KX

WHITE PAPER

TO SPOOF OR NOT TO SPOOF, THAT IS THE QUESTION

Author: **Steve Wilcockson**

To Spoof or Not to Spoof, that is the Question

Think carefully, someone could be watching!

By Steve Wilcockson

Regulators have been busy. Organizations such as SMBC Nikko, JP Morgan, Atlantic Trading, Natwest, and Robinhood have all recently incurred one or more significant fines, reputational damage and criminal prosecution – courtesy of the power and tenacity of regulators. A particular regulator focus has been the illegal activity of spoofing, with an increasing variety of asset classes where spoofing activities are being detected and prosecuted. This means regulated organizations must be more alert across all their markets.

Those regulated entities face other pressures too, for example ensuring rigour around their compliance processes, as in the case of Robinhood's cryptocurrency unit. Individual liability, enforced through the likes of the Dodd-Frank Act, MAR and SMCR, mean risks are personal as well as corporate. Prison terms frequently result, such as the case of Michael Coscia of Panther Energy Trading in 2015, the first criminal case to use the anti-manipulation authority provided in Section 747 of the Dodd-Frank Act to lay a charge of spoofing in the context of commodities transactions.

What is spoofing?

Spoofing is an illegal form of market manipulation in which a trader places a large series of orders to buy or sell a financial asset, such as a stock, bond or futures contract, with no intention of executing them. Rather, the trader–or "the spoofer"– misguides market participants on true levels of supply or demand on an instrument or instruments with a view to profiting from the resulting price movements. Several high-profile cases have come to the attention of regulators recently. Some were spoofing in its traditional form:

- At JPMorgan Chase & Co., precious-metals traders consistently manipulated the gold and silver market over a period of seven years and lied about their conduct to regulators who investigated them. Penalties are in the order of a billion USD

- Atlantic Trading incurred an Administrative Monetary Penalty Payment Order for Market Manipulation in JGB Futures to the tune of 42,850,000 Yen.

- Natwest, fined $35m, where one trader took advantage of the close correlation between U.S. Treasury securities and U.S. Treasury futures contracts. The trader then engaged in cross-market manipulation by placing spoof orders in the futures market in order to profit from trading in the cash market.

- SMBC Nikko has undergone an 18-month regulatory probe into spoofing activities undertaken by a prop trading desk to support the prices of stocks being traded as part of a block offer. It has led to massive reputational damage, arrests, suicide and scandal.

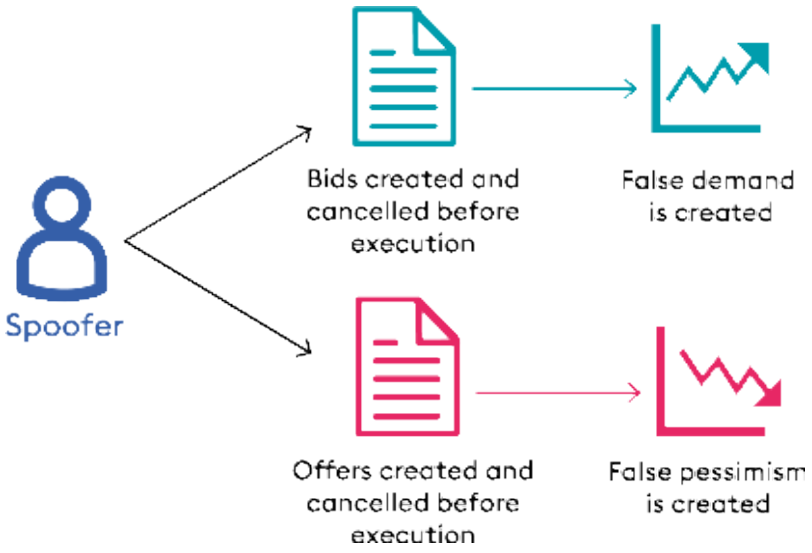

Fig 1: Characteristics of Spoofing

Why the Problem Remains

Such cases represent a clear failure to prevent instances of market abuse. How is that possible given recent investments in detection systems designed to help protect organisations from such activity?

Well, of course, technology on its own cannot solve ethics – personal ethics are, and always will be, an issue. However, technical evolution in how spoofing is conducted must be countered by similar evolution in how spoof detection systems operate. Traditional spoofing operates where the false, but manipulative, orders are placed on the same asset where the unlawful profits may ultimately be realised. Traditional systems may capture such instances well. Doing that one thing well does not protect from all outcomes, however. Systems can fail and have failed to identify more furtive manipulation, for example, realising the profit on a derivative by placing the spoofing orders on the underlying asset rather than directly on the contract itself.

In such cases, broader analysis and technologies must be applied to detection, for example those that look for correlations across assets, business units and markets. In itself, more monitoring means more onerous data and compute overhead, and team and workflow changes, but the problem is compounded by the network effect of so many potential combinations. Detection systems cannot be static. They need to be sufficiently agile, dynamic and flexible to handle greater data dimensionality – more data, more appropriate models. Robust statistics, machine learning and behavioural analytics can help quickly synthesize data, providing early indicators of suspicious activity to investigate further – and quickly eliminating false positives – and can add rigour to the actual event investigation. Static systems are not enough. Systems - and their owners in financial entities – need dimension-busting algorithms that can work with ever-increasing volumes and complexity of data, many needing to deliver compliance insights real and near-real time.

Scalable Analytics

Let's explore scalability requirements further as they apply to data and analytics. Data scalability requires the capture, processing and storage of vast and ever-increasing amounts of data across more assets and greater market depths, from level 1 to level 3. This gets augmented by internal and other forms of data, for example NLP analysis of unstructured phone records, or "behavioural" data sets surrounding trading (and other) activities of fund managers and traders. Systems must adapt to the evolving needs of the market, different data types and most importantly, scale for sheer volume, including in real-time.

Time-series data – collections of data organized through time - is the best base unit to capture that efficiency. Time-series data must be acquired and consolidated from multiple locations and in multiple formats, enabling ready processing to seek correlations, anomalies, and patterns. For example, when looking for spoofing and "layering" specifically, internal order/quote actions and trades are compared to market quotes, not just top of book but also their depth, and consolidated trades. The combination helps determine if deceptive orders and cancellations that form part of the strategy were actually marketable (i.e., likely to execute) at the time of transaction. This can consist of hundreds of millions of records when focused on equities or foreign exchange, yet options market analysis can increase the volume by orders of magnitude. The North American futures industry, for instance, generates over 100 billion order messages each day and the securities markets billions more.

The Value of Machine Learning, But Remember the Compliance Needs

The other challenge is that with so much data, you need techniques to make that data meaningful, to comprehend dimensionality and identify significant features. In plain terms, when looking for a needle in a haystack, select the right haystack to start with, and then minimize the disruption in the search. In such cases, machine learning can deliver more efficacy over such high dimensional data than rules-based solutions.

Yet for reasons of regulatory compliance and governance arising from the sector's misuse of complex models leading to the 2007/2008 Global Financial Crisis, rightly or wrongly, machine learning models have tended to only augment – not replace - audit and validation-friendly rules-based processes in financial services, including in surveillance. Rules-based models tend to be easier to replicate and explain, while "black box" data-driven machine learning models are harder to validate and audit, particularly those that deliver different outputs in response to the same inputs over time (as the model is retrained with new data).

It is, therefore, a fine line between model efficiency on one hand and need for regulatory validation and audit on the other. Machine learning efficacy is pretty much undisputed, including in surveillance situations. Machine learning techniques can compute over as many axes as there are useful features, easily. Their output can signal normal and abnormal patterns over high dimensions. However, simplifying those outputs as simple visualization and in an explainable and reproducible way for third parties can be non-trivial, hence regulatory-driven model risk concerns.

One popular machine learning method deployed across many industries and applications – from police surveillance to cybersecurity, from search engine recommendations to predictive healthcare and financial surveillance – is a Support Vector Machine (SVM). This is a great algorithm to identify and score features - measurable pieces of data– colurs and distances for example on an image, or, in the financial world, trade characteristics and trading patterns including fraudulent features across different data sets.

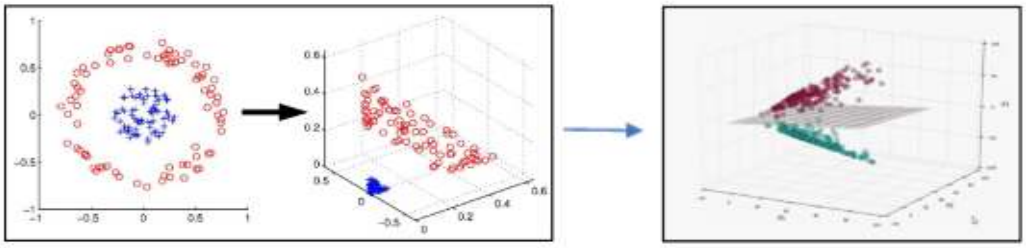

Fig 2: Support Vector Machines: Two classes of data that are non-linearly separable in two dimensions, but separable once they are transformed to a three-dimensional space, a hyperplane separating two classes of data (potentially normal and fraudulent transaction types) in three dimensions.

An SVM can identify a cluster of activity from a particular trader or group of traders demonstrating spoofing behavioural intent, e.g., patterns of bid cancellations before execution, that may be worthy of further investigation.

Many other algorithms and tests apply in addition to SVMs. Whatever the model approach - clustering or regression, linear or nonlinear, machine or deep learning algorithms - their parameterization is invaluable in financial surveillance. For spoofing, they can navigate well the frontiers and layers of normal and abnormal market activities, and assess balanced and unbalanced markets, where liquidity might be illusory or volume artificial.

The KX Surveillance Solution

Many current and legacy technologies struggle to process even a subset of the data at these volumes, let alone offer the flexibility to easily incorporate analytics methods that can help manage that dimensionality. Not so kdb+, on which the KX Surveillance solution is based, and which has been used for the past 20 years in processing huge financial datasets that can typically exceed petabytes (a petabyte is 1024 terabytes or approximately 1 billion gigabytes). KX processes the required volumes with ease in ways that more fixed, static solutions just can't, and what's more it centres on time-series, the essence of financial data relationships. Let's explore.

Pattern identification needs supporting metrics to further investigate and qualify the underlying activity. With spoofing and layering, for example, the concept of a marketable quote is important, as an order placed too far from the top of book would have little likelihood of being executed, and

> therefore unlikely to induce interest from other market participants. However, the marketability of a non-bona fide quote is not a static calculation as it varies from stock to stock and over time. The time-series nature of KX, and its programming language q, offers computational ability and agility, facilitating swift calculation of, for example, exponential moving averages and other time-based statistical benchmarks.

Temporal joins are another KX time-series feature particularly powerful in identifying various forms of manipulation. The first in-built temporal join, aj (asof), can be used to quickly identify the prevailing best bid/offer at the time of any internal order action. The second in-built temporal join, called the window join, is similarly invaluable in recreating a view of the market around the time of cancellations, at extremely granular time intervals (for example, one second windows over the six and a half hours of the open market). The flexibility to define and reconfigure such intervals in real time means compliance professionals can react swiftly to changing market conditions.

KX, through its time-series focus and its hyper-efficient language, q, adds algorithmic excellence and rigour, in ways that other more common programming languages cannot. This includes machine learning, with the Machine Learning Toolkit including a range of Clustering algorithms used to group data points and to identify patterns in their distributions. The algorithms make use of a k-dimensional tree to store points and scoring functions to analyse performance quality. Within the KX Surveillance offering, there is also a dedicated Support Vector Machine capability.

Conclusion

Spoofing is hard to detect. Its very existence relies on trades likely not being executed, and not executed across different markets and assets. As the examples indicated, regulators have shown that they have both teeth and tools to detect and punish such market abusing spoofers. Conduct ethics will forever challenge financial organizations and regulators, but dynamic technologies like the KX Surveillance solution can deliver data and algorithmic flexibility – and speed – to navigate highly complex data-sets, and massive volumes of them across fragmented organizations.

KX Surveillance arms organizations with powerful tooling not just for the now, but also for the future. With it, users can calibrate their parameters in real time to improve their detection quality and accuracy of live event monitoring, while the flexibility of the fastest, most powerful time-series database in the galaxy and time-busting utilities such as the replay engine eases retrospective investigation for all types of fraudulent behaviour and suspicious activity.

In this way, compliance professionals can reduce their organization's reputational risk, demonstrate good process, and avoid regulatory fines.

https://kx.com/solutions/surveillance

Steve Wilcockson **PRODUCT MARKETING AT KX**

Steve Wilcockson specialises in model-led and data-driven technologies in financial services. He has worked with quants, data scientists, data engineers, developers and business stakeholders on their research to production workflows on the buy-side, sell-side, front office, middle office, in insurance, and more.