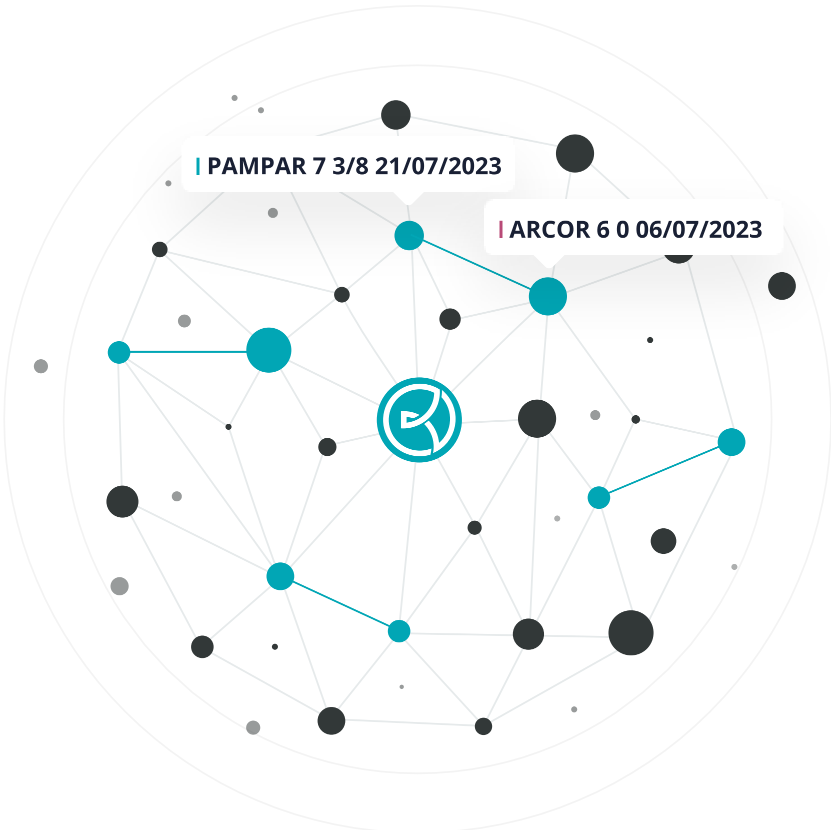
KATANA

Relative Value Trade Identification

September 2020

Executive summary

- Machine Learning algorithms in conjunction with big data technologies can perform calculations and analysis on millions of pairs, which is not feasible with spreadsheets and the typical technology stacks used in the fixed income space.
- The ability to analyse all possible swaps of a pair gives the confidence that the best swap can be found. More generally, with custom algorithms that search the entire space of relative values ideas, each investor/dealer can customize it to their mandate and strategy and still get an expanded view.
- For investors or dealers with quantitative teams already in place, the outputs of models can be used as input features on other internal models.
- On a timescale of 30 days, 91% of Katana's trade ideas result in compressions, with a median return of -11 bps.



Value is always relative

Every investment decision is anchored in a relative value view, whether this is the explicit intention or not.

When buying or selling a financial asset, the context of the risk-reward that alternative assets offer is always present. However, defining which are the relevant alternatives is not a trivial problem and when there are multiple assets to consider and a large number of candidates for alternatives the combinatorics grows exponentially. The reward for solving this problem is substantial, as identifying dislocations across multiple assets provide robust insight into the significance of price movements.

In its simplest formulation, relative value looks at a pair of assets. A pair is defined as two assets whose prices have historically moved together. When there is a significant change in the relative value of a pair, it's a signal to switch from the expensive to the cheaper asset, provided that the abnormal price relationship is not driven by a change in the fundamental risk factors of the two assets. Performing the inverse trade once the difference has reverted to its historical mean then yields a profit.

Pair trading strategies are known to be profitable in a wide variety of market contexts, with several studies showing that relative value as a trading signal produces significant reliable returns in fixed income [1].

The relative value approach can be valuable in the bond market as well. The bond market differs from other asset classes in terms of its size. The number of outstanding bonds exceeds 100,000 instruments. A company can have one stock but can have multiple bonds issued with different characteristics. Identifying which pair is “relevant” and which two assets can be paired is also non-trivial. Firstly, we want to eliminate any spurious correlations and secondly, we want to consider both static and time-series information in the modelling. Lastly, we want to ensure that the methodology is robust and does not lead to overfitting. Taking a systematic approach with no a priori assumptions makes this a computationally intensive problem where Machine Learning (ML) techniques have been shown to perform well.

[1] Jefferson *Duarte, “Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller?,” The Review of Financial Studies, Volume 20, Issue 3, May 2007, Pages 769–811, 06 July 2006. https://doi.org/10.1093/rfs/hhl026*

Inside Katana's machine learning algorithm

Katana uses its proprietary machine learning driven algorithm, which analyzes each possible pair of bonds in a given universe and delivers a comprehensive overview of the most interesting signals from nearly 200-million bond pairings. With an accuracy of 91% Katana provides relative value insights and trade ideas for bonds across North America, Europe, APAC & Emerging Markets. Portfolio managers and traders can monitor their holdings or watchlists to identify actionable opportunities and see broader trends in the market faster and more precisely.

The focus areas of Katana's ML algorithm are:

1. Relevant pair identification - Systematic identification of relative value pairs considering all possible pairs in a universe based on bond reference data and historical price movements.
2. Signals of Relative value Dislocations - Algorithmic discovery of market dislocations of all identified relative value pairs based on historical price movements of the pair of bonds.

The algorithm combines both all the available price history of bonds and the bond reference data to identify the relevant value pairs with the highest dislocations. The algorithm does not make any a priori assumptions and therefore searches all possible combinations. Users on the platform can customize the results so it matches their mandate and views.

Data and Pre-processing steps

Katana's ML algorithm uses end of day Z-spread-mid and bond reference data from IHS Markit. Pre-processing and cleaning data steps are applied prior to modelling to ensure outliers and possible errors are eliminated from the daily feeds. All available overlapping historic data in a pair is used in the modelling stage. Katana's algorithm reports signals every day, defined as relevant pairs that are significantly dislocated.

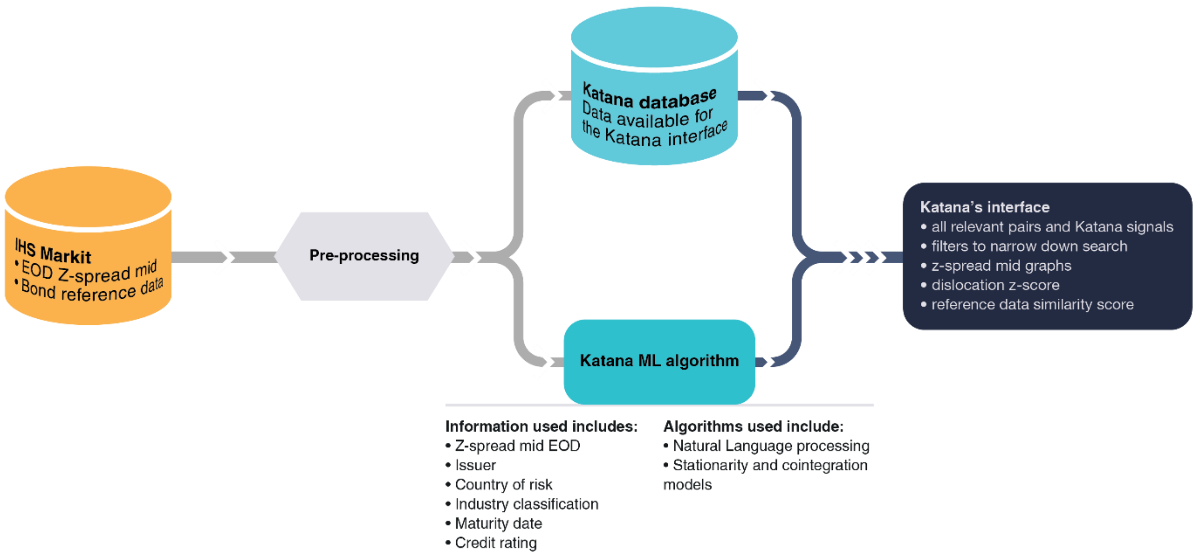


Figure 1: Data pipeline for the Katana platform and algorithm. Data is ingested once a day from IHS Markit, pre-processed and provided as an input in the Katana algorithm which uses reference and price data for the modelling step. Data and algorithmic output is provided to the users through the Katana platform.

Model training of ML algorithm

The subsequent step of the pipeline is to convolve both static and dynamic bond attributes into a machine learner which outputs daily signals of Relative Value pair dislocations as well as a list of all relevant pairs. The main advantage of ML algorithms over statistical models is that ML does not require to set assumptions and thus can adapt to the insights from the data.

Relevant pair identification

Spurious correlations are well known to arise between variables even when there is no relationship between them [2]. In fact, even two independent, stationary, autoregressive processes, regressed against each other, can often be found to have a high correlation when using ordinary least-square statistics, even though there's absolutely no relationship between them.

To address spurious correlations, a machine learner is trained to reliably identify attractive bond pairings based on weighted correlations of a bond's reference data, such as the bonds' issuer, country of risk, etc. The static information is selected with the collaboration of area experts of how they use and make decisions. Similarity scores are defined for each of the selected reference data using different approaches based on the data type, distribution and importance of each attribute including Natural Language Processing and one-hot encoding techniques. An overall similarity score is then constructed, aggregating the similarity scores of the different reference data. The reference data are convolved with each bonds' historic z-spread data, measuring both the cointegration and stationarity of each pair, to produce an initial subset of relevant pairs.

Signals of Relative value Dislocations

The second step of the algorithm evaluates whether the subset of relevant pairs selected on the previous step is dislocated. A signal is defined when a pairs' dynamics evolve beyond a threshold that is defined by an envelope using the past historical information (see Figure 2). The algorithm utilizes up to 5 years of overlapping price history between the bonds in any given pair. Over this period, a wide variety of statistical measures is calculated in multiple windows and over a range of horizons of different durations. These metrics are aggregated across all intervals in order to construct a holistic measure of how any given pair of bonds have performed in tandem throughout their history. The algorithm considers both of the absolute z-score dislocation of the pair, as well as its rate of change, to then identify the potential top-performing ideas. Katana provides both the daily signals, defined as “top ideas” on the user interface as well as all relevant pairs.

[2] Granger, C. W. J., and Newbold, P., “Spurious regression in econometrics,” Journal of Econometrics, 2, 111–120., 1974.

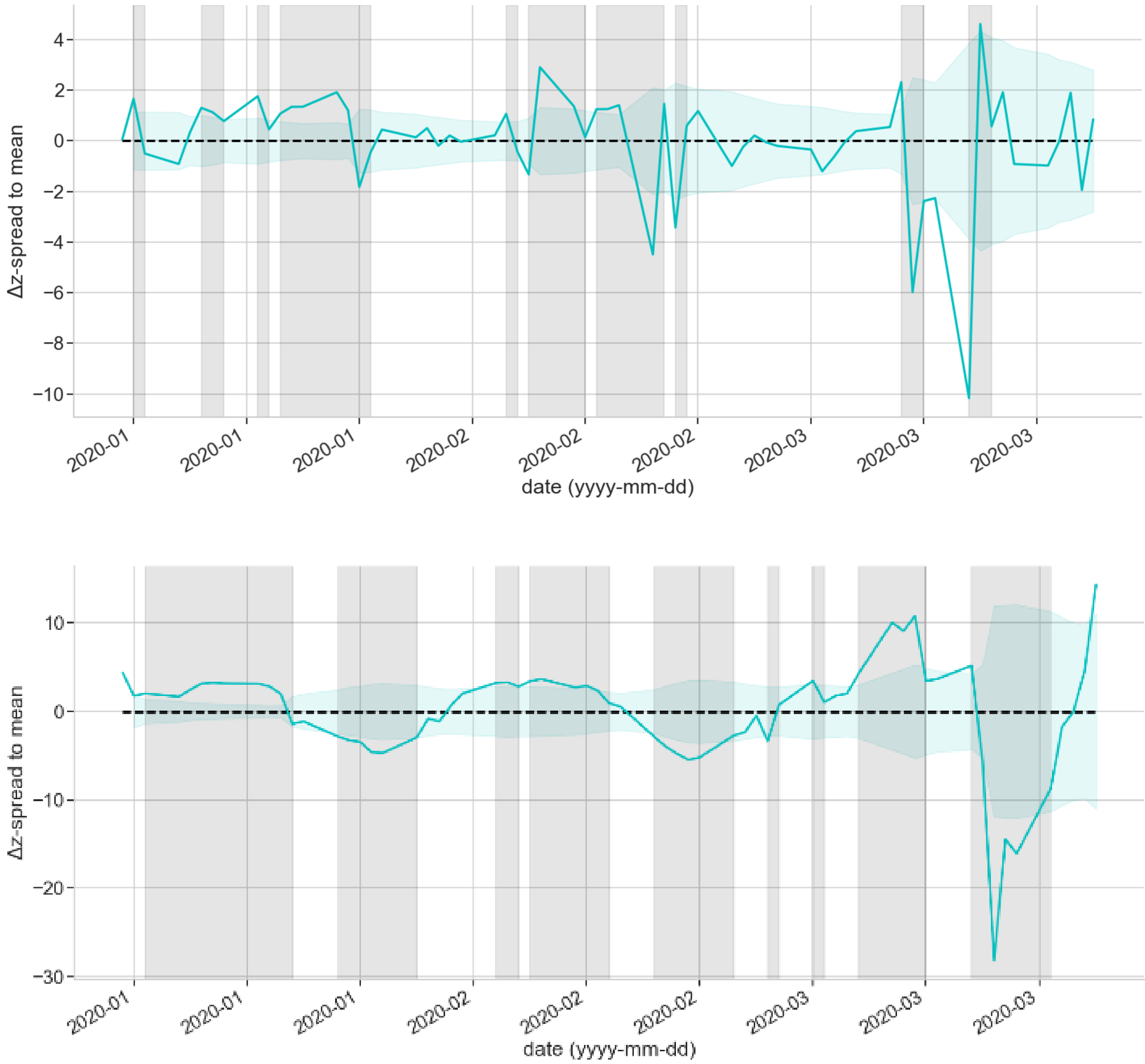


Figure 2. Example of two relative value pairs. The y-axis is the z-spread to mean difference between the two bonds and x-axis denotes time. The green shaded region tracks the z-spread difference history and denotes the weighted variance which can be used as a threshold to measure the typical duration of reversion cycles for that pair. The grey areas are periods identified by the model as dislocations and reversion events.

Signals performance

Performance Metrics

For each pair, the machine learning algorithm outputs a set of relative value signals at time t. To measure the performance of the signals generated, the compression on different time horizons (10, 30 and 60 days) is calculated. Compression is defined as the z-spread difference between the two bonds at time t+h, where h is horizon days, minus the z-spread difference between the two bonds at time t. When the compression is negative, it indicates a profit.

Backtesting framework

A robust backtesting framework evaluates the short- and long-term correlations and stationarity of two assets' histories over multiple windows of varying length. The purpose is not to imply the predictability about any two assets', but to reduce the likelihood that a recommended pair is the result of spurious correlation or bulk market movement that dominates the individual series' evolution over any particular short term period.

Evaluation is made on pairs generated on several distinct timescales, using data from 2018 to 2020. By combining the performance across timescales the effects of day-to-day market volatility that may push the perceived performance in one direction or another can be ameliorated, thereby providing a comprehensive and robust picture of the model's performance for ideas generated at any given time and which are expected to exhibit reverting behavior on varying timescales. The distribution of compressions of the Katana signals are compared against a selection of random pairs, looking at both median compression and percentage of signals with negative compressions.

Backtesting Performance

Overall, the Katana algorithm outperforms randomly selected pairings of bonds by a consistent margin across a wide variety of timescales (see Table 1 and Figure 3). On a timescale of 10 days, 88% of ideas generated by our model result in compression with a median return of -8.7 bps, compared to just 53% of ideas with a median return of -1 bps on an equivalent number of randomly selected pairs. On a timescale of 30 days, 91% of the trade ideas result in compressions, with a median return of -11 bps, compared to 56% of the randomly selected pairings which exhibit a median return of just -3 bps. On a timescale of 60 days 90% of the ideas exhibit compression, with a median performance of -11 bps, compared to 57% of the randomly selected pairings which exhibit a median return of -4.6 bps of compression.

Time Horizon	Random signals		Katana signals	
	Median compression	% negative compression	Median compression	Percentage of signals with negative compression
h=10	-1.0 bps	53%	-8.7 bps	88%
h=30	-3 bps	56%	-11 bps	91%
h=60	-4.6 bps	57%	-11 bps	90%

***Table 1**: Comparison of the median and percentage of pairs that have negative compression between random signals and Katana signals. The comparison is done looking at different time horizons (h), h measured in days since signal. Horizons of 10,30 and 60 days are displayed.*

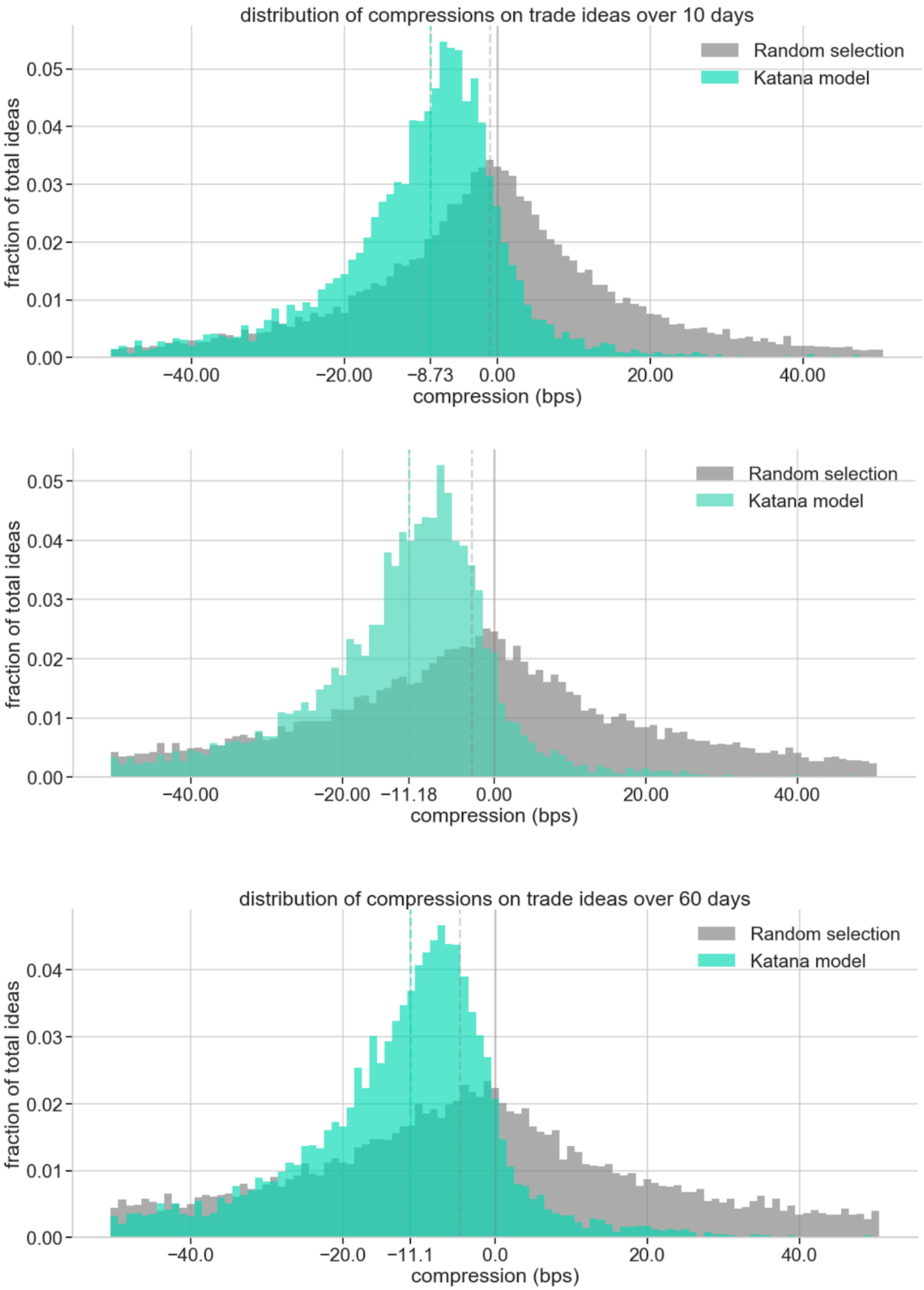


***Figure 3**: Comparison of the distributions of a random selection and the Katana signals. The three distributions look at different time horizons (h), h measured in days since signal. Horizons of 10,30 and 60 days are displayed.*

Disclaimer

Katana Labs Ltd. is not a regulated entity, registered as a broker-dealer or investment adviser in any jurisdiction. The information contained herein is not, and should not be construed as, an offer to sell or the solicitation of an offer to buy any securities. In addition, nothing contained herein is intended to be, nor shall it be construed as, investment advice, nor is it to be relied upon in making any investment or other decision. Prior to making any investment decision, you are advised to consult with your broker, investment adviser, or other appropriate tax or financial professional to determine the suitability of any investment. Different types of investments involve varying degrees of risk, and there can be no assurance that the future performance of any specific investment or investment strategy will be profitable or equal any historical performance level(s).

Neither Katana Labs Ltd., or any of its affiliates, nor each of their respective officers, directors, members, agents, representatives, employees, or contractors, shall be responsible or have any liability for investment decisions based upon, or the results obtained from, the information systems discussed herein. You assume all risks of loss resulting, directly or indirectly, from the use of such information systems, and all risks of loss from acting or refraining from acting as a result of any of the material contained in the foregoing presentation.

Katana Labs Ltd., a company incorporated under the laws of England & Wales with company registration number 12294570, whose registered office is at 12 New Fetter Lane, London, United Kingdom, EC4A 1JP.