CRACKING THE CODE

MACHINE LEARNING STRATEGIES FOR MINIMIZING INFORMATION LEAKAGE IN ALGORITHMIC TRADING

CORTEX EQUITIES

BNP PARIBAS

The bank for a changing world

CRACKING THE CODE

Machine Learning Strategies for Minimizing Information Leakage in Algorithmic Trading

When executing sizable orders, information leakage through trading operations is on every trader's mind. Different actions taken by execution algorithms leak information to various degrees. For orders that last many hours, the information leaked early on can have significant adverse effects on the overall execution quality. By using differentiated machine learning methods we are able to estimate the amount of information leakage that occurs during the execution of algorithmic orders and make real-time decisions, allowing us to reduce our overall market footprint while improving execution quality.

CORTEX EQUITIES

EVOLUTION OF MACHINE LEARNING IN EXECUTION

The use of sophisticated machine learning models inside execution algorithms has grown substantially over the last decade. These models generally provide predictions and analytics that guide algorithms through their decision-making process. At BNP Paribas, we have generally seen our machine learning models take on a few common trends in recent years.

COMMON TRENDS FOUND IN MACHINE LEARNING MODELS

PREDICTION UPDATING FREQUENCY

LOW → HIGH

A decade ago, algorithms often relied on predictions that were updated once a day, or even less frequently. Today our execution algorithms use a significant number of predictive analytics which typically are updated in near real-time, some even on each market quote update.

Such prediction models provide the most up-to-date information regarding the market conditions at present and are generally significantly more accurate than low-frequency models.

AMOUNT OF DATA USED TO CALIBRATE MODELS

SMALL → LARGE

The amount of data we collect and make available for the purpose of building quantitative models has grown steadily over time. We now have access to a large amount of market, order, and diverse alternative data. With the rise of both data volume and sophisticated data analytics platforms, data is more accessible and simpler for the end user.

At BNP Paribas, we place a high emphasis not only on data quality but also on data readiness. It is now crucial to be able to instantly bring together data from different sources to reconstruct a complete and exhaustive market and order view with fine granularity down to historical point in time. With this capability, machine learning models that consume a large number of different inputs can be easily trained, tested and cross-validated on datasets with millions of real-world samples.

MODEL PARAMETER COUNT AND COMPLEXITY

LOW → HIGH

With an increased amount of high-quality data, we have been able to grow model sizes significantly leading to enhanced predictability. This has largely mirrored the AI advancements over the past decade which can be characterized as a race to build increasingly larger models (with more parameters) with larger data sets (i.e. big data). Unlike the experience from a decade ago when more complex models could typically better fit the training data, but ultimately performed worse on unseen data than a simple regression model, recent results have shown that once the data hurdle is overcome, larger models show much better out-of-sample performance.

From our experience large machine learning models, based on either decision trees or neural networks, work very well in this setting and are playing a key role within BNP Paribas' execution algorithms.

PREDICTION TARGETS

FEW → DIVERSE

Not long ago, most models focused on predicting price returns and volume changes due to the fact this data tended to be the most accessible. Today we have models that can predict a large number of different targets, with some examples including:

- real-time fill probability of passive orders
- opportunity cost for posting passively
- hidden liquidity available on exchanges

This growing list of machine learning models provides guidance with verifiable efficacy to execution algorithms and continues to replace ad-hoc and handcrafted logic, which is becoming increasingly obsolete due to its time-consuming and error-prone nature.

UTILIZING MACHINE LEARNING TO ANALYZE INFORMATION LEAKAGE

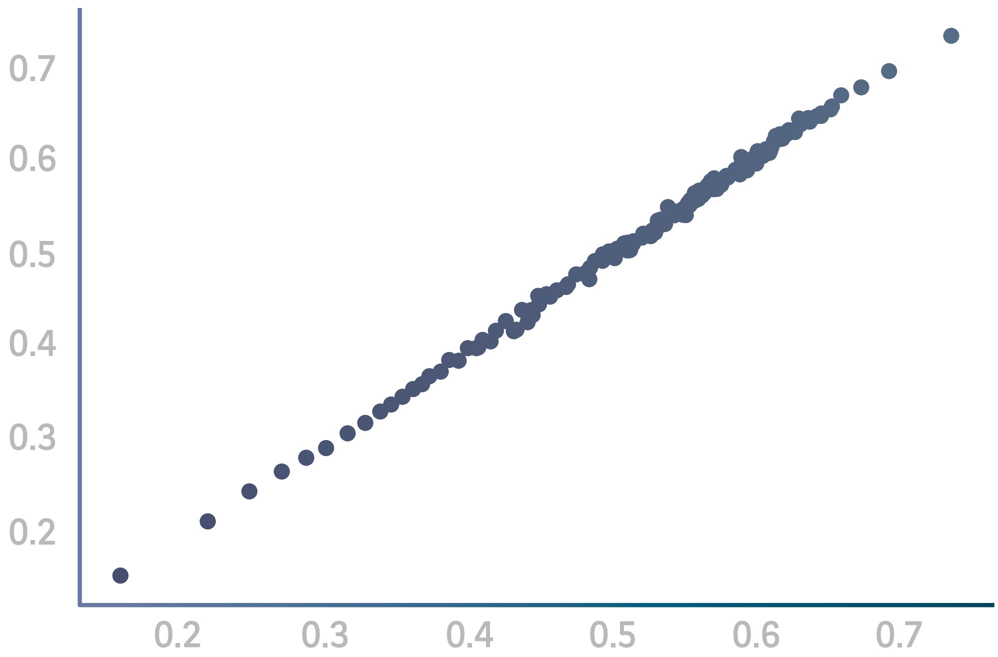
Imagine you are a buyer of a stock and during the life of your order you see the stock moves up by over 1% on relatively low volume. You most likely are left wondering how much of the move has to do with information leakage due to your trading activities and how much of that you could reduce. We are going to walk through this scenario and describe how we use machine learning models to answer these questions.

We start by looking at this problem systematically with the help of our machine learning toolkit. Our goal is to build a machine learning model that can predict the likelihood of the presence of an algorithmic order. Depending on how well this model works, and what information is most important for its performance, we may be able to statistically determine if information leakage has occurred, and provide guidance on how to reduce it in the future.

To start building the machine learning model, we will take a large number of BNP Paribas' own algorithmic orders that have a duration of more than 30 minutes. Leveraging our state-of-the-art analytics platform, we recreate the market condition and order status over the course of multiple points in time for each order within our sample set. This will become our positive sample set. Then we add an equal number of data points, which do not correspond to any live algorithmic orders, forming our negative sample set. We combine both the positive and negative sample sets and create a large number of input features to feed a machine learning model using a decision tree-based method. The result will then show that this model makes reasonably accurate predictions about the presence of algorithmic orders, and typically the predictions range between 30% - 70%. This result suggests that there is information leakage when executing a long-duration order since without that an overall correct model can only reliably predict 50% all the time.

> *OUR GOAL IS TO BUILD A MACHINE LEARNING MODEL THAT CAN PREDICT THE LIKELIHOOD OF THE PRESENCE OF AN ALGORITHMIC ORDER.*

Predicted vs. Actual Presence of Algo Orders



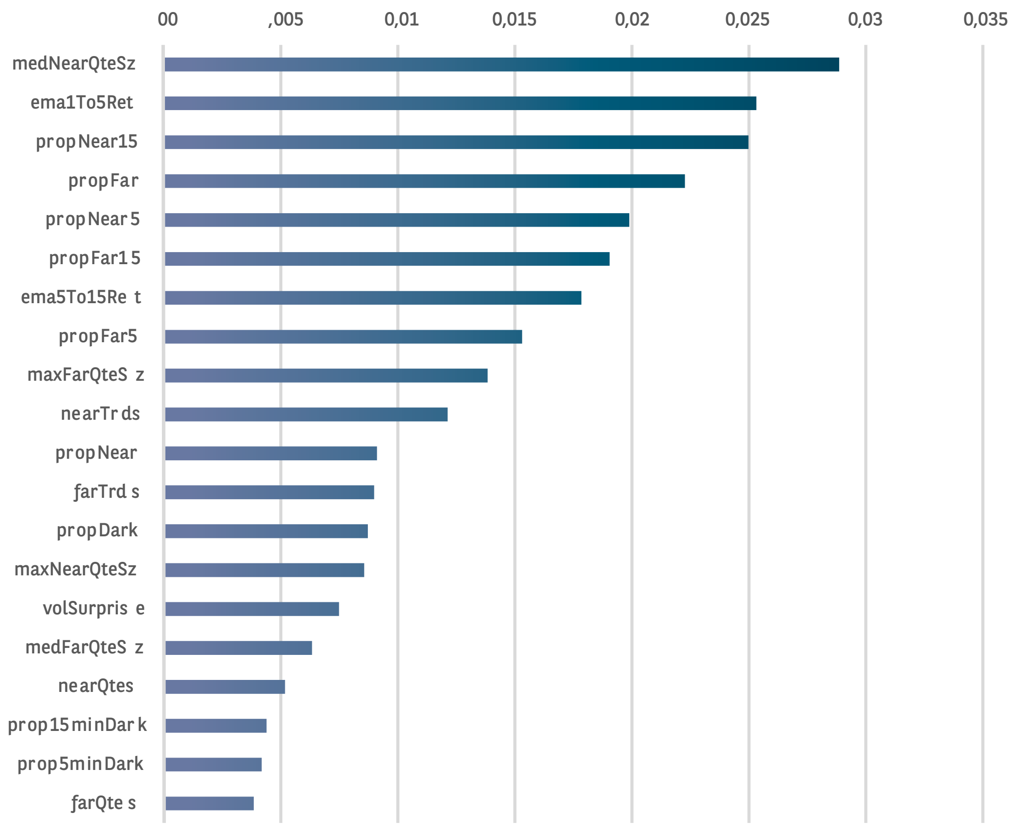
Source: Bloomberg, BNP Paribas

Considering that our model predicted the presence of information leakage, we can begin to look into how the information was leaked. First, we will look into our trained model to see which input features the model considers to be the most relevant indicator for the likely presence of a live algorithmic order.

Factors such as how a person trains a model with a different dataset can lead to a wide range of varying outcomes between us and our peers. Some of the most important input features may be linked to a particular signature leaked by a specific execution algorithm.

Redesigning the algorithm by adding some randomization or avoiding a particular action sequence may quickly reduce the leakage in such case. However, even after fixing such issues, the model will likely have sufficient predictive power, giving us a case to analyze the remaining features that are proving to be the most effective.

Sample Importance Ranking of Top Features Based on Contribution to the Model Prediction



Source: Bloomberg, BNP Paribas

We see some very intuitive features which are correlated with large recent price returns, newly established near touch prices, removal of substantial far side liquidity, and importantly, large near side posted size and order count. These findings point to a few of the inevitable actions an algorithm trading on an exchange will have to perform, such as post passively and take aggressively.

As explained earlier, we can more readily deal with the less intuitive features which would often indicate an unexpected signature leaked by an algorithm, and we can expect that techniques like randomization should help us reduce our market footprint. The bigger challenge is that no algorithm can avoid interacting with the market by posting and taking, and we have to ask ourselves what can be done to reduce the information leakage by leveraging our understanding of those more intuitive features and optimizing their footprints.

USING MACHINE LEARNING TO OPTIMIZE POSTING VS. TAKING & REDUCE INFORMATION LEAKAGE

We view the problem of optimizing instantaneous posting and taking decisions in the context of reducing the overall slippage of an entire algorithmic order (parent order). During the life of such a parent order, we need to continuously make posting and taking decisions. Imagine in the middle of executing the parent order, and as shown in the image below, we have executed some amount of this parent order and now we have to decide how to execute the next portion, which is the current active slice of the parent order. There is still some amount of the parent order left after the active slice is done.

It is worth noting that during the execution of the active slice, the accumulation of slippage at the parent order level mainly comes from **two sources:**

01 The average execution price for the active slice may turn out to be different than the current fair price of the stock

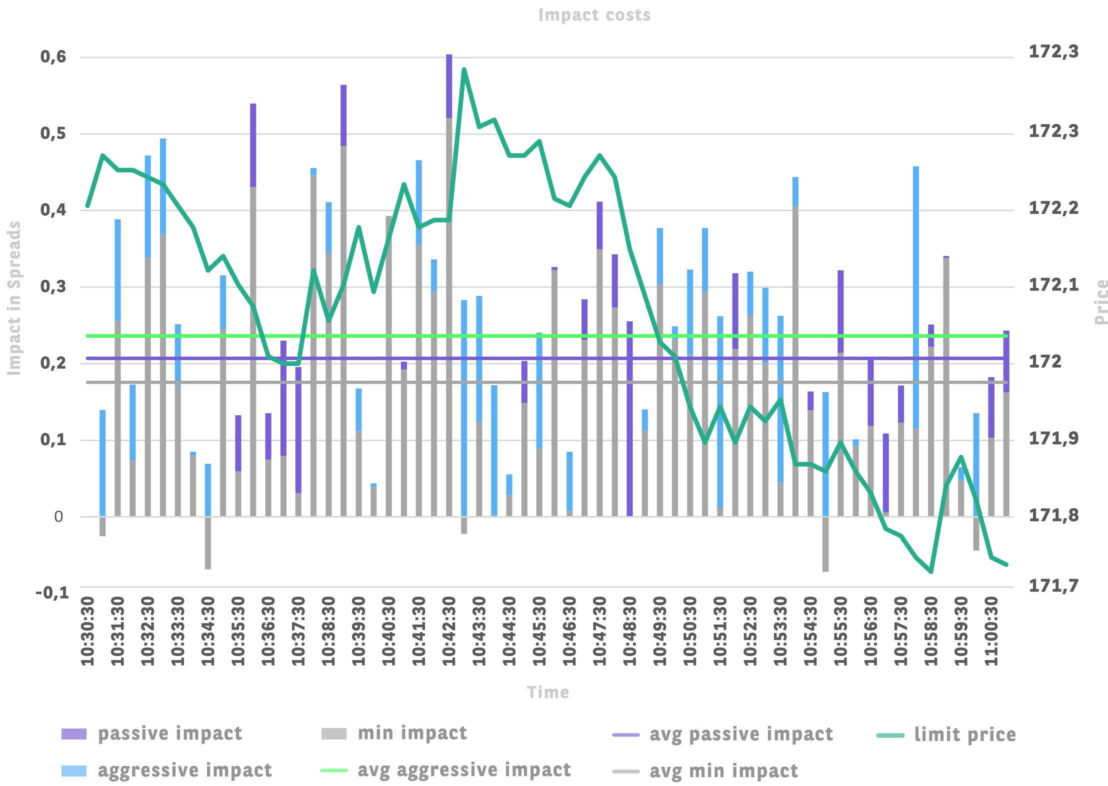
02 The stable price of the stock after we complete the current slice may turn out to be different than the current fair price of the stock

The relative importance of these two sources of slippage depends on the size of the active slice and the remaining quantity to be executed later. Depending on the actions we decide to take, such as posting or taking, and going to dark or lit venues, we tend to generate different expected results for each source of slippage. **To truly optimize our decision making process, we need to be able to predict both sources of slippage under each possible action. This is where machine learning comes into play.**

We first query our data analytics platform for a large number of algorithmic orders, pulling in both parent and child order information. We enrich the data set with market data observations, stock characteristics, child order action type (passive, aggressive, lit, dark, etc.) and add lookback windows to incorporate past decisions and market conditions. We then train the model to predict the actual slippage broken down into the two sources, previously mentioned. The result is a model that is able to make dynamic predictions on the ramification of posting and taking liquidity at any given time based on the market and order history up to that point in time.

We observe that the model is able to switch between posting and taking liquidity based on real-time market condition, with the goal of minimizing the parent order slippage. Overall we see a significant estimated reduction in slippage by taking the optimal set of decisions vs. either a post-only or take-only strategy.

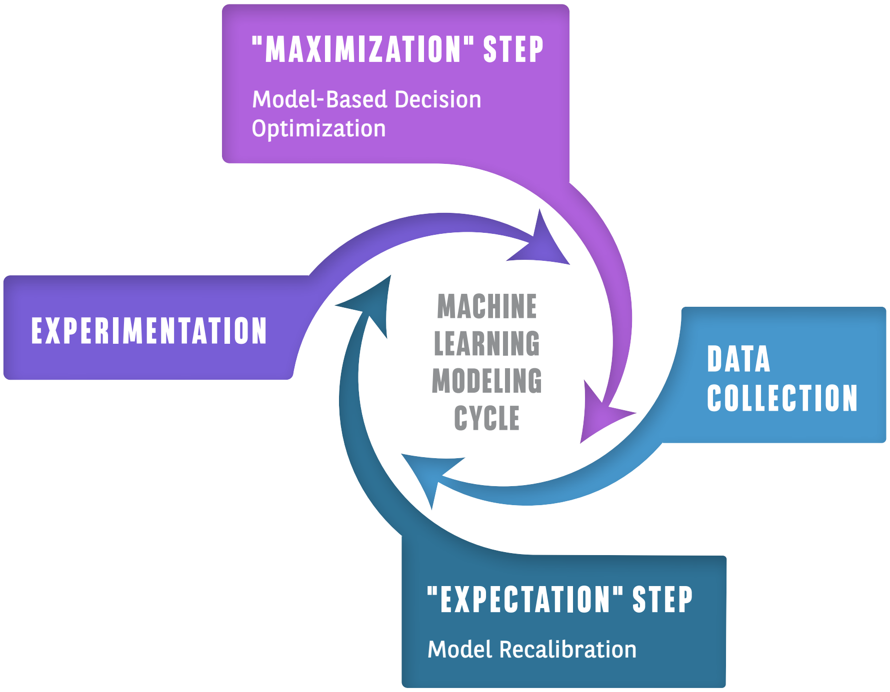
Passive and Aggressive Slippage Predictions during the Life of an Algorithmic Order



Source: Bloomberg, BNP Paribas

Machine Learning Modeling Cycle

At BNP Paribas, machine learning models are constantly evolving, within established controlled parameters, as they drive algorithmic decision making and new decision patterns that require model recalibrations. **The only way to keep models and algorithms both accurate and effective is through continuous integration of models and algorithms.** Having increased the link between a few traditionally separate functions such as quantitative research, product development, analytics, and data management, BNP Paribas' algorithmic execution business efficiently manages this machine learning modeling cycle by having all functions working "under the same roof" and within a single book of work. We believe this organizational advantage is the key to staying one-step ahead during the current machine learning and AI boom.



Summary

- BNP Paribas continues to refine and deploy machine learning models into our execution stack to provide real-time analytics and improve execution quality
- We analyze various sources of algorithmic information leakage with the support of machine learning techniques
- We have the ability to switch between passive and aggressive trading based upon the dynamic model predictions

FOR ADDITIONAL INFORMATION, PLEASE CONTACT

Cortex Equities Americas
Email: dl.cortex.equities.amer@bnpparibas.com
Phone: + 1 212-841-3600

FOLLOW US

CORTEX EQUITIES

BNP PARIBAS

The bank
for a changing
world