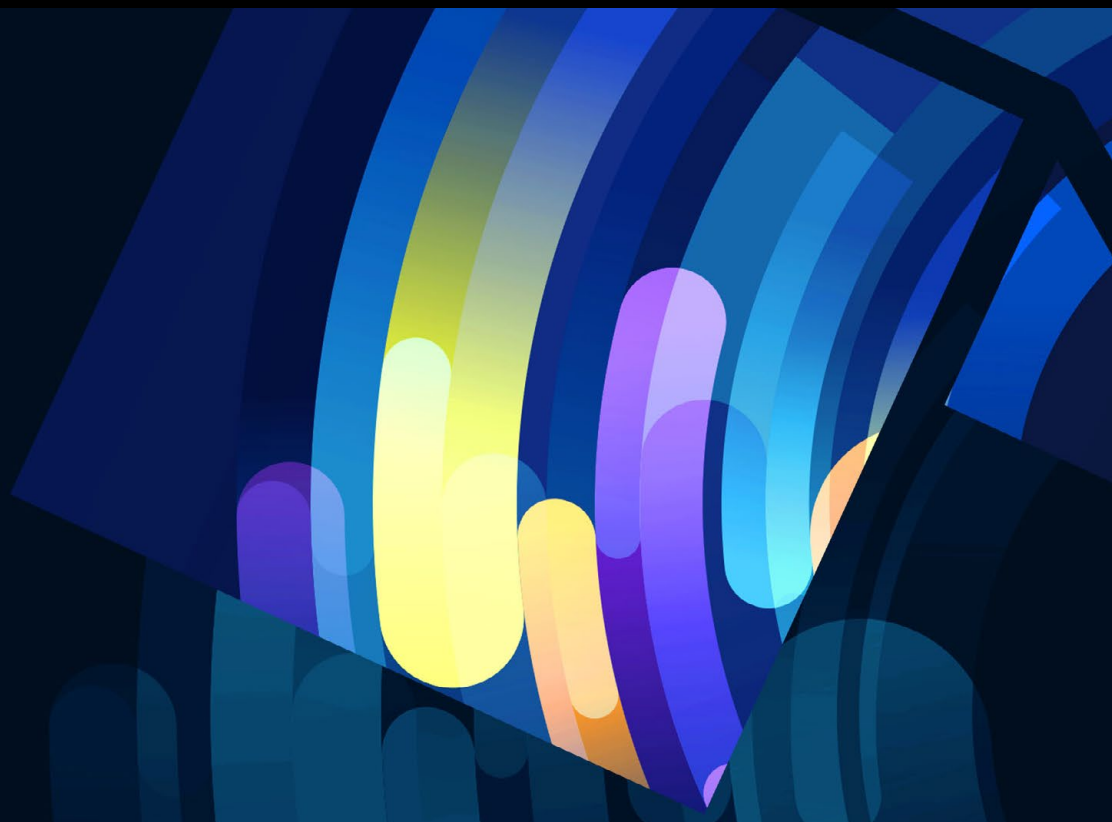


Shining Light into the Machine Learning Black Box

August 2023

Time to read: 12 minutes

While machine learning models are increasingly being used to predict returns, little progress has been made to leverage models for factor portfolio attribution, a critical component in analysing systematic portfolio returns. What are the limitations of existing methodologies, what is Shapley value, how is it applied and how can a SHAP based performance attribution framework be combined with non-linear models to interpret black box models and to better understand factor contributions?

Author

Wanjun Jin, CFA
Senior Portfolio Manager,
Man Numeric

Introduction

Modern portfolio management has increasingly embraced machine learning (ML) models to predict returns due to their ability to capture complex interactions between factors. The drawback is that the end result is often something close to a black box model with highly optimised outputs. This means it is often challenging to understand a model's predictions and decision-making. To counter this, model interpretation or attribution techniques are used to attempt to explain the rationale behind model predictions and to uncover the features that contribute most to the outcome. However, little progress has been made so far to leverage ML models for factor portfolio attribution, which is a critical component of systematic portfolio investment. Without this evolution, it is difficult to accurately understand which factors are affecting portfolio returns. While ML models may have improved returns, investors are currently somewhat blind to where those returns are coming from.

"While machine learning models may have improved returns, investors are currently somewhat blind to where those returns are coming from."

Existing linear factor attribution methodologies suffer from limitations such as a lack of ability to capture local interaction effects and the implied assumption of a singular global beta. Instead, we would argue that systematic investors need to look beyond the existing linear attribution models to find granular, local explanations of performance.

One solution is to use Shapley value. In this paper, we delve into what Shapley value is, how it can be applied to explain model outputs, and how we compute Shapley values using SHapley Additive exPlanations (SHAP) – a specific implementation of Shapley value. We also explain how a SHAP based performance attribution framework can be used for local and global portfolio attribution and introduce an innovative portfolio attribution system which uses Shapley value and SHAP to explain both the decision-making process and cross-sectional return variation at a local and global level. We also demonstrate the enhanced explanatory power of SHAP attribution by incorporating non-linear ML models such as XGBoost.

Why Change? The Limitations of Existing Linear Factor Attribution Methodologies

If ML models continue to provide reasonable returns, why worry about refining attribution methodologies? In short, because they are inadequate. Existing factor attribution methodologies such as time series regression, cross-sectional return attribution, and holdings-based attribution are based on linear models, making them unable to capture local interaction effects with the assumption of global linear beta.

For example, time series regression is limited by the dimensionality problem and the assumption of constant beta throughout time, making it less useful for dynamic portfolio management. Conversely, cross-sectional return attribution with a set of risk factors, as commonly used by risk model vendors, assumes that return generation can be attributed to a linear global factor model. Its close cousin, holdings-based attribution, estimates the exposure of the portfolio holdings to a set of custom factor portfolios. Although all three methodologies are based on the same linear factor return structure, they differ in terms of sophistication and customisation flexibility. However, these methodologies are not capable of capturing interaction effects due to the non-linear relationship between those independent variables.

"The Shapley value is a concept from cooperative game theory that measures the contribution of each player to a coalition game's payout."

The Solution: Introducing SHAP Portfolio Attribution

The Shapley value is a concept from cooperative game theory that measures the contribution of each player to a coalition game's payout. The four axioms of Shapley value[1] ensure that the payout distribution is fair when players can form coalitions and payout depends on the coalition's performance. Shapley value is the only payout method that satisfies these four axioms. Payout distribution is calculated based on the marginal contribution of a player by permutating through all combinations of the players.

1. The four axioms include: efficiency, nullity, symmetry and additivity.

> SHAP attribution can capture local interaction effects and other non-linear relationships that are beyond the reach of linear models employed in existing attribution methodologies.

The basic idea behind SHAP attribution is to explain every security's output (weight and return) as the sum of the contribution of each factor (aka feature in ML language). The set of factors used is defined by users. Examples include fundamental factors such as Barra factors, model scores or any metrics that can be used as inputs to a model that help predict the output. The SHAP value for a feature is the change in expected value brought by including that feature. This approach allows us to decouple the attribution from the underlying models used to explain portfolio holdings or stock returns, providing the flexibility to use any model we see fit to explain portfolio weight and cross-sectional returns.

With such an approach, we should be able to explain the decision-making process and cross-sectional return variation for every security in a portfolio i.e. a local explanation. SHAP attribution can therefore capture local interaction effects and other non-linear relationships that are beyond the reach of linear models employed in existing attribution methodologies.

Computing an exact Shapley value is computationally expensive and intractable when the number of features is large. Therefore, an approximate solution is necessary. One such way to compute an approximate Shapley value is through statistical sampling. SHAP is a popular implementation of the Shapley value and provides several approximation algorithms. It includes Kernel SHAP, a model-agnostic implementation of Shapley value and a few fast, efficient model-specific algorithm such as TreeSHAP for computing Shapley values. For all our studies, we use the SHAP package with Python API to compute the approximate Shapley value. Specifically, we use TreeSHAP when underlying models are decision tree based.

SHAP Portfolio Attribution Framework

We propose a portfolio attribution framework based on SHAP implementation of Shapley value. Attribution is the process of explaining the performance of a portfolio. Performance is the sum of product of each security's weight and return held in the portfolio. Investors have control over the portfolio weights but no control over the return. Therefore, portfolio attribution needs to explain both how investment decisions are made and what drives the returns. In general, investors use their proprietary models and a set of factors (style, industry, or country factors) to explain investment portfolio weights and cross-sectional return variation.

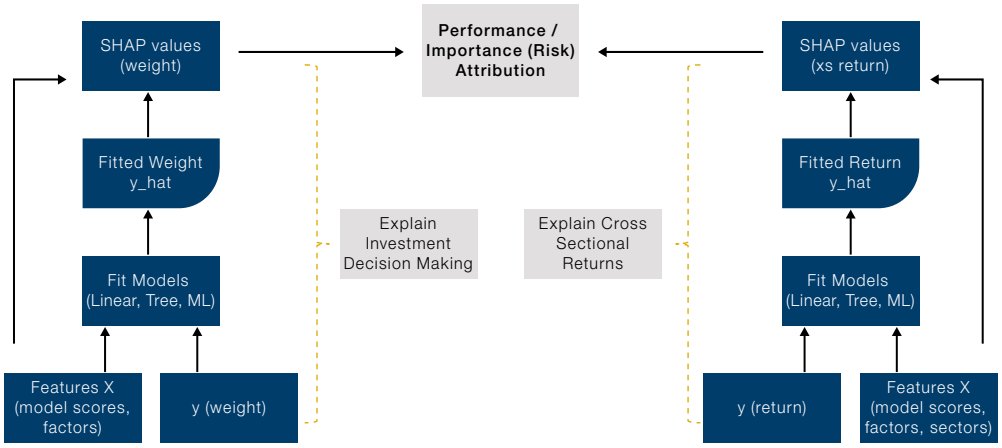
The model-agnostic property of Shapley value allows us to separate the models used to predict the output and the attribution of the factors used to explain the predicted value. Non-linear ML models can be used for explaining both decision-making and cross-sectional return sources. Linear models impose a strict global structure and embed a causality assumption, while non-linear ML models only require correlation and association.

For global interpretation at a portfolio or group level, we can aggregate SHAP values bottom-up from the security's Shapley values. In addition to consistency between local and global explanation, this approach offers flexibility and customised aggregation.

Illustration of SHAP Attribution Framework

The following diagram shows the framework for SHAP attribution. We fit two models based on user supplied factors, one to explain the decision-making process (i.e. the weight of a security in the portfolio) and one to explain cross-sectional return sources. Empirically tree-based models offer much better explanation, consistent with our observations that non-linear interaction between factors plays a significant role in both investment decision making (as the portfolio is subject to various investment and liquidity constraints even when return forecast comes from linear factor models) and cross-sectional return sources.

Figure 1. Performance Attribution as the Sum of Security Weight Multiplied by Return



Source: Man Numeric. For illustrative purposes.

Portfolio return is simply the sum of each security's weight multiplied by its return. Figure 1 shows that the model-fitted weight can be written as the sum of SHAP values of factors used to fit the weight model. Similarly, model predicted return can be written as sum of SHAP values of factors used to fit the return. Therefore, the return attribution for each security is simply the security's weight SHAP multiplied by its return SHAP – this is a face splitting product in matrix terms. Global performance attribution is done by aggregating SHAP values from security level. It is worth noting Shapley value measures the marginal contribution, which implies that SHAP value for the return is an attribution of excess return when the model is fitted over a broad universe; and for weight, active weight is preferred as it is more consistent with the marginal concept.

Interpretation and Connection to Existing Linear Factor Attribution

In the previous section, we demonstrate a general performance attribution framework using SHAP values to attribution portfolio weight and security return. This leads to the following equation at portfolio level:

$$\text{performance attribution (marginal)} = \sum_i \text{face splitting product } (\text{weight\_shap}_i, \text{return\_shap}_i)$$

If we use **m** features to explain weight and **k** features to explain return, there will be **m*k** SHAP values to explain the performance of each stock as a result of the full expansion. If the residual, which is the difference between the model-explained value and the true value, is also included in the calculation, then there are **(m+1)*(k+1)** performance attribution items for each security. This can become unwieldy and uninterpretable, even with a small number of features, and therefore some aggregation is needed. This framework offers users full control of how data should be aggregated and interpreted. We suggest two intuitive approaches and highlight their connection to existing linear factor portfolio attribution (expanded on in the Technical Appendix):

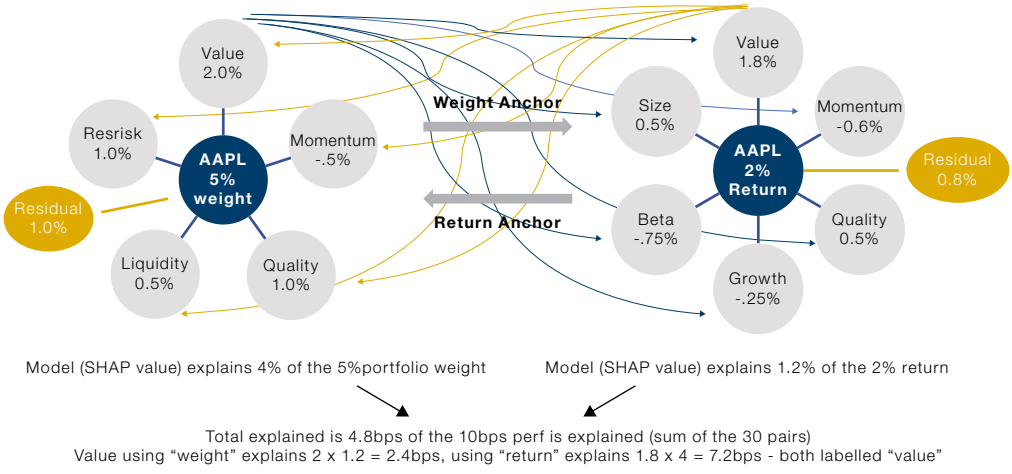
1. Aggregating from the weight side - for each feature in the m features used to explain the weight of a security, the **k** values used for returns are added up. With this aggregation method, there are **m** SHAP values to explain the performance of each security.
2. Aggregating from the return side - for each feature in the k features used to explain the return of a security, the **m** SHAP values used for weight are added up. With this aggregation method, there are **k** SHAP values to explain the performance of each security.

One benefit of the full expansion of weight and return SHAP values is to answer questions such as, for stock A, how much contribution comes from the weight driven by the Momentum factor and the return of the stock due to the Value factor? This is stock A's weight SHAP for Momentum multiplied by its return SHAP for Value. We can roll up this term across all securities to get a portfolio level answer.

Figure 2 illustrates the two methods. In this stylised example, Value is used to explain both weight and return, and the aggregation is done from either weight (grey lines) or return (yellow lines) side. We end up with very different explanations for how much performance we can attribute to Value. We would like to point out that attribution from the weight side is equivalent to the holdings-based attribution (HBA) when the underlying weight model is linear (with residual in security return included). Attribution from the return side is equivalent to the cross-sectional return-based attribution when the underlying return model is linear (with residual in weight included). Conceptually, the SHAP attribution we propose here is a more general framework that can accommodate the current factor attribution methods.

In this general framework, we have reconciled the often puzzling empirical observation that the same factor can have very different attribution using HBA versus returns-based attribution. Contrary to some claims, there is no inherent advantage of HBA over return-based attribution. They are both global linear model-based explanations with different perspectives; one uses factors to explain weight decision and ignores the return sources and the other approaches from the return side and ignores the drivers for investment decisions. Depending on the use case, one of the two aggregation methods may be preferred over the other, or sometimes interaction terms may also be of interest – as demonstrated by our previous example using Momentum as the weight source and Value as the return source. We can see in the diagram that the total "explained" performance remains unchanged no matter from which side we aggregate. We leave out the residual which is the "unexplained" part, and thus is one of the insights offered by the SHAP attribution framework.

Figure 2. Two Methods of Performance Attribution



Source: Man Numeric. For illustrative purposes.

"Combining powerful non-linear ML models such as XGBoost with SHAP based attribution improves portfolio performance attribution."

Enhanced Explanatory Power with SHAP and XGBoost

As previously mentioned, existing factor attribution methodologies are not capable of capturing interaction effects between different independent variables and the non-linear relationships among them. Our case study shows that combining powerful non-linear ML models such as XGBoost with SHAP based attribution improves portfolio performance attribution by capturing interaction effects between variables and allowing us to quantify the 'black box' non-linear relationship between factors. In our tests of attributing the performance of a diversified global portfolio with a few hundred securities to a set of features including our model scores and some common market variables, we can achieve R-squared of over 95% using tree-based models such as

XGBoost versus about 60% using linear models to explain portfolio weight. Similarly, tree-based models have consistently achieved much higher R-squared in explaining cross-sectional returns with a limited number of features compared to linear regression. The improvement comes mostly from being able to capture local interaction effects and other non-linear patterns at security level.

It should be noted that tree-based models have the tendency to "overfit" the data compared to linear regression, so regularisation is important to balance the additional explanatory power against overfitted, nonsensical results. Overall, the enhanced explanatory power and flexibility of SHAP attribution enable us to improve our understanding of portfolio management and investment decision-making at the most granular level.

Conclusion

Portfolio performance attribution is crucial in understanding and trusting the investment process. It is a challenging task for systematically managed portfolios as the "black box" nature of the models and the highly optimised portfolio construction make the output less interpretable. Shapley values combined with non-linear ML models are powerful tools for interpreting black box models and understanding model or factor contributions, providing insights into the decision-making processes, and realised return sources. Finally, the technique used here may offer potential research areas beyond performance attribution, such as the possible construction of risk models with embedded non-linearity using SHAP values for cross-sectional returns.

"Portfolio performance attribution is crucial in understanding and trusting the investment process."

Technical Appendix

Connection between HBA (holdings-based attribution) and SHAP attribution

$w_{shap,\ i,k}$ is the SHAP value for feature k and security i. The following is the SHAP value for security i's weight:

$$w_i = \underbrace{w_{pred.\ mean} + \sum_k w_{shap,\ i,k}}_{\text{Model predicted value for i: } \hat{w}_i} + \underbrace{\varepsilon_{w,i}}_{\text{residual: } w_i - \hat{w}_i} \quad (1)$$

For the linear model, the SHAP value for feature k is $\beta_k * (x_k - x_{k,\ mean})$. For de-meaned input, this is $\beta_k * x_k$ For HBA, weight regression we get the same β_k, if we use a score-weighted portfolio x_k as a factor portfolio. We drop the subscript i as the linear model has the same beta across all securities. Then HBA return attribution to feature k is (*exposure * factor portfolio weight * stock return*), as per the following:

$$\beta_k * x_k * r = w_{shap,\ k} * r \qquad \text{(holdings based attribution)}$$

For weight perspective SHAP attribution with a linear model, attribution is:

$$w_{shap,\ k} * \sum_k r_{shap,\ k} = w_{shap,\ k} * (r - \varepsilon_r)$$

The two differ only by a residual and would be the same if we include the residual. We choose not to include the residue focusing only on explained return. We can easily plug in the residual in the SHAP framework. HBA can be considered a special case of this framework with underlying linear model and residual return included.

Connection Between Cross Sectional Regression Attribution and SHAP Attribution

$r_{shap,\ i,k}$ is the SHAP value for feature k and security i. The following is the SHAP value for security i's return:

$$r_i = \underbrace{r_{pred.\ mean} + \sum_k r_{shap,\ i,k}}_{\text{Model predicted value for i: } \hat{r}_i} + \underbrace{\varepsilon_{r,i}}_{\text{residual: } r - \hat{r}_i} \quad (2)$$

For cross-sectional return-based attribution, performance attribution to feature k is portfolio weight multiplied by regression beta, which is basically $r_{shap,\ k}$ when a linear model is used in (2) assuming all features are centred. We have the following:

$$w * \beta_k * x_k = w * r_{shap,\ k} \qquad \text{(cross-sectional attribution)}$$

w and $\sum_k w_{shap,\ k}$ used for "return" perspective. SHAP attribution differs only by the weight residual ε_w. It is trivial to add back the residual so the two approaches are fully equivalent. Here we also drop the subscript i as beta from regression is constant across securities with a linear model.

Author

Wanjun Jin, CFA
Senior Portfolio Manager, Man Numeric

Wanjun Jin is a senior portfolio manager and researcher at Man Numeric. Wanjun joined Man Numeric in 2005 as a quantitative research associate. Prior to joining Man Numeric, she worked as a risk analyst for the gas trading desk at Sprague Energy, a Portsmouth, New Hampshire based oil and gas marketing company. Previously, Wanjun worked as a financial analyst for a leading private oil trading company in Houston, Texas. Wanjun received a Bachelor of Arts degree in economics from Peking University, a master's degree in financial mathematics from University of Chicago and a master's degree in economics from Tulane University. Wanjun is a CFA charterholder.

Important Information

This information is communicated and/or distributed by the relevant Man entity identified below (collectively the 'Company') subject to the following conditions and restriction in their respective jurisdictions.

Opinions expressed are those of the author and may not be shared by all personnel of Man Group plc ('Man'). These opinions are subject to change without notice, are for information purposes only and do not constitute an offer or invitation to make an investment in any financial instrument or in any product to which the Company and/or its affiliates provides investment advisory or any other financial services. Any organisations, financial instrument or products described in this material are mentioned for reference purposes only which should not be considered a recommendation for their purchase or sale. Neither the Company nor the authors shall be liable to any person for any action taken on the basis of the information provided. Some statements contained in this material concerning goals, strategies, outlook or other non-historical matters may be forward-looking statements and are based on current indicators and expectations. These forward-looking statements speak only as of the date on which they are made, and the Company undertakes no obligation to update or revise any forward-looking statements. These forward-looking statements are subject to risks and uncertainties that may cause actual results to differ materially from those contained in the statements. The Company and/or its affiliates may or may not have a position in any financial instrument mentioned and may or may not be actively trading in any such securities. Past performance is not indicative of future results.

Unless stated otherwise this information is communicated by the relevant entity listed below.

Australia: To the extent this material is distributed in Australia it is communicated by Man Investments Australia Limited ABN 47 002 747 480 AFSL 240581, which is regulated by the Australian Securities & Investments Commission (ASIC). This information has been prepared without taking into account anyone's objectives, financial situation or needs.

Austria/Germany/Liechtenstein: To the extent this material is distributed in Austria, Germany and/or Liechtenstein it is communicated by Man (Europe) AG, which is authorised and regulated by the Liechtenstein Financial Market Authority (FMA). Man (Europe) AG is registered in the Principality of Liechtenstein no. FL-0002.420.371-2. Man (Europe) AG is an associated participant in the investor compensation scheme, which is operated by the Deposit Guarantee and Investor Compensation Foundation PCC (FL-0002.039.614-1) and corresponds with EU law. Further information is available on the Foundation's website under www.eas-liechtenstein.li. This material is of a promotional nature.

European Economic Area: Unless indicated otherwise this material is communicated in the European Economic Area by Man Asset Management (Ireland) Limited ('MAMIL') which is registered in Ireland under company number 250493 and has its registered office at 70 Sir John Rogerson's Quay, Grand Canal Dock, Dublin 2, Ireland. MAMIL is authorised and regulated by the Central Bank of Ireland under number C22513.

Hong Kong SAR: To the extent this material is distributed in Hong Kong SAR, this material is communicated by Man Investments (Hong Kong) Limited and has not been reviewed by the Securities and Futures Commission in Hong Kong. This material can only be communicated to intermediaries, and professional clients who are within one of the professional investors exemptions contained in the Securities and Futures Ordinance and must not be relied upon by any other person(s).

Japan: To the extent this material is distributed in Japan it is communicated by Man Group Japan Limited, Financial Instruments Business Operator, Director of Kanto Local Finance Bureau (Financial instruments firms) No. 624 for the purpose of providing information on investment strategies, investment services, etc. provided by Man Group, and is not a disclosure document based on laws and regulations. This material can only be communicated only to professional investors (i.e. specific investors or institutional investors as defined under Financial Instruments Exchange Law) who may have sufficient knowledge and experience of related risks.

Switzerland: To the extent the material is made available in Switzerland the communicating entity is:

- For Clients (as such term is defined in the Swiss Financial Services Act): Man Investments (CH) AG, Huobstrasse 3, 8808 Pfäffikon SZ, Switzerland. Man Investment (CH) AG is regulated by the Swiss Financial Market Supervisory Authority ('FINMA'); and
- For Financial Service Providers (as defined in Art. 3 d. of FINSA, which are not Clients): Man Investments AG, Huobstrasse 3, 8808 Pfäffikon SZ, Switzerland, which is regulated by FINMA.

United Kingdom: Unless indicated otherwise this material is communicated in the United Kingdom by Man Solutions Limited ('MSL') which is a private limited company registered in England and Wales under number 3385362. MSL is authorised and regulated by the UK Financial Conduct Authority (the 'FCA') under number 185637 and has its registered office at Riverbank House, 2 Swan Lane, London, EC4R 3AD, United Kingdom.

United States: To the extent this material is distributed in the United States, it is communicated and distributed by Man Investments, Inc. ('Man Investments'). Man Investments is registered as a broker-dealer with the SEC and is a member of the Financial Industry Regulatory Authority ('FINRA'). Man Investments is also a member of the Securities Investor Protection Corporation ('SIPC'). Man Investments is a wholly owned subsidiary of Man Group plc. The registration and memberships described above in no way imply a certain level of skill or expertise or that the SEC, FINRA or the SIPC have endorsed Man Investments. Man Investments, 1345 Avenue of the Americas, 21st floor, New York, NY 10105.

MKT008895/ST/GL/W