High dimensional data visualisation for systematic FX trading

Evaluating the effect of training period length on the performance of machine learning algorithms

Does more data mean better models? Maybe not for neural network systematic FX trading strategies. In this paper, we explain a qualitative approach to estimating the optimal look-back window size to create training sets for deep neural network-based FX strategies.

Mehryar Emambakhsh
Vice President,
Senior Research
Scientist

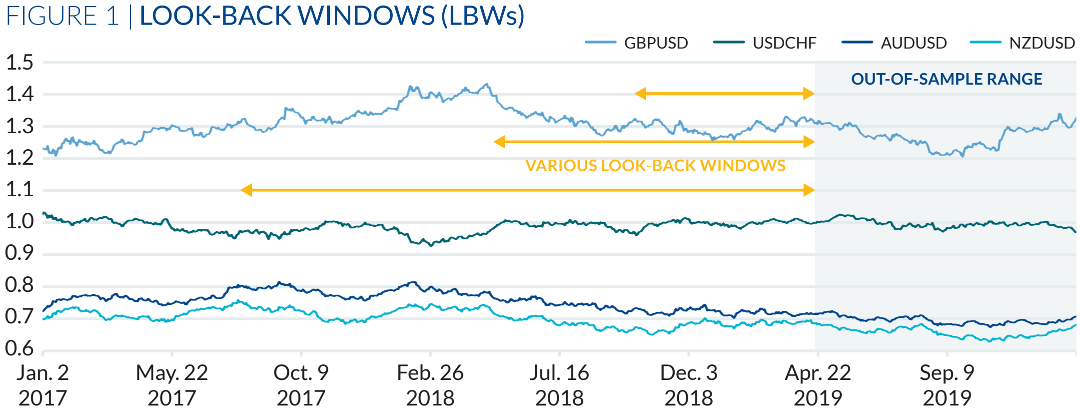
Richard Turner
Managing Director,
Head of Research

Explaining the problem: class over-representation vs. lack of data dilemma

Machine learning-based systematic FX trading algorithms utilise training sets to learn optimal model parameters. These training sets are usually constructed from various time series, sometimes with different granularities, and the main training set is usually split into several subsets to perform training, validation, model selection, and (simulated) out-of-sample evaluation.

A common approach to extracting training samples from raw sequential data is to allocate a (temporal) look-back window (LBW) length. To capture the latest market trends, these LBWs are usually selected to be close in time to the out-of-sample period. Different LBW sizes generate different training sets with their own specific statistical features. In Figure 1 we show daily spot FX rates for four (example) currency pairs. While the gray area shows the out-of-sample (test) range, various LBWs (shown as yellow arrows) can be used to construct different training sets.

FIGURE 1 | **LOOK-BACK WINDOWS (LBWs)**



Source: Mesirow

The captured samples within the selected LBW range are then pre-processed and prepared for the next steps. Stationarity analysis may be necessary if the statistics of the signal change over time. Another step could be normalisation: the input signal to a machine learning pipeline is usually normalised, which can be an extremely important part of its learning performance.

CONSEQUENCES OF LBWS THAT ARE TOO LONG

The LBW used to construct a training set can have a significant effect over the above processes. Windows that are too long can create less stationarity, the calculated normalisation parameters over such long train sets can be irrelevant to the recent data, and thus the performance of the following processes in a machine learning pipeline can then be severely affected.

For systematic FX trading classification tasks, where each sample is labelled as going either long or short, an LBW that is too long can increase overlap between classes or, rather, overrepresent classes as a result of the intrinsic high unpredictability of a financial time series. Longer LBWs greatly increase the probability of having similar samples labelled to different classes as it is extremely difficult for a machine learning task to classify similar samples correctly. The resulting FX trading strategies can then underfit and generate low training accuracy because they have been unable to learn much from the data (a machine learning algorithm under-fits when it totally fails to model even the training data).

CONSEQUENCES OF LBWS THAT ARE TOO SHORT

A solution for the above issues is to select shorter LBWs, for which the resulting training sets can be more stationary and generate more consistent normalisation parameters. As shorter LBWs correspond to more recent data, they enable more immediate market trends to be captured. This, however, comes with a price; selecting LBWs that are too short reduces the training set size and the risk of overfitting increases.

When the number of training samples is significantly lower than the model complexity (which can roughly be estimated as the number of free parameters in a machine learning model), our overfitting risk increases. A model overfits when there is a significant drop in performance between the training vs. out-of-sample evaluation and it fails to generalise unseen data. Example of complex machine learning models are deep neural networks, which can easily have tens of thousands of parameters and can quickly overfit when trained on small datasets.

Therefore, to sum up:

1. An LBW that is too long can create more training samples, but, with higher non-stationarity and class over-representation, the machine learning models can potentially underfit;
2. An LBW that is too short can generate more stationary training sets with lower class over-representation, but can also output a much smaller training set, and the machine learning models can potentially overfit.

In this paper we address these issues and detail our approach to finding the optimal LBW size, utilised as part of Mesirow Currency Management's machine learning pipelines. We will explain our "temporal sample weighting" technique to push for an even larger LBW, while alleviating class over-representation. We first explain how the input data is prepared for the systematic FX trading task, we then detail our high dimensional FX time series visualisation technique, and, finally, we explain how a temporal sample weighting can effectively reduce class over-representation when using larger LBW sizes.

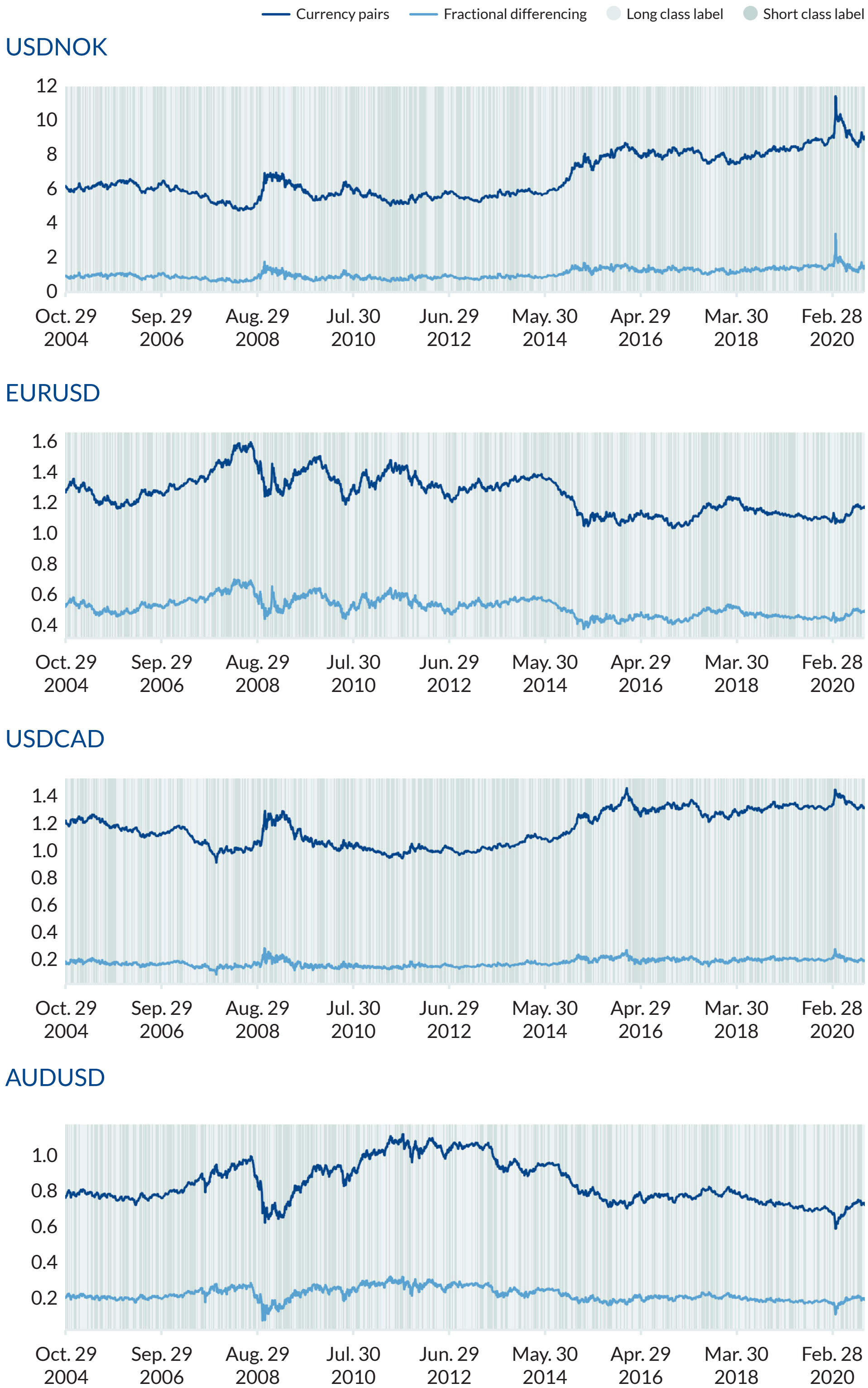
Data preparation: pre-processing and labelling

The performance of most machine learning algorithms is highly affected by the stationarity level of the input signals. First, to make the data stationary, a controlled fractional differencing operator[1] is applied to each time series. As shown in Figure 2 for USDNOK, XEUUSD, USDCAD and AUDUSD currency pairs, the fractionally differenced signals are significantly more stationary than the input FX rate.

We then perform a *z*-score normalisation. To facilitate for the trade sizing classification, the triple barrier method[2] is used to label each data sample into going long or short (class labels are shown as vertical lines in Figure 2).

Throughout this paper, we assume the samples are $1\times d$ vectors; $d=N*T$ is the data dimensionality, N is the number of input currencies, and T is the number of previous days for each sample. For M samples, each labelled as long or short, this process constructs an $M\times d$ matrix. This matrix represents our d-dimensional feature space.

FIGURE 2 | INPUT DAILY FX RATES AND THEIR FRACTIONAL DIFFERENCING



Source: Mesirow

Data visualisation via t-SNE

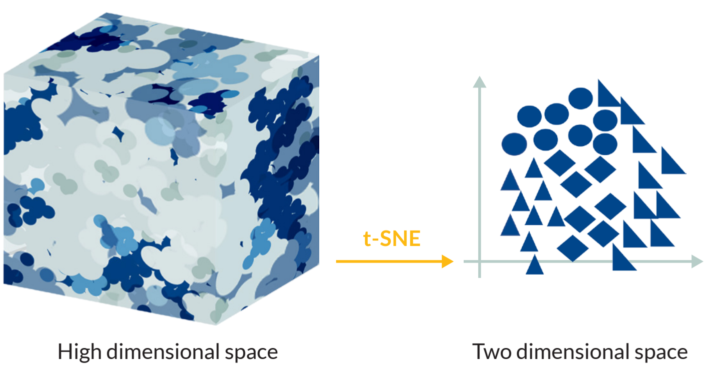
WHAT IS t-SNE?

Our goals are to visualise the feature space explained in the previous section and to reduce the dimensionality from d (which can be very high) to 2, thus enabling illustration of the M samples over a 2-dimensional plane.

One way to perform such visualisation would be to use principal component or linear discriminant analysis (PCA/LDA), but, unfortunately, these techniques can fail to separate classes as the generated samples from an FX sequence are not linearly separable. Also, due to high overlap between classes in FX trading problems, non-linear kernel-based PCA/LDA methods significantly merge samples into one cluster, failing a successful visualisation in 2 dimensions.

Alternatively, we can use t-SNE data visualisation. t-SNE is a recursive approach to reduce the data dimensionality.[3] It can map the data from a very high dimensional space (which is impossible for us to visualise) onto a lower dimensional space, facilitating data visualisation on a two-dimensional plane (Figure 3). The basic idea behind t-SNE is that similarity between the samples in any dimensionality must be preserved. In other words, similar points should always reside close to each other, independent of what dimensionality we are in. Therefore, at every iteration, t-SNE maps the data onto a lower dimensional space, such that the intra-similarity among points is preserved.

FIGURE 3 | **MAPPING**



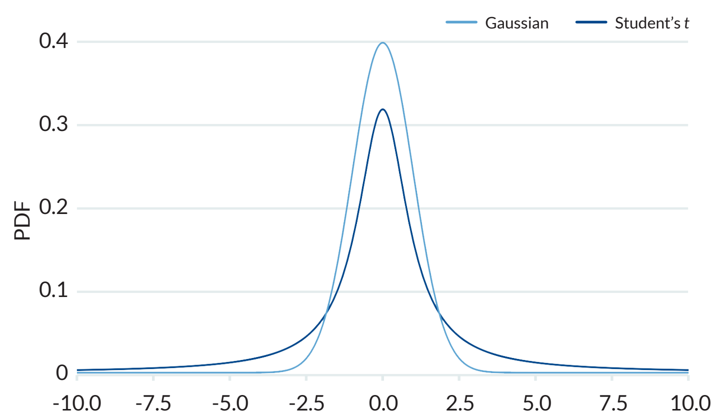
t-SNE performs mapping from high dimensional space onto lower dimensions, enabling a 2-dimensional data visualisation. Each class is illustrated using different markers. For a binary classification systematic FX trading problem, we would have only two markers, each representing going long or short.

Source: Mesirow

t-SNE estimates the between-points similarity in a high dimensional space as a Gaussian distribution, while in a lower dimensional space, it is quantified using a Student's t distribution with one degree of freedom (Figure 4). The key objective of t-SNE is to make these two distributions as alike as possible, such that the local structure of the data is preserved. If two points are too close to each other in a high dimensional space, we would want them to stay close in lower dimensional space as well.

This iterative process is performed via an optimisation procedure. Using the gradient descent algorithm, the Kullback–Leibler divergence (a method of computing dissimilarity between two probability density functions (PDFs)), constructed using the Gaussian and the Student's t distributions, is minimised. The use of the Student's t distribution, which is more heavy-tailed than the Gaussian distribution as shown in Figure 4, facilitates assigning higher scores to dissimilar points when mapped onto the lower dimensional space, which places them even farther from each other. This results in a more distinctive and clearer visualisation.

FIGURE 4 | **GAUSSIAN AND STUDENT'S T DISTRIBUTIONS**



t-SNE uses Gaussian and Student's t (with one degree of freedom) distributions to calculate between-points similarity in higher and lower dimensions at each algorithm iteration. As Student's t is more heavy-tailed than a Gaussian distribution, it can map dissimilar points farther from each other, facilitating a clearer visualisation.

Source: Mesirow

Unlike PCA or LDA, t-SNE is a non-linear technique. Therefore, it can preserve possible non-linearities in the final visualisation.

t-SNE visualisation of systematic FX trading datasets

The t-SNE algorithm explained in the previous section is used for visualising our FX trading dataset. We construct several training sets using various LBWs, then, to qualitatively evaluate the effect of training sample size over training samples' distribution, t-SNE is applied to these training sets. The visualisation results are shown in Figure 5. The dark and light blue circles represent samples going short and long.

This figure clearly shows the lack of training data vs. class over-representation dilemma. When the length of the LBW is short (6 months, 1 year or 2 years in Figure 5-a, -b and -c), there is significantly less class overlap, which makes it easier for a classification algorithm (such as the neural networks) to converge to an optimal loss. This, however, comes with the price of having fewer training samples, which risks the classification model to overfit.

On the other hand, while longer LBWs (4, 8, and 16 years) give a greater number of samples, they totally overlap the samples from different classes, causing class over-representation (Figure 5-d, -e and -f). Using such training sets can make it difficult for machine learning algorithms to find classification boundaries and can cause them to underfit.

In the next section, we explain our solution to this dilemma, in order to find the optimal LBW value.

FIGURE 5 | LACK OF TRAINING SAMPLES VS. CLASS OVER-REPRESENTATION DILEMMA

Figure 5 shows t-SNE visualisation of our systematic FX trading feature space labelled via the USDNOK currency pair, for various LBW lengths (we have used 5000 iterations and 55 perplexity within the t-SNE algorithm to generate these results). Longer LBWs (d, e, f) create higher class overlap, which can potentially result in under-fitting, while shorter LBWs generate fewer number of training samples (a, b, c), increasing the risk of over-fitting.

Source: Mesirow

Our solution: temporal sample weighting

From the results in the previous section, we can conclude that longer LBWs create lower class separability and shorter LBWs create higher class separability. While the latter seems to be a better choice for training a classifier, due to its low sample numbers, it is more likely to overfit the model.

To solve this dilemma between class separability and lack of data, for a given currency pair, we first perform similar t-SNE visualisation in Figure 5 by varying the LBW size. From the results, the largest LBW that shows the best class separability is selected, and, at the next step, we apply a sample weighting process to the resulting training set.

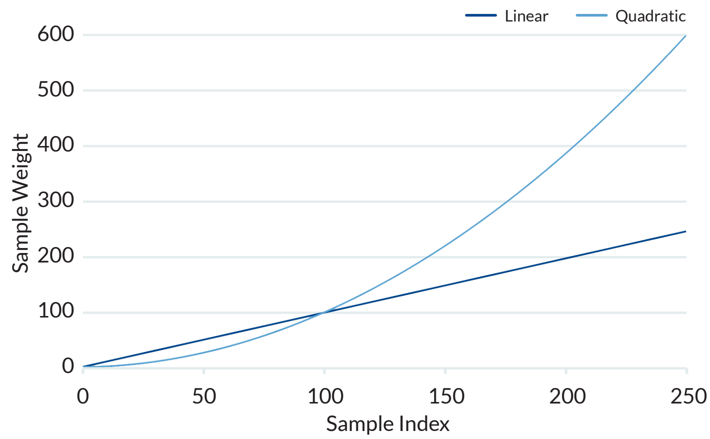
In machine learning and statistics, sample weighting is a technique to assign variable importance to each data point. A sample weighted dataset prioritises and/or deprioritises some samples over the others, by assigning higher or lower importance level to them.[4] In our work, we apply sample weighting such that more recent samples have higher priority. In other words, we assign the importance level to each sample based on its start time, and hence, we call it temporal sample weighting.

Using these two steps (selecting the longest LBW with highest class separability from t-SNE results and temporal sample weighting), we can not only make sure the classifier receives sufficient samples to minimise the over-fitting risk, but we can also capture the latest market trends by assigning higher weights to recent samples.

In our experiments, we use two methods to perform the temporal sample weighting: linear and quadratic. These are illustrated in Figure 6. Linear sample weighting (shown in dark blue) increases the weights based on their index with a constant slope. The quadratic sample weighting, however, assigns lower weights to farther samples (lower indexed samples), while allocating significantly higher weights to more recent samples (higher indexed samples).

To generate this figure, we have assumed we have 250 samples, where higher sample indexes correspond to more recent data points.

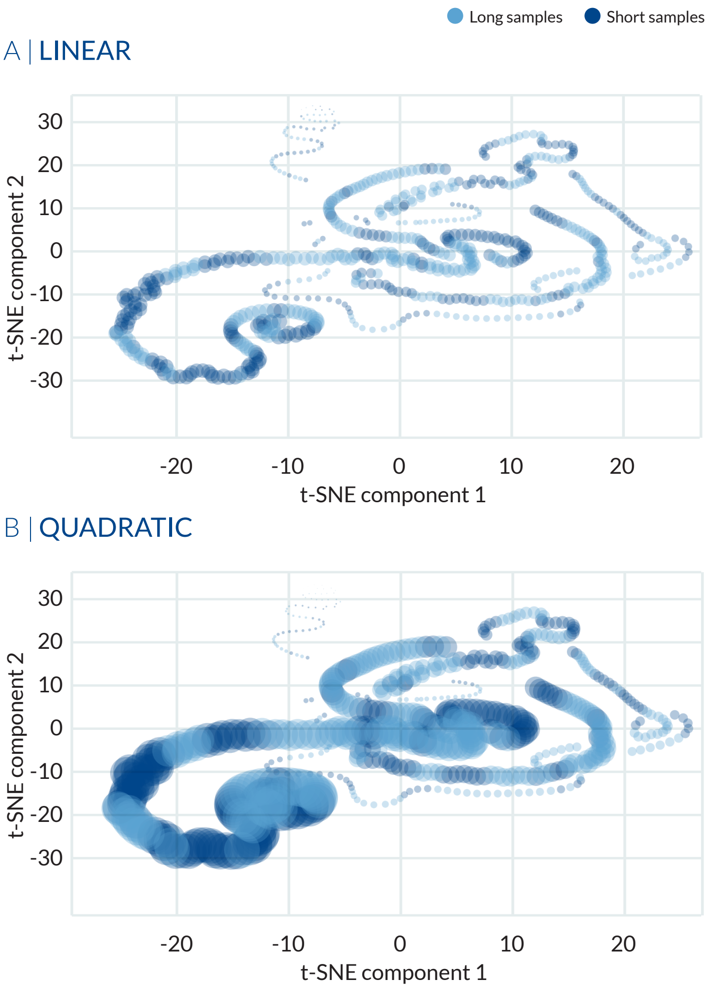
FIGURE 6 | **TEMPORAL SAMPLE WEIGHTING**



Source: Mesirow

As an example, Figure 7 shows the t-SNE visualisation of a linear (Figure 7-a) and quadratic (Figure 7-b) temporally sampled training set, labelled using the USDNOK currency pair. For both figures, the marker area is proportional to its sample weight. Both techniques assign higher importance to more recent samples and allocate lower weights to older samples. This way, we can utilise a relatively long LBW to create large enough training sets while reducing class overlap by assigning higher weights to more recent samples.

FIGURE 7 | t-SNE VISUALISATION



t-SNE visualisation of temporally weighted training sets with 4 years LBW via: (a) linear and (b) quadratic sample weighting. The marker area is proportional to its corresponding sample weight. More recent samples have higher weights and are depicted as larger.

Source: Mesirow

Conclusions

In this paper, we examined a qualitative approach to estimating an optimal LBW size for generating training sets for deep neural network-based systematic FX trading strategies. First, we explained the lack of data vs. class over-representation dilemma. Then, after briefly explaining t-SNE (a well-known high dimensional data visualisation technique), we explained how this dilemma can be visualised over daily FX data. Finally, we proposed a temporal sample weighting solution to not only minimise the effect of class over-representation, but also to emphasise more recent incoming data, capturing the latest market trends.

This work is totally data driven and easily extendable to other financial time series, such as stock market data. In addition to FX trading, the proposed approach can also be applied to other tasks, such as regression analysis over multi-modal financial data and machine learning-based portfolio allocation tasks.

The methodology to find the optimal LBW length is currently being used by the deep neural network systematic FX trading strategy at Mesirow Currency Management.

About Mesirow

1. Clive WJ Granger and Roselyne Joyeux. "An introduction to long-memory time series models and fractional differencing." Journal of Time Series Analysis 1.1 (1980): 15-29.

2. Marcos Lopez De Prado. Advances in Financial Machine Learning. John Wiley & Sons, 2018.

3. Maaten, Laurens van der and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of Machine Learning Research (2008): 2579-2605.

4. Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. "What are we weighting for?." Journal of Human Resources 50.2 (2015): 301-316.

Mesirow Currency Management ("MCM") is a division of Mesirow Financial Investment Management, Inc. ("MFIM") a SEC registered investment advisor. The information contained herein is intended for institutional clients, Qualified Eligible Persons and Eligible Contract Participants and is for informational purposes only. This information has been obtained from sources believed to be reliable but is not necessarily complete and its accuracy cannot be guaranteed. Any opinions expressed are subject to change without notice. It should not be assumed that any recommendations incorporated herein will be profitable or will equal past performance. Mesirow Financial does not render tax or legal advice. Nothing contained herein constitutes an offer to sell or a solicitation of an offer to buy an interest in any Mesirow Financial investment vehicle(s). Any offer can only be made through the appropriate Offering Memorandum. The Memorandum contains important information concerning risk factors and other material aspects of the investment and should be read carefully before an investment decision is made.

This communication may contain privileged and/or confidential information. It is intended solely for the use of the addressee. If this information was received in error, you are strictly prohibited from disclosing, copying, distributing or using any of this information and are requested to contact the sender immediately and destroy the material in its entirety, whether electronic or hardcopy.

Certain strategies discussed throughout the document are based on proprietary models of MFCM's or its affiliates. No representation is being made that any account will or is likely to achieve profits or losses similar to those referenced.

Australian Investors: The information contained herein is intended for Wholesale Clients only and is for informational purposes only. This document is not a prospectus or product disclosure statement under the Corporations Act 2001 (Cth) (Corporations Act) and does not constitute a recommendation to acquire, an invitation to apply for, an offer to apply for or buy, an offer to arrange the issue or sale of, or an offer for issue or sale of, any securities or investment service in Australia, except as set out below. The strategy has not authorised nor taken any action to prepare or lodge with the Australian Securities & Investments Commission an Australian law compliant prospectus or product disclosure statement. Accordingly, this strategy and document may not be issued or distributed in Australia other than by way of or pursuant to an offer or invitation that does not need disclosure to investors under Part 6D.2 or Part 7.9 of the Corporations Act, whether by reason of the investor being a 'wholesale client' (as defined in section 761G of the Corporations Act and applicable regulations) or otherwise. This document does not constitute or involve a recommendation to acquire, an offer or invitation for issue or sale, an offer or invitation to arrange the issue or sale, or an issue or sale, of any strategy or investment service to a 'retail client' (as defined in section 761G of the Corporations Act and applicable regulations) in Australia.

Canadian Investors: The information contained herein is intended for Permitted Clients only and is for informational purposes only. This confidential material pertains to the offering of the currency strategies described herein only in those jurisdictions and to those persons where and to whom they may be lawfully offered for sale, and only by persons permitted to sell such strategies. This material is not, and under no circumstances is to be construed as, an advertisement or a public offering of the strategies described herein in Canada. No securities commission or similar authority in Canada has reviewed or in any way passed upon this document or the merits of the strategies described herein, and any representation to the contrary is an offence.

EU Investors: The information contained herein is intended for Professional Clients as the term is defined by MiFID II and is for informational purposes only. Recipients that are classified under MiFID II as retail clients must opt up to Professional Clients before receiving any services from Mesirow Currency Management.

Japanese Investors: Mesirow Currency Management provides discretionary investment management services to managed accounts held on behalf of qualified investors only. MCM will not act as agent or intermediary in respect of the execution of a discretionary investment management agreement. Please note that this presentation is intended for educational purposes and solely for the addressee and may not be distributed.

Hong Kong Investors: The contents of this document have not been reviewed by any regulatory authority in Hong Kong. You are advised to exercise caution in relation to the contents of this document. You should obtain independent professional advice prior to considering or making any investment. The investment is not authorized under Section 104 of the Securities and Futures Ordinance of Hong Kong by the Securities and Futures Commission of Hong Kong. Accordingly, the distribution of this Presentation Material and discretionary management services in Hong Kong are restricted. This Presentation Material is only for the use of the addressee and may not be distributed, circulated or issued to any other person or entity.

South Korean Investors: Upon attaining a client, Mesirow Financial Investment Management, Inc. ("MFIM") will apply for the appropriate licenses and retain the services of a local licensed intermediary (a Korean financial investment company). In the interim, MFIM will rely on and sub-delegate to Mesirow Advanced Strategies, Inc. ("MAS").

This material is not intended for investment nor distribution purposes, but is being provided to the addressee for educational purposes only. This book may contain privileged and/or confidential information and is intended solely for the use of the addressee. If this information was received in error, you are strictly prohibited from disclosing, copying, distributing or using any of this information and are requested to contact the sender immediately and destroy the material in its entirety, whether electronic or hardcopy. Nothing contained herein constitutes an offer to sell or a solicitation of an offer to buy an interest in any Mesirow investment vehicle or to invest in any strategy. Any offer can only be made to Qualified Professional Investors through the appropriate Investment Management Agreement or Offering Memorandum, which contains important information concerning risk factors and other material aspects of the investment and should be read carefully before an investment decision is made.

Mesirow Financial Investment Management, Inc. ("MFIM") is not making any representation with respect to the eligibility of any potential investors or recipients of this material to invest in or acquire any interests therein under the laws of Korea, including but without limitation the Foreign Exchange Transaction Act and Regulations thereunder. The investment or interests may only be offered to Qualified Professional Investors, as such term is defined under the Financial investment Services and Capital Markets Act, and no investment or any of the interests may be offered, sold or delivered, or offered or sold to any person for re-offering or resale, directly or indirectly, in Korea or to any resident of Korea except pursuant to applicable laws and regulations of Korea.

Singapore Investors: Mesirow Currency Management provides discretionary investment management services to managed accounts held on behalf of qualified investors only. MCM will not act as agent or intermediary in respect of the execution of a discretionary investment management agreement. Please note that this presentation is intended for educational purposes and solely for the addressee and may not be distributed.

Swiss Investors: Services are only offered to Regulated Qualified Investors, as defined in Article 10 of the Swiss Collective Investment Scheme Act. There can be no guarantee investment advice will be profitable or meet its investment objectives.

United Kingdom Investors: In the United Kingdom, this communication is directed only at persons who fall within the definition of: (i) "investment professionals" as defined in COBS 4.12 and Article 14 of the Financial Services and Markets Act 2000 (Promotion of Collective Investment Schemes) (Exemptions) Order 2001 (the "PCISE Order"); or (ii) "high net worth companies, unincorporated associations etc" as defined in COBS 4.12 and Article 22(2)(a) to (d) of the PCISE Order (all such persons together being referred to as "Relevant Persons"). This communication must not be acted on or relied on by persons who are not Relevant Persons. Any investment or investment activity to which this communication relates is available only to Relevant Persons and will be engaged in only with Relevant Persons.

Additional Information: Currency strategies are only suitable and appropriate for sophisticated investors that are able to lose all of their capital investment. This communication may contain privileged and/or confidential information. It is intended solely for the use of the addressee. If this information was received in error, you are strictly prohibited from disclosing, copying, distributing or using any of this information and are requested to contact the sender immediately and destroy the material in its entirety, whether electronic or hardcopy.

Certain strategies discussed throughout the document are based on proprietary models of MCM's or its affiliates. No representation is being made that any account will or is likely to achieve profits or losses similar to those referenced.

Performance pertaining to the Currency Alpha and Macro strategies may be stated gross of fees or net of fees. Performance information that is provided net of fees reflects the deduction of implied management and performance fees. Performance information that is provided gross of fees does not reflect the deduction of advisory fees. Client returns will be reduced by such fees and other expenses that may be incurred in the management of the account. Advisory fees are described in Part II of Form ADV of Mesirow Financial Investment Management, Inc. Simulated model performance information and results do not reflect actual trading or asset or fund advisory management and the results may not reflect the impact that material economic and market factors may have had, and can reflect the benefit of hindsight, on MCM's decision-making if MCM were actually managing client's money in the same manner. Performance referenced herein for Currency Alpha and Macro strategies prior to October 1, 2018, the date that the Currency Alpha and Macro Strategies team joined Mesirow, occurred at prior firms. Any chart, graph, or formula should not be used by itself to make any trading or investment decision. Any currency selections referenced herein have been included to illustrate the market impact of certain currencies over specific time frames. The inclusion of these is not designed to convey that any past specific currency management decision by MCM would have been profitable to any person. It should not be assumed that currency market movements in the future will repeat such patterns and/or be profitable or reflect the currency movements illustrated above.

Comparisons to any indices referenced herein are for illustrative purposes only and are not meant to imply that a strategy's returns or volatility will be similar to the indices. The strategy is compared to the indices because they are widely used performance benchmarks.