Cross Asset Strategy

Feature exploration via Boruta and Principal Components Analysis

- **One basic question we are trying to ask is: 'What's driving the market now?' Given the vast number of factors, we turn to machine learning techniques for help.**
- **We make a number of unconventional modifications to canned models (random forest, Boruta, principal components) to produce a custom feature selection process with a broad variety of cross asset applications.**
- **We illustrate the effectiveness of this process in a number of ways: simulation, a security replication exercise, and a cross asset tactical study. Generally, the results point to improved security selection and dynamic risk budgeting, with less turnover.**

We thank Kumar Vemuri and Ian Johnston of the APAC Applied AI & Machine Learning team for their critical contributions to this research note.

Global Cross Asset Strategy

Thomas Salopek [AC]
(1-212) 834-5476
thomas.salopek@jpmorgan.com
J.P. Morgan Securities LLC

Federico Manicardi [AC]
(44-20) 7742-7008
federico.manicardi@jpmorgan.com
J.P. Morgan Securities plc

See page 28 for analyst certification and important disclosures.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Abstract

In statistical applications for finance, we often want an equation that explains one variable as a function of other variables. An example is wanting an equation that tells us how to replicate the returns of an index using a weighted combination of other assets or factors. Although machine learning techniques such as random forest, gradient boosting trees, neural network, etc., can learn more complex mapping of input variables to an output variable, a simple equation lets us not only measure the relative importance of each input variable but also easily describe their relationships with the output variable.

The standard approach to obtain this equation is to run a multiple regression analysis, but this relies on assumptions that are not valid in all use cases and is not always robust due to multicollinearity. We propose a pipeline that combines ideas from machine learning and statistics to help practitioners enhance their multiple regression analyses and obtain robust equations after an initial round of feature selection. We demonstrate the utility of the pipeline in a simulation study and then apply it in a handful of relevant case studies.

Feature Exploration Pipeline

Let $\boldsymbol{X}$ be an n-by-p matrix of input data for n samples and p variables and $\boldsymbol{y}$ be an n-by-1 vector of output data. We want an equation that explains $\boldsymbol{y}$ as a linear combination of the columns of $\boldsymbol{X}$. Multiple regression analysis sets up this problem as $y = x_1B_1 + x_2B_2 + \ldots + x_pB_p +$ error where the error is assumed to be normally distributed with mean zero and constant variance, and the weights $\{B_1, B_2, \ldots, B_p\}$ are unknown values to be estimated.

The well-known solution to this problem is $\boldsymbol{B} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$, but the matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ is numerically unstable or undefined if $\boldsymbol{X}$ suffers from multicollinearity or other issues, e.g., if $p > n$ *(i.e., the number of variables is greater than the number of samples)*. In these cases the estimates of $\boldsymbol{B}$ are at best highly volatile and at worst impossible to calculate, so we must take action to address these issues.

To help enhance the standard analysis, our pipeline thus consists of two steps:

1) Data-driven feature selection using a custom model within the Boruta algorithm [1]

2) Principal component regression to robustly explain $\boldsymbol{y}$ using the refined set of features

Conducting a round of feature selection before fitting the model and leveraging principal component analysis to de-correlate the features are popular ideas that can be helpful in general because we want our equation to be parsimonious and to include only relevant features.

Feature Selection

Consider the, arguably, simplest and fastest way to do feature selection by fitting a simple linear regression model for each feature independently of the others and then retaining the features with statistically significantly results. This will not detect non-

linear effects or interaction effects, but it will detect significant main effects, even weak ones, with a sufficiently large sample size.

However, to apply this idea the user must decide on an initial threshold of statistical significance, e.g., the standard value of 5%, and then correct for the fact that conducting multiple hypothesis tests on the same underlying dataset will lead to an increased probability of a false positive.

It is common for feature selection algorithms like this one to rely on a threshold or hyper-parameter to distinguish between important and unimportant features. Different thresholds may be appropriate for different use cases, and so ideally we want a way to pick a suitable threshold in a data-driven way.

One such idea exists in a feature selection algorithm called Boruta [1]. This algorithm first augments $\boldsymbol{X}$ with a set of noisy features that by design have no relationship with $\boldsymbol{y}$ and then iteratively updates a list of selected features by fitting a model, calculating feature importance values, and using the most important noisy feature to derive a threshold for feature selection.

The original algorithm suggests using a random forest for this process because "the random forest … is relatively quick, can usually be run without tuning of parameters, and it gives a numerical estimate of the feature importance." However, there will be cases where an untuned random forest is not able to find important features, e.g., if the trees are not allowed to grow sufficiently deep, then they could miss features that only have weak effects on the target variable.

We propose a hybrid approach that uses Boruta but replaces the random forest model with a bootstrap version of the simple linear regression model where the median squared test statistic of the slope parameter for each feature is used as the measure of importance. This approach combines the best of both ideas and allows us to quickly refine the initial set of features into a filtered set that have been determined to be important and relevant for predicting $\boldsymbol{y}$ without having to manually specify a threshold of statistical significance.

In addition to the model used to calculate feature importance values, Boruta has other degrees of freedom that researchers might consider, but we believe that the default idea of using the most important noisy feature as a benchmark for feature selection is reasonable. We confirmed that this idea is highly effective in a simulation study that is detailed in a later section.

Principal component regression

A standard way to de-correlate the features in a dataset is to apply principal component analysis to obtain a set of linear combinations of the original features that are orthogonal to each other. This means that we obtain a new matrix $\boldsymbol{P} = \boldsymbol{XL}$ where $\boldsymbol{L}$ is a p-by-c matrix of loadings or weights of each original feature onto each of c components. If we then use this new matrix in a regression model, we obtain a new set of weights $\boldsymbol{W}$ that explain $\boldsymbol{y}$ in terms of $\boldsymbol{P}$, i.e., $\boldsymbol{y} = \boldsymbol{PW}$, where $\boldsymbol{W}$ is a c-by-1 vector of parameters. We can rewrite this in terms of the original features as $\boldsymbol{y} = \boldsymbol{XLW}$ and note that $\boldsymbol{LW} = \boldsymbol{B}_{PCA}$ is a p-by-1 vector of weights for the columns of $\boldsymbol{X}$.

In addition to de-correlating the features, the PCA transformation has the added benefit of helping to de-noise the feature matrix. The first principal component is the projection of $\boldsymbol{X}$ with the highest variance, and each subsequent component is the projection of $\boldsymbol{X}$ with the highest variance that is also orthogonal to all prior projections.

It is usually the case that we can explain most of the variation in $\boldsymbol{X}$ with a number of components $c < p$, and by doing so we can help the model focus more on the salient signals in $\boldsymbol{X}$. This is akin to regularization where the number of components helps control for the complexity of the model with larger values allowing for more features in $\boldsymbol{X}$ to contribute to the column space of $\boldsymbol{P}$.

We propose exploring the full "path" of regression models for each possible value of c and using the results to decide on a parsimonious and robust set of weights to explain $\boldsymbol{y}$ as a function of $\boldsymbol{X}$. Each value of c leads to a different $\boldsymbol{L}$ that explains some amount of variation in $\boldsymbol{X}$, results in some set of weights $\boldsymbol{B}_{PCA}$ and some coefficient of determination R^2 in the regression model for $\boldsymbol{y}$. Based on the progression of these values for increasing values of c, the user can decide on an appropriate choice of c, e.g., by looking for a point of diminishing returns in R^2.

The default ordering of the principal components of $\boldsymbol{X}$ is the one such that the incremental amount of variance in $\boldsymbol{X}$ explained by each component is maximized and each component is orthogonal to all previous components. This ordering does not take $\boldsymbol{y}$ into account, and so it may be the case that a feature is important in predicting $\boldsymbol{y}$ but does not explain so much of the variation in $\boldsymbol{X}$ and thus primarily contributes to later components in the default ordering.

This is noteworthy because the ordering of the components will affect the "path" of each feature's coefficient, and so in our pipeline we also allow for the reordering of the principal components of $\boldsymbol{X}$ in descending order of the strength of their correlation to $\boldsymbol{y}$. We demonstrate the effect of this optional reordering in a case study that is detailed in a later section, and refer the interested reader to other recent work that explores a similar idea [2].

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Simulation Study

To investigate the effectiveness of our feature selection component, we simulated data for $\boldsymbol{X}$, $\boldsymbol{B}$ and $\boldsymbol{y}$ in a number of different scenarios, ran a set of feature selection algorithms and compared the performance of their results.

For each value of n in {100, 1000, 5000, 10000}, and each value of p in {10, 50, 100} (where $p < n$), we simulated $\boldsymbol{X}$ by drawing n samples from a p-dimensional multivariate normal distribution with a zero mean vector and an identity matrix as the covariance matrix.

We simulated initial values of $\boldsymbol{B}$ by drawing p values independently from a standard normal distribution. To evaluate the different feature selection models we then forcibly set a proportion of the elements of $\boldsymbol{B}$ to zero so that we had a clear set of important and unimportant features.

We considered proportions of r in {0.1, 0.5, 0.9}, and in each case to decide which elements of $\boldsymbol{B}$ to set to zero, we simulated a vector $\boldsymbol{T}$ of p independent random variables from a Bernoulli distribution with probability of success equal to r, i.e., we randomly generated p values that are equal to one with probability r and zero with probability 1-r.

We then revised the initial values of $\boldsymbol{B}$ by taking the element-wise multiplication with $\boldsymbol{T}$ so that on average pr of the elements of $\boldsymbol{B}$ became zero.We required that at least one element of $\boldsymbol{B}$ be non-zero during this process. A perfect feature selection model would thus select only the features that have corresponding elements of $\boldsymbol{T}$ equal to one.

Finally, for each value of s in {1, 5, 10}, we simulated $\boldsymbol{y}$ by adding n independent and normally distributed errors with zero mean and standard deviation equal to s to the vector of $\boldsymbol{XB}$. For each scenario we simulated 10 replicates of $\boldsymbol{X}$, $\boldsymbol{B}$ and $\boldsymbol{y}$ and then fit a set of models to each dataset and compared the list of selected features to the ground truth and calculated precision, recall, and the F-score, which is the harmonic mean of precision and recall.

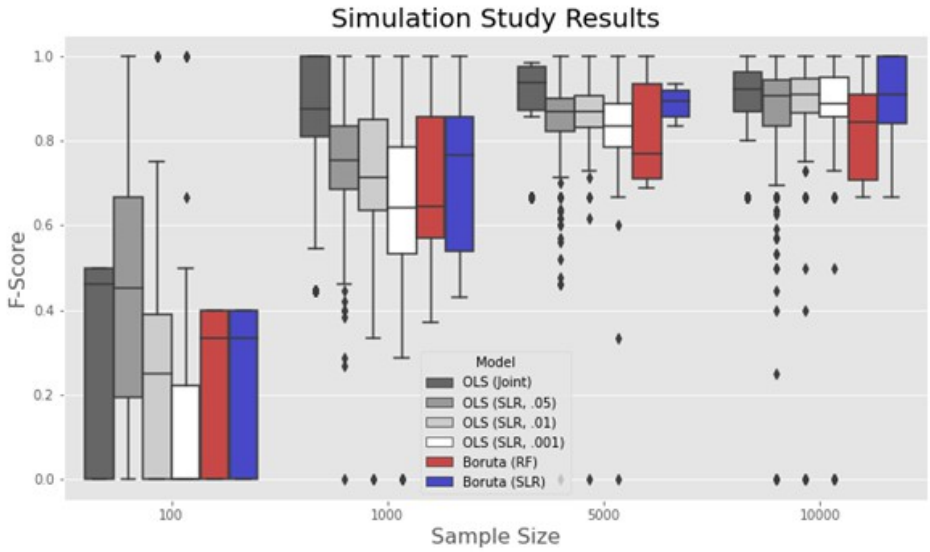
Figure 1 summarizes the results of this simulation study for the different values of n.

Each model is described as follows:

- OLS (Joint): Fit a multiple regression model using all features and select features with p-values less than .05
- OLS (SLR, *threshold*): Fit a simple linear regression model for each feature independently and select those with p-values less than *threshold*
- Boruta (RF): Apply the standard Boruta feature selection algorithm with an untuned random forest regressor model that has the following hyper-parameters in scikit-learn [3]:
  {“n_jobs” = 1, “max_depth” = 5, “min_samples_leaf” = 5, “max_leaf_nodes” = 50}
- Boruta (SLR): Apply the Boruta feature selection algorithm with a custom model that instead applies the bootstrap algorithm to fit a simple linear regression model for each feature independently a certain number of times and uses the median squared value of each feature’s test statistic as the measure of importance

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 1: Model performance by size for 'kitchen sink' vs 'p-hacking' vs Boruta



Source: J.P. Morgan

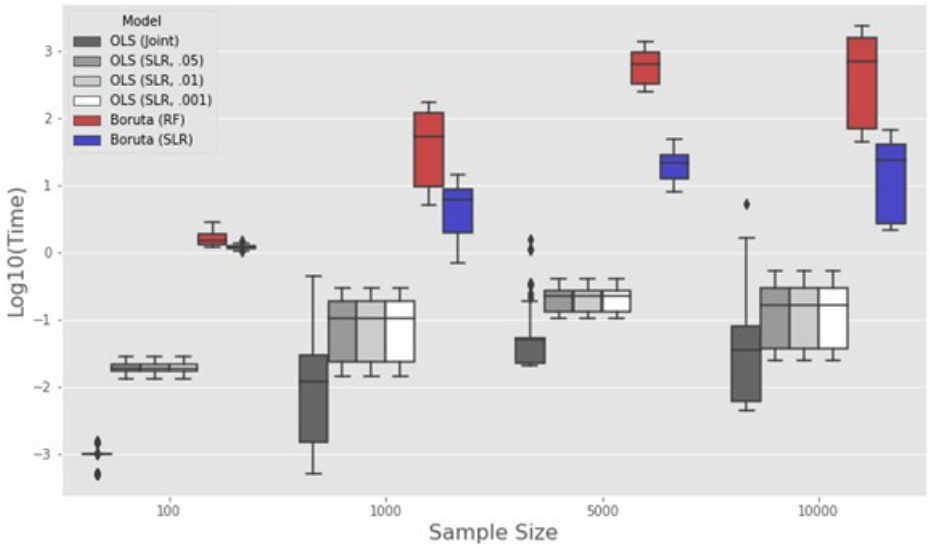
For relatively smaller sample sizes, we see that there is generally a degradation in performance as we move from the OLS (Joint) model to the OLS (SLR, .001) model. The OLS (Joint) model is the actual underlying model that was used to generate the simulated dataset, so it should intuitively consistently have relatively strong performance. The degradation in performance as we apply varying thresholds to the simple linear regression models for each feature independently reveals how the simple idea breaks down, e.g., false positives for thresholds that are too large and false negatives for thresholds that are too small. These models generally stabilize for larger sample sizes, and we see that the distribution of F-scores for the simple linear regression models become less volatile but have more extreme values than that of the OLS (Joint) model.

In terms of the Boruta models, we see that the untuned random forest model and the bootstrap simple linear regression model are comparable for relatively smaller sample sizes with performance that is typically somewhere between the OLS SLR ideas and the OLS (Joint) model. As the sample size increases, we see that the Boruta model that utilizes the bootstrap simple linear regression model has a higher median F-score with smaller variation in its distribution. This is likely due to the aforementioned situation for which an untuned random forest model struggles to select features that have relatively weak effects on the target variable.

A more thorough study might reveal other pros and cons of these models in the presence of more realistic datasets and other issues such as multicollinearity, and it also may be the case that other models including a tuned random forest could achieve greater performance. However, our intention at this point is simply to demonstrate the value of substituting an untuned random forest model with a simpler idea that not only does not require any tuning but can also achieve similar or greater performance across a variety of scenarios.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 2: Time to run each model



Source: J.P. Morgan

Figure 2 shows the distribution of time spent in seconds on the log scale for each model. The simulation study was conducted on a machine with an Intel Xeon Processor E3-1585L v5 (8M Cache, 3.00 GHz) and only a single job was used to fit the random forest models. The OLS SLR models have the same distribution because these differ only in terms of the threshold that is used on the p-values to select features. As the sample size grows we see a clear separation in the time required to run the Boruta feature selection algorithms with significantly more time needed for the ones that use random forest models.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Case Studies

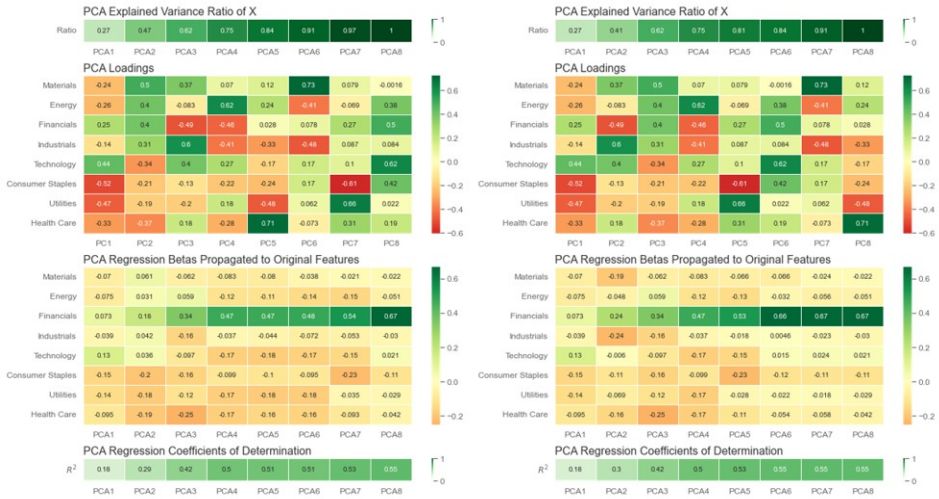
1) Building an equation to explain and replicate the returns of a single stock (sample US bank stock)

We test our pipeline on real financial time series data by applying it on the available data in Bloomberg between 2000-01-01 and 2016-12-31 to build an equation that explains the daily excess returns of the sample US bank stock relative to the SPX Index in terms of the excess returns of a collection of sector indices and a random noise variable that has no relationship with the target variable. Boruta removes some of the features including the noise term, and then the PCA Regression reveals that as we increase the number of components, we see that (as we might expect) the Financials sector index gradually gains in importance until it is the dominant factor in a full model with all components.

We consider the default ordering of the principal components as well as the descending order in terms of the magnitude of correlation to ***y***. Figure 3 shows a side-by-side comparison of the results of this analysis.

Figure 3: Benefit of re-ordering the Principal Components

Left: default ordering of PCs, **right:** rearranged PCs based on strength of correlation to ***y***



Source: J.P. Morgan, Bloomberg Finance L.P.

Each heat map in Figure 3 has a number of columns equal to the total number of possible components, 8, which is also the number of features that were selected by Boruta. From the top to the bottom, the first shows the percent of variation of ***X*** explained for each number of components, and then the second shows the loadings of each component. The third shows the PCA Regression betas propagated back to the original features, and the final one shows the R^2 values for each possible model.

In this case we see that the terminal values of the coefficients for each feature, i.e., the last column of the third heat map in each chart, are the same, but the "paths" to reach these values are slightly different. By rearranging the principal components we attempt to prioritize explaining the variation in ***y*** instead of the variation in ***X*** as we increase the

number of components.

The R^2 values for the model with no rearrangement plateaus with 4 components, and so we consider stopping there and using the resulting weights to explain $\boldsymbol{y}$ in terms of $\boldsymbol{X}$. For the models that use rearranged components, we reach a plateau of around 6 components instead. We arrived at these values programmatically by searching for the first point at which the series of R^2 values, scaled by the final value so that the final value is always 1, does not increase by more than .01.

If we were to simply fit an OLS model to all data, we would obtain the coefficients shown in the final column of the third heat map. It is noteworthy that in some places these are quite different from the ones shown in the column for the model with 4 components and the default ordering of components. In both cases by adding more components, we increase the magnitude of the weight on the Financials and decrease the magnitude of the weight on some other features including the Technology and the Utilities Indices.

To test the quality of the equation and the effect of using either model as opposed to the full set of features, we first convert each set of weights into normalized versions that sum to one in magnitude and then take corresponding long and short positions in the underlying indices from 2018-01-01 to 2023-03-03 in an attempt to recreate the return stream of a sample bank stock.

Note: While it may not be possible to invest in many indices directly, one can realize similar returns by investing in Index mutual funds and ETFs.

We compare the resulting portfolios with each other and with a benchmark strategy wherein we simply buy and hold the sample US bank stock in the same time period. This is a simplified example for which we assume a constant allocation to each asset over time, i.e., a daily rebalancing, and do not account for transaction costs and so on. In practice, portfolio managers could potentially re-compute weights using the most up-to-date historical return data for all constituents of interest for use during the periodic, e.g., monthly, rebalancing of the passively managed portfolio. We obtain the following results for a subset of metrics of interest:

Table 1: PCA Performance results

		Relative to Benchmark (except Turnover)		
Metric	Benchmark	PCA (Default)	PCA (Rearranged)	PCA (All)
Sharpe	1.00	1.15	1.15	1.15
Max Drawdown	1.00	0.82	0.93	0.91
Volatility	1.00	0.75	0.81	0.81
Sortino	1.00	1.08	1.08	1.08
CAGR	1.00	1.14	1.15	1.17
R (Correlation)	1.00	0.83	0.9	0.89
R^2	1.00	0.69	0.8	0.9
Turnover		1.66	1.01	1.00

Source: J.P. Morgan

Note that the turnover is calculated for each strategy independently as the total number of dollars bought or sold (divided by 2) over time and then scaled to be relative to the PCA (All) strategy. All other metrics are shown relative to the benchmark buy and hold

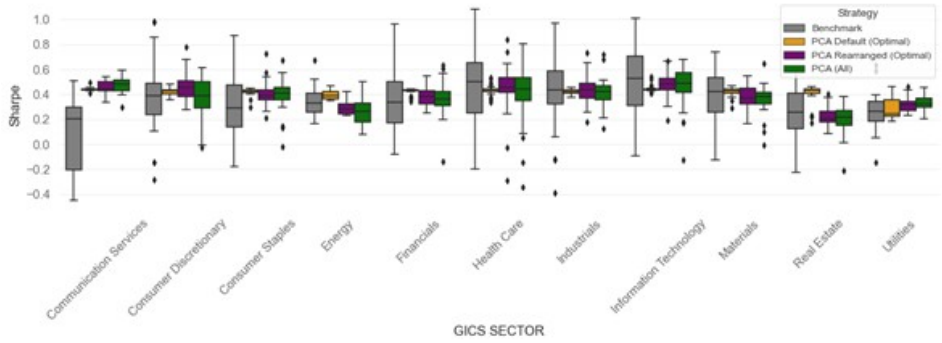
Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

strategy.

The PCA strategies show modest improvements to these metrics, but the PCA (Default) strategy is slightly better in terms of increasing Sharpe and Sortino ratios and minimizing volatility and max drawdown. However, the effect of rearranging the PCs in this case is that the selected model produces a series of daily returns with higher correlation to the actual values (0.90 compared with 0.83) with much smaller turnover.

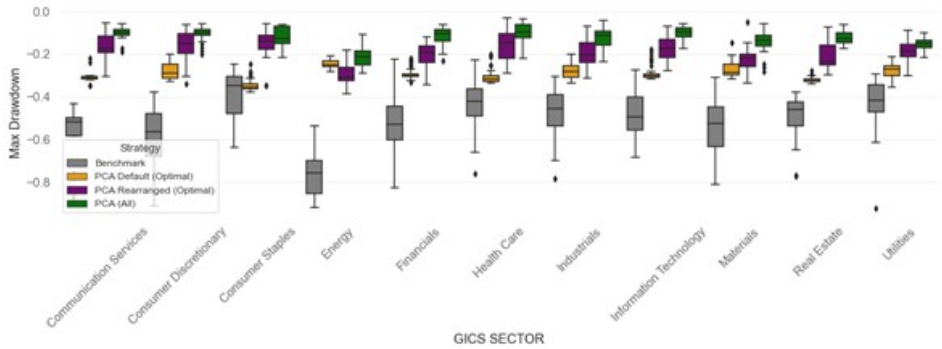
To help provide more context to these numbers and to understand if and how these findings generalize to other stocks, we replicated this kind of analysis for 377 members of the SPX Index as of 2023-03-03 with sufficient data coverage. Figure 4 through Figure 11 show the results of this study summarized by GICS sector. For this broader analysis we used a set of 60 indices that capture a variety of factors of interest in addition to the sector indices from above. We show 4 sets of results: the benchmark model, the optimal PCA model for each type of PC ordering, and the PCA (All) model that is consistent across ordering types.

Figure 4: Sharpe ratio



Source: J.P. Morgan, Bloomberg Finance L.P.

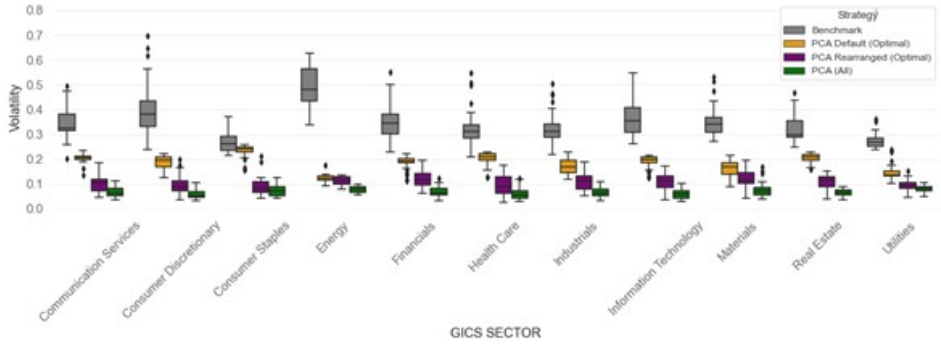
Figure 5: Max Drawdown



Source: J.P. Morgan, Bloomberg Finance L.P.

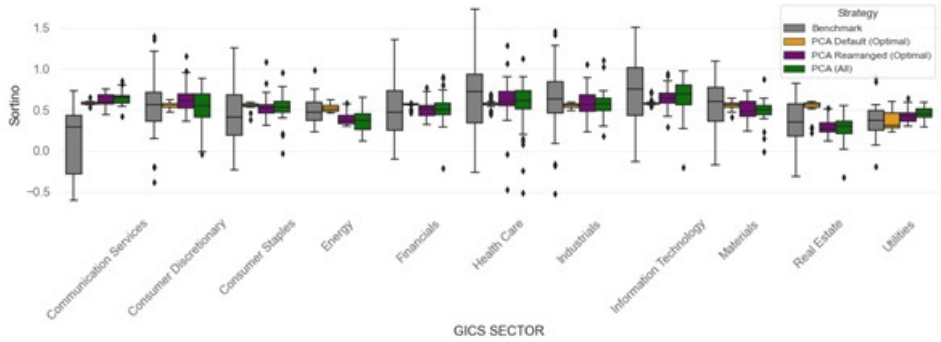
Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 6: Volatility



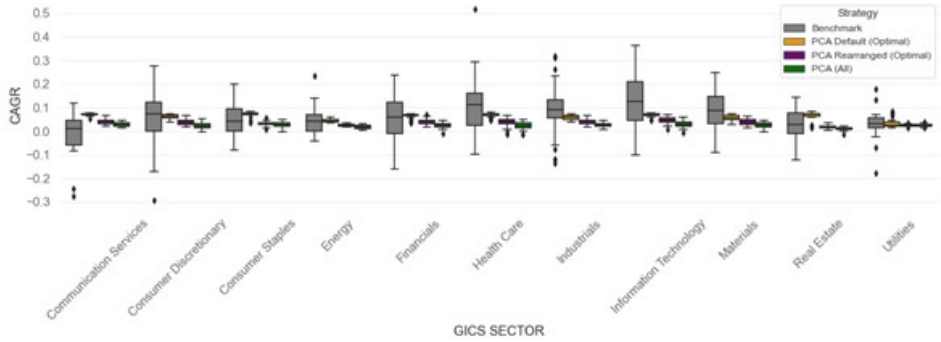
Source: J.P. Morgan, Bloomberg Finance L.P.

Figure 7: Sortino



Source: J.P. Morgan, Bloomberg Finance L.P.

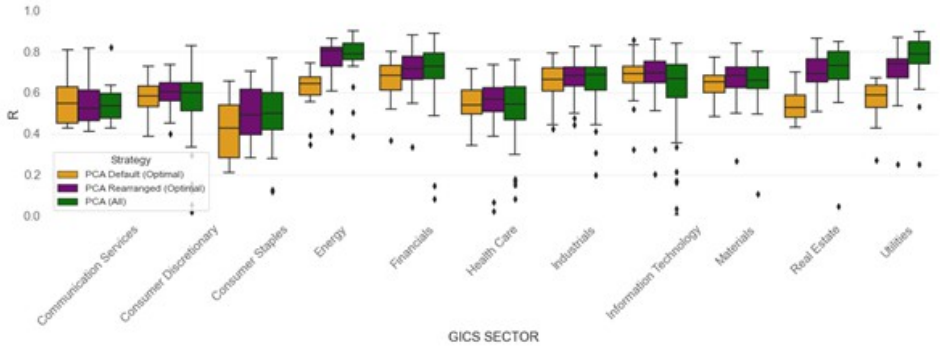
Figure 8: CAGR



Source: J.P. Morgan, Bloomberg Finance L.P.

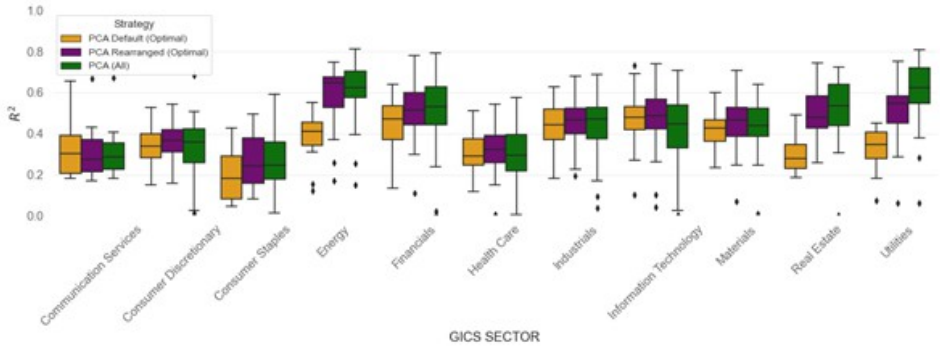
Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 9: R



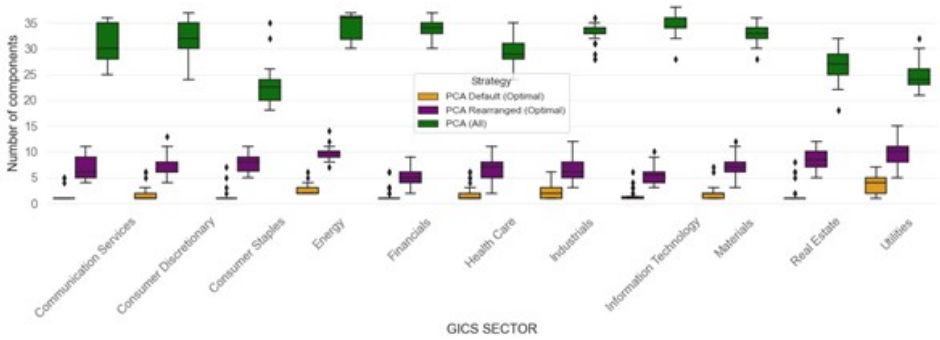
Source: J.P. Morgan, Bloomberg Finance L.P.

Figure 10: R-squared



Source: J.P. Morgan, Bloomberg Finance L.P.

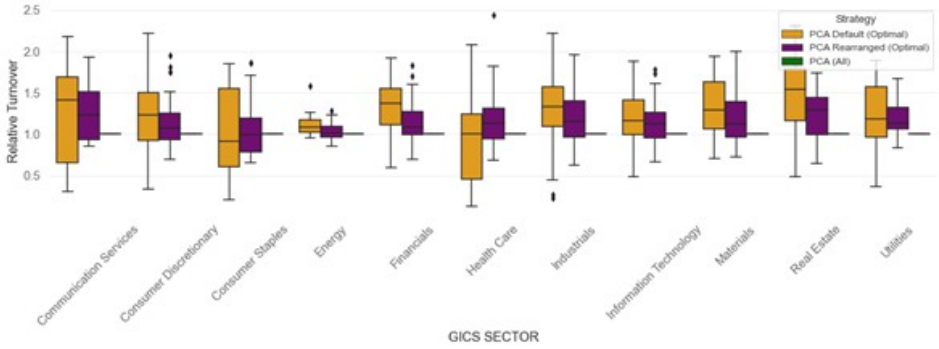
Figure 11: Number of Components



Source: J.P. Morgan, Bloomberg Finance L.P.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 12: Relative turnover



Source: J.P. Morgan, Bloomberg Finance L.P.

Note that Figure 12 shows the turnover relative to the PCA (All) strategy.

The results confirm that the observed patterns in the sample US bank stock case study are consistent across a variety of stocks. The PCA Regression–based replication portfolios generally achieve the median Sharpe ratio for each sector or greater with uniformly lower volatility and reduced max drawdowns. Moreover, there is not much difference between the distributions of the simpler PCA models with a subset of rearranged components than the PCA models with all components; however, for some metrics there is a significant difference between these models and the PCA models that use a subset of the components in their default ordering.

We see in Figure 11 that Boruta typically filters the initial set of 60 indices down to around 30 or so, i.e., the number of components in the PCA (All) models. From there if we use the default ordering of the principal components along with the stopping rule mentioned previously, then we usually pick less than 5 components for the model. However, if we rearrange the principal components in descending order of the strength of their correlation with $\boldsymbol{y}$ and use the same stopping rule, then we instead pick between 5 and 10 components. The PCA Rearranged models thus strike a good compromise between the simple PCA Default models and the complex PCA (All) models, and this is reflected in various metrics, e.g., volatility, max drawdown, etc.

The variance in the distribution of each metric for the PCA-based models is typically small compared with the benchmark portfolios. One reason for this is that the unexplained idiosyncratic component of the variation in each stock is usually somewhat large when building models only based on these indices. Figures 9 and 10 confirm this by showing the relatively modest values of the correlation coefficient and the coefficient of determination between each PCA model's return streams and the benchmark strategy of simply buying and holding each asset.

Figure 12 shows that the turnover for the models that use a subset of the components is generally higher than the models that use all principal components and moreover that the models that use the default ordering of components tend to have the highest turnover. We discuss some possible reasons for this in a later section.

In practice there are ways to try to improve the accuracy of this kind of replication, e.g., use a larger set of features based on user preference and investment interest and/or change the input dataset so that it includes more recent information.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

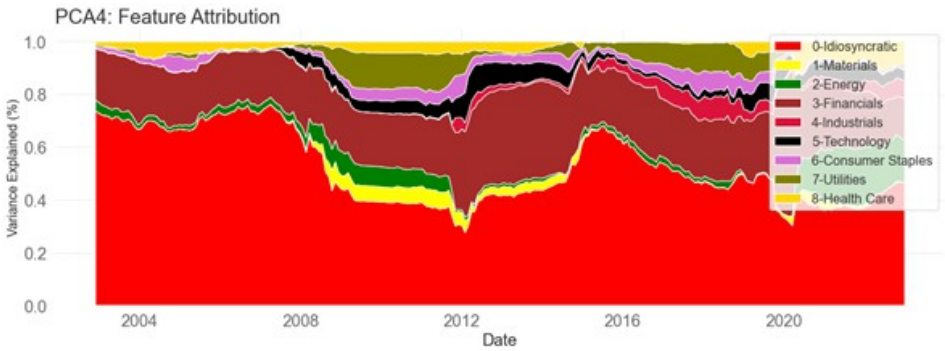
2) Building multiple equations to reveal how relationships to features evolve over time

By applying the pipeline to a rolling window of data, we can also explore how the importance of each feature in predicting the target variable changes over time. We use the squared value of the beta for each feature from the regression model as a proxy for its importance and then calculate the relative importance as the ratio of importance to the overall sum of importance values. We then attribute a portion of R^2 to each feature by scaling R^2 by the relative importance values, and finally take $1 - R^2$ to be the idiosyncratic variation in $\boldsymbol{y}$ that is unexplained by the model.

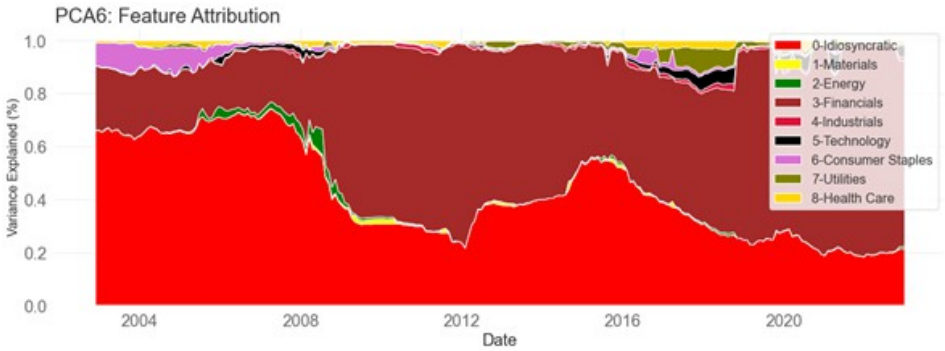
In the case of sample US bank stock from the earlier simplified example, with a rolling window of three years, we can see in Figure 13 that an R^2 value of 0.5 or greater when using $c = 4$ components in the default ordering is generally only achieved during extreme periods including financial crises as well as the COVID pandemic. These represent periods in which many assets become strongly correlated such that their movements are well captured by the top principal components. Higher R^2 values are generally achieved by the model that uses $c = 6$ components in the rearranged ordering, and we see that the Financials sector contributes the most to the model's explanatory power. This result is very similar to the model that uses all possible components as well.

Figure 13: Optimal model for a given number of components

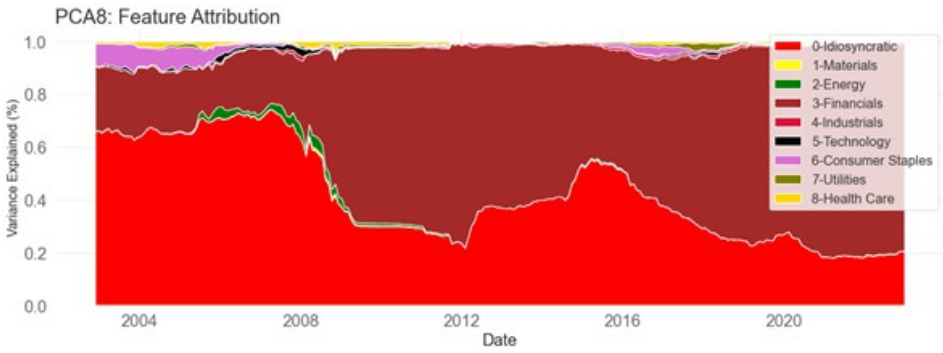
Optimal model with default PC ordering (4 components)



Optimal model with rearranged PCs (6 components)



Model that uses all principal components (8 components)



Source: J.P. Morgan, Bloomberg Finance L.P.

3) *Q-Boruta*

We present another application of the Boruta algorithm we call Q-Boruta, for quant research. The feature selection part of our pipeline replaces Boruta's default model with a custom one to try not to miss important features with relatively weak effects on the response variable. With this, we obtain an equation that explains one variable in terms of other variables, but some practitioners may instead be interested in selecting a set of assets for which a given trading strategy works well.

To this end, we replace Boruta's default model with one that no longer tries to predict $\boldsymbol{y}$ in terms of $\boldsymbol{X}$ and calculate feature importance values, but that ignores $\boldsymbol{y}$ and simply applies a trading strategy to each feature in $\boldsymbol{X}$ and computes a corresponding performance metric of interest like a Sharpe ratio.

We again use the bootstrap algorithm to build a distribution of these metrics by selecting many random contiguous subsets of time from the dataset, running the strategy on each one, and finally calculating the median Sharpe ratio for each asset. In this process we set negative Sharpe ratios to zero so that Boruta only ever operates on non-negative feature importance values.

Boruta then selects the set of assets in $\boldsymbol{X}$ for which the typical performance of the strategy is significantly greater than that of the best shadow asset. There are other ways to conduct back-tests to check the effectiveness and applicability of a strategy to a particular set of assets, so this may serve as a convenient complementary analysis to other techniques like using random portfolios to build a benchmark distribution of performance metrics.

To illustrate Q-Boruta we built a basic momentum strategy as a custom model that calculates the 12-month momentum for an asset, i.e., the percent change in price from 1 year prior to now, and invests when that value is positive. The decision to invest on any given day is based on the prior day's available information, with an investment holding period of 1 year. As before, trading costs, etc., are ignored. We consider around 100 assets from two asset classes (Equities and FX) with over 15 years of available data. Boruta supplements these with an additional 100 shadow assets it uses during the selection process.

We focus on these asset classes for simplicity, but others can be used as well. Not all asset classes may be appropriate for a given strategy, e.g., Q-Boruta might determine that a momentum strategy works well for commodities (e.g., trained during a super-

cycle "up phase"), but later that strategy may fail (out of sample) if the market environment changes.

Training period (2004-2022)

We apply Q-Boruta to the period from roughly 2004 to 2022 and find 15 assets for which the 12-month momentum strategy works well. Figure 14 shows a comparison of the average cumulative profit and loss (PnL) paths for two simple portfolios that buy and hold an equal amount of the selected and unselected assets, respectively. Figure 15 shows the paths for each asset individually.

Figure 14: Average Cumulative PnL: Buy and Hold Strategy (training period)

Source: J.P. Morgan

Figure 15: Cumulative PnL: Buy and Hold Strategy (training period)

Source: J.P. Morgan

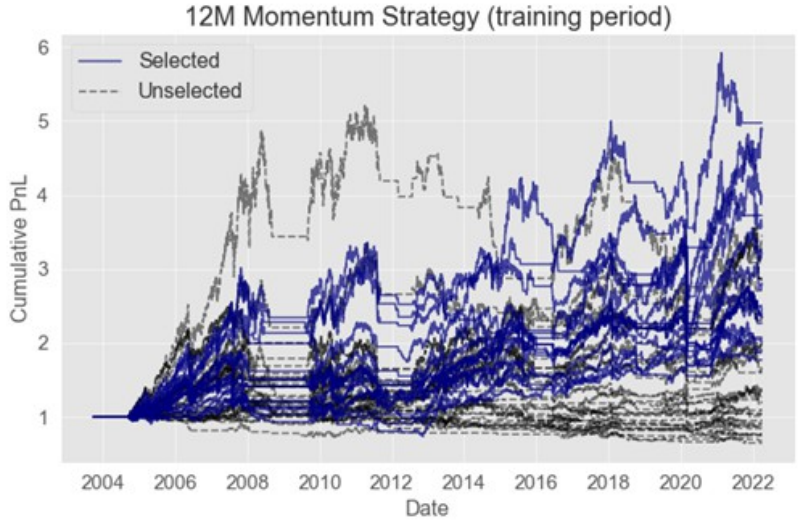
Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 16 and Figure 17 respectively show the average and individual cumulative profit and loss paths for the momentum strategy instead.

Figure 16: Average Cumulative PnL: 12M Momentum Strategy (training period)

Source: J.P. Morgan

Figure 17: Cumulative PnL: 12M Momentum Strategy (training period)



Source: J.P. Morgan

For both the buy and hold benchmark strategy and the 12-month momentum strategy, we see a widening gap between the performance of the portfolios for the selected versus unselected assets. The buy and hold benchmark achieves a higher cumulative PnL for the selected assets; however, the 12-month momentum strategy is able to avoid some large market drawdowns, e.g., around 2008 to 2010.

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

An index that tracks the Latin American market is an unselected asset with a relatively high cumulative PnL trajectory in the 2000s but then a stagnant and declining trajectory in the 2010s and beyond. This reflects the commodities boom and subsequent disappointing economic growth in Latin American countries around that time [4]. Q-Boruta avoids this asset and instead selects the ones for which the momentum strategy tends to work consistently well.

Hold-out period (2022-2023)

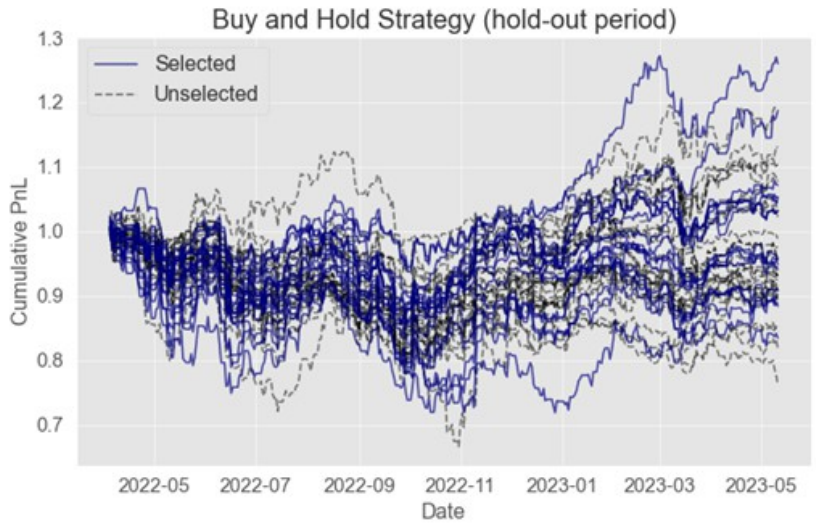
Figure 18 and Figure 19, respectively, show the average and individual cumulative PnL paths of the buy and hold strategy for each asset in the hold-out period (April 4th, 2022 to May 11th, 2023).

Figure 18: Average Cumulative PnL: Buy and Hold Strategy (hold-out period)

Source: J.P. Morgan

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 19: Cumulative PnL: Buy and Hold Strategy (hold-out period)



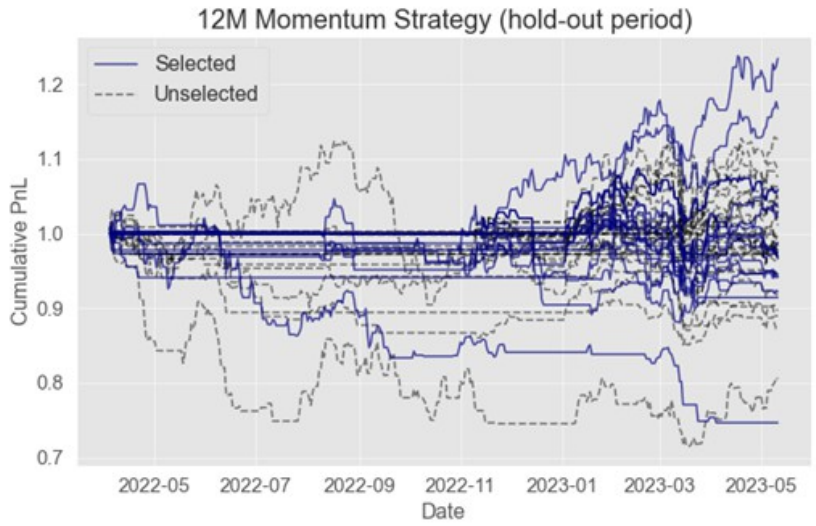
Source: J.P. Morgan

We see that the hold-out period has generally not been a good time to invest in these assets. The average cumulative PnL paths for both selected and unselected assets end slightly below 1. Figure 20 and Figure 21 show the average and individual cumulative PnL paths of the momentum strategy during this time period instead.

Figure 20: Average Cumulative PnL: 12M Momentum Strategy (hold-out period)

Source: J.P. Morgan

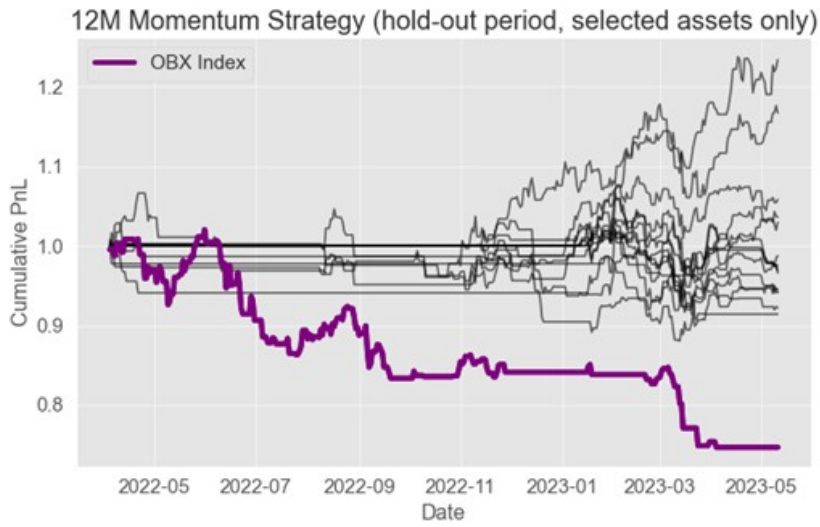
Figure 21: Cumulative PnL: 12M Momentum Strategy (hold-out period)



Source: J.P. Morgan

We see that both strategies perform similarly and somewhat poorly. Interestingly, there are even some small periods during which the unselected assets portfolio slightly outperforms the selected assets one. This is because a selected asset that tracks the Oslo Stock Exchange (OBX Index) significantly pulls down the average performance of the selected asset portfolio. Figure 22 highlights this asset in a view that shows the cumulative PnL paths for the selected assets only.

Figure 22: Cumulative PnL: 12M Momentum Strategy (hold-out period, selected assets only)

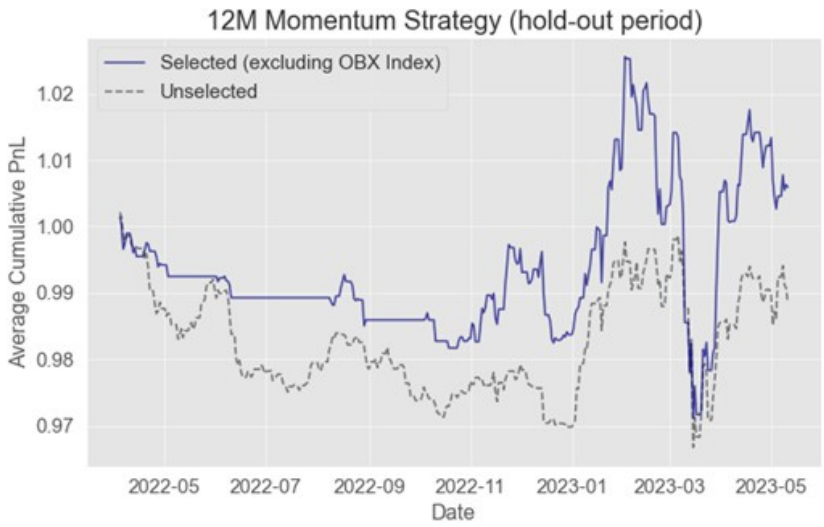


Source: J.P. Morgan

Figure 23 shows an updated view of the average cumulative PnL paths if we exclude OBX Index from the set of selected assets. Here we see that the revised selected assets portfolio consistently outperforms the unselected assets one. This revised selected assets

portfolio also ends with a slightly positive overall profit.

Figure 23: Average Cumulative PnL: 12M Momentum Strategy (hold-out period)



Source: J.P. Morgan

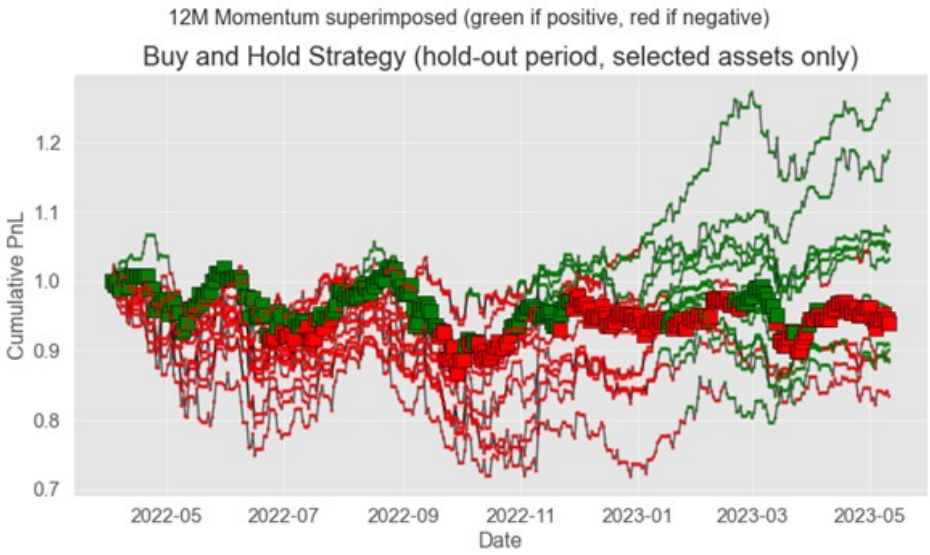
Although the 12-month momentum strategy worked well enough for the OBX Index during the training period for Q-Boruta to select it, market conditions have changed subsequently for this to no longer be the case. Figure 24 highlights the breakdown by superimposing the 12-month momentum for each selected asset on top of its cumulative PnL path during the hold-out period. The color of the marker for each day indicates whether the momentum strategy would have invested (green) or not (red) in a particular asset, and OBX Index is the only asset shown with large square markers.

There are several stretches of time during 2022 when OBX Index is the only asset with positive 12-month momentum. Figure 25 shows a view with a longer history, and it appears that the 12-month momentum of the OBX Index started decoupling from the rest of the selected assets sometime during late 2021 or early 2022.

Several factors may have contributed to this decoupling including but not limited to: interest rate hikes by Norges Bank throughout the year, rising energy costs, softening demand for freight and a new aquaculture tax proposed in September [5].

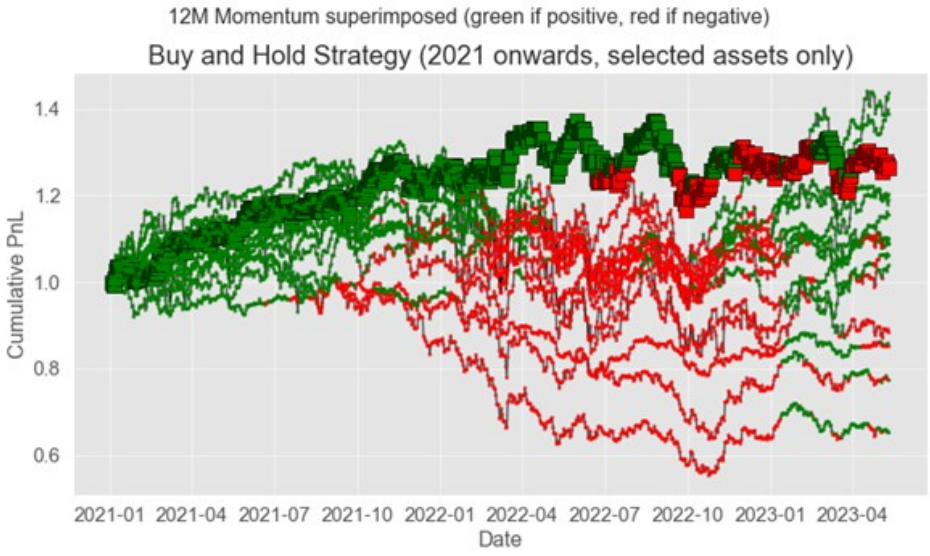
Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

Figure 24: Cumulative PnL



Source: J.P. Morgan

Figure 25: Cumulative PnL: 2021 onwards



Source: J.P. Morgan

Table 2 shows some portfolio summary statistics for a buy and hold strategy on all assets, and the 12-month momentum strategy on the selected (with and without OBX Index) and unselected assets. The initial portfolio that applies the momentum strategy to all selected assets performs poorly, but after dropping OBX Index it achieves the best performance overall.

Table 2: Portfolio summary statistics

Metric	Benchmark	12M Momentum Strategy		
	Buy and Hold (All)	Selected	Selected (no OBX)	Unselected
Sharpe	-0.1	-0.22	0.11	-0.19
Max Drawdown	-14.70%	-5.47%	-5.29%	-3.53%
Volatility	0.09	0.04	0.04	0.03
Sortino	-0.13	-0.29	0.15	-0.24
CAGR	-1.74%	-1.24%	0.49%	-0.83%

Source: J.P. Morgan

The anomaly we observe with the OBX Index reminds us that an assumption for Q-Boruta and similar methods is that the patterns observed during the training period will be representative of the ones that will be observed in the future, but of course this is not necessarily the case in the real world.

A particular strategy may work for an asset in certain situations and not others, and so another application of our pipeline that we leave to future work is to let $\boldsymbol{y}$ be a binary indicator for whether a given start date for a strategy works well or not for an asset and then to let $\boldsymbol{X}$ be a corresponding set of macroeconomic indicators that describe the state of the world on each day. By replacing linear regression with logistic regression and running our pipeline with this setup, we could explain regimes or situations in which a strategy is effective for an asset of interest.

The selection from Q-Boruta is based on a large historical window for a particular simple momentum strategy as an example. Depending on the strategy you analyze with the framework, how you set and tune model parameters and hyperparameters (historical window size, etc.), and the universe on which you run the model, your results and mileage may vary.

There are many factors that affect the performance of instruments in financial markets, and past performance is not always indicative of future results.

Discussion

It would seem that for the task of replicating a single stock's return stream as a linear function of the return streams of a set of indices, the model that uses a subset of the principal components in a rearranged order strikes a compromise between using a subset of the components in their default ordering and using all components.

The model that uses all components uses all available information, and we have seen in individual case studies, e.g., the sample US bank stock, that this leads to most of the final regression betas being concentrated on a handful of stocks with other betas being close to zero. Betas close to zero naturally lead towards lower turnover because there is less of a need to rebalance for those assets. If there is no redundant information in the feature matrix, then we should use all available information to achieve the best equation that maps inputs to the output.

On the other hand, the model that uses a small number of components in their default ordering only has access to the components that are the most helpful in explaining the variation in the feature matrix. Those may or may not also be helpful in explaining the variation in the target variable. Moreover, these models tend to assign betas with

relatively larger magnitudes to the features in part because with relatively limited data they are not yet able to distinguish which features are redundant or unnecessary once accounting for the full set of features. This can lead to higher turnover and if the first few components are not also helpful in explaining the variation in the target variable then this also leads to suboptimal equations.

The model that uses a small number of components in their rearranged ordering tends to move us closer to the model that uses all components so that we begin to lessen the loadings on unimportant features and increase the loadings on the important ones. However, we still are not using all available data, and so there is a loss of some information that is perhaps noise but also possibly signal. As a result we strike a compromise between the previous models and the turnover tends to be somewhere in between.

Perhaps the models that use rearranged components are able to do better than their counterparts that use all components in situations where for a particular target variable there is redundancy in the feature matrix, and so the simplified equation obtained by the PCA-based strategy is able to hone in on the signal and remove some of the noise.

Limitations

The pipeline does not account for any possible non-linear relationships or interaction effects between the input variables and the output variable. Moreover, it does not test any of the assumptions of the multiple regression model or try to identify and resolve any possible data quality issues, e.g., outliers or missing data. We consider the former as possible next steps and the latter as tasks and decisions for the practitioner to undertake based on each use case.

Next Steps

The pipeline could be improved by adding a step to test and account for non-linear relationships or interaction effects between the input variables and the output variable. To test for non-linear relationships, we could use correlation coefficients such as the predictive power score or the Chatterjee correlation coefficient, that measure the strength of meaningful and potentially non-linear relationships between two variables. If meaningful non-linear relationships are detected, then we could apply a suitable flexible transformation to a feature (e.g., natural cubic spline). We could test for interaction effects using Friedman's H-statistic [6] and include significant transformed features in the model.

Regardless of the ordering of principal components in our pipeline, we sometimes observe irregularities in the "paths" of coefficients, e.g., reversals in direction, due to multicollinearity in the original feature matrix. We could further enhance the pipeline by surfacing these issues to the user if necessary and applying an additional filtering step in the analysis. For example, we could show the user a clustering of features based on their pair-wise correlations and ask the user to pick one feature to retain from each cluster. We could also introduce a regularization component to the model-fitting process that embeds the prior beliefs of the user in terms of which features are more important *a priori*, e.g., via a technique such as Bayesian variable selection.

There are trade-offs to these ideas, but a common key point is to keep a human (expert) in the loop when making these decisions. Our goal is to enable the user to reach the

desired equation of interest in a way that encourages decisions based on testing preconceived hypotheses. This convenient-to-use framework should not be employed for "p-hacking."

Cross Asset Strategy series

- *With a mixed macro environment and range-bound markets, we upgraded our PCA tool to screen dislocations, from May 2023*
- *Meet CACTI, Our Cross Asset Crisis Tracking Indicator* from Apr 2023
- *3Y to 5Y Capital Market Assumptions: Stick to the late cycle playbook amid noisy macro and range-bound markets* from Mar 2023
- *Capital Market Assumptions: Robust modelling of 3Y to 5Y expected returns – Bayesian vs Traditional* from Dec 2022
- *Applied Portfolio Construction Research: Can you time your optimization technique?* from Oct 2022
- *What can save the 60/40?* from Jul 2022
- *What if the mean reverts?* from May 2022
- *Mind the gap in vol* from Mar 2022
- *The dislocations you shouldn't miss* from Feb 2022
- *Rising Rates and the 60/40* from Jan 2022
- *Systematic TAA based on market dislocations* from Dec 2021
- *Volatility risk premia and the 60/40* from Sep 2021

References

1. Kursa MB, Rudnicki WR (2010). “Feature Selection with the Boruta Package.” Journal of Statistical Software, 36(11), 1–13.
2. Bonaccolto, Giovanni and Caporin, Massimiliano, On the Ordering of Dynamic Principal Components and the Implications for Portfolio Analysis (April 26, 2022). Available at SSRN: https://ssrn.com/abstract=4094084 or http://dx.doi.org/10.2139/ssrn.4094084
3. Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., at al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830
4. https://www.americasquarterly.org/article/latin-americas-decade-long-hangover/
5. https://www.efeedlink.com/contents/09-29-2022/a5585075-1235-4b78-8a97-bf1d45488eaa-0001.html
6. J.H. Friedman and B.E. Popescu (2005). “Predictive Learning via Rule Ensembles.” Section 8.1

Thomas Salopek
(1-212) 834-5476
thomas.salopek@jpmorgan.com

J.P.Morgan

Appendix (Some additional notes)

“Term Structure” of Forecasts

In this paper we illustrate how an input feature set can be used with our pipeline to relate a set of input features after careful feature selection and principal component regression. We show with examples how our framework can be employed both on a one-off, and on a rolling basis, for explaining contemporaneous returns and for making return forecasts.

We could go a step further and build out a full "term structure" of forecasts (contemporaneous, +1D, +1W, +2W, +1M, +3M, +6M, +12M) and show heatmaps for a subset of factors to demonstrate how and where any relationships deteriorate. Of course, this would entail re-running the model for different horizons, and would give different regression equations for each.

The choice of factors will also vary by forecast horizon as well as by time period. There is combinatorial explosion in the number of possible experiments that can be run given the researcher degrees of freedom and the dimensionality of the problem specification, so we leave our analysis here and encourage the reader to explore further pathways on their own.

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 the Research Analyst's 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, if applicable, they also certify, as per KOFIA requirements, that the Research Analyst's analysis was made in good faith and that the views reflect the Research Analyst's own opinion, without undue influence or intervention.

All authors named within this report are Research Analysts who produce independent research unless otherwise specified. In Europe, Sector Specialists (Sales and Trading) may be shown on this report as contacts but are not authors of the report or part of the Research Department.

Important Disclosures

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, and certain non-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.

Explanation of Equity Research Ratings, Designations and Analyst(s) Coverage Universe:
J.P. Morgan uses the following rating system: Overweight [Over the next six to twelve months, we expect this stock will outperform the average total return of the stocks in the analyst's (or the analyst's team's) coverage universe.] Neutral [Over the next six to twelve months, we expect this stock will perform in line with the average total return of the stocks in the analyst's (or the analyst's team's) coverage universe.] Underweight [Over the next six to twelve months, we expect this stock will underperform the average total return of the stocks in the analyst's (or the analyst's team's) coverage universe.] Not Rated (NR): J.P. Morgan has removed the rating and, if applicable, the price target, for this stock because of either a lack of a sufficient fundamental basis or for legal, regulatory or policy reasons. The previous rating and, if applicable, the price target, no longer should be relied upon. An NR designation is not a recommendation or a rating. In our Asia (ex-Australia and ex-India) and U.K. small- and mid-cap equity research, each stock's expected total return is compared to the expected total return of a benchmark country market index, not to those analysts' coverage universe. If it does not appear in the Important Disclosures section of this report, the certifying analyst's coverage universe can be found on J.P. Morgan's research website, https://www.jpmorganmarkets.com.

J.P. Morgan Equity Research Ratings Distribution, as of April 01, 2023

	Overweight (buy)	Neutral (hold)	Underweight (sell)
J.P. Morgan Global Equity Research Coverage*	47%	38%	15%
IB clients**	47%	44%	34%
JPMS Equity Research Coverage*	46%	41%	13%
IB clients**	66%	65%	53%

***Please note that the percentages might not add to 100% because of rounding.**
****Percentage of subject companies within each of the "buy," "hold" and "sell" categories for which J.P. Morgan has provided investment banking services within the previous 12 months.**
For purposes only of FINRA ratings distribution rules, our Overweight rating falls into a buy rating category; our Neutral rating falls into a hold rating category; and our Underweight rating falls into a sell rating category. Please note that stocks with an NR designation are not included in the table above. This information is current as of the end of the most recent calendar quarter.

Equity Valuation and Risks: For valuation methodology and risks associated with covered companies or price targets for covered companies, please see the most recent company-specific research report at http://www.jpmorganmarkets.com, contact the primary analyst or your J.P. Morgan representative, or email research.disclosure.inquiries@jpmorgan.com. For material information about the proprietary models used, please see the Summary of Financials in company-specific research reports and the Company Tearsheets, which are available to download on the company pages of our client website, http://www.jpmorganmarkets.com. This report also sets out within it the material underlying assumptions used.

A history of J.P. Morgan investment recommendations disseminated during the preceding 12 months can be accessed on the Research & Commentary page of http://www.jpmorganmarkets.com where you can also search by analyst name, sector or financial instrument.

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.

Registration of non-US Analysts: Unless otherwise noted, the non-US analysts listed on the front of this report are employees of non-US affiliates of J.P. Morgan Securities LLC, may not be registered as research analysts under FINRA rules, may not be associated persons of J.P. Morgan Securities LLC, and may not be subject to FINRA Rule 2241 or 2242 restrictions on communications with covered companies, public appearances, and trading securities held by a research analyst account.

Other Disclosures

J.P. Morgan is a marketing name for investment banking businesses of JPMorgan Chase & Co. and its subsidiaries and affiliates worldwide.

UK MIFID FICC research unbundling exemption: UK clients should refer to UK MIFID Research Unbundling exemption for details of JPMorgan's implementation of the FICC research exemption and guidance on relevant FICC research categorisation.

All research material made available to clients are simultaneously available on our client website, J.P. Morgan Markets, unless specifically permitted by relevant laws. Not all research content is redistributed, e-mailed or made available to third-party aggregators. For all research material available on a particular stock, please contact your sales representative.

Any long form nomenclature for references to China; Hong Kong; Taiwan; and Macau within this research material are Mainland China; Hong Kong SAR (China); Taiwan (China); and Macau SAR (China).

J.P. Morgan Research may, from time to time, write on issuers or securities targeted by economic or financial sanctions imposed or administered by the governmental authorities of the U.S., EU, UK or other relevant jurisdictions (Sanctioned Securities). Nothing in this report is intended to be read or construed as encouraging, facilitating, promoting or otherwise approving investment or dealing in such Sanctioned Securities. Clients should be aware of their own legal and compliance obligations when making investment decisions.

Any digital or crypto assets discussed in this research report are subject to a rapidly changing regulatory landscape. For relevant regulatory advisories on crypto assets, including bitcoin and ether, please see https://www.jpmorgan.com/disclosures/cryptoasset-disclosure.

The author(s) of this research report may not be licensed to carry on regulated activities in your jurisdiction and, if not licensed, do not hold themselves out as being able to do so.

Exchange-Traded Funds (ETFs): J.P. Morgan Securities LLC ("JPMS") acts as authorized participant for substantially all U.S.-listed ETFs. To the extent that any ETFs are mentioned in this report, JPMS may earn commissions and transaction-based compensation in connection with the distribution of those ETF shares and may earn fees for performing other trade-related services, such as securities lending to short sellers of the ETF shares. JPMS may also perform services for the ETFs themselves, including acting as a broker or dealer to the ETFs. In addition, affiliates of JPMS may perform services for the ETFs, including trust, custodial, administration, lending, index calculation and/or maintenance and other services.

Options and Futures related research: If the information contained herein regards options- or futures-related research, such information is available only to persons who have received the proper options or futures risk disclosure documents. Please contact your J.P. Morgan Representative or visit https://www.theocc.com/components/docs/riskstoc.pdf for a copy of the Option Clearing Corporation's Characteristics and Risks of Standardized Options or http://www.finra.org/sites/default/files/Security_Futures_Risk_Disclosure_Statement_2018.pdf for a copy of the Security Futures Risk Disclosure Statement.

Changes to Interbank Offered Rates (IBORs) and other benchmark rates: Certain interest rate benchmarks are, or may in the future become, subject to ongoing international, national and other regulatory guidance, reform and proposals for reform. For more information, please consult: https://www.jpmorgan.com/global/disclosures/interbank_offered_rates

Private Bank Clients: Where you are receiving research as a client of the private banking businesses offered by JPMorgan Chase & Co. and its subsidiaries ("J.P. Morgan Private Bank"), research is provided 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 Global Research division.

Legal entity responsible for the production and distribution of research: The legal entity identified below the name of the Reg AC Research Analyst who authored this material is the legal entity responsible for the production of this research. Where multiple Reg AC Research Analysts authored this material with different legal entities identified below their names, these legal entities are jointly responsible for the production of this research. Research Analysts from various J.P. Morgan affiliates may have contributed to the production of this material but may not be licensed to carry out regulated activities in your jurisdiction (and do not hold themselves out as being able to do so). Unless otherwise stated below, this material has been distributed by the legal entity responsible for production. If you have any queries, please contact the relevant Research Analyst in your jurisdiction or the entity in your jurisdiction that has distributed this research material.

Legal Entities Disclosures and Country-/Region-Specific Disclosures:
Argentina: JPMorgan Chase Bank N.A Sucursal Buenos Aires is regulated by Banco Central de la República Argentina ("BCRA"- Central Bank of Argentina) and Comisión Nacional de Valores ("CNV"- Argentinian Securities Commission" - ALYC y AN Integral N°51). **Australia:** J.P. Morgan Securities Australia Limited ("JPMSAL") (ABN 61 003 245 234/AFS Licence No: 238066) is regulated by the Australian Securities and Investments Commission and is a Market, Clearing and Settlement Participant of ASX Limited and CHI-X. This material is issued and distributed in Australia by or on behalf of JPMSAL only to "wholesale clients" (as defined in section 761G of the Corporations Act 2001). A list of all financial products covered can be found by visiting https://www.jpmm.com/research/disclosures. J.P. Morgan seeks to cover companies of relevance to the domestic and international investor base across all Global Industry Classification Standard (GICS) sectors, as well as across a range of market capitalisation sizes. If applicable, in the course of conducting public side due diligence on the subject company(ies), the Research Analyst team may at times perform such diligence through corporate engagements such as site visits, discussions with company representatives, management presentations, etc. Research issued by JPMSAL has been prepared in accordance with J.P. Morgan Australia's

Research Independence Policy which can be found at the following link: J.P. Morgan Australia - Research Independence Policy. **Brazil**: Banco J.P. Morgan S.A. is regulated by the Comissao de Valores Mobiliarios (CVM) and by the Central Bank of Brazil. Ombudsman J.P. Morgan: 0800-7700847 / ouvidoria.jp.morgan@jpmorgan.com. **Canada**: J.P. Morgan Securities Canada Inc. is a registered investment dealer, regulated by the Investment Industry Regulatory Organization of Canada and the Ontario Securities Commission and is the participating member on Canadian exchanges. This material is distributed in Canada by or on behalf of J.P.Morgan Securities Canada Inc. **Chile:** Inversiones J.P. Morgan Limitada is an unregulated entity incorporated in Chile. **China:** J.P. Morgan Securities (China) Company Limited has been approved by CSRC to conduct the securities investment consultancy business. **Dubai International Financial Centre (DIFC)**: 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 - The Gate, West Wing, Level 3 and 9 PO Box 506551, Dubai, UAE. This material has been distributed by JP Morgan Chase Bank, N.A., Dubai Branch to persons regarded as professional clients or market counterparties as defined under the DFSA rules. **European Economic Area (EEA):** Unless specified to the contrary, research is distributed in the EEA by J.P. Morgan SE ("JPM SE"), which is subject to prudential supervision by the European Central Bank ("ECB") in cooperation with BaFin and Deutsche Bundesbank in Germany. JPM SE is a company headquartered in Frankfurt with registered address at TaunusTurm, Taunustor 1, Frankfurt am Main, 60310, Germany. The material has been distributed in the EEA to persons regarded as professional investors (or equivalent) pursuant to Art. 4 para. 1 no. 10 and Annex II of MiFID II and its respective implementation in their home jurisdictions ("EEA professional investors"). This material must not be acted on or relied on by persons who are not EEA professional investors. Any investment or investment activity to which this material relates is only available to EEA relevant persons and will be engaged in only with EEA relevant persons. **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 J.P. Morgan Broking (Hong Kong) Limited (CE number AAB027) is regulated by the Securities and Futures Commission in Hong Kong. JP Morgan Chase Bank, N.A., Hong Kong Branch (CE Number AAL996) is regulated by the Hong Kong Monetary Authority and the Securities and Futures Commission, is organized under the laws of the United States with limited liability. Where the distribution of this material is a regulated activity in Hong Kong, the material is distributed in Hong Kong by or through J.P. Morgan Securities (Asia Pacific) Limited and/or J.P. Morgan Broking (Hong Kong) Limited. **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 the 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 and the Bombay Stock Exchange Limited (SEBI Registration Number – INZ000239730) and as a Merchant Banker (SEBI Registration Number - MB/INM000002970). Telephone: 91-22-6157 3000, Facsimile: 91-22-6157 3990 and Website: http://www.jpmipl.com. JPMorgan Chase Bank, N.A. - Mumbai Branch is licensed by the Reserve Bank of India (RBI) (Licence No. 53/ Licence No. BY.4/94; SEBI - IN/CUS/014/ CDSL : IN-DP-CDSL-444-2008/ IN-DP-NSDL-285-2008/ INBI00000984/ INE231311239) as a Scheduled Commercial Bank in India, which is its primary license allowing it to carry on Banking business in India and other activities, which a Bank branch in India are permitted to undertake. For non-local research material, this material is not distributed in India by J.P. Morgan India Private Limited. Compliance Officer: Spurthi Gadamsetty; spurthi.gadamsetty@jpmchase.com; +912261573225. Grievance Officer: Ramprasadh K, jpmipl.research.feedback@jpmorgan.com; +912261573000.

Investment in securities market are subject to market risks. Read all the related documents carefully before investing. Registration granted by SEBI and certification from NISM in no way guarantee performance of the intermediary or provide any assurance of returns to investors.

Indonesia: PT J.P. Morgan Sekuritas Indonesia is a member of the Indonesia Stock Exchange and is registered and supervised by the Otoritas Jasa Keuangan (OJK). **Korea**: J.P. Morgan Securities (Far East) Limited, Seoul Branch, is a member of the Korea Exchange (KRX). JPMorgan Chase Bank, N.A., Seoul Branch, is licensed as a branch office of foreign bank (JPMorgan Chase Bank, N.A.) in Korea. Both entities are regulated by the Financial Services Commission (FSC) and the Financial Supervisory Service (FSS). For non-macro research material, the material is distributed in Korea by or through J.P. Morgan Securities (Far East) Limited, Seoul Branch. **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 holds a Capital Markets Services License issued by the Securities Commission in Malaysia. **Mexico**: J.P. Morgan Casa de Bolsa, S.A. de C.V.and J.P. Morgan Grupo Financiero are members of the Mexican Stock Exchange and are authorized to act as a broker dealer by the National Banking and Securities Exchange Commission. **New Zealand**: This material is issued and distributed by JPMSAL in New Zealand only to "wholesale clients" (as defined in the Financial Markets Conduct Act 2013). JPMSAL is registered as a Financial Service Provider under the Financial Service providers (Registration and Dispute Resolution) Act of 2008. **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. **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. **Russia**: CB J.P. Morgan Bank International LLC is regulated by the Central Bank of Russia. **Singapore**: This material is issued and distributed in Singapore by or through J.P. Morgan Securities Singapore Private Limited (JPMSS) [MCI (P) 060/08/2022 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), 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 material in Singapore are to contact JPMSS or JPMCB Singapore in respect of any matters arising from, or in connection with, the material. As at the date of this material, 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 material. Arising from its role as a 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. **South Africa**: J.P. Morgan Equities South Africa Proprietary Limited and JPMorgan Chase Bank, N.A., Johannesburg Branch are members of the Johannesburg Securities Exchange and are regulated by the Financial Services Board. **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. Material 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 that may give rise to conflicts of interests, unless otherwise disclosed in the "Important Disclosures" in this material. **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. **UK:** Unless specified to the contrary, research is distributed in the UK by J.P. Morgan Securities plc ("JPMS plc") which 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. JPMS plc is registered in England & Wales No. 2711006, Registered Office 25 Bank Street, London, E14 5JP. This material is directed in the UK only to: (a) persons having professional experience in matters relating to investments falling within article 19(5) of the Financial Services and Markets Act 2000 (Financial Promotion) (Order) 2005 ("the FPO"); (b) persons outlined in article 49 of the FPO (high net worth companies, unincorporated associations or partnerships, the trustees of high value trusts, etc.); or (c) any persons to whom this communication may otherwise lawfully be made; all such persons being referred to as "UK relevant persons". This material must not be acted on or relied on by persons who are not UK relevant persons. Any investment or investment activity to which this material relates is only available to UK relevant persons and will be engaged in only with UK relevant persons. Research issued by JPMS plc has been prepared in accordance with JPMS plc's policy for prevention and avoidance of conflicts of interest related to the production of Research which can be found at the following link: J.P. Morgan EMEA - Research Independence Policy. **U.S.**: J.P. Morgan Securities LLC ("JPMS") is a member of the NYSE, FINRA, SIPC, and the NFA. JPMorgan Chase Bank, N.A. is a member of the FDIC. Material published by non-U.S. affiliates is distributed in the U.S. by JPMS who accepts responsibility for its content.

General: Additional information is available upon request. The information in this material has been obtained from sources believed to be reliable. While all reasonable care has been taken to ensure that the facts stated in this material are accurate and that the forecasts, opinions and expectations contained herein are fair and reasonable, JPMorgan Chase & Co. or its affiliates and/or subsidiaries (collectively J.P. Morgan) make no representations or warranties whatsoever to the completeness or accuracy of the material provided, except with respect to any disclosures relative to J.P. Morgan and the Research Analyst's involvement with the issuer that is the subject of the material. Accordingly, no reliance should be placed on the accuracy, fairness or completeness of the information contained in this material. There may be certain discrepancies with data and/or limited content in this material as a result of calculations, adjustments, translations to different languages, and/or local regulatory restrictions, as applicable. These discrepancies should not impact the overall investment analysis, views and/or recommendations of the subject company(ies) that may be discussed in the material. J.P. Morgan accepts no liability whatsoever for any loss arising from any use of this material or its contents, and neither J.P. Morgan nor any of its respective directors, officers or employees, shall be in any way responsible for the contents hereof, apart from the liabilities and responsibilities that may be imposed on them by the relevant regulatory authority in the jurisdiction in question, or the regulatory regime thereunder. Opinions, forecasts or projections contained in this material represent J.P. Morgan's current opinions or judgment as of the date of the material only and are therefore subject to change without notice. Periodic updates may be provided on companies/industries based on company-specific developments or announcements, market conditions or any other publicly available information. There can be no assurance that future results or events will be consistent with any such opinions, forecasts or projections, which represent only one possible outcome. Furthermore, such opinions, forecasts or projections are subject to certain risks, uncertainties and assumptions that have not been verified, and future actual results or events could differ materially. The value of, or income from, any investments referred to in this material may fluctuate and/or be affected by changes in exchange rates. All pricing is indicative as of the close of market for the securities discussed, unless otherwise stated. Past performance is not indicative of future results. Accordingly, investors may receive back less than originally invested. 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. This material may include views on structured securities, options, futures and other derivatives. These are complex instruments, may involve a high degree of risk and may be appropriate investments only for sophisticated investors who are capable of understanding and assuming the risks involved. The recipients of this material must make their own independent decisions regarding any securities or financial instruments mentioned herein and should seek advice from such independent financial, legal, tax or other adviser as they deem necessary. J.P. Morgan may trade as a principal on the basis of the Research Analysts' views and research, and it may also engage in transactions for its own account or for its clients' accounts in a manner inconsistent with the views taken in this material, and J.P. Morgan is under no obligation to ensure that such other communication is brought to the attention of any recipient of this material. Others within J.P. Morgan, including Strategists, Sales staff and other Research Analysts, may take views that are inconsistent with those taken in this material. Employees of J.P. Morgan not involved in the preparation of this material may have investments in the securities (or derivatives of such securities) mentioned in this material and may trade them in ways different from those discussed in this material. This material is not an advertisement for or marketing of any issuer, its products or services, or its securities in any jurisdiction.

Confidentiality and Security Notice: This transmission may contain information that is privileged, confidential, legally privileged, and/or exempt from disclosure under applicable law. If you are not the intended recipient, you are hereby notified that any disclosure, copying, distribution, or use of the information contained herein (including any reliance thereon) is STRICTLY PROHIBITED. Although this

transmission and any attachments are believed to be free of any virus or other defect that might affect any computer system into which it is received and opened, it is the responsibility of the recipient to ensure that it is virus free and no responsibility is accepted by JPMorgan Chase & Co., its subsidiaries and affiliates, as applicable, for any loss or damage arising in any way from its use. If you received this transmission in error, please immediately contact the sender and destroy the material in its entirety, whether in electronic or hard copy format. This message is subject to electronic monitoring: https://www.jpmorgan.com/disclosures/email

MSCI: Certain information herein ("Information") is reproduced by permission of MSCI Inc., its affiliates and information providers ("MSCI") ©2023. No reproduction or dissemination of the Information is permitted without an appropriate license. MSCI MAKES NO EXPRESS OR IMPLIED WARRANTIES (INCLUDING MERCHANTABILITY OR FITNESS) AS TO THE INFORMATION AND DISCLAIMS ALL LIABILITY TO THE EXTENT PERMITTED BY LAW. No Information constitutes investment advice, except for any applicable Information from MSCI ESG Research. Subject also to msci.com/disclaimer

"Other Disclosures" last revised May 13, 2023.

Completed 19 May 2023 02:40 PM EDT

Disseminated 19 May 2023 02:40 PM EDT