



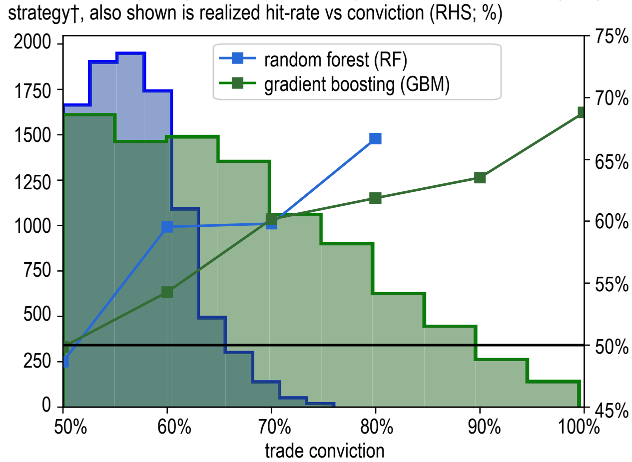
* RF and GBM classifiers cross-validated on 10-year Treasury performance (daily trades, 1-week holding) from 2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 1 for more details.

† 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

Exhibit 4: … and this was true for both balanced and non-balanced buy/sell class weighting† approaches

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



Source: J.P. Morgan

But we can also ask: given a classifier's confidence at trade inception, how well does it perform after the fact? In other words, **we can ask how self-aware the classifier is of its own abilities on a day-to-day basis**. **Within cross-validation, we found an encouragingly direct relationship between conviction in the trade and subsequently realized hit rate**. This behavior is true of both balanced and non-balanced sample weighting schemes for both GBM and RF approaches (**Exhibits 3 and 4**). While the conviction/hit-rate behavior is quite similar across techniques, the frequency of high-conviction trading days is much higher for GBM. If GBM were

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

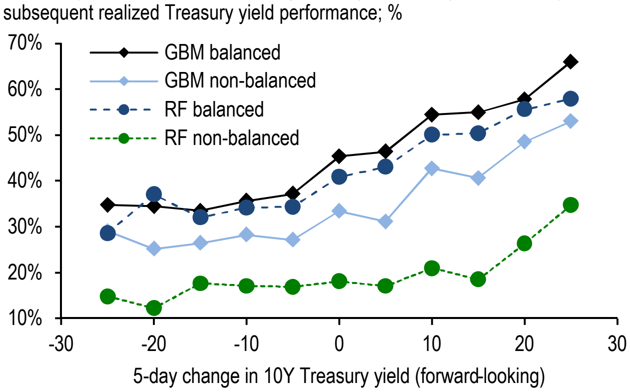
Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

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

always more confident, this would have zero impact on observed average returns and variance (you're simply executing the strategy in larger size). But rather GBM's confidence spans a broader dynamic range, often trading in small size when it's unsure of how to proceed, but also quite often trading in larger size when it grows confident it's seen this kind of market before. This behavior lies at the heart of its higher returns/higher variance performance compared to RF—GBM is at times a riskier player. It's important to note the above plots are shown for the top five hyperparameter choices, which we ranked in terms of Sharpe ratio. **Whereas the confidence distribution varies only subtly across all RF hyperparameters, the distribution can vary quite starkly for GBM, as demonstrated in the variance range in Exhibit 2**.

Exhibit 5: Both GBM and RF classifiers enhance sell fraction, on average, ahead of selloffs, though behavior is still overly biased towards buying bonds ...

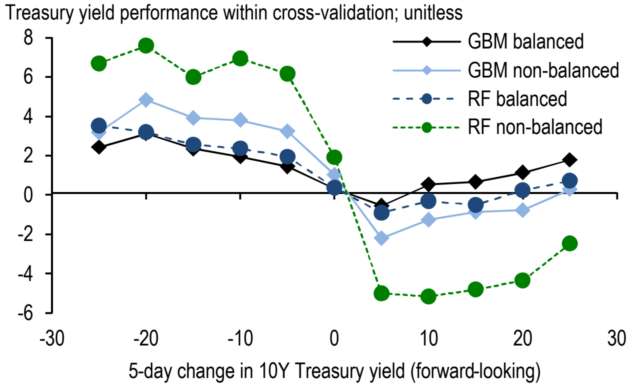
Frequency with which ML classifiers go short (sell fraction) broken out by subsequent realized Treasury yield performance; %



Note: See Exhibit 1 for details on the classifiers and trade construction.
Source: J.P. Morgan

Exhibit 6: ... and this behavior translates into stronger outperformance in a rally than underperformance in a selloff, with the asymmetry greatest among "balanced" class-weighted classifiers

Annualized Sharpe ratio across ML classifiers broken out by subsequently realized Treasury yield performance within cross-validation; unitless



Source: J.P. Morgan

Up to now we've focused exclusively on global behavior (e.g., overall returns and sell fraction) but have not addressed how these classifiers perform in various market scenarios. **An obvious breakout to explore is bull vs bear markets. In particular, we can ask: do our classifiers correctly position short ahead of selloffs and long ahead of rallies? And how does their performance vary in these situations? Exhibit 5** answers the first question, illustrating that our classifiers do in fact tend to sell bonds more frequently ahead of selloffs (identified simply as rising 10Y yields). For our five-day hold period, carry is not particularly large (though we do include this in our full P/L simulation), and thus an ideal classifier would have a sell fraction of 0% when yields are about to fall and 100% when they're about to rise. A classifier that avoids losses, on average, in both rallies and selloffs would at the very least have a sell fraction below 50% in rallies and above 50% in selloffs. Our classifiers are not quite that good: **in rallies they indeed buy bonds more often than they sell, but only ahead of very large selloffs (in excess of 20bp) do our classifiers manage to get short a majority of the time**. In other words, our ML schemes are overly biased towards buying bonds, which is perhaps unsurprising given the evolution of 10Y yields over the past decade. The situation is particularly acute for the non-balanced RF scheme, whereas balanced RF and both flavors of GBM do a much better job approaching the proper sell fraction behavior. **The classifier that performs best in this sense is balanced GBM, which was found to sell more often than it buys, on average, ahead of selloffs in excess of 10bp in cross-validation**.

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

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

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

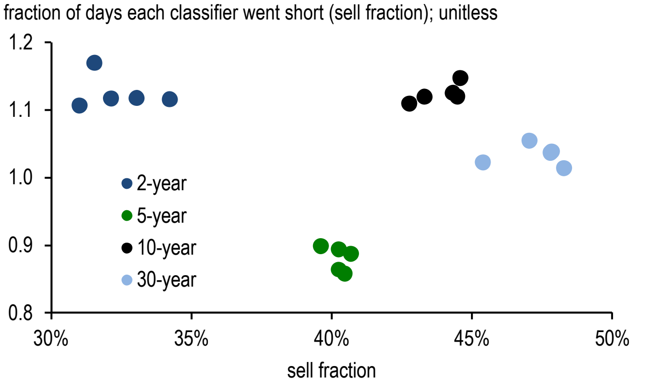
How does this translate into realized performance? **Exhibit 6 shows risk-adjusted returns broken out by degree of rally and selloff**. Given the all-long bias, it's not too surprising to learn the strongest performance comes in strong rallies, with performance dropping precipitously as the market moves into selloffs. That said, **the behavior is fortuitously asymmetric: losses in a selloff are far more subtle than gains in a rally, particularly for GBM and balanced RF**. For balanced GBM, positive risk-adjusted returns were achieved, on average, across a broad range of market environments, with the notable exception of mild selloffs.

Trading performance across the curve[1]

Before exploring recent performance of the classifiers described above on "quarantined" data from 2017-present, we introduce results from training our ML trading strategies on other benchmark tenors. **In particular, we present the cross-validation performance for 2-, 5-, 10-, and 30-year maturities**.

Exhibit 7: GBM classifiers achieved Sharpes in excess of unity across most of the curve in cross-validation, with sell fraction rising steadily from the front end to the long end

Annualized Sharpe ratio within our post-crisis cross-validation period for the five top-performing GBM learners across some benchmark Treasury tenors versus fraction of days each classifier went short (sell fraction); unitless



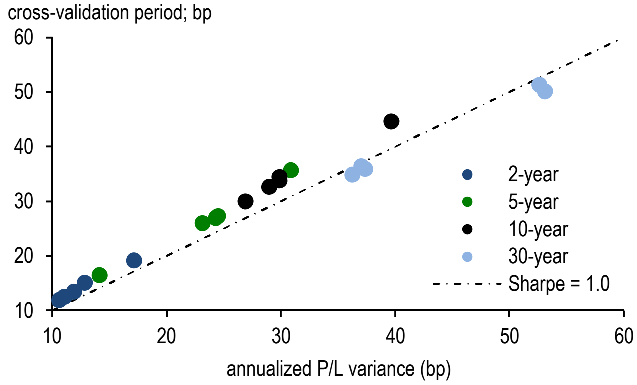
Notes: Positions were taken daily throughout the test period, holding the then-on-the-run Treasury note/bond. Trades were sized based on the gradient boosting machine (GBM)'s level of conviction, following the Kelly Criterion assuming a symmetric payout distribution, e.g. S = 2*P – 1, where P is between 50% and 100%.

Our classifiers' predictors were trained on data beginning in mid-2008 and tested out-of-sample beginning in 2009. 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 2017-18 was held in "quarantine" and not under consideration.

Source: J.P. Morgan

Exhibit 8: Returns were least variable—and lowest—in 2Y notes, rising substantially further out the curve

Average annualized return versus annualized variance across tenors within our cross-validation period; bp



Source: J.P. Morgan

We found much of the classifiers' behavior described above generalizes well across tenors, including a) the monotonic rise in hit rate as trade conviction rises, b) the systematic bias towards buying bonds, particularly for non-balanced sample weighting, and c) strong performance in a rally and mild losses in a selloff. Perhaps most important of all, across tenors we achieve comparable Sharpes in cross-validation in excess of the passive benchmark and often in excess of unity for the balanced GBM scheme (**Exhibit 7**). **One interesting trend we found was a rising sell-fraction among top-performing classifiers further out the curve, with learners trained on 2s selling 30-35% of the time whereas those**

[1] PnL in both cross-validation and later quarantine assumes overnight repo financing incorporating specialness for that particular issue.

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

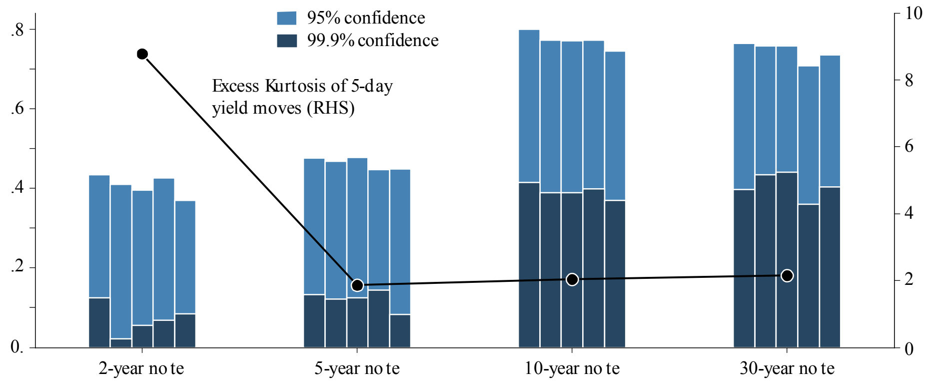
Joshua Younger
(1-212) 270-1323
joshua.d.younger@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

trained on 10s and 30s sold roughly 45% of the time. In terms of P/L variance, top-performing GBM classifiers trained to 10Y notes and 30Y Bonds (and to a lesser extent 5Y notes) tended to exhibit notably higher volatility of returns than 2Y notes perhaps unsurprisingly given short-term yield moves are more volatile in the long end whereas the front end is more anchored to Fed expectations (**Exhibit 8**).

Exhibit 9: GBM (and RF) learners outperform a "passive," uniform buying benchmark to very high confidence, across benchmark tenors we tested, within our 2008-16 cross-validation period

Residual Sharpe* (a measure of benchmark outperformance) of GBM learners† within our 2008-16 cross-validation period for strategies trading 2-, 5-, 10-, and 30-year Treasuries (positive values denote statistically significant outperformance to the indicated confidence level) and excess kurtosis of 5-day yield moves (RHS); 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 95th and 99.9th 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.

† All predictors cross-validated on daily trades from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 1 for more details.

Note: Excess kurtosis is a measure of the extent to which yield moves in a given tenor are distributed with "fat tails" ... i.e. the prevalence of large moves that would be judged outliers when assuming normally distributed returns

Source: J.P. Morgan

To judge the statistical significance of our outperformance across the curve, we employed a Monte Carlo approach, as described in our original write-up (see *Androids*) shuffling the daily buy/sell decisions of each classifier (which range from -1 to 1 continuously, representing a high-conviction sell to a high-conviction buy) and re-computing risk-adjusted returns many, many times. This thus provided a distribution of Sharpe ratios achievable by random luck, allowing us to judge how confident we were in the algorithm's outperformance—a strategy akin to a t-test, though designed to be immune to simplifying assumptions about our data (e.g., normally distributed returns). **Exhibit 9** summarizes the results of this exercise in cross-validation by subtracting the 95th and 99.9th percentile Sharpes achieved in the random shuffling of decision making from the actual performance of the classifiers. **Thus, when the remaining "residual Sharpe" is positive, we can say with high confidence that our results are highly inconsistent with outperforming by "dumb luck."** For the 10- and 30-year Treasury tenors, we find the outperformance is very substantial at even the 99.9% confidence level, thus assuring us that, given the classifier's sell frequency and sizing decisions, it's most likely adding substantial value beyond simply buying bonds. In the case of 5- and 2-year notes, the outperformance is again significant to the 99.9% level, though not overwhelmingly so. This seeming underperformance at the front end is the product of a two-fold disparity in the performance of yields: first, a passive 5-day hold strategy in 2-year notes yielded a substantially higher Sharpe (over 1.0) for 2s and marginally higher for 5s, than for 10-year notes and the bond. And second, the

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

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

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

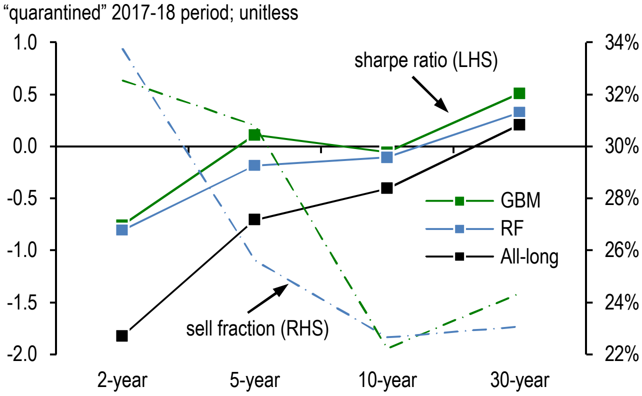
observed kurtosis of yields is markedly higher at the front end than further out the curve (e.g. the prevalence of "tail" events is higher in 2s, see again Exhibit 9).

Performance on 'quarantined' 2017-18 data[2]

In our previous work, we chose to exclude the final year's data entirely from the learning process. This data was used neither to train nor cross-validate the learners, and it was likewise withheld from all design and engineering exercises until the moment of final judgement. **Ten months later, and armed with another successful ensemble learning technique (GBM), we revisit performance in "quarantine," now across benchmark tenors**.

Exhibit 10: Nearly two years into 'quarantine', amidst a sustained selloff across the curve, our ensemble classifiers have managed to tread water in most tenors, outperforming most in the front-end …

Sharpe ratio and sell fraction of GBM and RF learners as measured in the "quarantined" 2017-18 period; unitless

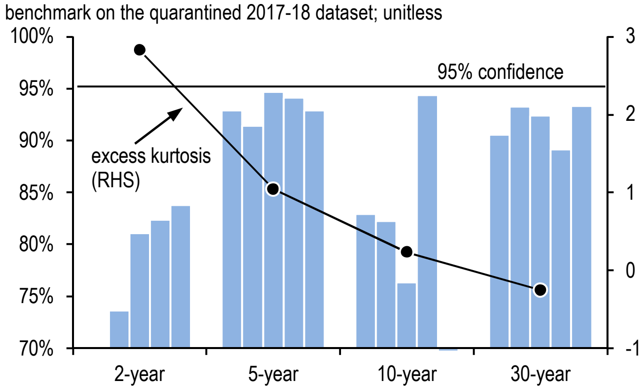


All predictors cross-validated on daily trades from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit 1 for more details. For each predictor we pre-selected the 5 top candidates from each technique before setting it loose on the 'quarantined' 2017-18 data. Absolutely no information from 2017-18 was used while training and vetting these predictors.

Source: J.P. Morgan

Exhibit 11: … and our Monte Carlo framework suggests the out-performance of all-long is unlikely to have occurred by random chance, though we fail to reject the null hypothesis at the 95% level

Statistical confidence that our candidate GBM learners outperformed the passive benchmark on the quarantined 2017-18 dataset; unitless



Source: J.P. Morgan

For the quarantined data from 2017-present, we took our five top-performing classifiers from both balanced GBM and balanced random forest, trained them over the full cross-validation time period (2008-16) and then set them loose from 2017-present, neither re-training the classifiers nor fiddling with design decisions as returns began to roll in. Armed with these results, we then compared each ML strategy's performance against our benchmark of daily uniform buying. **Exhibit 10** summarizes the results of this exercise for both balanced GBM and RF techniques, across our four benchmark bond tenors. For this nearly 2-year period, we found GBM and RF outperformed the benchmark, by a varying margin, across tenors, with outperformance optically most robust at the front end and narrowest in 30Y Bonds. The past year saw a strong selloff in rates at the front end and into intermediates, with uniform buying of 2-year notes for a 1-week hold period producing a Sharpe of nearly -2.0, including transaction costs. **Our 2Y GBM and RF classifiers were unable to weather this storm with positive returns**—not at all surprising to us given their unfortunate bias towards buying. In total, the classifiers

[2] We incorporate transaction costs using BrokerTec interdealer electronic Treasury trading data for each tenor, taking the time-weighted average width of the top queue position during the New York trading session, averaged for each calendar year. In practice this is roughly 0.4 bp running for 2s, 0.2-0.3 bp for 5s and 10s, and 0.1-0.2 bp for 30s (with 30s in particular biased lower over the past few years).

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.com

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

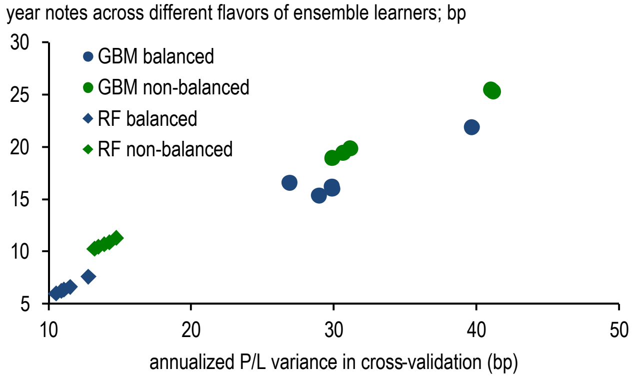
tended to sell only 30-35% of days. That said, **they still outperformed the passive strategy by a wide margin, producing a Sharpe of roughly -0.8 with transaction costs**. Further out the curve, GBM managed to simply tread water in 5- and 10-year notes, posting neither large losses nor gains, while the passive strategy again took losses on average. In 30Y bonds, the classifiers achieved the highest headline Sharpe, of nearly 0.5, though the passive approach here was hot on their heels, as 30Y bonds have rallied, on net, over the full period. **Narrowly outperforming passive longs—and even taking mild losses—was not unheard of behavior during the cross-validation period (see Exhibit 10 of *Androids*), and we don't view this recent rout as reason alone to scrap such an experiment**.

Exhibit 11 shows the result of this exercise, where we find confidence in our classifiers' outperformance ranges from roughly 70% in the 2Y and 10Y sectors (depending on which top five classifier you look at) to *nearly* 95% for the 5Y note. In the scientific community a confidence of 95% is seen as the classic standard beyond which you "reject your null hypothesis"—i.e. at such a threshold you attain confidence that the signal is almost certainly not a fluke of noisy data, and that a real effect is at play. **While recent performance has not afforded us such luxuries, we nonetheless believe the jury is still out and that a much longer time series will be needed to conclude whether or not our simple ML approach adds value**.

One point of detail on the above figure: **It may seem odd that in the 2Y sector confidence in outperformance was quite low, given the 2Y classifiers outperformed passive longs by the widest margin, even on a risk-adjusted basis**. However, our Monte Carlo framework penalized 2s more heavily since the 2Y sector has markedly higher kurtosis (fat-tailed returns) than further out the curve (Exhibit 11, RHS axis), a situation consistent with performance in cross-validation, but here the penalty is much more severe given the substantially smaller number of trading days (i.e., smaller sample size).

Exhibit 12: Learners that exhibited more volatile returns in cross-validation continued to demonstrate comparatively higher vol in quarantine, with GBM returns and vol elevated compared to RF...

Annualized variance of returns in quarantine compared to in cross-validation for 10-year notes across different flavors of ensemble learners; bp

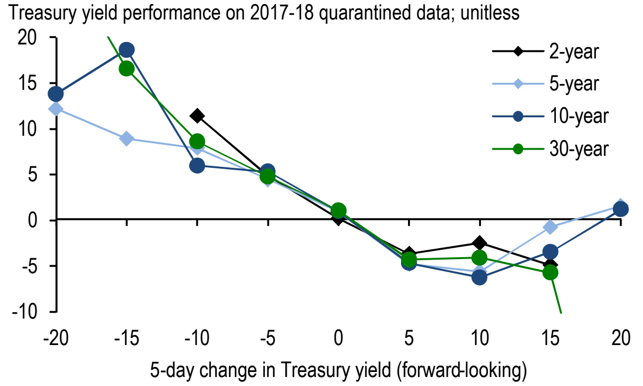


All predictors cross-validated on daily trades from mid-2008-16; trades sized with the Kelly Criterion, see notes of Exhibit GB1 for more details. For each predictor we pre-selected the 5 top candidates from each technique before setting it loose on the 'quarantined' 2017-18 data. Absolutely no information from 2017-18 was used while training and vetting these predictors.

Source: J.P. Morgan

Exhibit 13: ... and classifiers continued to demonstrate a fortuitous asymmetry in quarantine between strong returns in rallies and subtler losses in selloffs

Annualized Sharpe ratio across ML classifiers broken out by subsequently realized Treasury yield performance on 2017-18 quarantined data; unitless



Source: J.P. Morgan

In cross-validation, we found the intriguing result that while risk-adjusted returns (e.g., Sharpe ratios) were quite comparable between the best RF and GBM classifiers, with GBM incrementally outperforming, average return and variance differed dramatically between the two, with RF classifiers showing much subtler returns and lower variance than GBM, which in contrast spanned quite a broad range of behaviors. **One important question is if this behavior between techniques and among classifiers within the same technique is consistent over time**. To judge this, we looked at the relationship between variance in cross-validation and quarantine, shown in **Exhibit 12**, where we find relative variance remains quite steady between classifiers over the two periods. This suggests an important application of these ML trading schemes could potentially be converting/tailoring the variability and average returns of a particular asset or index to an investor's risk tolerance, with GBM providing a much broader dynamic range of outcomes to choose from.

Finally, we verify whether or not another attractive feature of the ML schemes holds between cross-validation and quarantine: namely, the algorithms' ability to avoid large losses in selloffs, while enjoying strong returns in rallies. **Exhibit 13** confirms that this is indeed the case, with a well-behaved, asymmetric returns profile versus yield performance across tenors, the exception being some sharp drawdowns in the 30Y sector. While this behavior is somewhat encouraging, it is still far from ideal—ideally machine learning would deliver a trading strategy that made money, on average, in *both* rallies and selloffs, immune to the directionality of the market. Indeed, the admittedly tepid headline performance of our trading schemes in recent months is fundamentally a symptom of this shortcoming: over the past year, across the curve, rates have sold off; and any trading scheme that wishes to weather such situations with robust performance ought to have a buy/sell ratio that is not overtly biased towards buying bonds. We explore possible remedies to this problem in subsequent work.

Munier Salem
(1-212) 270-0317
munier.salem@jpmorgan.com

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

Jimmy Huang
(1-212) 834-2096
guanjie.huang@jpmchase.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

Private Bank Clients: Where you are a client of the private banking businesses offered by JPMorgan Chase & Co. and its subsidiaries ("J.P. Morgan Private Bank"), research is issued to you by J.P. Morgan Private Bank and not by any other division of J.P. Morgan, including but not limited to the J.P. Morgan corporate and investment bank and its research division.

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) 099/04/2018 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 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. J.P. Morgan’s research coverage universe spans listed securities across the ASX All Ordinaries index, securities listed on offshore markets, unlisted issuers and investment products which Research management deem to be relevant to the investor base from time to time. J.P. Morgan seeks to cover companies of relevance to the domestic and international investor base across all GIC sectors as well as across a range of market capitalisation sizes. **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**: Research relating to equity securities is issued and distributed in Taiwan by J.P. Morgan Securities (Taiwan) Limited, subject to the license scope and the applicable laws and the regulations in Taiwan. 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.

"Other Disclosures" last revised June 30, 2018.