CITADEL | Securities

SEPTEMBER 2021

Market Lens

Exchange Best Practices for Reducing Operational Risk at Broker-Dealers

Citadel Securities in Collaboration with the New York Stock Exchange (NYSE), Nasdaq, Miami International Securities Exchange (MIAX), Members Exchange (MEMX), and BOX Options (BOX)

EXECUTIVE SUMMARY

In this Market Lens paper, Citadel Securities, NYSE, Nasdaq, MIAX, MEMX, and BOX synthesize lessons learned from managing operational risks that arise in the interactions between exchanges and their participants and identify two broad categories of operational risk which have the potential to disrupt trading. We then propose several key changes that exchanges can implement to help their participants to manage, mitigate, or remedy these risks. We believe that incorporating these best practices from the US exchanges, who are collaborating with this paper and manage operational risk well, and applying them to other markets and geographies can improve the stability of the markets ecosystem.

Risk #1: An individual firm's people, systems, or tools act or malfunction in a way that creates substantial errors while trading on an exchange

- Proposed remediation:
  - Build latency-neutral exchange controls
  - Enable granular risk controls below the firm level
  - Increase transparency and functionality of exchange-level controls

Risk #2: Firms' risks can interact with and be complicated by exchange architecture decisions, impacting exchange or participant stability

- Proposed remediation:
  - Enact policies, procedures and design principles that minimize or discourage excessive and/or improperly formatted messages
  - Institute self-match prevention controls

FOREWORD

Exchanges sit at the heart of our capital markets, facilitating price discovery, liquidity formation, and risk transfer. Given their central role in our markets, exchanges also can play a critical role in helping their participants manage and reduce operational risks. This paper will focus on key operational risks that can arise in trading, and how measures at exchanges can help Broker-Dealers manage such risks and reduce other operational burdens.

The first and perhaps most familiar risk is of an individual firm, which trades on an exchange, and whose people, systems, or tools act or malfunction in a way that triggers or creates substantial errors. A second risk arises from technology decisions that may increase the operational burden of firms, which could impact exchange or participant stability. Each type of risk has the potential to cause damage not only to the specific market participant affected but also to the wider financial system.

Following some high-profile incidents, greater attention has been paid in recent years to implementing controls that prevent trading system malfunction risk at the individual trading firms who own this risk. But capital markets globally have focused less on safeguarding the interface between individual firms and exchanges in the event an error at an individual firm does still arise, which plays an important role in stopping damaging and disruptive events before they occur or limiting the effects of those events.[1] Exchanges here have a vital role to play in helping their participants manage these risks.

Citadel Securities has written this paper in collaboration with several major exchanges, including NYSE, Nasdaq, MIAX, MEMX, and BOX. The paper lays out two key operational risks for Broker-Dealers related to trading system malfunction risk and risks arising from the interaction with exchange technology, as well as several best practices that the exchange collaborators to this paper have developed to help their participants manage these operational risks. We believe that exchanges and their regulators in other markets and geographies can enhance their capabilities based on these best practices, thereby strengthening the robustness and resiliency of the global financial system.

[1]Additional information on how exchanges can respond to their internal operational risks can be found in Citadel Securities' Market Lens: Recommended Best Practices to Enhance Exchange Recovery & Mitigate the Impact of Outages

OVERVIEW OF BEST PRACTICES FOR MANAGING KEY OPERATIONAL RISKS

The following table summarizes best practices for exchanges to support their participants in managing operational risks.

Best practice	Why it is important	Important features
No latency penalty for using risk controls	• If there is any latency disadvantage for using an exchange-based risk control, participants may not use it for fear of being at a competitive disadvantage	• Controls should be integrated into the workflow regardless of use of the control • Controls should be rigorously tested to ensure that there is no latency penalty in practice
Granular controls / risk limits	• Firms typically trade from multiple desks or strategies but controls are typically only applied at an aggregate firm-level • Similarly, brokers provide market access to multiple clients and need to manage risk on a client-by-client basis • A firm-level risk limit is therefore difficult to quantify and to use	• Firms should be able to set controls at the appropriate level for their activity: on a firm, desk, client, strategy, etc., level
User-friendly and transparent controls	• To increase control adoption, participants must be able to understand and manage controls offered by exchanges for each connectivity arrangement (e.g., session)	• A controls user interface should identify all controls on an exchange • Participants should be able to manage controls and reset them if required (with appropriate safeguards in place)
Functionality to identify all orders and cancel if required	• Participants who are not aware of their current market positions or outstanding orders (e.g., due to a technology issue) face significant risk	• Participants should be able to immediately evaluate all open outstanding orders (individually or in aggregate) and cancel all orders if required
Session-dependent risk management	• Liquidity levels differ materially intra-day, making controls that lack time-based configuration granularity potentially less effective	• Exchanges should work with their participants to develop practices and capabilities to better manage the risks of different sessions and conditions
Controls to reduce excess and improperly formatted messages	• Exchange technology decisions may lead to an incentive for market participants to send additional messages or improperly format messages to gain a latency advantage	• Exchanges should have policies, procedures, and design principles that mitigate, minimize, or otherwise address excessive messaging as well as avoid processing of improperly formatted messages
Self-match prevention (SMP)[2]	• Market participants may unintentionally trade with themselves and be subject to fines or potential bans • Where available, SMP is often offered on a broker basis (meaning that self-matching can occur when clients use multiple brokers)	• SMP protocols that align with local regulations, practices, and participant needs should be available and configurable on an ID, sub-ID, or field level

[2] Exchanges are at different stages in SMP development with many offering some form of SMP. Many may require significant development to offer SMP across different MPIDs and solutions should be prioritized based on local market and participant needs.

INTRODUCTION

Financial markets have undergone a sea change over the past few decades, as technological advances and increased competition have reduced costs for both investors and issuers. These advancements have consequently shifted risk from humans to systems, thereby increasing the risk of technological errors, glitches, and system failures that can harm both individual market participants and the broader financial ecosystem. Such incidents can affect securities prices, including official closing-auction prices, which are relied upon as references for an array of economic activities (e.g., investment-fund flows, corporate mergers). They also can affect the solvency of major trading firms, which may lead to counterparty defaults that clearing and settlement providers will need to resolve.

Perhaps the best-known operational risk incident is the 2012 near-failure of Knight Capital Group, after a faulty code change for routing retail customer orders caused the firm to enter a flood of erroneous orders into the market shortly after the trading day began. This incident caused disruption in the pricing of many stocks and racked up large losses for Knight. There have been many other, typically smaller, operational risk events at trading firms. A firm that trades options, for example, sent erroneous trades to the market, resulting in substantial losses for the firm. Other well-known incidents have included the recent exchange outages experienced by multiple exchanges over the last three years.

Episodes like these have prompted many trading firms and exchanges to improve internal risk controls. Regulators, too, have stepped up oversight and rulemaking around operational risk events, like the SEC's implementation of Rule 15c3-5 (the "Market Access Rule"). The rule requires Broker-Dealers to have controls to limit exposure and help ensure compliance with regulatory requirements for both their own systems, as well as for systems accessing the market through direct or sponsored access. Similarly, many exchanges have made significant improvements and investments to promote exchange stability and risk management, but this has been uneven across the exchange landscape. The authors of this paper believe that exchanges globally should adopt — and many have adopted — a range of best practices, including:

- Implementing exchange-level risk controls that act as a secondary layer of protection beyond firm-level safeguards
- Ensuring that exchange technology decisions — how gateways, networks and matching engines handle orders and implement risk controls — minimize unnecessary message traffic
- Adopting policies and procedures for internal regulations and updates that take into account member feedback, minimize differential treatment, and allow time for preparation and testing by market participants

The remainder of this paper will review these key forms of operational risk and how exchanges can help their participants minimize them. These recommendations were developed based on the experience of liquidity providers in collaboration with exchanges. While the recommendations have broad applicability across different markets, they may not be relevant for all markets and all participants.

OPERATIONAL RISK GOES BEYOND INDIVIDUAL FIRMS

As discussed above, poorly controlled trading and systems at an individual firm can cause serious damage not only to that firm but also to the broader marketplace.

Trading System Malfunction Risk

On August 1, 2012, Knight Capital introduced faulty code for order routing, causing Knight to accidentally flood the exchange with orders.[3] Some 140 stocks were affected,[4] but only six moved by 30% or more, the threshold established after the 2010 "Flash Crash" for "clearly erroneous" transactions. Knight incurred significant losses, which threatened its ability to settle the trades. Only after Knight sold these positions at a $440 million loss,[5] and received a separate $400 million capital injection in which it ceded majority control of itself, was the situation resolved. If Knight had been unable to pay for the trades, the industry, through DTCC, would have been forced to honor the trades.

The Knight episode was a wake-up call for many trading firms, which soon began taking risk controls far more seriously. But even robust firm-level risk controls may not have prevented the ripple effects of Knight's code failure. The volume of trading on August 1 amounted to approximately one-third of Knight's average US equity trading volume, so likely would not have triggered a firm-wide risk limit. Only more granular risk limits would have helped identify and stop the responsible code. Exchanges providing participants a way to quickly identify all outstanding orders, combined with "kill switches" to immediately cancel all orders, also could have stopped much of the erroneous trading by enabling Knight to more readily understand outstanding orders and stopping them while it investigated the issue.

[3] *Capital Fiasco.* CIO Magazine, August 14, 2012
Knight Shows How To Lose $440 Million in 30 Minutes. Bloomberg News, August 2, 2012

[4] *Error by Knight Capital Rips Through Stock Market.* Reuters, August 1, 2012

[5] *Goldman Sachs Priced Knight Unwind at $440 Million.* CNBC, August 3, 2012

Risk Arising from Interaction with Exchange Technology
The second major form of operational risk is driven by the way exchange participants' systems interact with exchange technology decisions. The way exchanges configure their order-entry gateways, networks, matching engines and other infrastructure can — sometimes unintentionally — increase the operational requirements on the exchanges and their participants through higher message traffic and other burdens. This can increase the likelihood of errors at the Broker-Dealer or exchange level, as well as the potential impact to the broader market ecosystem. Indeed, while many exchanges, including the collaborators to this paper, were highly successful in navigating the elevated volumes that occurred in the wake of COVID-19, other exchanges globally have experienced outages over the last three years which are often driven by capacity constraints. The increased messaging traffic that can result from exchange technology decisions can only heighten this risk.

There are three major forms of this type of exchange technology risk: excessive messaging, message configuration, and the risk of a firm relying exclusively on internal systems to prevent self-matching. Exchange technology decisions can result in excess messaging by providing a latency advantage to participants who are more active. Some traders may, for example, intentionally increase the number of messages they use per transaction in order to gain an advantage. These scheduling advantages incentivize market participants to keep ports "warm" but the resulting traffic could strain the capacity of an exchange and potentially lead to outages, particularly during times of volatility and market stress, if the exchanges do not have the capacity to handle the additional message traffic. The behavior can also be self-perpetuating: if one participant gains a latency advantage by increased messaging, others will follow suit to remain competitive, potentially leading to unnecessary increases in overall message traffic. The second exchange technology risk that results in excess messaging is from participants configuring their messages in a way that allows them to gain a latency advantage.

The last key exchange technology decision that can drive significant operational risk and burdens is the risk that a firm unintentionally trades with itself. Not all exchanges offer basic self-match protection by automatically cancelling orders which would result in a self-match. Other exchanges do offer self-match prevention, but only at a member or single participant ID (MPID) level. While self-matching is primarily a compliance risk, the operational complexity required to prevent self-matching without exchange protections is significant. Because most modern exchange participants do not trade from a single system, attempting to track potential self-matches across systems adds substantial operational complexity and risk. This makes exchange-level self-match prevention capabilities a key control to help manage the operational risk that exchange participants face.

HOW EXCHANGES CAN HELP PARTICIPANTS MINIMIZE OPERATIONAL RISKS

As outlined above, risk controls implemented by individual firms to govern their own trading activity only go so far to address industry-wide operational risks. Exchanges can provide an important, additional layer of protection beyond what their member firms implement themselves. The following are recommended best practices — falling into two categories, exchange-level controls and technological hygiene that exchanges can adopt to supplement market-participant efforts and help ensure a fair and equal market for participants.

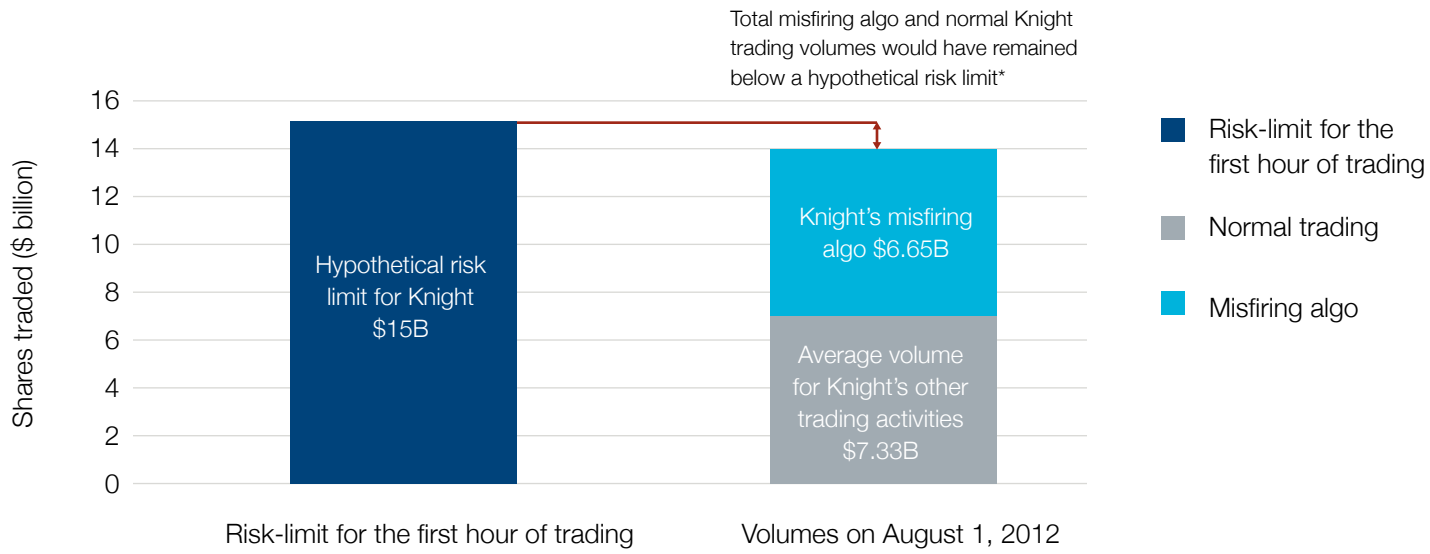
Ensure that exchange-level risk controls are truly latency-neutral for participants. Exchange-level controls often feature multiple options and settings. Insufficiently well-designed and tested controls can create what amount to penalties, driven by the time and computational power required to perform various stages of checks, if applied only to participants who opt-in to their use. This could produce incentives for *all* firms to avoid using *any* controls, for fear of suffering a competitive disadvantage. One way to address this, while maintaining choice for member firms, is to ensure orders follow the same order processing logic regardless of which options or features are enabled[6] — similar to how all co-located servers in an equalized data center incur the same cabling distance to the matching engine, regardless of their physical proximity to it. Additionally, exchanges should vigorously test controls to ensure no latency penalty exists in practice. Exchanges should actively publicize the net-neutral risk controls.

Allow and encourage members to tailor risk controls more-finely than at only the firm level. As discussed, firm-wide risk controls can be ineffective in stopping runaway algorithms or other errors restricted to a single strategy or trading desk. A more specific limit at Knight likely would have stopped its algorithm before it erroneously amassed nearly $7 billion in positions, but a firm-level one may not have helped. Trading firms make a significant portion of their profits on the most volatile days of the year, when volume can routinely be more than double that of an average session. Consequently, firm-level risk limits are typically set

[6] In rare instances, it may not be possible to create absolute latency neutrality because it would result in significant latency being added.

at significantly higher-than-average volume levels so they do not trip during these active days. In Knight's case, a hypothetical firm-wide limit could have been approximately $60 billion on a given trading day (which is roughly double Knight's 2011 average volumes). During the first hour of the day, which typically accounts for about 20-25 percent of total volume,[7] a proportional limit would have been $15 billion. On August 1, 2012, Knight executed transactions worth $6.65 billion within the first hour of trading. Assuming the rest of the firm was having an average day, it would have traded $7.3 billion in the first hour (based on the same 25% assumption). This totals $14.0 billion in volume — still short of the $15 billion hypothetical firm-wide risk limit.

Example: A Firm-level Risk Control Likely Would Not Have Identified Knight's Misfiring Algo



* Hypothetical risk limit is defined as 2x of Knight's average daily activity in 2011 ($29.3B). Expected volume during the first hour is ~25% of daily volume or $15B. Assuming other strategies were operating on an average day, they would have traded $7.33B in shares coupled with the misfiring algo's $6.65B in shares.

Such safeguards could be used not only by proprietary trading firms, but also by banks and agency brokers at the individual-desk level. Big banks, for example, may have different exchange connections for their electronic agency business, portfolio trading, and cash or high-touch desks. Banks that provide market access to clients could similarly supplement the individual risk limits they set for those customers with exchange-level controls.

Improve transparency of exchange-level controls so that configurations are clear and controls can be rapidly enabled, disabled, and adjusted across all activity. In many cases, exchanges' risk-mitigation functionality is not transparent to users, who cannot always tell whether individual controls and settings are active. Accordingly, exchanges should allow members to view and manage all risk control settings as appropriate. Firms should be able to set specific exposure levels and be alerted when nearing and hitting those levels. Tripping a threshold should lead to defined actions for each control, which can include canceling all orders, stopping new orders, etc. Firms also need to be able to rapidly reset controls as required and with appropriate oversight if the firm wants to continue trading after hitting its limits.

This would encourage participants to use the controls by giving them more discretion and could prevent firms from relying on human intervention to manage the controls. It could also enable participants to manage the outcome of controls on a control-by-control basis. For example, while in many cases a participant may wish to cancel all outstanding orders and exit the market if a significant limit was breached, there are cases where this could expose the firm to additional risk (if the exited firm was not flat upon order cancellation for example) and another response would be appropriate and preferred. The appropriate response to tripping a given control is a worthwhile discussion between an exchange and its participants. Access to such a system should be strictly controlled with appropriate security measures.[8]

[7] Sifma Equity Market Structure Primer. https://www.sifma.org/wp-content/uploads/2018/07/SIFMA-Insights-EMS-Primer_FINAL.pdf

[8] Exchanges also would do well to employ similar security governing firms' ability to create new order-entry ports.

Develop functionality that enables each firm to immediately evaluate the number of orders by session and — if required by a loss of connectivity or control — cancel all live orders. Such capability will minimize the significant operational and financial risk associated with market participants being unaware of positions and/or outstanding orders, such as during exchange outages or other operational issues.[9] The evaluation function could be as simple as the aggregate number of open orders for a session at a given moment.

Work with market participants to align on an approach to better manage different intraday liquidity conditions (such as pre-market and after-hours). Today, most risk protections are based on regular hours trading dynamics. After-hours conditions, however, are far different, with volume and liquidity often dramatically lower. Accordingly, erroneous order entry during periods of relatively small volume can have an outsized influence on pre-market and after-hours sessions. The industry has not yet reached a consensus on the best approach to manage these dynamics. Leading exchanges are partnering with their participants to develop the capabilities and practices to mitigate these risks.

Exchanges should have policies, procedures, and design principles that mitigate, minimize, or otherwise address excessive messaging. While market makers cancel messages in the healthy functioning of markets,[10] checks for cancellation rates, completeness, and quality will help exchanges identify and correct any participants who are sending excessive numbers of cancelled messages or improperly formatted messages.

Institute self-match prevention (SMP) controls to address self-matching. Exchanges should implement SMP protocols that align with local regulations, market practices, and participant needs. Protocols should be clear and transparent. To address markets where clients are trading through multiple brokers, protocols can be on an MPID or configurable-field level, with options for “cancel newest,” “cancel oldest,” and “cancel all.”

CONCLUSION

The electronification of markets has brought massive efficiency benefits to investors and issuers, and electronic markets have generally proven to be very stable and resilient. For example, trading, clearing and settlement infrastructure proved extraordinarily resilient in the US during the unprecedented volatility and volume that accompanied the onset of the COVID-19 pandemic in early 2020. Nevertheless, electronic markets also bring the risk of technology glitches and outages that can not only damage individual firms but also produce market-wide implications.

Individual firms, regulators, and exchanges have done much in recent years to improve industry protections against such incidents. The lessons learned from the world’s leading exchanges can and should be applied broadly to provide a critical additional layer of safeguards, both by implementing exchange-level risk controls and ensuring that order-entry, execution, networking and communications architecture discourage excessive messaging and put participants on a level playing field. Many exchanges around the world have already adopted some of the best practices identified in this paper for achieving this secondary layer of operational-risk protection. Nevertheless, further efforts are warranted to minimize operational risk incidents and better respond to those that will still occur.

[9] A recent whitepaper by Jane Street, *Dead Man’s Switch: Making Options Markets Safer Through Active Quote Protection*, may be instructive here.

[10] Citadel Securities published a Market Lens paper on healthy order cancellation in markets: “Why do Electronic Traders Cancel Orders?”