EXEQUTION
Analytics

TIME SERIES SIMILARITY SEARCH : PREDICTING INTRADAY PRICE MOVEMENTS

TIME SERIES SIMILARITY SEARCH

EXEQUTION Analytics

The financial markets are characterised by their dynamic nature and rapid price fluctuations, presenting a significant challenge for traders and analysts aiming to predict short-term price movements. This report seeks to evaluate the use of time series similarity search algorithms as an approach to enhance intraday price prediction in equities markets.

In recent years, time series similarity search has emerged as a powerful tool for analysing sequential data, offering flexibility and effectiveness in identifying patterns and anomalies across potentially large datasets. Unlike traditional statistical methods that focus on overall trends and distributions, time series similarity search excels at identifying specific patterns within data sequences.

We explore the use of this technique to enhance intraday price forecasting using equities tick data in Japan. We seek to understand if price movements in realtime can be used to identify similar historical patterns and whether this information can be effectively modelled to predict price direction over the rest of the day.

RESEARCH OBJECTIVES

EXEQUTION Analytics

The Japan Stock Exchange starts continuous trading at 9am and closes at 3pm with an hour break between 11:30 and 12:30. Our goal is to investigate if applying time-series similarity search to a vector of prices observed during the morning session can improve our prediction of how stock prices will move later in the day. We generate a vector of prices for a given stock using 1 minute bins between 09:00 and 11:30 and compare this vector to equivalent price vectors observed across a similar universe of stocks over the previous two months of data, stored in a vector database. This is then used to generate a time-series of predicted price movements for the afternoon session between 12:30 and 15:00.

Our goal is to differentiate between price continuance and price reversion. We test the sensitivity of the predictive model by varying the underlying universe of stocks as the basis of our search database, varying the length of the training period in the opening session and the length of the predictive period in the afternoon session.

For the purposes of our report, we use KDB.AI, a powerful time-based vector database from KX to power our research.

KDB.AI

At ExeQution Analytics, we are huge fans of KX technology, with decades of experience using kdb+/q to power timeseries analysis at speed and scale. We were excited to explore and experiment with KDB.AI, a powerful time-based vector database designed to store and search vector embeddings quickly and easily.

We installed a KDB.AI server trial licence within our environment allowing us unlimited access to cores and storage and ensuring no data leakage. We were impressed with the speed of search; after the vector embeddings are inserted into the database, similarity searches covering large datasets take mere milliseconds.

KDB.AI supports a variety of indexing methods and similarity metrics, enabling a flexible approach to data storage and search algorithms based on project requirements. We found the inbuilt embedding layer TSC (time series compression) very useful, but appreciated the flexibility to use multiple embedding methods.

EXEQUTION Analytics

METHODOLOGY

Indexing Method

After comparing a number of indexing methods, we settled on using Hierarchical Navigable Small Worlds graph (HNSW). HNSW is a proximity graph which stores vectors as nodes in layers with connections to the nearest vectors in the same layer.

The advantages of this method are speed and extensibility : it was much faster to search, without sacrificing accuracy, and suitable for larger and extensible data sets. We rejected the IVFPQ method as it does not scale to large datasets as efficiently, due to the requirement to train for clusters. We also rejected the Flat indexing method as the retrieval performance was significantly slower.

Similarity Metric

Cosine Similarity compares the similarity between two vectors using the angle between them and ignores the magnitude of that direction. We chose this algorithm for our similarity metric for its superiority in matching the overall direction and movement of a trend.

Cosine Similarity also works well with windows of same size or constant length vectors.

Other choices evaluated but ruled out include L2 (Euclidean distance), which is better suited for anomaly detection, and Dot Product which places more importance on magnitude rather than direction.

Vector Embeddings

Transformed Temporal Similarity Search is a powerful compression tool for implementing vector searches with time-series data, but by the nature of compression and dimensionality reduction, some information is lost from the original data. TSS handles this by choosing to be lossy in noisy sections of time-series windows.

We chose this as a method of reducing the vector of prices into an 8 dimensional embedding to be stored in our database.

TSC is a recent introduction in KDB.AI and serves as a method to reduce storage space and decrease insertion time whilst preserving the pattern.

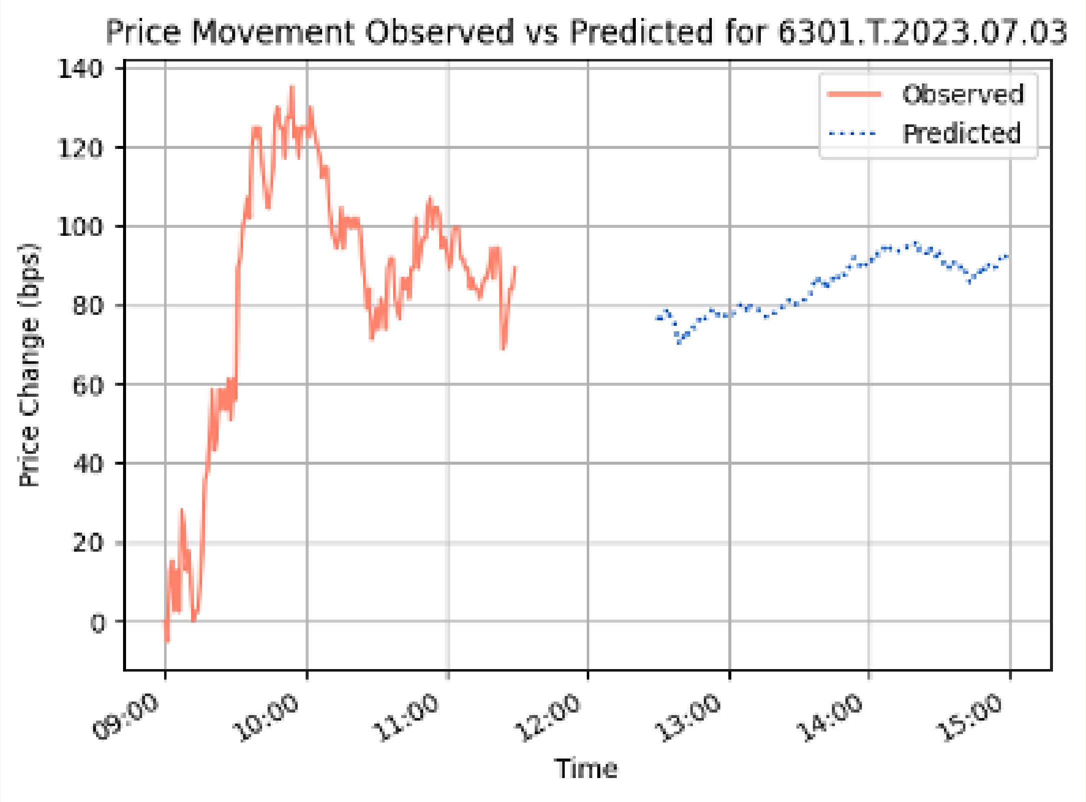
PRICE PREDICTION EXAMPLES

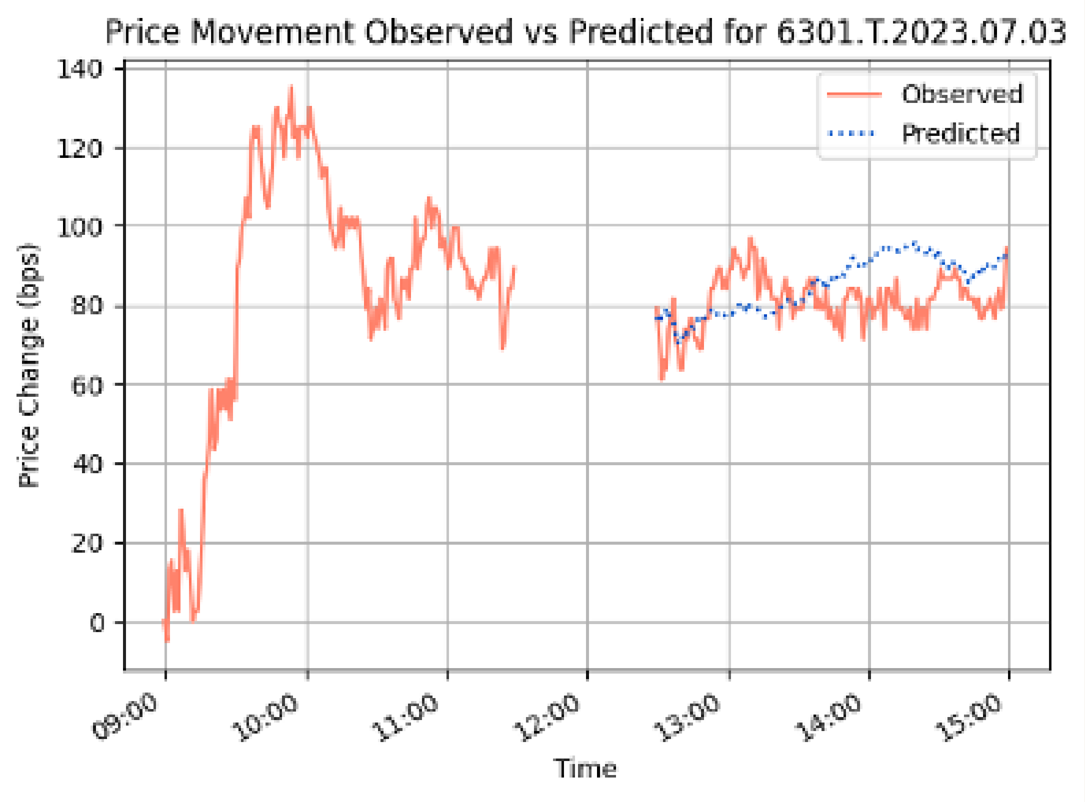
The graphs below demonstrates an example of our approach. We measure price changes in basis points in one minute bins between 09:00 and 11:30 as represented by the red line.

We search our historical database to identify similar patterns in price movements observed over the last two months and pull out the nearest 150 vectors as described in the methodology on the previous page.

We apply a model to these vectors to predict price movements for the afternoon session as represented by the dotted blue line.

The first graph shows the prediction for the afternoon made at 11:30 whereas the second graph compares observed price movements versus our prediction.





CONTINUANCE VS REVERSION

We measure the predictive power of our model very simply by evaluating if the direction predicted at 11:30 matches the observed price direction at end of day.

Because the Japan Stock Exchange closes for an hour between 11:30 and 12:30, this gives us a natural window to split the day up. We investigate how the performance of our model changes as we vary the length of the trading vector in the morning and the time to predict in the afternoon.

Our results show that the last thirty minutes of trading before the lunch break return the best results for predicting the direction of price movement in the afternoon, with the accuracy of the prediction increasing as the length of the prediction time increases.

Length of Training Vectors

Time to Predict	09:00-11:30	10:00-11:30	11:00-11:30
12:30 - 13:00	40.3%	42.4%	56.5%
12:30 - 14:00	50%	45%	58%
12:30 - 15:00	59%	58%	63%

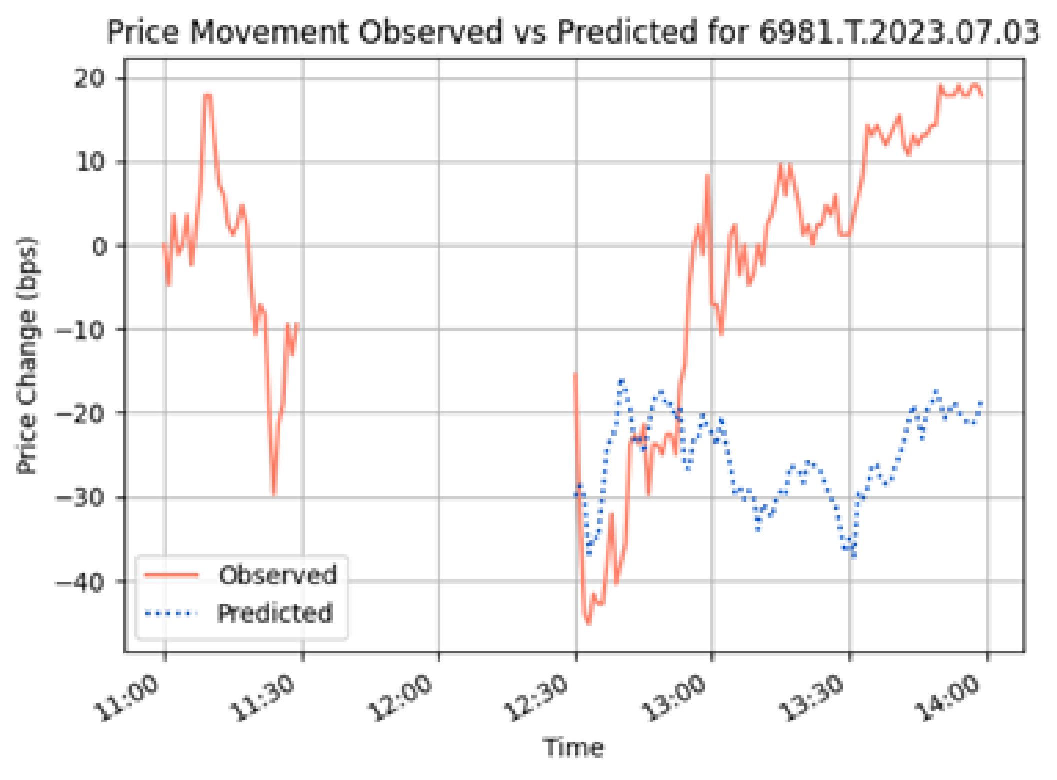
Within the universe of TOPIX 100 Stocks

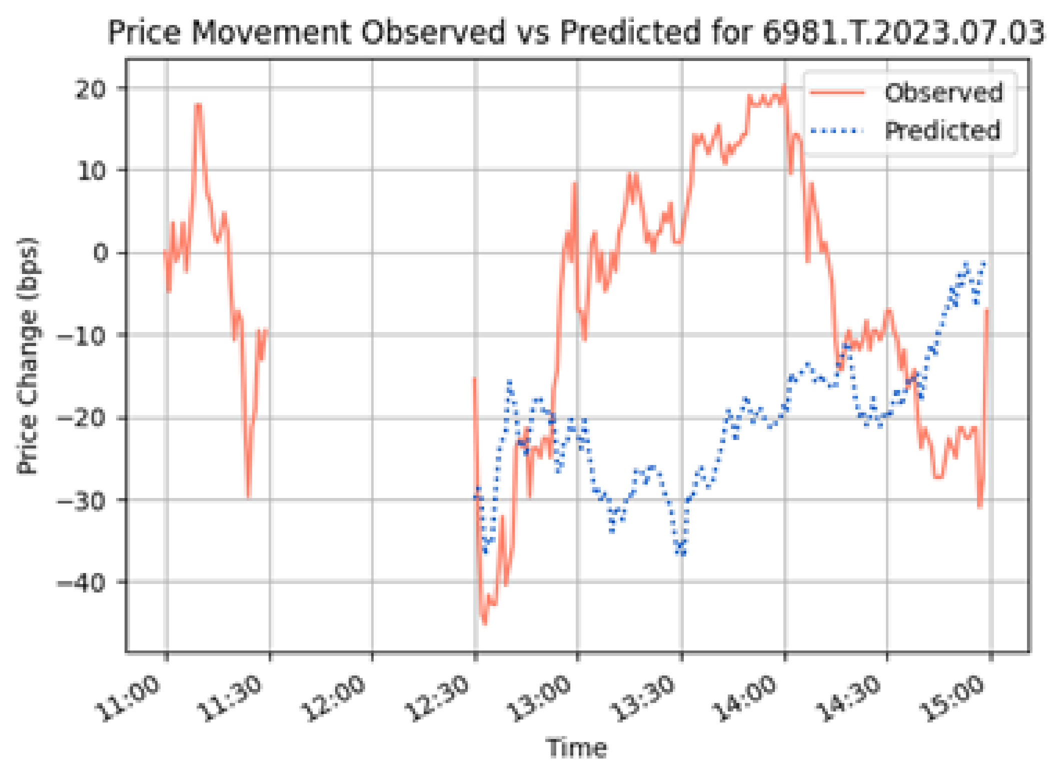
PREDICTIVE POWER

The graphs below show an example of where the predictive power of our model improves as the time to prediction increases.

If we evaluated performance of our prediction at 14:00, the price appears to be trending away from its intraday low, observed soon after the afternoon session began. A review at 15:00 tells a very different story, with the stock reverting to a level that is very close to our original prediction.

This is an interesting finding and counter intuitive to our expectations. In our experience using traditional statistical methods, longer term directional movements are typically more difficult to predict than short term movements.





STOCK CLASSIFICATIONS

To understand the sensitivity of our model to different classifications of stocks, we rerun our back tests against three different universes :

- Large Cap Stocks : Members of the TOPIX 100 Index
- Mid Cap Stocks : Members of the TOPIX 500 Index
- Small Cap Stocks : Members of the TOPIX Index (excluding the top 500 names above)

Whilst the accuracy of our model increases for large cap names as the length of the prediction period increases, this trend is harder to see in mid-cap and small-cap stocks.

Stock Classifications

Time to Predict	Large Cap	Mid Cap	Small Cap
12:30 - 13:00	56.5%	51.7%	44.5%
12:30 - 14:00	58%	48.5%	45.4%
12:30 - 15:00	63%	52.52%	45.4%

As our dataset includes only two months of data, it is natural that the more stable stocks have higher predictive accuracy and seems likely that the predictive power of less liquid names could be improved with a longer dataset.

CONCLUSIONS

In this report, we explored the efficacy of time series similarity search algorithms in predicting intraday price directional movements using equities tick data in the Japan Stock Exchange. Our research shows that by leveraging advanced similarity search techniques, we can identify short-term patterns and trends in price movements. This pattern identification can enhance the predictive power of intraday trading models, thereby improving decision-making and trading performance.

The analysis highlights a key differentiating feature of time series similarity search over traditional statistical methods, leading to improved accuracy for longer term intraday predictions versus shorter term horizons.

Future research could focus on exploring hybrid models that combine time series similarity search of price vectors with other market microstructure characteristics to further enhance predictive accuracy.

Overall, the application of time series similarity search in predicting intraday price movements presents a powerful tool for traders and analysts, offering a sophisticated approach to navigating the complexities of the equity markets.

EXEQUTION
Analytics

At ExeQution Analytics, we pride ourselves on our ability to harness the power of technology and market data to offer detailed analysis on the ever-evolving world of market microstructure and trading behaviours. Analysing financial trading data presents many hurdles and opportunities. The sheer volume of data, coupled with its granularity and complexity poses challenges to overcome but each piece of information holds a wealth of insights waiting to be unlocked. Amidst the chaos and noise, lies the potential of discovery and innovation. Every tick, every trade, and every market event offers an opportunity to uncover hidden patterns, exploit inefficiencies, and gain a competitive edge. We strive to identify these opportunities and transform data into actionable strategies.