International High-Frequency Arbitrage for Cross-Listed Stocks

Cédric Poutré[1], Georges Dionne†[2], Gabriel Yergeau[2]

15 March 2022

Abstract

We explore mean-reverting arbitrage activities for international cross-listed stocks and develop a methodology to study the effect of information latency in high-frequency trading. The high-frequency strategy is a hybrid between triangular arbitrage and pairs trading. The strategy can be generalized to multiple cross-listed stocks environments without additional restrictions. Market frictions such as trade costs, inventory control, and arbitrage risks are considered. We test the strategy with cross-listed stocks involving three exchanges in Canada and the United States in 2019. The annual net profit with the limit order strategy is around US$6 million. International latency arbitrage with market orders is not profitable with our data.

Keywords: Latency arbitrage, high-frequency trading, cross-listed stocks, mean-reverting arbitrage, international arbitrage, supervised machine learning.

JEL codes: G02, G10, G11, G14, G15, G22.

Acknowledgements

Financial support from SSHRC Canada, Canada Foundation for Innovation, Canada First Research Excellence Fund, and IVADO is acknowledged. We thank Claire Boisvert and Mohamed Jabir for their competent support to this research project.

† Corresponding author: Canada Research Chair in Risk Management, HEC Montréal, 3000, Chemin de la Côte-Sainte-Catherine, room 4.454, Montreal (Qc), Canada, H3T 2A7. Phone: 514-340-6596. georges.dionne@hec.ca.

[1] Université de Montréal, [2] HEC Montréal

International High-Frequency Arbitrage for Cross-Listed Stocks

15 March 2022

Abstract

We explore mean-reverting arbitrage activities for international cross-listed stocks and develop a methodology to study the effect of information latency in high-frequency trading. The high-frequency strategy is a hybrid between triangular arbitrage and pairs trading. The strategy can be generalized to multiple cross-listed stocks environments without additional restrictions. Market frictions such as trade costs, inventory control, and arbitrage risks are considered. We test the strategy with cross-listed stocks involving three exchanges in Canada and the United States in 2019. The annual net profit with the limit order strategy is around US$6 million. International latency arbitrage with market orders is not profitable with our data.

Keywords: Latency arbitrage, high-frequency trading, cross-listed stocks, mean-reverting arbitrage, international arbitrage, supervised machine learning.

JEL codes: G02, G10, G11, G14, G15, G22.

1. Introduction

We study the profitability of mean-reverting arbitrage activities of international cross-listed stocks on two stock exchanges and a derivatives exchange which we apply to North American markets. The strategy is generalizable to a broader cross-listed universe. Our main research question is as follows: Is high-speed arbitrage profitable for High-Frequency Traders (HFTs) under strong competition, when all potential arbitrage costs and risks are considered?

Stock exchanges in different countries often use distinct market microstructures, whereas many large public firms employ cross-border listing to reduce their cost of capital and increase their access to liquidity. The current market structure of stock exchanges in North America and Europe is very competitive, fragmented, and fast (Biais and Woolley, 2011; Jones, 2013; Goldstein et al., 2014; O'Hara, 2015; Wah, 2016). Changes in regulation, particularly the Regulation NMS in the US and the IIROC rules in Canada[1], led to an increase in the number of trading venues, thus further fragmenting financial markets (Garriott et al., 2013; Chao et al., 2019). In 2019, there were more than twenty designated exchanges in North America. Further, competition related to trading fees, rebates, and colocation fees has increased significantly in recent years (Thomson Reuters, 2019).

The existence of multiple venues means that the price of a given asset need not always be the same across all venues for a very short period, opening the door to high-speed arbitrage across markets (O'Hara, 2015; Foucault and Biais, 2014). Given that this form of arbitrage can be done by creating portfolios that result from spatial arbitrage, traders must appraise intra-market liquidity

[1] Regulation NMS in the US: SEC Exchange Act Release No. 34-51808 (June 9, 2005). IIROC rules in Canada: CSE Trading Rules and the *Universal Market Integrity Rules*, of the Investment Industry Regulatory Organization of Canada (IIROC, 2015). See also The MiFID Directive in Europe: Directive 2004/39/EC of the European Parliament and of the Council of April 21, 2004 on markets in financial instruments.

and analyze the assets' serial correlation. Nonetheles, serial correlation dissipates very quickly, which further increases the possibility of high-speed spatial arbitrage (Budish et al., 2015).

In a market fragmentation context, traders need to search for liquidity across many venues in the same country or across countries. High speed can be crucial when there is strong competition. The ability of HFTs to enter and cancel orders very rapidly makes it hard for many traders to discern where liquidity really exists, which creates more opportunities for HFTs to exploit profitable trading opportunities.

International latency arbitrage opportunities may also arise because of different market models used in local exchanges, different regulations, transient supply and demand shocks, and the arrival of new local information that generates asynchronous adjustments in asset prices. These additional arbitrage possibilities terminate either when an arbitrageur exploits the new opportunity or when market makers update their quotes to reflect the new information (Foucault et al., 2017). However, local market makers are not always harmonized in real time. High-speed international arbitrage may then benefit all market participants (those with and without high speed) by reducing inter-market bid-ask spreads, a measure of market quality (Hendershott et al., 2011; Riordan and Storkenmaier, 2012). As a result, HFTs may even become inter-market makers who provide liquidity with their arbitrage activities, as we demonstrate with our mean-reverting strategy.

Whereas arbitrage forces should drive prices to attain an equilibrium, an exchange that acts as a price leader could attract a significant portion of order flow if the adjustment takes time. In this case, it is reasonable to assume that price discovery will tend to occur primarily in the original stock exchange of a cross-listed stock. For example, empirical evidence suggests that prices on Canadian and U.S. exchanges mutually adjust for Canadian-based cross-listed stocks (Eun and Sabherwal, 2003; Chouinard and D'Souza, 2003).

Considering a cross-country environment, we revisit latency arbitrage strategies, and propose a new model of international mean-reverting arbitrage with FX rate hedging. The present study is the first to examine stocks' cross-country mean-reverting arbitrage with FX rate hedging. We adopt the perspective of a unique temporal frame of reference, which means that we synchronize the data feeds of exchange venues and explicitly consider the latency that comes from the transmission of information between them and the data processing time. This approach, coupled with the inclusion of trading costs and trading risks in our methodology, generates more realistic results than those obtained in previous studies.

Our strategy is a hybrid between triangular arbitrage and pairs trading. It signals when the prices of cross-listed stocks deviate enough from their relative equilibrium that an economically viable arbitrage opportunity occurs. We construct a portfolio of synthetic instruments from pairs of cross-listed stocks of the same company traded on two exchanges and compute their relative spread (SPRD), defined as the ratio of the stock prices (our synthetic future) and a hedging position in the equivalent currency futures. The relative spread deviation resulting from a variation between the synthetic instrument and the hedging instrument is expected to be mean reverting. We analyze this intraday reverting behavior in detail for each pair of stocks between exchanges. Economically significant deviations of the relative spread from its target value could lead to arbitrage opportunities. We develop different arbitrage strategies to exploit these deviations and to demonstrate the potential profitability of mean-reverting arbitrage opportunities that exist between international exchanges.

According to Foucault and Moinas (2019), empirical studies must consider the effect of trading speed on each component of bid-ask spreads separately. These components are adverse selection costs, inventory costs, and order processing costs. We consider adverse selection costs via non-execution risk. Inventory costs are minimized by applying restrictions on the quantities traded and by precluding overnight positions. Order processing costs are considered via infrastructure and trading platform costs, and fees and rebates are also explicitly quantified. We then consider overnight positions to evaluate their effects on our results.

HFT technologies provide speed and information superiority (Biais et al., 2015; Foucault and Moinsa, 2019), but they introduce various costs such as high technology costs, trading fees and colocation fees (Bongaerts and Achter, 2021; Andonov, 2021; Baron et al., 2019, and Shkilko and Sokolov, 2020). Potential important arbitrage profits or realized opportunity costs described in the literature are often based on strong (and sometimes unrealistic) assumptions about the functioning of financial markets. The most prevalent costs are latency costs, direct trading fees, rebates on trading fees, and trading platform, colocation and proprietary data feed costs. Moreover, the closing of positions is not always coherent with the market reality. Mean-reversion risk, execution risk, and non-execution risk are additional cost components that may affect arbitrage profits. We propose a methodology to introduce all the costs and adjust our algorithm performance accordingly.

Given that high-frequency trading is very fast and competitive, the risk that the market will move between the time of observing an arbitrage opportunity and the time of the exchange receiving orders sent by a trader's algorithm (i.e., execution risk when using market orders, non-execution risk when using limit orders) is very hgh. Latency costs for the transmission and the processing of information may matter when exchanges are distant and assets quoted in different currencies are present. Moreover, because gains per trade for high-frequency traders are relatively small given their short holding periods, trading costs and rebates may be significant in the computation of net profits, especially when considering the enormous quantity of trades per day that HFTs perform. The colocation and the proprietary data feed costs are also significant at many exchanges, although they have decreased due to recent competition between exchanges. The fact that all these potential costs were overlooked may have generated an undeniable overestimation of the latency arbitrage profitability presented in the literature (Wah, 2016, Budish et al., 2015, Tivnan et al., 2019 and Dewhurst et al., 2019, among others).

As Chen, Da and Huang (2019) assert, the understanding of arbitrage activity in the empirical research is still limited. To our knowledge, we are the first to quantify the importance and the economic value of providing liquidity in the context of arbitrage while considering the limit order book (LOB) queue positions and limit orders instead of market orders exclusively. Our approach is consistent with the revisited HFT market maker definition proposed by O'Hara (2015): "HFT market making differs from traditional market making in that it is often implemented across and within markets, making it akin to statistical arbitrage."[2] Our mean-reverting strategy is a form of statistical arbitrage.

We test the model across three North American exchanges during the first six months of 2019: the New York Stock Exchange (NYSE) and the Chicago Mercantile Exchange (CME) in the United States, and the Toronto Stock Exchange (TSX) in Canada. We also discuss how the strategy is generalizable without additional restrictions to a much larger trading universe. As Gagnon and Karolyi (2010) note, over 3,000 companies had two or more listings in 2008, highlighting the importance of international arbitrage in market equilibrium. Our results report a net annual profit of about C$8 million (US$6 million) for 2019 for this international arbitrage activity, with 36 profitable cross-listed stocks that can be managed by one trader in a large trading firm. The 36 profitable pairs of stocks were selected from the 74 potential cross-listed stocks by using a dynamic

[2] See also Rein et al. (2021) and Krauss (2017) on statistical arbitrage.

decision tree model from machine learning. The gross annual profit was about C$19 million, and the main difference between the gross and the net annual profits is explained by latencies in the transmission and processing of information, and the non-execution risk because we used limit orders. Trading fees were consequently not important, yet rebates were significant. We also show that international arbitrage opportunities with market orders are not profitable mainly due to transaction fees and the execution risk associated with latency.

The rest of our paper is organized as follows. Section 2 presents the literature on arbitrage trading with high-frequency data. Emphasis is put on empirical studies that have estimated the profitability of this trading activity in an HFT environment. Section 3 outlines our strategy based on a mean-reverting model of arbitrage that can be executed with market orders or limit orders. We show the main differences between the two approaches with an emphasis on trading cost and rebates. Section 4 presents the methodology used to study the effect of information latency in HFT and how we consider the multiple arbitrage costs and risks associated with high-frequency arbitrage. Section 5 details the data from TSX, NYSE and CME and how it is managed. It also documents the real latency costs, as well as the trading fees and rebates, the colocation and the proprietary data feed costs at the TSX, i.e. the trading location used in the application of this paper. Section 6 is dedicated to our empirical results and Section 7 discusses the performance of our arbitrage strategy. Section 8 concludes the paper.

2. Related literature

Two main issues are at the heart of research on high-frequency trading (HFT): profitability and fairness in trading. Both are interconnected and require appropriate research approaches that are fundamental to understanding the behavior of trading participants and making adequate policy recommendations when necessary. The structure of exchange markets has been radically transformed by new technology over the last 25 years. HFT is executed by extremely fast computers, and software programming for trading is often strategic.

Liquidity and price discovery now arise in a more complex way, often owing to high speed. These changes have affected the market microstructure and the formation of capital in financial markets. They may also have reduced fairness between market participants, warranting new regulatory rules. However, conclusions on the private net benefits of high-frequency trading and its fairness are not always based on solid academic research, according to O'Hara (2015) and Chen et al (2019). In fact, the debate about the high-frequency trading arms race is still open (Foucault and Moinas, 2019; Aquilina et al, 2022).

Academic interest in latency arbitrage is a relatively recent phenomenon, and available studies have investigated it from different angles. The idea that price dislocations exist in fragmented markets is not new. In fact, contributions from the 1990s highlighted the issue in the US, even when market fragmentation was not as prevalent as it is today (Blume and Goldstein, 1991; Lee, 1993; Hasbrouck, 1995). More recent studies on that matter include Shkilko et al (2008) and Ding et al (2014). Soon after, other articles began mentioning the possibility for high-speed traders to exploit these market anomalies. Foucault and Biais (2014) and O'Hara (2015) both mention that HFTs can capitalize on latency arbitrage opportunities but they conclude that strong empirical evidence is still necessary.

Hasbrouck and Saar (2013) are among the first to investigate trading activities within the millisecond environment. Menkveld (2014, 2016) analyzes the behavior of a HFT who is a market maker. He shows that the market maker reduces price variations for the same stock on different exchanges by doing arbitrage activities across trading venues. Budish et al (2015) document the latency arbitrage opportunities between the CME and the NYSE from 2005 to 2011. They demonstrate that correlation between a pair of related assets breaks down as speed between quotes increases. They show that these breakdowns roughly yield an average of US$75 million a year from a simple latency strategy of arbitraging the spread of one pair of highly correlated assets: the S&P 500 exchange traded fund (ticker SPY) traded in New York and the S&P 500 E-mini futures contract (ticker ES) traded in Chicago. That pair of instruments had an average of 800 daily arbitrage opportunities during that period, and the authors notice that the arbitrage frequency tracks the overall volatility of the market, with a higher number of opportunities during the financial crisis in 2008, the Flash Crash on May 6, 2010, and the European crisis in summer 2011.

Budish et al (2015) also find that the median ES-SPY arbitrage opportunities duration declines drastically from 97 milliseconds in 2005 to 7 milliseconds in 2011, which is explained by the high-speed arms race led by HFT firms. The median profits per arbitrage opportunity remain relatively constant over time even though competition clearly reduced the duration of arbitrage opportunities. Budish et al (2015) mention the latency issue, but in a rather incomplete fashion. Their approach does not consider latencies such as the real information transportation cost between the two exchanges nor the information treatment time of a round trip. They may have overestimated the real profits generated by their trading strategy and underestimated the execution risk since they used market orders in their application. In their study, around 85% of latency arbitrage opportunities had a duration of less than 10 milliseconds in 2011. It is possible that this proportion has grown since then, given the technology developments since 2011. This emphasizes the importance of including new latency assumptions for our more recent period of analysis. Finally, as they mention, their strategy only considers bid-ask spread costs, whereas a richer estimate of arbitrage opportunities must also include, at least, exchange fees, and all latency costs.

Wah (2016) examines latency arbitrage opportunities on a larger scale for cross-listed stocks of the S&P 500 in eleven US stock exchanges in 2014. The strategy uses crossed market prices (i.e., when the bid price in an exchange is higher than the ask price in another exchange for the same stock) to locate arbitrage opportunities documented in MIDAS trades and quotes data from the SEC.[3] Considering one infinitely fast arbitrageur operating on these eleven markets, the author estimates that arbitrage opportunity profits were US$3.03 billion in 2014 for the S&P 500 tickers alone. However, round trip information transportation and information treatment time are not considered in the profitability of the strategy, nor are the other trading costs (except for the bid-ask spread cost, due to the use of market orders).

Tivnan et al (2019) and Dewhurst et al (2019) also examine latency arbitrage on cross-listed stocks in the US National Market System, but with data in 2016 from MIDAS. These two studies consider actionable dislocation segments in their computations, i.e., latency arbitrage opportunities that last longer than the two-way travel time for a fiber optic cable between exchanges' servers. At this trading speed, the transportation time assumption is especially important, even more so when exchanges are far apart, as in our application. Tivnan et al (2019) and Dewhurst et al (2019) have a more realistic approach when compared with Wah (2016) but they do not consider information treatment time, nor trading costs.

3. Methodology

3.1 Arbitrage process

We propose an innovative hybrid approach involving pairs trading and triangular arbitrage for cross-listed stocks between two exchanges with differing currencies. In its simplest form, this

[3] MIDAS is the US Securities and Exchange Commission's Market Information Data Analytics System.

approach is based on the identification of mean-reverting arbitrage opportunities from a basket of equities traded on their home exchange (noted as Exchange 1), their cross-listed peers at another exchange (noted as Exchange 2), and the currency-futures contract between the two currencies (noted as Currency 1 and Currency 2) for hedging purposes. This strategy also encompasses the simpler case where the two exchanges are using the same currency. That particular application does not require currency hedging, but still relies on the formulations provided in this paper. We will also discuss how the proposed stragtegy can be generalized to more than two exchanges and two currencies, thus expanding the overall tradeable universe.

We first compute a synthetic instrument calculated as the ratio of the stock's simultaneous prices at Exchange 2 and at Exchange 1 (the synthetic, henceforth) obtained from the combination of opposite positions of the same stock being traded on both exchanges. As for internationally cross-listed stocks, the stock prices share two underlying factors: the firm's fundamental value and the exchange rate (Scherrer, 2018). Given that we use the same stock in the two exchanges, the idiosyncratic differences are minimal and should not affect the convergence in pairs trading, contrary to what is often observed with different stocks in the literature (Frazzini et al, 2018; Engelberg et al, 2009; Pontiff, 2006).

Second, we hedge the synthetic with an opposite position in the currency future. Defining the relative spread (SPRD) as equal to the ratio of the synthetic over the currency future, we must test for the SPRD stationarity, a *sine qua non* condition for mean-reverting strategies. At equilibrium, SPRD must converge to a value close to 1.0 for each pair in all trading days, with very few exceptions. Spot and futures prices should diverge slightly, only by the basis value, which accounts for maturity differences in the two instruments.

As a distance criterion, we propose a non-parametric threshold rule adjusted for the strategies' net costs in order to uncover economically relevant opportunities. This is an alternative to standard deviation multiples (Stübinger and Bredthauer, 2017; Gatev et al, 2006). The chosen distance approach is simple and transparent, and allows for large-scale empirical applications (Krauss, 2017).

As market makers on both exchanges might not be perfectly integrated, we have to consider the differences between the functioning of the microstructures. These sources of divergences may influence limit order books (depth, granularity, imbalance, and bid-ask spread) and marketable orders (trade intensity and potential directional or bouncing behavior).

Data from geographically distant exchanges may be asynchronous. We propose a synchronization procedure to replicate an arbitrageur's information processing lag. We implement a two-regime shift incurred by transport delays of information to and from the exchange servers, and we correct the timestamps for the exchanges' processing time and matching delays. The synchronization is effective at Exchange 1's colocation server.

Our strategy does not hold overnight positions[4]. This prevents hedging overnight gap risk and tying up capital due to end-of-day margin requirements (Menkveld, 2014). This also avoids being forced to unwind positions due to margin squeezes (Brunnermeier and Pedersen, 2008). We use the exchanges' appropriate trading fees and rebates to evaluate net arbitrage performances, as well as colocation and trading platform expenses. Details on these costs are provided in Table 1 of Section 6.

[4] In our application, we also considered not closing opened positions at market close. But, because of the fast mean-reversion time of the signals, and the fact that we stop opening new positions 15 minutes before market close (see Online appendix B for more practical considerations), overnight positions were very rare and small in volume. This modification did not significantly modify the strategy's performance and is not further analyzed.

3.2 Relative spread

Arbitrage opportunities are identified by constructing a relative spread (SPRD) equal to the ratio of the synthetic spread to the hedging instrument, the currency futures:

$$\gamma_t \equiv \frac{S_{2,t}/S_{1,t}}{r_t},$$

where γ_t is the mathematical notation for SPRD value at time t, $S_{1,t}$ and $S_{2,t}$ are the cross-listed stock values at Exchange 1 and Exchange 2, and r_t is the exchange rate computed from the currency hedging instrument's value. We define simultaneous prices as prices from a unique time frame of observation that considers the information transportation and treatment time between trading venues, which is known as latency.

We write:

$$\gamma_t^{Short} = \frac{S_{2,t}^{Bid}/S_{1,t}^{Ask}}{r_t^{Ask}} \text{ and } \gamma_t^{Long} = \frac{S_{2,t}^{Ask}/S_{1,t}^{Bid}}{r_t^{Bid}}$$

as the time series of the short and long relative spreads, where the exponents Bid and Ask are the stock prices on the bid and ask side.

3.3 Market order arbitrage strategy

A potential arbitrage opportunity arises when the synthetic is not in equilibrium with the observable exchange rate at time t, that is when:

$$\gamma_t^i \neq \tau^i,\ i \in \{Short, Long\},$$

where τ^i is the mean equilibrium value expected from the mean-reverting processes. The arbitrage opportunity ends when the equilibrium is restored at time $t' > t$ where t' is defined as:

$$t' \quad \equiv argmin_{s>t}\{s \mid \gamma_s^i = \tau^i\},\ i \in \{Short, Long\}.$$

The synthetic is potentially overvalued when:

$$\gamma_t^{Long} = \frac{S_{2,t}^{Ask}/S_{1,t}^{Bid}}{r_t^{Bid}} > \tau^{Long}.$$

In that case, since Γ_t^{Long} is assumed to be mean-reverting. This mispricing can be exploited by shorting $1/\tau^{Long}$ shares of Exchange 2 stock, taking a long position of one share in Exchange 1 counterpart (which means that we short the synthetic), and taking a long position in the currency future of the same value as the Exchange 2 stock position in order to hedge our position, all transactions at time t. Then, we must revert the three positions at time t' using market orders to lock the profit per Exchange 1 stock ($P_{t'}$) in Currency 1 at time t':

$$P_{t'} = \frac{1}{\tau^{Long} r_{t'}^{Ask}}\left(S_{2,t}^{Bid} - S_{2,t'}^{Ask}\right) + \left(S_{1,t'}^{Bid} - S_{1,t}^{Ask}\right) + \frac{S_{2,t}^{Bid}}{\tau^{Long} r_{t'}^{Ask}}\left(\frac{r_{t'}^{Bid}}{r_t^{Ask}} - 1\right) - c_{t'}^{Long},$$

where $c_{t'}^{Long}$ measures the trading costs in Currency 1. Assuming a liquid currency future with a low bid-ask spread, we can use the following approximation for the profitability of the positions:

$$P_{t'} \approx \frac{1}{\tau^{Long} r_{t'}^{Bid}}(S_{2,t}^{Bid} - S_{2,t'}^{Ask}) + (S_{1,t'}^{Bid} - S_{1,t}^{Ask}) + \frac{S_{2,t}^{Bid}}{\tau^{Long} r_{t'}^{Bid}}\left(\frac{r_{t'}^{Bid}}{r_t^{Ask}} - 1\right) - c_{t'}^{Long} \tag{1}$$

where we have substituted $r_{t'}^{Ask}$ with $r_{t'}^{Bid}$. Supposing a perfect hedge, we only buy a fraction of the currency futures of nominal N_{FX} (in Currency 2) that equals the amount invested in Exchange 2 stock at time t. So only a fraction of the constant futures' trading price is paid on this cost-per-share basis. The trading costs paid for opening and closing our positions in Currency 1 at time t', $c_{t'}^{Long}$, are approximated by:

$$c_{t'}^{Long} \approx 2c_1 + 2\frac{c_2}{\tau^{Long} r_t^{Bid}} + 2\frac{c_{FX}}{N_{FX}} \cdot \frac{S_{2,t}^{Bid}}{\tau^{Long}}$$

where c_1 and c_2 are the constant per-share trading fees for market orders on Exchange 1 (in Currency 1) and Exchange 2 (in Currency 2) respectively, and c_{FX} is the per-contract trading costs (in Currency 1) with nominal N_{FX}.

When the three instruments return to equilibrium, the definition of t' implies that:

$$\frac{S_{2,t'}^{Ask}/S_{1,t'}^{Bid}}{r_{t'}^{Bid}} = \tau^{Long} \Rightarrow \frac{S_{2,t'}^{Ask}}{\tau^{Long} r_{t'}^{Bid}} = S_{1,t'}^{Bid}.$$

Using this last equality in (1), we get:

$$P_{t'} \quad \approx \frac{S_{2,t}^{Bid}}{\tau^{Long} r_t^{Ask}} - S_{1,t}^{Ask} - c_{t'}^{Long},$$

which means that to generate a positive profit at time t', we at least need to have:

$$P_{t'} > 0 \Leftrightarrow \quad \frac{S_{2,t}^{Bid}}{\tau^{Long} r_t^{Ask}} - S_{1,t}^{Ask} > c_{t'}^{Long}$$

and we can rewrite the last inequality as

$$\gamma_t^{Long} \frac{r_t^{Bid} S_{1,t}^{Bid}}{S_{2,t}^{Ask}} \frac{S_{2,t}^{Bid}}{\tau^{Long} r_t^{Ask}} - S_{1,t}^{Ask} > c_{t'}^{Long}$$

$$\gamma_t^{Long} > \tau^{Long} \quad \underbrace{\frac{r_t^{Ask} S_{2,t}^{Ask}}{r_t^{Bid} S_{2,t}^{Bid}} \frac{S_{1,t}^{Ask} + c_{t'}^{Long}}{S_{1,t}^{Bid}}}_{>1, \text{in normal market conditions}} \equiv \kappa_t^{Over}. \quad (2)$$

Equation (2) gives us a dynamic upper non-parametric threshold κ_t^{Over} indicating when a short position in our relative spread (SPRD) is profitable because it is overvalued considering trading costs and bid-ask spreads when only market orders are used. This profitability holds when there is a return to equilibrium to close the positions. The same logic with opposite positions also holds when the synthetic is potentially undervalued, or when:

$$\gamma_t^{Short} = \frac{S_{2,t}^{Bid}/S_{1,t}^{Ask}}{r_t^{Ask}} < \tau^{Short}.$$

This results in a dynamic lower non-parametric threshold at which a long position in the synthetic is profitable considering trading costs and bid-ask spreads when market orders are used:

$$\gamma_t^{Short} < \tau^{Short} \underbrace{\frac{r_t^{Bid} S_{2,t}^{Bid}}{r_t^{Ask} S_{2,t}^{Ask}} \frac{S_{1,t}^{Bid} - c_{t'}^{Short}}{S_{1,t}^{Ask}}}_{<1, \text{in normal market conditions}} \equiv \kappa_t^{Under} \tag{3}$$

where $c_{t'}^{Short} \approx 2c_1 + 2\frac{c_2}{\tau^{Short} r_t^{Bid}} + 2\frac{c_{FX}}{N_{FX}} \cdot \frac{S_{1,t}^{Ask}}{\tau^{Short}}$.

Once again, the profitability of the strategy holds when there is a return to equilibrium to close the long position of SPRD.

From equations (2) and (3), we have a set of two signals, γ_t^{Long} and γ_t^{Short}, where $\gamma_t^{Long} > \gamma_t^{Short}\ \forall t$ (which implies that $\tau^{Long} > \tau^{Short}$) in normal market conditions and with their respective dynamic non-parametric thresholds, κ_t^{Over} and κ_t^{Under}, where $\kappa_t^{Over} > \tau^{Long} > \tau^{Short} > \kappa_t^{Under}\ \forall t$.

The arbitrage strategy can be summarized as follows:

- When γ_t^{Long} crosses κ_t^{Over} from below: short $1/\tau^{Long}$ shares of $S_{2,t}$, long $S_{1,t}$ and long the currency future for the same value as the one invested in Exchange 2 stock,
- When γ_t^{Short} crosses κ_t^{Under} from above: long $1/\tau^{Short}$ shares $S_{2,t}$, short $S_{1,t}$ and short the currency future for the same value as the one invested in Exchange 2 stock,
- Close the positions when the equilibrium is restored at t'.
- Repatriate the profits generated at Exchange 2 to Exchange 1 whenever they cross N_{FX}.

Finally, we add the per-share fixed colocation cost and proprietary data feed cost to compute net profit on a given period.

3.4 Limit order arbitrage strategy

We now switch to limit orders, as paying the bid-ask spread on the three instruments can be very costly. The strategy remains the same as with market orders. The main difference is in the profitability equation used to find entry thresholds. The relative spread is potentially overvalued when:

$$\gamma_t^{Short} = \frac{S_{2,t}^{Bid}/S_{1,t}^{Ask}}{r_t^{Ask}} > \tau^{Short}.$$

In that case, we short SPRD at time t and revert the three positions when the equilibrium of Γ_t^{Short} is restored at time t'. This results in a profit in Currency 1 of:

$$P_{t'} = \frac{1}{\tau^{Short} r_{t'}^{Ask}} \left(S_{2,t}^{Ask} - S_{2,t'}^{Bid}\right) + \left(S_{1,t'}^{Ask} - S_{1,t}^{Bid}\right) + \frac{S_{2,t}^{Ask}}{\tau^{Short} r_{t'}^{Ask}} \left(\frac{r_{t'}^{Ask}}{r_t^{Bid}} - 1\right) - \tilde{c}_{t'}^{Short} \quad (4)$$

per Exchange 1 stock, where $\tilde{c}_{t'}^{Short}$ has the same formulation as $c_{t'}^{Short}$, but instead of c_1 and c_2 being the per-share trading-costs for market orders, they are now per-share trading fees (or trading rebates) for using limit orders.

Employing the same logic as previously used to obtain the non-parametric entry thresholds κ_t^{Over} and κ_t^{Under}, we find that the dynamic upper threshold indicating a profitable short position in our relative synthetic spread using limit orders is given by:

$$\gamma_t^{Short} > \tau^{Short} \underbrace{\frac{r_t^{Bid} S_{2,t}^{Bid}}{r_t^{Ask} S_{2,t}^{Ask}} \frac{S_{1,t}^{Bid} + \tilde{c}_{t'}^{Short}}{S_{1,t}^{Ask}}}_{\text{multiplicative term}} \equiv \tilde{\kappa}_t^{Over}, \quad (5)$$

and the dynamic lower non-parametric threshold for long positions in our relative synthetic spread using limit orders is given by:

$$\gamma_t^{Long} < \tau^{Long} \underbrace{\frac{r_t^{Ask} S_{2,t}^{Ask}}{r_t^{Bid} S_{2,t}^{Bid}} \frac{S_{1,t}^{Ask} - \tilde{c}_{t'}^{Long}}{S_{1,t}^{Bid}}}_{\text{multiplicative term}} \equiv \tilde{\kappa}_t^{Under}. \quad (6)$$

Notice that the term multiplying the equilibrium level in equation (2) is always greater than the multiplicative term in equation (5). This means that arbitrage opportunities are available at a lower level of γ_t^{Short} with limit orders, and thus should be more frequent. This is true since limit orders greatly reduce the costs related to the strategy. The same observation can be made for the long position non-parametric thresholds of equations (3) and (6): limit orders push the entry thresholds to a more easily attainable level compared with market orders.

From equations (5) and (6), we have a set of two signals, γ_t^{Short} and γ_t^{Long} with their respective dynamic non-parametric thresholds, $\tilde{\kappa}_t^{Over}$ and $\tilde{\kappa}_t^{Under}$. The arbitrage strategy can be summarized as follows:

- When γ_t^{Short} crosses $\tilde{\kappa}_t^{Over}$ from below: short $1/\tau^{Short}$ shares of $S_{2,t}$, long $S_{1,t}$ and long the currency future for the same value as the one invested in Exchange 2 stock,
- When γ_t^{Long} crosses $\tilde{\kappa}_t^{Under}$ from above: long $1/\tau^{Long}$ shares $S_{2,t}$, short $S_{1,t}$ and short the currency future for the same value as the one invested in the Exchange 2 stock,
- Close the positions when the equilibrium is restored at t'.
- Repatriate the profits generated at the Exchange 2 to the Exchange 1 whenever they cross N_{FX}.

3.5 Strategy at the portfolio level and aggregate hedging

Consider a universe Ω of N cross-listed stocks on Exchange 1 and Exchange 2, $|\Omega| = 2N$. We wish to execute the cross-listed stocks arbitrage strategy defined in the previous sections, on every pair contained in that universe. This extension is applicable with both market orders and limit orders and is important for the application of the two previous strategies.

Due to the development of our strategy, aggregating every position in a single portfolio offers a built-in hedging effect against movements of the exchange rate whenever positions are opened in both Γ_t^{Short} and Γ_t^{Long}, because the aggregated position in Exchange 2's market is reduced compared to the sum of the absolute position of every independent portfolio for each pair. The hedge can be optimized with the use of currency futures. This section explores that extension.

Let us define $v_{1,t}^{(n)}, v_{2,t}^{(n)} \in \mathbb{R}, n \in \{1, \dots, N\}$ the size of the position in the cross-listed stock n in both markets at time t. A position is long when the size is positive, a position is short when the size is negative, and the size is zero when no position is opened in the asset. Let us also define the total non-repatriated profits, in their respective currency, generated at Exchange 2 and the FX Exchange at time t respectively by $G_{2,t},\ G_{FX,t} \in \mathbb{R}$. Hence, the portfolio's exposures in Currency 1 at Exchange 1, Exchange 2 and FX Exchange at time t are respectively given by:

$$\begin{aligned} V_{1,t} &= \sum_{n=1}^{\mathrm{N}} v_{1,t}^{(n)} S_{1,t}^{(n)}, \\ V_{2,t} &= \sum_{n=1}^{\mathrm{N}} v_{2,t}^{(n)} \frac{S_{2,t}^{(n)}}{r_t} + \frac{G_{2,t}}{r_t}, \end{aligned}$$

$$V_{FX,t} = \frac{v_{FX,t}^{*} N_{FX}}{r_t} + G_{FX,t},$$

where $v_{FX,t}^{*} \in \mathbb{R}$ is the optimal position size in the currency futures at time t that we are trying to obtain. The total value of the portfolio in Currency 1, V_t, is given by:

$$V_t = V_{1,t} + V_{2,t} + V_{FX,t}.$$

By taking a position in the currency future that is the inverse of the position in Exchange 2, we obtain:

$$V_{FX,t} = -V_{2,t} \Leftrightarrow v_{FX,t}^{*} = -r_t \frac{V_{2,t} + G_{FX,t}}{N_{FX}}, \tag{7}$$

which results in a neutral aggregated position in Exchange 2's market: $V_{2,t} + V_{FX,t} = 0$. The portfolio's value is now simply given by $V_t = V_{1,t} \Longrightarrow \frac{\mathrm{d}V_t}{\mathrm{d}r_t} = \frac{\mathrm{d}V_{1,t}}{\mathrm{d}r_t} = 0$. The last equality supposes the mathematical independence of Exchange 1 stocks' prices and the exchange rate. In the universe Ω, a portfolio invested in cross-listed stock pairs that follows the proposed strategy for every pair achieves an optimal hedge against currency risk at any time t when that portfolio has a neutral aggregated position in Exchange 2's currency. If the aggregated position in Exchange 2 stocks is not neutral, we can take a position of $v_{FX,t}^{*}$ contracts in the currency future to get a perfect hedge.

The hedging of the portfolio is done by rebalancing our position in the currency future to the optimal value, if necessary, whenever we open or close positions in pairs of cross-listed stocks, compared with the pair-wise strategy that requires taking the inverse position taken at Exchange 2 at every arbitrage opportunity.

3.6 Generalization of the strategy beyond two stock exchanges and a single exchange rate

The proposed strategy and the formulated arbitrage signals can be applied to more general trading environments. Indeed, the arbitrage signals γ formulated in the last section can be computed for any cross-listed stock pair between any two stock exchanges and any currency for both stocks (shared or not) without any modification. The global tradeable universe for which the proposed strategy can be applied to is thus quite large, as discussed in the introduction. We now present different additional trading environements where the strategy can be applied.

The first additional trading environment is when there are two stock exchanges with a single currency for the cross-listed stock's pair. This can be done by setting $r_t = r_t^{Bid} = r_t^{Ask} = 1, \forall t$ and ignoring the currency hedging instrument. The signals are thus solely based on the equilibrium between the two microstructures, which corresponds to the model of Budish et al. (2015): whenever a sudden jump occurs in one of the two stocks, the correlation between the stocks breaks down and an arbitrage opportunity potentially opens up. In our case, the arbitrage signals consider both the closing conditions and the trading costs associated with sending orders to seize the arbitrage opportunity.

The second trading environment is when there are more than two stock exchanges and a single currency for the cross-listed stocks. Once again, this can be done by using the same constraint on r_t and ignoring currency hedging as previously discussed. But, a second constraint needs to be put in place to select which arbitrage opportunity to capture whenever multiple opportunities occur at the same time for the same stock and exchange. This is necessary since each stock can be part of more than two exchanges, so multiple cross-listed pairs can contain the given stock. In that case, only the cross-listed pair with signal γ that is the farthest from equilibrium τ is executed (i.e. the pair with the maximum expected profitability). This relates closely to the model of Wah (2016), but the author did not consider latency, inventory management, nor any trading cost.

The final case is when there is more than two stock exchanges and multiple exchange rates hosted by any number of exchanges. The trading signals γ can be computed for every combination of cross-listed stocks pair and their applicable exchange rate. As in the previous case, multiple arbitrage opportunities can happen at the same time for the same stock at a single exchange. Again, only the pair with signal γ that is the farthest for its equilibrium τ is executed for that particular stock. To the best of our knowledge, this has not been studied in the litterature yet.

Overall, by adding simple contraints to the proposed strategy, either on the obervable exchange rate r_t, currency hedging, or on the selection of arbitrage opportunities computed by our signals γ, the strategy can be applied to any stock pair.

4. Latencies, arbitrage costs, and arbitrage risks

4.1 Latencies and arbitrage costs

A factor of interest in this contribution is latency. In trading terms, latency refers to the time it takes for an agent to interact with the market. We follow closely Hasbrouck and Saar's (2013) measure of latency based on three components: the time it takes for a trader to learn about an event, generate a response, and have the exchange act on that response. (See also Foucault and Moinas, 2019). We split that definition into two separate quantities so that we can have more granularity on the impact of latency on the high-frequency trading strategies.

The first quantity of importance is the latency of a message from any exchange to Exchange 1, which includes the one-way transportation time of the information to Exchange 1, and the information treatment time needed by the agent's servers collocated at Exchange 1 and having access to a proprietary data feed. The second quantity of importance is the latency of a message from Exchange 1 to another exchange, which is comprised of the one-way transportation time of information from Exchange 1 to the receiving exchange, and the matching engine delay of that last exchange.

Information treatment time refers to the timespan required to receive and analyze incoming information from the exchanges, followed by the decision to trade or not. Exchange server procedure considers information reception at the exchange gates, limit order book (LOB) positioning or matching of an incoming limit order (with the LOB) and issuing traders' confirmation to the server gates. Round-trip latency measures the total latency delay for a message between two exchanges.

We apply a two-regime model associated with normal and extreme market conditions based on quote and trade message intensity. The regime shifts, from the normal state to the extreme one, are often due to bursts in the events stream, phenomena well documented in the literature (Friederich and Payne, 2015; Menkveld, 2016; Egginton, et al., 2016; Dixon et al., 2019; Shkilko and Sokolov, 2020). To help us recreate this behavior, we use a latency regime variable that varies depending on the number of messages a certain exchange received in the last millisecond on a per-asset basis. This quantity is a good proxy of an exchange's server traffic, which has a positive relationship with computational delays occurring during the information treatment time and the matching engine time components of latency. The normal regime generates a minimal, baseline, value of the latency that exists between two exchanges and a bonus on that minimal latency is added for the extreme regime.

The latency regime variable for a given asset remains in its normal state up to a certain static threshold for the number of messages in a single millisecond for that asset, which we set as the 95th percentile of its empirical distribution.

Let us define $q^i_{95\%}$ as the 95th percentile of the empirical distribution of the number of messages in one millisecond for asset i and define q^i_j the number of messages during the millisecond preceding and ending at message $j \in [1, N^i]$ where N^i is the total number of messages for asset i during the full period. Let us also define $L^i_j \in \{normal, extreme\}$ the latency regime of asset i at message j. Then, its value is computed as follows:

$$L^i_j = \begin{cases} normal & if\ q^i_j < q^i_{95\%} \\ extreme & if\ q^i_j \geq q^i_{95\%} \end{cases} \quad \forall i, j.$$

By adding the corresponding latency to the original exchange timestamp of every message, we can approximately synchronize the data feeds of geographically distant exchanges into a single point of observation (e.g. Exchange 1) as they would be in practice because of the natural and technological limits of information propagation. Our methodology emulates that relativistic effect so that what is observed by the trading algorithm at any point is a past state of markets. The same idea applies when the algorithm sends an order to a given exchange. We add the corresponding latency so that the agent does not interact immediately with that exchange. This makes it possible to study the influence of latency on the performance of high-frequency trading strategies.

4.2 Arbitrage risks

4.2.1 Execution risk

The choice between limit and market orders relies, in part, on the difference between non-execution risk and execution risk (Mavroudis, 2019; Dugast, 2018; Liu 2009; Kozhan and Tham, 2012; Brolley, 2020). To empirically solve this trade-off, we first evaluated our algorithm's performance using market orders exclusively. As we will see, using only market orders leads to a negative economic value with our data in the sense that the cost of immediacy (conceding the bid-ask spread) cannot be borne by the arbitrageur in the vast majority of trades. This high cost also results in a very low number of potential arbitrage opportunities, since the divergence of SPRD is rarely large enough to compensate it. This means that traders must always control for market conditions (Foucault and Moinas, 2019). We then constrained our algorithm to limit orders, except for the liquidation of positions to avoid overnight exposures. We also use marketable limit orders to offset unexecuted legs. There remain two additional risks.

4.2.3 Non-execution risk

We evaluate non-execution risk costs by managing the LOB queuing priorities. We mitigate the risk of non-execution by keeping dynamically our limit orders to the LOB's level one. This is implemented conditional on the persistence of an expected profitable arbitrage. Otherwise, we liquidate positions, if any, by issuing marketable limit orders (Dahlström and Nordén, 2018).

4.2.4 Mean-reversion risk

Mean-reversion risk arises after initial positions are taken. It materializes when the circuit breaker timer is triggered. All arbitrage legs are then liquidated via marketable limit orders. As we will see, this risk is very low in our data since the processes Γ_t^{Short} and Γ_t^{Long} are stationary for almost all stocks and trading days.

5. Data, data synchronization, trading and quoting emulator, empirical latencies, and other trading costs

5.1 Data

We use LOB level one data and trade data that we obtained from: the TAQ NYSE OpenBook and the TAQ NYSE Trades historical data timestamped to the microsecond, the CME Market Depth FIX Canadian Dollar Futures historical data timestamped to the nanosecond, and Trades and Quotes Daily historical data from TMX Group timestamped to the nanosecond. All the data was timestamped at the respective exchanges, and span from January 7th, 2019 to June 28th, 2019, inclusively. We only selected dates where the three exchanges were opened, meaning that we eliminated every holiday from our sample.[5] The timestamps were truncated and rounded to the nearest millisecond above so that potential microscopic errors in the timestamps do not affect the results.

Overall, there are 120 trading days in our data set. We have access to 74 pairs of cross-listed stocks that were listed on both the TSX and the NYSE during at least two weeks of that period. Pairs where one of the stocks got de-listed from an exchange at any point were kept in the sample, but the strategy was only applied to periods where both stocks of the pair were listed and active. All cross-listed S&P/TSX 60 stocks are present in our sample during the six months. Table A2 of Online appendix A describes every available pair and Table A3 includes their aggregated statistics during the period of analysis.

The time series of daily number of trades and quotes in the two exchanges for some pairs of stocks of interest are presented in Figure D1 of Online appendix D. The four rows of graphs in

[5] TSX: February 18th: Family Day; April 19th: Good Friday; May 20th: Patriot's Day.
NYSE and CME: January 21st: Martin Luther King Jr. Day; February 18th: President's Day; April 19th: Good Friday; May 27th: Memorial Day.

Figure D1 present the trades and quotes data of some of the most often selected stocks for arbitrage. We do not observe any pattern between the number of trades and the number of quotes. The main differences seem to be related to the type of industry.

We use the quarterly C/US futures listed on CME: 6CH9 expiring March 19, 2019; 6CM9 expiring June 18, 2019; and 6CU9 expiring September 17, 2019. We do not use monthly futures because they have a smaller open interest. The continuous futures contract is created by concatenating the three futures' data and by adjusting the LOB level one and trade prices of the consecutive contracts so that no jumps are artificially created. The concatenation dates are determined based on the daily transaction volume of consecutive futures. That is, whenever the futures contract with the farthest expiration date generates a significantly higher daily transaction volume than its predecessor and remains more actively traded, we switch to that futures' trades and quotes for the continuous futures that we use in the strategy. In order to have a greater hedge, we employ the Micro C/US futures contract with a nominal of C$10,000, which we approximate by dividing the prices of our continuous futures by 10, because of its nominal of C$100,000.

5.2 Data synchronization

The strategy is launched each week, from Monday to Friday, starting at 9:30 am and ending at 4:00 pm Eastern Time when the three exchanges are all opened to continuous trading. Both the TSX and NYSE are in the Eastern Time zone, but the CME is in the Central Time zone, one hour behind. Hence, we add an hour to the time stamps of the CME data to synchronize the three exchanges' clocks.

5.3 Trading and quoting emulator

Our methodology and the different trading strategies are implemented in Deltix's QuantOffice, a trading system suite used by multiple traders, which brings us closer to real trading practice. The Deltix trading suite allows us to replay the synchronized events of the three stock markets (level one LOB and trades) as they were obtained in streaming by traders. By handling these events and following our orders position in the queues, we can determine as realistically as possible the real-time performance that would have been obtained with our strategies. Note that a single ex-ante set of parameters was tested. This implementation makes it possible to consider trading fees and rebates, latency costs and other trading risks and costs presented above. It confirms the order status (creation, cancelation, or execution) just as it would have happened in streaming trading considering market frictions and ever-changing market states. Standard reports, such as a trade report and a performance report, are generated at the end of a strategy's execution and these are used to compute our results.

Moreover, we can manage the individual and aggregated positions, and calculate the respective Profit and Loss Reports (PnL) altogether with performance statistics. These PnLs represent the economic value of our arbitrage opportunities. Using our performance as a benchmark, we can evaluate the economic impact of latency risk by varying the aforementioned latency parameters. The general rules of the trading and quoting emulator on LOB level one data and information on how executions and non-executions occur are presented in Online appendix F.

5.4 Empirical latencies and other costs

Table 1 documents the 2019 latency costs, trading costs, rebates, colocation costs, and proprietary data feed (including a trading platform) costs used in this study. Orders and positions are managed at TSX's colocation premises in Toronto (TSX, 2020). Information comes from the

TSX, the NYSE, and the CME. We address asynchronicities by adjusting the TSX timestamps based on round-trip transportation time, arbitrageur information processing delays, and exchanges matching engine delays presented in Table 2. Table 1 also documents the positive trading fees for the removers of liquidity and rebates for the providers. Colocation costs in Toronto are considered in our monthly portfolio performance estimations, as well as proprietary data feeds which enable some trading firms to receive updates from the exchange faster than other traders who do not pay for this service.

Table 1. Arbitrage costs

Definition	Description	Measurement	In Deltix
Information transportation time between exchanges	Transportation time details: Toronto – Chicago: Fiber paths Toronto – New York: Microwave path (regular) Fiber path (extreme situations)	See Table 2	Adjusted raw dataset timestamp fed to Deltix
Information treatment time	Timespan required to receive and analyze incoming information from the exchanges, followed by the decision to trade or not.	See Table 2	Adjusted raw dataset timestamp fed to Deltix
Exchange trading fees	TSX member trading fees per share[1] NYSE Type A stocks per share[2] CME Globex C/US FX futures per contract[3]	Removing: \$0.0015 Providing: (\$0.0011) Removing: \$0.00275 Providing: (\$0.00120) \$100k notional value: \$0.32 \$10k notional value (e-micro): 0.04\$	Applied to matched orders
Colocation cost	Colocation with exchange connectivity rates	Half cabinet (21U, 3 kw maximum): \$5,250 monthly Initial set-up fee: \$5,250 one-time	Included in monthly portfolios performance

Proprietary data feed	TSX & Venture level 1 Distribution Trading use case license	$4,000 monthly	Included in monthly portfolios performance

[1] *https://www.tsx.com/resource/en/1756/tsx-trading-fee-schedule-effective-june-4-2018-en.pdf*

[2] *https://www.nyse.com/markets/nyse/trading-info/fees*

[3] *https://www.cmegroup.com/company/clearing-fees.html*

For both latency regimes, the latency to and from TSX is set as the sum of the intervals' center of each of their components found in Table 2, for the respective market condition. We round latencies up to the closest integer.

Table 3 details the empirical latencies used. Following the methodology introduced in Section 4, our estimation of the empirical distribution of messages per millisecond used a random sample of six weeks, where each sampled week came from a different month contained in our data.

Table 2. Latencies[1]

	TRANSPORTATION ONE WAY TO TSX		ARBITRAGEUR	TRANSPORTATION ONE WAY FROM TSX		MATCHING ENGINE	TOTAL
Market condition	Exchanges from–to	Transportation time	Information treatment	Exchanges from–to	Transportation time	Exchange server	Round-trip latency
Normal	TSX–TSX	+ 5 µs	+ 10–70 µs	TSX–TSX	+ 5 µs	+ 100–300 µs	120–380 µs
	NYSE–TSX	+ 2.4 ms	+ 10–70 µs	TSX–NYSE	+ 2.4 ms	+ 100–300 µs	4.91–5.17 ms
	CME–TSX	+ 5 ms	+ 10–70 µs	TSX–CME	+ 5 ms	+ 1–5 ms	11.01–15.07ms
Extreme	TSX–TSX	+ 5–10 µs	+ 200–500 µs	TSX–TSX	+ 5–10 µs	5–10 ms	5.21–10.52 ms
	NYSE–TSX	+ 4.8–9.6 ms	+ 200–500 µs	TSX–NYSE	+ 4.8–9.6 ms	5–10 ms	14.80–29.7 ms
	CME–TSX	+ 5–10 ms	+ 200–500 µs	TSX–CME	+ 5–10 ms	50–100 ms	60.20–120.50 ms

[1] Latencies are obtained following discussions with a major Canadian financial institution trading actively in Canada and in the United-States. ms: millisecond; µs: microseconds.

Table 3. Latencies used when testing the strategies, depending on the latency regime, the origin of the message and the exchange where the message is sent.

Latency regime	Exchanges from–to	Latency	Exchanges from–to	Latency
Normal	TSX–TSX	1 ms	TSX–TSX	1 ms
	NYSE–TSX	3 ms	TSX–NYSE	3 ms
	CME–TSX	6 ms	TSX–CME	8 ms
Extreme	TSX–TSX	1 ms	TSX–TSX	8 ms
	NYSE–TSX	8 ms	TSX–NYSE	15 ms
	CME–TSX	8 ms	TSX–CME	83 ms

6. Empirical results

We now present the statistical results of our study in three steps. We first compute the performance of the trading strategy of Budish et al (2015) applied to our data.[6] The goal of this exercise is to isolate the importance of considering latencies, execution risk, and trading costs when evaluating the benefits of HFT arbitrage. It also serves as a benchmark to compare our trading strategies and test how previously proposed arbitrage strategies are profitable with our more recent data.

We then present the results from our strategies. We show that arbitrage with market orders is not profitable, while arbitrage with limit orders provides positive profits when latencies, rebates, exchange fees, and non-execution risk are considered. Other conclusions are discussed.

6.1 Budish et al (2015) contribution

This contribution examines arbitrage opportunities between the two largest financial instruments that track the S&P 500 index, the SPDR S&P 500 exchange traded fund (ticker SPY) and the S&P 500 E-mini futures contract (ticker ES), using millisecond-level direct feed data from different stock exchanges and the Chicago Mercantile Exchange. The application is consequently very different from arbitrage trading of the same stock in two different exchanges but some comparisons with our research are important given that this article suggests strong modifications to the functioning of continuous HFT trading. The authors first demonstrate that the high correlation between the two securities observed from the bid-ask midpoints breaks down at very high-frequency time. This correlation breakdown creates technical arbitrage opportunities

[6] See Online appendix E for the analysis of the results obtained with Wah (2016) strategy.

estimated at approximately US$75 million of gross profit per year for the two securities alone on all markets where the SPY is traded (not only at the NYSE). Their period of analysis includes many high volatility periods such as the 2007-2009 financial crisis. For a more regular year like 2005, the total gross profit is US$35 million.[7] Verifying from Bloomberg that the share of the NYSE for this market is 25%, the annual gross profit for 2005 is US$8.75 million for the NYSE alone. These numbers represent gross profits because trading fees are not considered, nor are latencies and exchange fees. Only bid-ask spread costs are computed.

The above numbers come from the following market environment: there is no arbitrageur entry in the market over the period considered and trading firm observes variations in the signal (perfectly correlated with the fundamental value of the stock) on the stock price with zero-time delay. There is zero latency in sending orders to the exchange and receiving updates from the exchange. This is a pure continuous trading environment with no asymmetric information and inventory costs where opened positions at an exchange can be immediately closed at another exchange with a different asset.

The strategy of Budish et al (2015) is first implemented with their theoretical settings and minor modifications to adapt it to our data. In that sense, prices at the NYSE are continuously transferred to C$ following the C/US futures observed at the CME. In addition, we used two hypotheses employed in their model: there is an absence of latency and opened positions at an exchange can be immediately closed at another exchange, resulting in a trade. Table 4 Panel A presents the results obtained with our data and using the arbitrage strategy presented in Online appendix A.2 of their article with market orders only. The second column of Table 4 Panel A

[7] The CBOE Volatility Index (VIX) of the average closing price was equal to 12.81 in 2005, 32.69 in 2008, and 15.39 in 2019.

presents the results that are obtained following as closely as possible their theoretical framework. In the next two columns, latency is considered.

Table 4. Panel A. Budish et al (2015) model with our 2019 data

1 Model	2 Budish Original	3 Budish Original - With 1x Latency	4 Budish Original - With 3x Latency
Latency multiplier	0	1	3
Pair selection	No	No	No
Gross profit	$1,421,685.23	$998,328.25	$1,116,673.07
Loss	$0.00	-$11,492.18	-$18,696.78
Trading fees	-$75,167.39	-$57,973.82	-$67,232.10
Trading rebates	$0.00	$0.00	$0.00
Total net profit	$1,346,517.84	$928,862.25	$1,030,744.19
Mean daily net profit	$11,811.56	$8,147.91	$9,041.62
Median daily net profit	$1,968.76	$1,189.76	$1,219.35
Mean daily net profit per pair, per day	$110.95	$76.54	$84.93
p-value Kolmogorov-Smirnov test[1]		1.00	0.65
1st most profitable day (date - profit)	2019/01/28 $184,196.22	2019/01/28 $121,108.28	2019/01/28 $127,578.22
5th most profitable day (date - profit)	2019/01/30 $66,060.79	2019/01/24 $47,904.13	2019/01/24 $50,816.97
1st most unprofitable day (date - profit)	2019/06/24 -$161.55	2019/06/24 -$450.32	2019/06/03 -$2,222.67
5th most unprofitable day (date - profit)	2019/05/31 -$77.85	2019/06/27 -$340.18	2019/06/24 -$681.72
Average time in trade[2]	00:00.0	00:00.0	00:00.0
# net profitable trades	31,762	23,313	29,226
# net unprofitable trades	1,176	1,336	1,817
# trades	32,938	24,649	31,043
% net profitable trades	96.43%	94.58%	94.15%
Average volume per trade	345.63	352.77	326.16
Average net profit per trade	$40.88	$37.68	$33.20
Average profit per net profitable trades	$42.75	$40.75	$36.32
Average profit per net unprofitable trades	-$7.97	-$15.91	-$16.99

[1] H0: F(x) <= G(x), H1: F(x) > G(x). F(x), G(x) = CDF of daily net profits for sample 1 and sample 2, respectively: p-value of 1.00 for no latency vs 1x latency and 0.65 for 1x latency vs 3x latency.

[2] HH: MM: SS. U: hours: minutes: seconds: fractions of a second.

We observe, in column 2 of Panel A, that gross profit is limited to C$1.4 million for six months of continuous trading or about C$2.8 million for a year, which is below the C$10.60 million (US$8.75) for the low volatility year of 2005 with their data. Many factors can explain the difference. The main difference is mostly related to the average daily trade volume of the assets. They document 800 daily arbitrage opportunities in their data, while in our data we have 200 daily arbitrage opportunities with their strategy for the 74 stocks.

We also observe that introducing trading fees does not significantly affect the profitability in the second column, but some opportunities do not cover the trading costs. The main difference in profitability is obtained when we introduce latency. This effect is observed in the next two columns where the total net profitability drops by more than 30% when latency is introduced. The daily net profitability is statistically greater when latency is ignored (see *p*-values). This is mainly explained by the fact that true cross-markets occasions observed at a single geographical point last a shorter amount of time and some are now inexistant compared to a latency-free environment, thus decreasing the number of trades by around 25%. Captured arbitrage opportunities are also less profitable. Comparing the net profitability of colum 2 with that in column 3, we can observe that profits where indeed inflated in column 2 because of a simplified market environement.

Another hypothesis was made in the strategy of Budish et al (2015): Exact opposite positions in different exchanges count as a trade and result in a null inventory in both accounts. The next panel of Table 4 (Panel B) does not use this simplified environment, meaning that positions can only be closed with an opposite position at the same exchange with the same stock. The second column does not include latency. The next two columns do.

Table 4. Panel B. Practical Budish et al (2015) model with our 2019 data

1 Model	2 Budish Practical	3 Budish Practical - With 1x Latency	4 Budish Practical- With 3x Latency
Latency multiplier	0	1	3
Pair selection	No	No	No
Gross profit	$779,282.29	$441,466.25	$666,886.91
Loss	-$789,845.78	-$456,295.76	-$695,876.53
Trading fees	-$11,686.80	-$6,957.61	-$11,089.11
Trading rebates	$0.00	$0.00	$0.00
Total net profit	-$22,250.29	-$21,787.12	-$40,078.73
Mean daily net profit	-$195.18	-$191.12	-$351.57
Median daily net profit	-$5.11	-$44.00	-$49.72
Mean daily net profit per pair, per day	-$1.83	-$1.80	-$3.30
p-value Kolmogorov-Smirnov test[1]		0.18	0.80
1st most profitable day (date - profit)	2019/06/27 $2,473.72	2019/06/28 $2,043.65	2019/06/28 $2,158.73
5th most profitable day (date - profit)	2019/06/20 $1,219.50	2019/06/21 $292.03	2019/06/20 $233.39
1st most unprofitable day (date - profit)	2019/05/15 -$9,698.68	2019/05/15 -$5,221.17	2019/06/03 -$7,570.36
5th most unprofitable day (date - profit)	2019/06/03 -$1,718.65	2019/05/21 -$1,294.97	2019/06/05 -$2,132.83
Average time in trade[2]	126.06:12:08	127.12:57:37	127.14:15:11
# net profitable trades	974	702	961
# net unprofitable trades	958	708	987
# trades	1,932	1,410	1,948
% net profitable trades	50.41%	49.79%	49.33%
Average volume per trade	585.56	477.63	551.3
Average net profit per trade	-$11.52	-$15.45	-$20.57
Average profit per net profitable trades	$796.17	$625.62	$690.11
Average profit per net unprofitable trades	-$832.69	-$651.09	-$712.53
Total Short Inventory Remaining @ Close (C$)	$354,467,602.46	$276,237,299.21	$309,494,680.19
Total Long Inventory Remaining @ Close (C$)	$271,097,081.28	$211,074,656.88	$236,477,971.72

[1] H0: F(x) <= G(x), H1: F(x) > G(x). F(x), G(x) = CDF of daily net profits without and with latency, respectively: 0.18 for no latency vs 1x latency and 0.8 for 1x latency vs 3x latency.

[2] D.HH: MM: SS. U: days.hours: minutes: seconds: fractions of a second.

We observe in Panel B that the strategy does not generate any net profit when we abandon the hypothesis of a trade occuring when exact opposite positions are taken in two different exchanges. The net profitability is even more statistically reduced when latency is considered. Column 3 of Panel B would be the closest results obtained by an HFT firm using the strategy during our data period. Another salient point is the very large accumulated inventory that needs to be managed. This is attribuable to the fact that price discovery primarily occurs on the Canadian exchange (Eun and Sabherwal, 2003; Chouinard and D'Souza, 2003). Coupled with a positive directional market like in our period, the jumps in prices happened most of the times on the bid side of the book for the Canadian stock first.[8] This resulted in taking the same short TSX positions and long NYSE positions repeatedly, thus rarely closing previous positions to generate a trade. This shows the importance of inventory management and currency hedging in an international arbitrage context. Overall, by not considering practical trading aspects such as latency or real market functioning, Budish et al (2015) inflated latency arbitrage profitability.

6.2 Our contribution with market orders

Using the Augmented Dickey-Fuller test for stationarity, we obtain that both $\{\gamma_t^{Short}\}_{t=1}^{T}$ and $\{\gamma_t^{Long}\}_{t=1}^{T}$ time series from January 7th 2019 to June 28th 2019 are stationary for almost all stocks in all trading days where the three exchanges are open at the same time, at a *p*-value of 1%, with continuous observation time. Details are presented in Table A.1 of Online appendix A. Given that the SPRD time series are stationary and exhibit strong mean-reversion, we define τ^i, $i \in \{Short, Long\}$ as the equilibrium level of the mean-reverting processes $\left\{\Gamma_t^i\right\}_{t=1}^{T}$ with observations $\{\gamma_t^i\}_{t=1}^{T}$.

[8] The only exception is TRQ which dropped by 25% in our period exhibiting an opposite trading behavior.

The main results from our strategy when using market orders are presented in Table 5. This strategy is not profitable because it is too expensive to obtain enough liquidity and orders are subject to execution risk (Loss line). Trading fees affect the profitability of this strategy because the arbitrageur consumes liquidity with market orders. Thus, following our theoretical strategy with market orders is hazardous, especially when latency is considered. Indeed, we also observe, in columns three and four, that increasing latency reduces the net profitability even more and this effect is largely significant in both columns (significant *p*-values). Finally, the utilization of future contracts increases the average trading time. Our small number of arbitrage opportunities, explained by the use of market orders, implies that intraday values of our realized profits do not vary sufficiently to modify our positions in the futures contracts that hedge these quantities. This results in positions in the futures that are only closed hours, or even days, after being opened.

6.3 Our contribution with limit orders

The most interesting results from our contributions are from limit orders where arbitrageurs mainly provide liquidity to the markets. In Table 6, we observe a gross profit of C$ 9.6 million with selected pairs of cross-listed stocks obtained with supervised machine learning from our universe of 74 possible pairs[9] (see Online appendix C), and for six months of trading. Adding latency in the next columns affects the profitability of our strategy by reducing the net profits by about 25%. However, the percentage of net profitable trades is rather constant between the three columns. The profitability (unprofitability) between days of trading is also quite stable and using futures contracts for hedging the exchange risk does not increase the average time of trade very much because of the constant movement of our realized profits that are repatriated at every C$10,000 of gain or loss. The average volume per trade is quite low and stable and is similar to

[9] This method of pair selection was also applied to market orders.

that in Budish et al (2015), as can be seen in the second column of Table 4. We could have used larger volumes with higher probability of non-execution risk. We chose to be conservative to minimize the impact on the price discovery process. The annual colocation cost and proprietary data feed total cost in Toronto is $116,250. Consequently, international arbitrage of cross-listed stock is profitable with our proposed limit order strategy even when all latencies, costs and risks are considered.

Therefore, the main question is the following: does a net annual profit of about C$8 million (US$6 million, column 3 Table 6, with real latencies and all costs) seem reasonable for this international arbitrage activity that can be managed by one trader in a large trading firm? Note that Budish et al's (2015) original model with market orders generated a gross annual profit of US$8.75 million from the NYSE in 2005 (C$10.60), in a year where the VIX was comparable to that of 2019. But their model made only about C$2 million of gross annual profit with our data in 2019 because the market activity is much less intense with our selected cross-listed stocks than with their two very liquid financial assets. Moreover, as they claimed, their trading model was quite simple and they predicted that a more sophisticated one should generate higher profits, which we demonstrated here in an international context with limit orders.

To eliminate the probability of back test overfitting (Bailey et al, 2014), we only tested one set of parameters for our strategies, which we deemed reasonable beforehand: $\beta = 0.05$ (See Online appendix B). It is applied to every pair and every day of our data. Of course, the probability that this set of parameters is the optimal one for any pair and any day is close to zero, and if we had back tested the strategies multiple times, we could have selected the set that generated the greatest profitability and performance metrics of our portfolio. However, by using a single set of parameters fixed before any testing, and reporting the results generated by it, we ensure that our findings are generalizable. Hence, the metrics that were shown in this section could be improved and our results thus offer a conservative, but reasonable, measure of the profitability of international arbitrage of cross-listed stocks between Canada and the US.

Table 5. Results with market orders

1 Model	2 Market orders	3 Market orders 1	4 Market orders 2
Latency multiplier	0	1	3
Pair selection	No	No	No
Gross profit	$38,660.35	$41,508.69	$41,620.24
Loss	-$58,361.15	-$96,751.29	-$128,442.17
Trading fees	-$17,890.26	-$22,121.43	-$31,985.04
Trading rebates	$0.00	$0.00	$0.00
Total net profit	-$37,591.06	-$77,364.03	-$118,806.97
Mean daily net profit	-$329.75	-$678.63	-$1,042.17
Median daily net profit	-$18,24	-$207.53	-$595.92
Mean daily net profit per pair, per day	-$4.46	-$9.17	-$14.08
p-value Kolmogorov-Smirnov test[1]		1.00	1.00
1st most profitable day (date - profit)	2019/03/06 $354.30	2019/05/31 $21.63	2019/05/31 $51.54
5th most profitable day (date - profit)	2019/06/21 $196.92	2019/06/17 -$2.54	2019/04/29 -$94.49
1st most unprofitable day (date - profit)	2019/01/30 -$4,053.94	2019/01/16 -$4,682.15	2019/05/16 -$4,692.79
5th most unprofitable day (date - profit)	2019/03/26 -$2,095.20	2019/01/29 -$3,504.39	2019/01/28 -$3,785.02
Average time in trade (excl. futures contracts)[2]	00:06:34.41	00:06:37.83	00:04:42.15
Average time in trade (incl. futures contracts)[2]	02:12:28.60	00:59:30.36	00:57:59.53
# net profitable trades	1,284	1,092	1,590
# net unprofitable trades	2,130	2,927	4,814
# trades	3,414	4,019	6,404
% net profitable trades	37,61%	27.17%	24.83%
Average volume per trade	1,529.78	1,592.15	1,449.57

1 Model	2 Market orders	3 Market orders 1	4 Market orders 2
Average net profit per trade	-$11.01	-$19.25	-$18.55
Average profit per net profitable trades	$26.46	$32.92	$21.99
Average profit per net unprofitable trades	-$33.60	-$38.71	-$31.94

[1] H0: F(x) <= G(x), H1: F(x) > G(x). F(x), G(x) = CDF of daily net profits for sample 1 and sample 2, respectively: p-value of 1.00 for no latency vs 1x latency and 1.00 for 1x latency vs 3x latency.

[2] HH: MM: SS. U: hours: minutes: seconds: fractions of a second.

Table 6. Results with limit orders

1 Model	2 Limit orders	3 Limit orders 1	4 Limit orders 2
Latency multiplier	0	1	3
Pair selection	Yes	Yes	Yes
Gross profit	$9,608,178.87	$8,641,338.63	$8,363,528.28
Loss	-$4,757,168.60	-$5,041,665.26	-$5,168,902.58
Trading fees	-$78,132.64	-$82,067.16	-$83,537.87
Trading rebates	$553,201.20	$476,071.01	$458,542.50
Total net profit	$5,326,078.83	$3,993,677.22	$3,569,630.33
Mean daily net profit	$46,719.99	$35,032.26	$31,312.55
Median daily net profit	$44,453.98	$33,756.44	$29,610.42
Mean daily net profit per pair, per day	$2,273.19	$1,704.51	$1,523.53
p-value Kolmogorov-Smirnov test[1]		1.00	1.00
1st most profitable day (date - profit)	2019/05/09 $100,142.51	2019/05/09 $82,330.71	2019/05/09 $77,292.31
5th most profitable day (date - profit)	2019/05/13 $78,509.62	2019/06/20 $58,157.95	2019/05/07 $53,633.28
1st most unprofitable day (date - profit)	2019/06/04 $15,061.17	2019/03/13 $12,210.91	2019/03/13 $9,130.81
5th most unprofitable day (date - profit)	2019/03/18 $22,810.62	2019/03/18 $15,997.46	2019/03/18 $13,349.39
Average time in trade (excl. futures contracts)[2]	00:01:29.51	00:01:39:10	00:01:41.22

1 Model	2 Limit orders	3 Limit orders 1	4 Limit orders 2
Average time in trade (incl. futures contracts)	00:01:46.55	00:01:56.61	00:01:58.19
# net profitable trades	1,063,897	930,388	892,772
# net unprofitable trades	325,351	322,230	327,096
# trades	1,389,248	1,252,618	1,219,868
% net profitable trades	76.58%	74.28%	73.19%
Average volume per trade	188.10	187.99	188.36
Average net profit per trade	$3.83	$3.19	$2.93
Average profit per net profitable trades	$9.51	$9.76	$9.84
Average profit per net unprofitable trades	-$14.71	-$15.78	-$15.94
% trade using marketable orders	16.42%	19.56%	20.50%

[1] H0: F(x) <= G(x), H1: F(x) > G(x). F(x), G(x) = CDF of daily net profits for sample 1 and sample 2, respectively: p-value of 1.00 for no latency vs 1x latency and 1.00 for 1x latency vs 3x latency.

[2] HH: MM: SS. U: hours: minutes: seconds: fractions of a second.

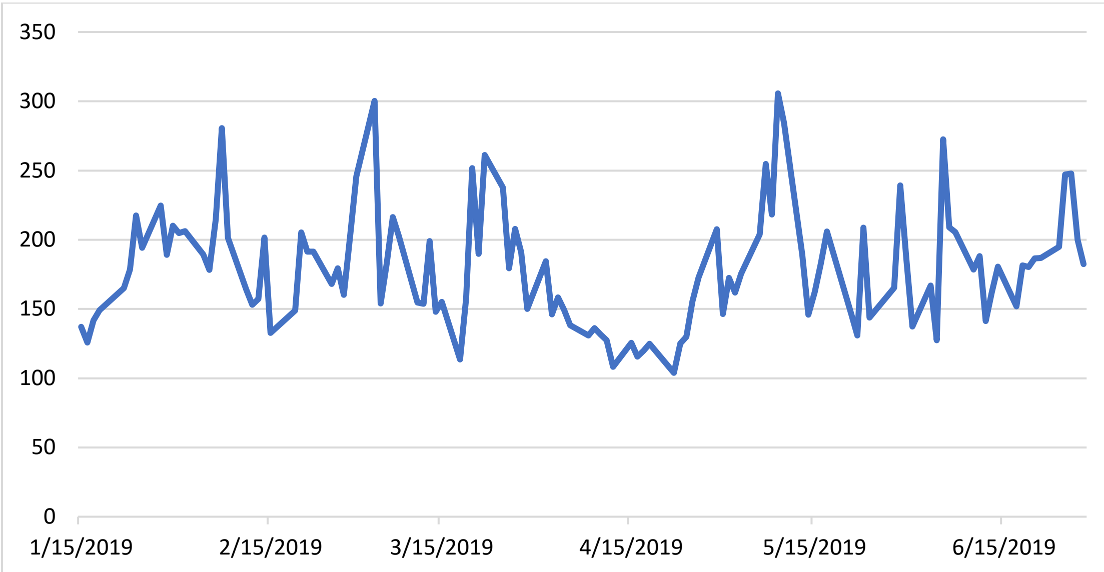
7. Trading strategy performance

7.1 Statistics

In this section, we present a more detailed view of the performance of the limit order mean-reverting strategy in the real latency setting, presented in column 3 of Table 6. We define a captured arbitrage opportunity as an opportunity where the positions in a pair at TSX and NYSE are both opened and closed with limit orders following the arbitrage strategy described in Section 3. This excludes arbitrage opportunities where a least one leg had to be closed by the stop-loss or the chronometer circuit breakers implemented for risk management.

Figure 1 shows the mean daily number of captured arbitrage opportunities per ticker and the mean duration of the positions behind these opportunities. The number of captured arbitrage opportunities (Panel 1a) exhibits some daily fluctuations, but the quantity remains stationary over the period. On average, there are 180 captured arbitrage opportunities per ticker per day. The mean duration, computed as the mean of the daily means of captured opportunity pairs, is about 122 seconds (Panel 1b), and is also stationary during our period of analysis. Note that both quantities are anticorrelated (Pearson correlation coefficient: -0.923). This is because the strategy does not enter a new position when the previous one is still opened, this condition avoids building huge inventories which would involve, among others, significant price impact when ending arbitrage activities. Thus, a longer time to close both legs of the strategy directly leads to a lesser number of potential arbitrage opportunities to be captured.

Panel 1a: Mean daily number of captured arbitrage opportunities per selected pair



Panel 1b: Mean duration in seconds per captured arbitrage opportunity over all selected pairs

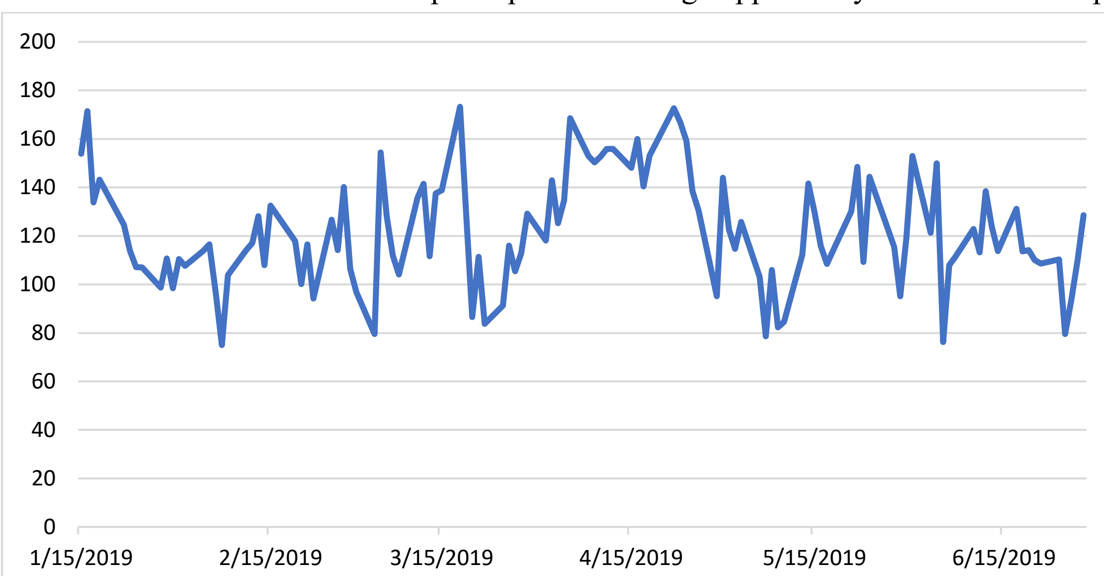
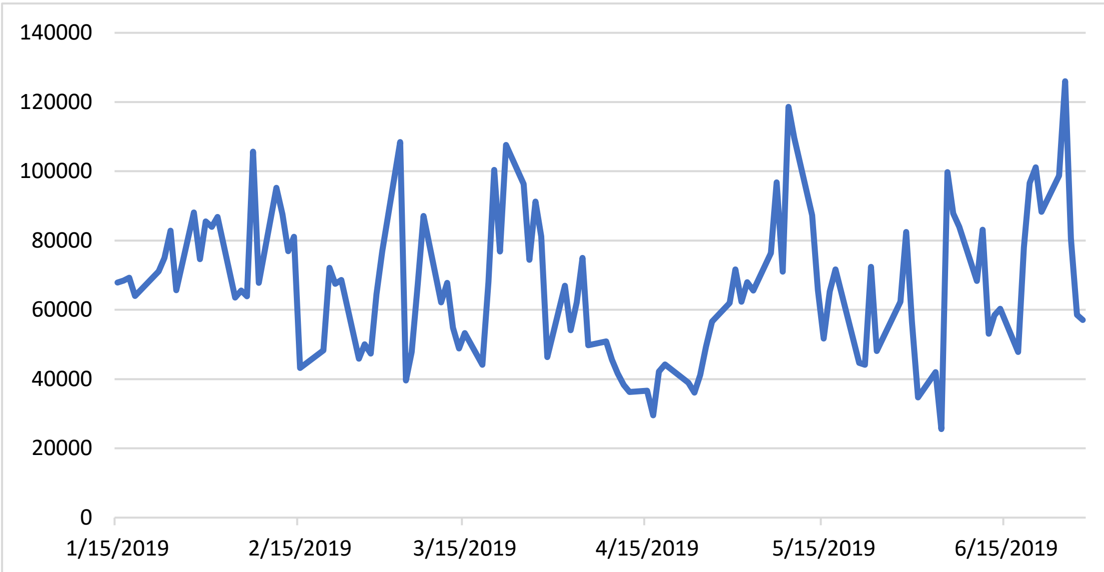


Figure 1. Captured arbitrage opportunities during the period of analysis

Figure 2 shows the daily net profit measured as the average per captured arbitrage opportunity as well as the total realized net profit per day over the selected assets in the first six months of 2019. The mean total daily realized net profit is C$67,369 (Panel 2a) and the mean net profit per captured arbitrage opportunity is around C$19 (Panel 2b), in line with the expected high- frequency quoting activities. Per ticker, the daily mean is equal to C$3,411.

Panel 2a: Total daily net profit from captured arbitrage opportunity over all selected pairs (C$)



Panel 2b: Daily net mean profit per captured arbitrage opportunity (C$)

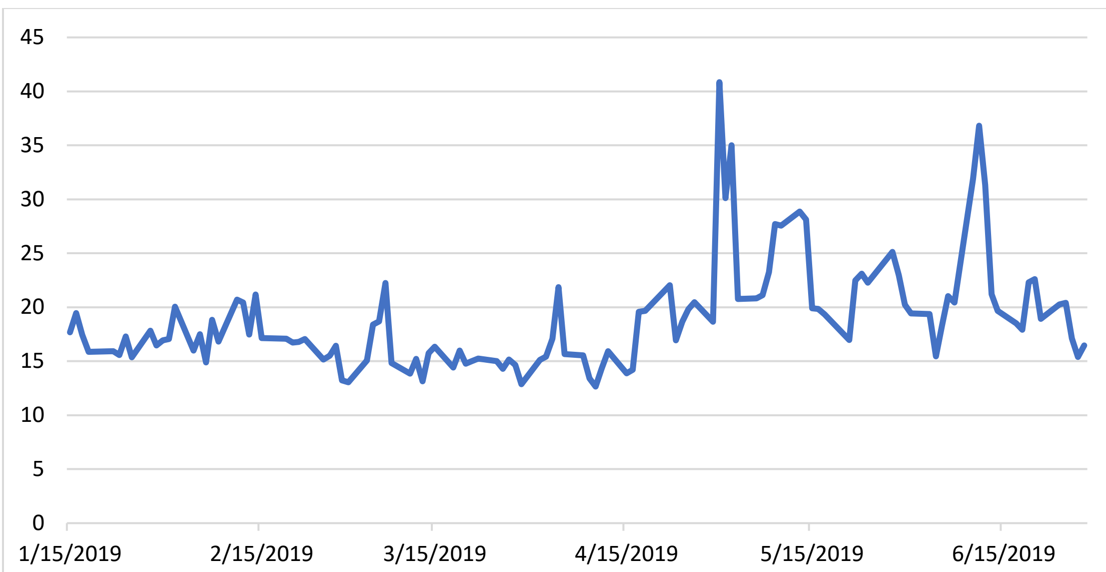


Figure 2. Profitability of captured arbitrage opportunities during the period of analysis

Figure 3 shows the empirical cumulative distribution function (CDF) of the net profit per captured arbitrage opportunity in C$. Based on this CDF, 99.7% of the captured arbitrage opportunities are profitable. The median is around C$11, and the 99 percentile is around C$110. This confirms the theoretical validity of the strategy, meaning that when an arbitrage opportunity is perfectly captured with limit orders, it is almost guaranteed to be profitable. The remaining 0.3% of unprofitable captured arbitrage opportunities are obtained because we cannot always close the positions at the exact equilibrium value, as explained in Section 4.

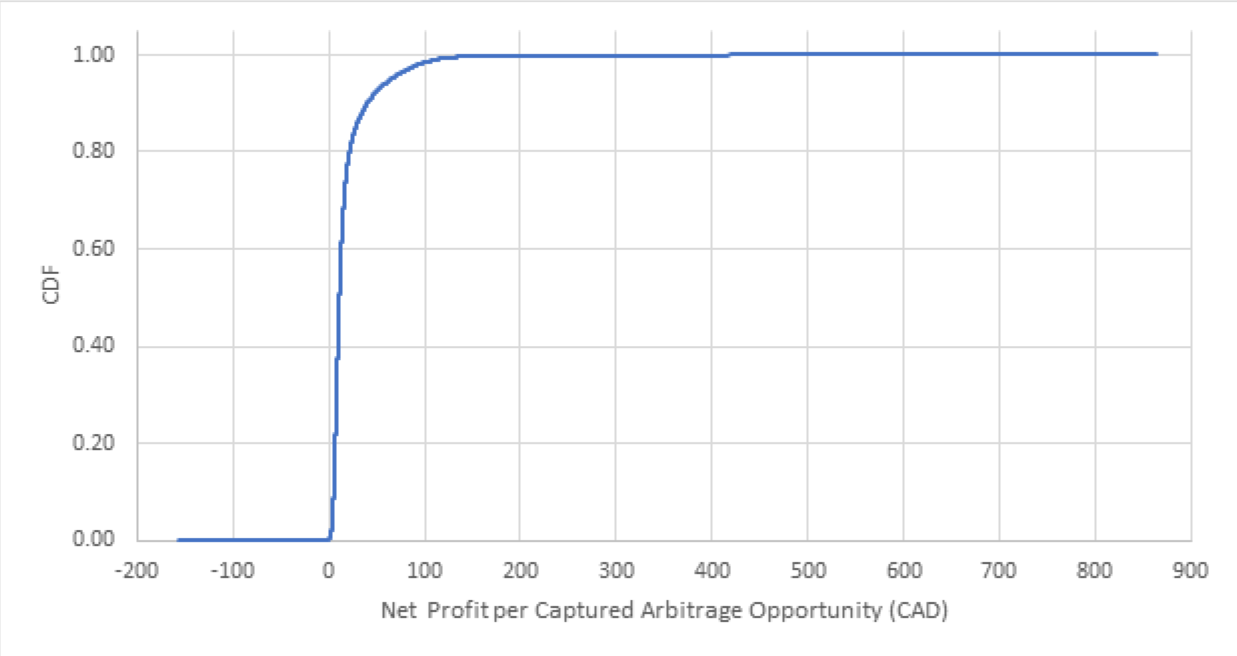


Figure 3. Empirical CDF of net profit per captured arbitrage opportunity (C$)

7.2 Regression analysis

To better understand the stylized facts affecting the daily net profitability of the strategy, we employ a regression analysis. Using standard variables such as the intraday volatility of the assets' mid-price traded at exchange $i \in \{TSX, NYSE, CME\}$ on day t ($vol_{i,t}$), the average bid-ask spreads ($spread_{i,t}$), the total trading volumes ($trade_{i,t}$), and the total quantity of messages resulting from the LOB level one updates ($messages_{i,t}$), all in their respective currency, we want to explain the average net profitability of the selected pairs on day t ($\overline{profits_t}$). We compute every variable with

$i \in \{TSX, NYSE\}$ as the weighted mean of the stock-level variable in the selected pair on day t, where the weight accorded to a specific stock is the proportion of its daily traded value compared with the total traded value for every stock of the same exchange in our portfolio on that day (all in C$). Table 7 reports the descriptive statistics of these variables, and Table A5 reports their Pearson correlation coefficients. All variables are described in Table A4.

Table 7. Descriptive statistics of variables used in the regression analysis to explain the daily net profit of the strategy with limit orders.

Variable	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max.
$\overline{profits}$	3,411	1237	1,636	2,543	3,201	4,002	8,471
vol_{TSX}	0.458	0.143	0.259	0.361	0.412	0.524	0.974
vol_{NYSE}	0.467	0.151	0.269	0.357	0.420	0.548	1.007
vol_{CME}	0.086	0.047	0.024	0.054	0.074	0.114	0.244
$spread_{TSX}$	5.791	1.267	3.567	4.916	5.688	6.213	1.097
$spread_{NYSE}$	6.854	1.279	4.800	5.835	6.732	7.385	1.079
$spread_{CME}$	0.576	0.020	0.542	0.566	0.576	0.584	0.715
$trade_{TSX}$	775,033	361,920	349,538	506,020	686,478	949,768	2,288,334
$trade_{NYSE}$	280,935	153,312	118,714	183,374	226,571	296,655	973,461
$trade_{CME}$	64,664	75,053	4,195	27,383	34,537	46,817	297,363
$messages_{TSX}$	59,136	13,474	35,229	48,619	58,494	67,034	94,736
$messages_{NYSE}$	53,719	13,253	32,036	43,001	51,791	62,492	101,549
$messages_{CME}$	192,958	55,922	66,246	150,944	186,874	223,187	346,698
Count			114				

The volatilities of the mid-price of cross-listed stocks have similar distributions on both stock exchanges. The same applies for the spread and the number of messages from LOB level one. On the other hand, the volume of trades at the TSX is almost three times greater than at the NYSE, which is expected from a portfolio composed entirely of Canadian stocks.

From Table A5, we observe a significant and positive relationship between the strategy's profitability and the volatility of the markets. The bid-ask spread of the stocks is the variable that is the most highly and positively correlated with the profitability of the strategy, which is expected since the strategy uses limit orders. Finally, the numbers of updates of LOB level one are all statistically and positively correlated to the strategy's profitability, which will be explained later in this section.

As expected, the pairs of same variables on the TSX and NYSE exchanges are highly correlated. To reduce potential multicollinearity, we combine each pair of equity variables into one variable by using the mean of the respective TSX and NYSE variable values, thus creating the variables $\overline{vol}_{stocks,t}$, $\overline{spread}_{stocks,t}$, $\overline{trade}_{stocks,t}$ and $\overline{messages}_{stocks,t}$. The linear regression model is written as follows, for day $t \in \{1,2,\dots,114\}$:

$$
\begin{aligned}
\overline{profits}_t = {} & b_0 + b_1 vol_{CME,t} + b_2 \overline{vol}_{stocks,t} + b_3 spread_{CME,t} + b_4 \overline{spread}_{stocks,t} \\
& + b_5 trade_{CME,t} + b_6 \overline{trade}_{stocks,t} + b_7 messages_{CME,t} \\
& + b_8 \overline{messages}_{stocks,t} + \varepsilon_t,
\end{aligned}
$$

where $\varepsilon_t \sim N(0,\sigma^2), \forall t$. The regression coefficients are obtained by ordinary least squares, and the covariance matrix is estimated with the heteroskedasticity and autocorrelation consistent approach of Newey-West. Table 8 summarizes the regression results.

Table 8. OLS linear regression for the average daily net profitability of the limit order strategy with Newey-West covariance matrix estimation

Variable	Coefficient	p-value
$intercept$	-3,823.011	0.202
vol_{CME}	-2,533.717	0.103
$\overline{vol}_{stocks}$	695.229	0.284
$spread_{CME}$	-1,817.241	0.723
$\overline{spread}_{stocks}$	791.142	0.000
$trade_{CME}$	0.001	0.518
$\overline{trade}_{stocks}$	-0.002	0.009
$messages_{CME}$	0.003	0.139
$\overline{messages}_{stocks}$	0.069	0.000
$Adj.\ R^2$	0.662	
$F\ stat$	22.570	

As the regression suggests, the number of LOB level one update messages, the size of the spread and the trading volume of the stocks contribute significantly to the daily net profits generated for our portfolio of cross-listed stock pairs. These results are consistent with our machine learning pair selection methodology (See Online appendix C for more details). A larger spread for the stocks is directly beneficial to our limit order strategy, which can be explained by equations (4), (5) and (6). Together, these equations tell us that a larger spread lead to a higher profit for any given arbitrage opportunity and that the profitable arbitrage opportunities are thus more frequent for days with larger spreads. As for the number of messages, the result is intuitive because a higher level one activity generally increases the likelihood of our active limit orders to be filled, or cancelled because of our risk management circuit breakers in the case where the prices deviate from our limit orders' prices. Hence, the more messages we observe, the faster our orders can be executed or canceled and the faster the strategy can move on to the next opportunity (which was observed in Figure 1), as opposed to days when markets are quieter and limit orders can remain in the LOB for longer periods of time. Lastly, a larger trading volume contributes negatively to our profitability, especially at the NYSE. The higher latency to that exchange prevents us from reacting very rapidly compared to other participants collocated at the NYSE. Thus trades occurring before our limit orders included in the LOB (or even before the information was analyzed by our algorithm) can cause the mispricing to dissipate.

7.3 Profitability

Figure 4 shows the net cumulated profits over the entire period on a trade basis. There is minimal intraday drawdown, and as was shown in Figure 2 (Panel a), the net daily profits are stationary, which explains the quasi linearity of the function in Figure 4.

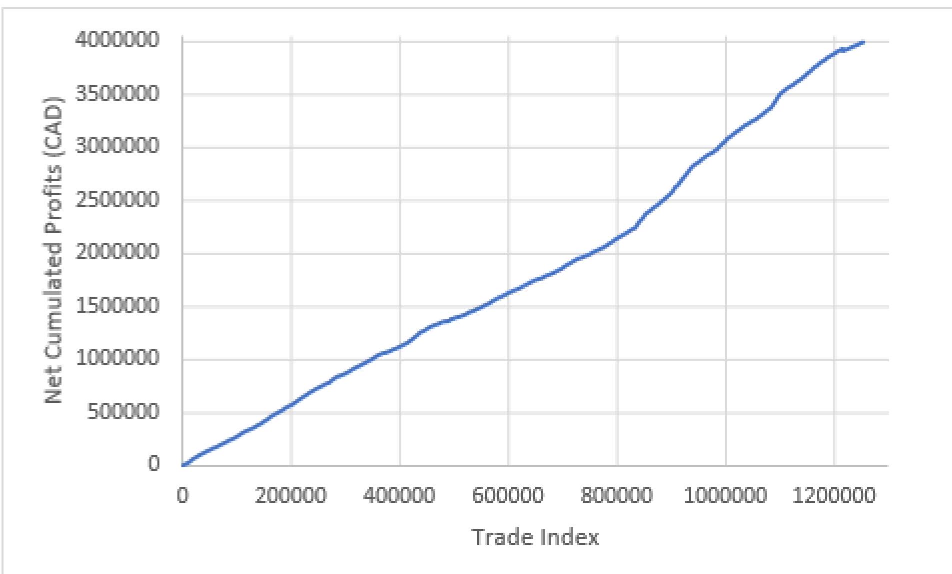


Figure 4. Net cumulated profits (C$) on a trade basis over the entire period

Figure 5 presents the daily maximum of net aggregated positions taken at each exchange for our portfolio of selected pairs. The maximum net open position in absolute value is around C$453,000 at the TSX, C$465,000 at the NYSE, and C$590,000 at the CME, meaning that an investment of C$1,000,000 to cover the margins is more than enough. Note that only a margin of

US$1,100 per C/US futures contract is needed at the CME. Given the annual net profit of C$8 million generated by the strategy in 2019, this results in an annual net return of 700%. When considering management fees of 5%, the annual net return is 660%.

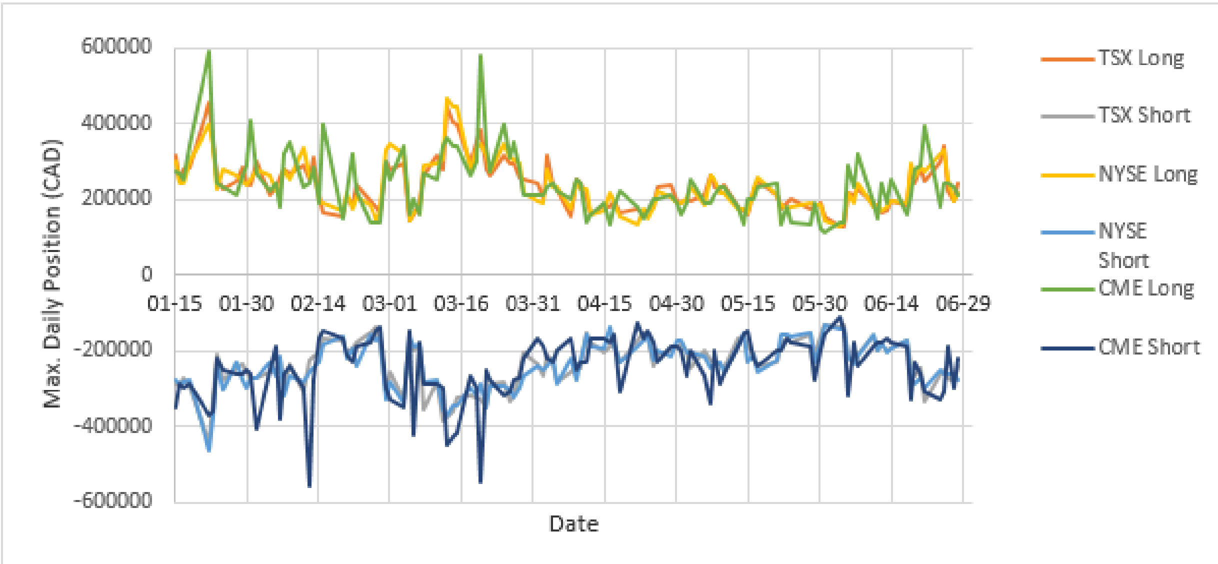


Figure 5. Maximum daily net aggregated long and short positions of the selected pairs portfolio at the three exchanges

Figure 6 shows the empirical CDF of the needed aggregated net margin in C$. This margin at time t can be expressed as follows:

$$M_t = \left|V_{TSX,t}\right| + \left|V_{NYSE,t} - \frac{G_{NYSE,t}}{r_t}\right| + \frac{1{,}100}{r_t}\left|V_{CME,t} - \frac{G_{CME,t}}{r_t}\right| /100{,}000\,.$$

where $V_{TSX,t}$, $V_{NYSE,t}$ and $V_{CME,t}$ are the portfolio exposure in C$ in their respective exchange.

Once again, we can see that a capital of C$1,000,000 always covers the margins in the three exchanges, while C$185,000 covers 80% of the needed margins at any time, meaning that the high levels of aggregated positions are transitory.

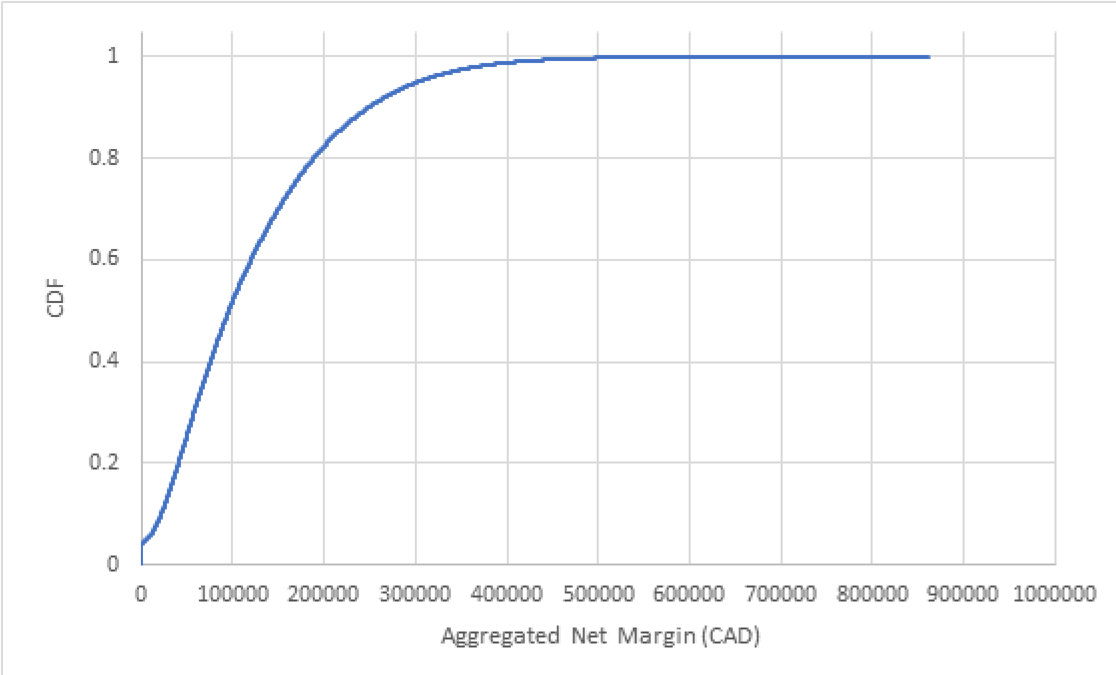


Figure 6. Empirical CDF of the needed aggregated net margin in C$

The annualized Sharpe ratio computed from the daily returns and the margin of C$1 million is 51.04 (48.5 when considering management fees). It is very high, but our daily profits are perfectly comparable to the trading profits of HFTs found in Baron et al (2014). Our result is mainly explained by the very low volatility of the profits as seen in Figures 3 and 5. Also the Deflated Sharpe Ratio proposed by Bailey and López de Prado (2014) is approximately equal to 1. This very high value is mainly explained by the fact that we did not resort to multiple back testing trials, generating an absence of variance across the trials and a quasi null likelihood of a false discovery.

8. Conclusion

We study the profitability of mean-reverting arbitrage activities of international cross-listed stocks on two stock exchanges and a derivatives exchange with a novel trading strategy that is generalizable to a broader cross-listed universe. The theoretical strategy signals when the prices of cross-listed stocks deviate enough from their relative equilibrium that an economically viable arbitrage opportunity occurs. We apply the model to North American markets during the first six months of 2019, namely to the New York Stock Exchange (NYSE) and the Chicago Mercantile Exchange (CME) in the United States, and the Toronto Stock Exchange (TSX) in Canada.

This paper is the first to examine stocks' cross-country mean-reverting arbitrage. We work with a unique temporal frame of reference, meaning that we synchronize the data feeds from the exchange venues by explicitly taking into account the latency that comes from the transmission of information between the exchanges and the information processing time. We also consider all potential arbitrage trading costs. We show that mean-reverting arbitrage is profitable with order book transactions and queuing priorities. We consider the obtained profits as reasonable when compared with previous contributions in the literature. In previous studies, the profitability of latency arbitrage is often overestimated by not considering both the practical aspects of arbitrage trading and the market frictions in their applications. International latency arbitrage with market orders is not profitable with our data.

Our original goal was not to contribute to the normative discussion about the effect of continuous HFT on the general welfare of financial markets. Rather, it was to replicate the precise behavior of a trading firm to provide a better estimate of the arbitrage market functioning with high-frequency trading. Our research highlights the high-frequency arbitrageur's economic incentive to act as a liquidity provider and the importance of considering real market frictions in HFT research. Our results could be useful to improve the understanding of the complex nature of high-frequency trading. Our model can be deployed in a real-time environment by institutional investors, professional arbitrageurs, market makers, hedgers, and regulators. Our approach provides a contemporary understanding of an economically viable arbitrage approach that helps restore equilibrium in financial markets.

Arbitrage activities are very useful to restore equilibria in markets when price distortions are observed. These activities are usually carried out by the largest traders under strong competition. These traders provide the markets with liquidity and are remunerated for this activity. Are the profits they earn too high? The results of this study do not provide a conclusive answer to this question, but we have demonstrated that large traders can make positive profits under fair trading conditions.

Another issue often discussed in the literature is the costly race for high-speed trading. This race is almost over because the realized observed speeds for information transmission between exchanges by the largest traders have reached their physical limits when compared with the speed of light (Anova, 2022). The same observation can be made for information processing, where the inter-server latency is converging to the propagation delay of light (Thomas et al., 2018). It is not clear how additional regulation that targets speed reduction could improve economic welfare in current markets.

Finally, do these arbitrage activities affect long-term investors who are not involved in arbitrage activities, which represent the vast majority of stock investors? We do not have sufficient data to answer this question, but our discussions with investors and traders seem to confirm that the effect is small. This issue warrants additional quantitative research.

References

Ait-Sahalia, Y., Saglam, M., 2017. High frequency market making. SSRN Electronic Journal.

Aquilina, M., Budish, E., O'Neill, P., 2022. Quantifying the high-frequency trading "arms race". The Quarterly Journal of Economics 137, 493–564.

Bailey, D.H., Borwein, J.M., Lopez de Prado, M., Zhu, Q.J., 2014. Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. Notices of the American Mathematical Society 61, 458-471.

Bailey, D. H., López de Prado, M., 2014. The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. Journal of Portfolio Management 40, 94–107.

Baron, M., Brogaard, J., Hagstromer, B., Kirilenko, A., 2019. Risk and return in high-frequency trading. Journal of Financial & Quantitative Analysis 54, 993-1024.

Baron, M., Brogaard, J., Kirilenko, A., 2014. The trading profits of high frequency traders. Unpublished working paper, Princeton University, University of Washington, Massachusetts Institute of Technology.

Biais, B., Foucault, T., Moinas, S., 2015. Equilibrium fast trading. Journal of Financial Economics 116, 292-313.

Biais, B., Woolley, P., 2011. High frequency trading. Working paper. Toulouse School of Economics, University of Toulouse. Toulouse, France.

Blume, M., Goldstein, M.A., 1991. Differences in execution prices among the NYSE, the regionals, and the NASD. Available at SSRN: https://ssrn.com/abstract=979072 or http://dx.doi.org/10.2139/ssrn.979072.

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., Sokolov, K., 2018. High frequency trading and extreme price movements. Journal of Financial Economics 128, 253-265.

Brolley, M., 2020. Price improvement and execution risk in lit and dard markets. Management Science 66, 863-886.

Brunnermeier, M. K., Pedersen, L. H., 2008. Market liquidity and funding liquidity. The Review of Financial Studies 22, 2201-2238.

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. The Quarterly Journal of Economics 130, 1547-1621.

Chao, Y., Chen, Y., Mao, Y., 2019. Why discrete price fragments U.S. stock exchanges and disperses their fee structures. The Review of Financial Studies 32, 1068-1101.

Chen, H., Chen, S., Chen, Z., Li, F., 2019. Empirical investigation of an equity pairs trading strategy. Management Science 65, 370-389.

Chen, Y., Da, Z., Huang, D., 2019. Arbitrage trading: The long and the short of it. The Review of Financial Studies 32, 1608-1646.

Chouinard, E., D'Souza, C., 2003. The rationale for cross-border listings. Bank of Canada Review, Winter, 23-30.

Conrad, J.S., Wahal, S., 2020. The term structure of liquidity provision. Journal of Financial Economics 136, 239-259.

Dahlström, H., Nordén, L.L., 2018. The determinants of limit order cancellations. Mimeo, Stockholm Business School, Sweden. 57 p.

DeltixLabs, 2020. The Deltix product suite: Features and benefits. Retrieved Feb. 25, 2021 at: https://www.deltixlab.com/products/.

Dewhurst, D.R., Van Oort, C.M., Ring IV, J.H., Gray, T.J., Danforth, C.M., Tivnan, B.F., 2019. Scaling of inefficiencies in the U.S. equity markets: Evidence from three market indices and more than 2900 securities. arXiv:1902.04691v2 [q-fin.TR].

Ding, S., Hanna, J., Hendershott, T., 2014. How slow is the NBBO? A comparison with direct exchange feeds. The Financial Review 49, 313-332.

Dixon, M.F., Polson, N.G., and Sokolov, V.O., 2019. Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading. Applied Stochastic Models in Business and Industry 35, 788-807.

Dugast, J., 2018. Unscheduled news and market dynamics. The Journal of Finance 73, 2537-2586.

Engelberg, J., Pengjie G., Jagannathan, R., 2009. An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity. Third Singapore International Conference on Finance.

Eun, C.S., Sabherwal, S., 2003. Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks. The Journal of Finance 58, 549-575.

Foucault, T. and Biais, B., 2014. HFT and market quality. Bankers, Markets & Investors 128, 5-19.

Foucault, T., Kozhan, R., Tham, W.W., 2017. Toxic arbitrage. The Review of Financial Studies 30, 1053-1094.

Foucault, T., Moinas, S., 2019. Is Trading Fast Dangerous? In: W. Mattli (Ed.) Global Algorithmic Capital Markets: High Frequency Trading, Dark Pools, and Regulatory Challenges, Oxford University Press.

Frazzini, A., Israel, R., Moskowitz, T.J., 2018. Trading costs. SSRN Electronic Journal. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3229719.

Friederich, S., Payne, R., 2015. Order-to-trade ratios and market liquidity. Journal of Banking & Finance 50, 214-223.

Gagnon, L.J., Karolyi, G.A., 2010. Do International Cross-Listings Still Matter? Evidence On Financial Globalization and Crises, Thorsten Beck, Sergio Schmukler, Stijn Claessens, eds., Elsevier North-Holland Publishers.

Garriott, C., Pomeranets, A., Slive, J., Thorn, T., 2013. Fragmentation in Canadian equity markets (white paper). Bank of Canada Review. Ottawa, Canada.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K. G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies 19, 797-827.

Goldstein, M., Kumar, P., and Graves, F.C., 2014. Computerized and high frequency trading. The Financial Review 49, 177-202.

Krauss, C., 2017. Statistical arbitrage pairs trading strategies: Review and outlook. Journal of Economic Surveys 31, 513-545.

Hasbrouck, J. and Saar, G., 2013. Low-Latency Trading. Journal of Financial Markets 16, 646-679.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. The Journal of Finance 50, 1175-1199.

Hendershott, T., Jones, C., and Menkveld, A., 2011. Does algorithmic trading improve liquidity? The Journal of Finance 66, 1-33.

Jones, C.M., 2013. What do we know about high-frequency trading? Columbia Business School Research Paper. No. 13-11.

Kozhan, R., Tham, W.W., 2012. Execution risk in high-frequency arbitrage. WBS Finance Group Research Paper no 179. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2030767.

Krauss, C., 2017. Statistical arbitrage pairs trading strategies review and outlook. Journal of Economic Surveys 31, 513-545.

Lee, C.M.C., 1993. Market integration and price execution for NYSE-listed securities. The Journal of Finance 48, 1009-1038.

Liu, W.M., 2009. Monitoring and limit order submission risks. Journal of Financial Markets, 12, 107-141.

Mavroudis, V., 2019. Market manipulation as a security problem. Working paper, University College London.

Menkveld, A.J., 2016. The Economics of high-frequency trading: Taking stock. Annual Review of Financial Economics 8, 1-24.

Menkveld, A.J., 2014. High-frequency traders and market structure. Financial Review 49, 333-344.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

O'Hara, M., 2015. High frequency market microstructure. Journal of Financial Economics 116, 257-270.

O'Hara, M., 2016. Something for Nothing: Arbitrage and Ethics on Wall Street. W.W. Norton et Company.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42, 35-52.

Rein, C., Rüschendorf, L, Schmidt, T., 2021. Generalized statistical arbitrage concepts and related gain strategies. Mathematical Finance 31, 563-594.

Riordan, R., Storkenmaier, A., 2012. Liquidity and price discovery. Journal of Financial Markets 15, 416-437.

Scherrer, C.M., 2018. Information processing on equity prices and exchange rate for cross listed stocks. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3247594.

Shkilko, A., Sokolov, K., 2020. Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs. The Journal of Finance 75, 2899-2927.

Shkilko, A.V., Van Ness, B.F., Van Ness, R.A., 2008. Locked and crossed markets on NASDAQ and the NYSE. Journal of Financial Markets 11, 308-337.

Stübinger, J., Bredthauer, J., 2017. Statistical arbitrage pairs trading with high-frequency data. International Journal of Economics and Financial Issues 7, 650-662.

Tassel, P.V., 2020. The law of one price in equity volatility markets. Federal Reserve Bank of New York, Staff Reports, no. 953.

Tivnan, B.F., Dewhurst D.R., Van Oort, C.M., Ring IV, J.H., Gray, T.J., Koehler, M.T.K., McMahon, M.T., Slater, D., Veneman, J., Danforth, C.M., 2019. Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30. arXiv:1902.04690v3 [q-fin.TR].

Thomson Reuters, 2019. https://www.reuters.com/article/us-usa-exchanges-idUSKBN25H23K.

TSX, 2020. Co-Location Fee Schedule. https://www.tsx.com/resource/en/1756/tsx-trading-fee-schedule-effective-june-4-2018-en.pdf.

Wah, E., 2016. How prevalent and profitable are latency arbitrage opportunities on U.S. stock exchanges? Available at SSRN: https://ssrn.com/abstract=2729109 or http://dx.doi.org/10.2139/ssrn.2729109.