Correlated High-Frequency Trading

Ekkehart Boehmer, Dan Li, and Gideon Saar*

This version: September 2016

Abstract

In this paper, we examine product differentiation in the high-frequency trading (HFT) industry by looking at the correlated behavior of HFT firms. Since the "product" of an HFT firm is a proprietary trading strategy, we use a principal component analysis to detect three underlying strategies that are common to multiple HFT firms. We show that the short-horizon volatility of most stocks loads negatively on the extent of market-wide competition between HFT firms, and document a negative relation between HFT competition and market concentration, presenting evidence that smaller trading venues are more viable when HFT competition is higher.

*Ekkehart Boehmer (eboehmer@smu.edu.sg) is from the Lee Kong Chian School of Business, Singapore Management University; Dan Li (lidan@hku.hk) is from the Faculty of Business and Economics, University of Hong Kong; and Gideon Saar (gs25@cornell.edu) is from the Johnson Graduate School of Management, Cornell University. We thank Matthew Baron, Evangelos Benos, James Brugler, Bjorn Hagstromer, Johan Hombert, Adam Joseph-Clark, Elvis Jarnecic, Boyan Jovanovic, Andrew Karolyi, Albert Menkveld, Sophie Moinas, Dermot Murphy, Maureen O'Hara, Mike Prior, Ioanid Rosu, Mehmet Saglam, Mehrdad Samadi, and Giorgio Valente for helpful comments, as well as conference and seminar participants at the 13th Annual IDC Conference, the 3rd Annual SEC/CFP Conference on Financial Market Regulation, Cornell University, the PBC School of Finance at Tsinghua University, Tel Aviv University, and VU Amsterdam. Dan Li thanks the Investment Industry Regulatory Organization of Canada (IIROC) for providing her with the data used in this study during her work at IIROC as part of a research agreement, and Maureen Jensen for her support and many helpful discussions. All conclusions are those of the authors and do not represent the views of IIROC. Part of this research was conducted while Gideon Saar was visiting the School of Management at Tel Aviv University.

Industry competitiveness is an important theme in the economics of industrial organization. It is of particular interest when the industry in question plays a central role in the performance of our securities markets, which are essential to information production and allocation of capital. In this paper, we take a closer look at product differentiation and intra-industry competition in the high-frequency trading (HFT) industry using unique regulatory data from Canada. HFT firms use computerized algorithms for proprietary trading, and they engage in electronic market making, cross-trading venue price arbitrage, short-term statistical arbitrage, and various other opportunistic strategies.[1] This subset of financial service firms is responsible for most of the order flow and about half of the trading in equity markets in both the United States and Canada. The dominant presence of these firms gave rise to concerns over both the fairness and stability of equity markets, and a contributing factor to these concerns has been the secrecy that surrounds their operations.

As in Hoberg and Philips (2016), our study is based on the concept that product similarity is germane to industry boundaries. Given that the "product" of an HFT firm is a proprietary (and secretive) trading strategy, we first grapple with how to define product categories and assess how many firms compete in each one. The number of direct competitors provides crucial information concerning the competitiveness of the industry. We then investigate how competition between HFT firms impacts the market, and focus on two questions. First, we study whether greater competition, manifested as more correlated HFT strategies, increases the short-horizon volatility of individual stocks. Second, we examine the relation between HFT competition and market concentration in today's fragmented equity market.

Our investigation of these issues is facilitated by data from the Investment Industry Regulatory Organization of Canada (IIROC), which is the national self-regulatory organization that oversees all equity trading venues in Canada. The unique features of these data include the ability to observe the HFT activity of both licensed dealers and HFT firms that use direct market access (DMA) arrangements, as well as the ability to track the activity of each HFT firm across all trading venues in Canada. Our sample consists of S&P/TSX 60 Index stocks, and we analyze activity from 30 days that represent bullish, bearish, and neutral trading environments during a period from June 2010 through March 2011.

[1] High-frequency trading therefore does not include agency algorithms that execute orders on behalf of investors; the term "HFT" is used only to denote proprietary trading operations conducted by standalone firms or the trading desks of larger financial firms.

We characterize 31 trading firms as "high-frequency traders," and perform a principal component analysis (PCA) to ascertain product differentiation in this industry. This data-driven methodology helps us detect underlying strategies that are common to multiple HFT firms and to identify the firms that follow each one. We find evidence of competition in at least three distinct underlying common strategies. While there are 17 HFT firms that do not appear to pursue one of these common strategies, the 14 firms that follow the three common strategies we identify represent most of the HFT activity: 96.21% of the messages that HFT firms send to the market and 78.97% of the volume they trade. Therefore, we find that competing HFT firms in three differentiated product categories generate most of the HFT activity in the Canadian market.

The PCA does not attach any economic interpretation to the underlying common strategies it identifies. We therefore analyze the principal component scores by regressing them on various attributes of the HFT firms and the market environment. The regression outputs could be interpreted to suggest that one of the principal components represents a cross-venue arbitrage strategy, another one involves market making, and a third is related to short-horizon directional speculation. While we stress that these labels are merely suggested by patterns of regression coefficients (they do not constitute a definitive characterization of these strategies), the important takeaway from our analysis is that there is clear product heterogeneity in this industry. Therefore, we believe it may be more constructive to discuss the impact of HFT on the market in the context of specific strategies rather than in the aggregate.

We identify product categories in this industry and characterize competition between HFT firms by looking at correlated HFT activity as it represents similarity in the firms' strategies and the manner in which they respond to market stimuli. If such competition means that HFT firms engage in strategies that are highly correlated across stocks, the central role they play in our markets could exacerbate return movements and increase stock volatility. Our investigation uncovers the opposite result: the short-horizon volatility of most stocks loads negatively on the extent of market-wide competition between HFT firms. This contrasts with the positive relations that both market volatility and the aggregate magnitude of HFT activity have with individual stock volatility, suggesting that the negative relation with HFT competition constitutes a separate and distinct effect. Our results show that the strongest driver behind this negative relationship is competition between HFT firms that follow the second principal component, which we (loosely) associate with a market-making strategy, suggesting that the reduction in volatility could stem in part from a reduction in transitory price movements (e.g., Ho and Stoll 1983). These results indicate that HFT competition may bring dual benefits to the market in the form of potential lower rents for the services HFT firms provide as well as lower volatility. We stress, however, that our analysis reflects normal market conditions and this conclusion may not hold during an intense market breakdown, such as the Flash Crash in May of 2010 in the United States.

We also investigate how competition between HFT firms relates to competition between trading venues in the market. One role for HFT firms in a fragmented market structure is to be the market consolidators that transform the environment into a virtual central electronic limit order book (from the perspective of most investors) by quickly moving orders across markets to ensure that prices are the same across trading venues and liquidity exists where it is needed. We show that concentration of trading in the market is negatively related to competition between HFT firms that follow two of the three common HFT strategies. We further study one driver of the relationship between HFT competition and market concentration by looking at whether higher correlation in HFT strategies on a specific trading venue increases the percentage of time that the venue features the best prices or the narrowest spreads, which we view as measures of the viability or competitiveness of the venue. We find that smaller trading venues are more viable while the dominant trading venue is less viable when HFT competition is more intense, which is likely one of the drivers behind the negative relation we document between HFT competition and market concentration.

We contribute to the literature by presenting new insights on competition in the HFT industry and how it impacts the market. Our most important contribution is the ability to define three product categories in the HFT industry, despite the complete lack of information about products in this space, and identify the firms that compete directly with one another in each category. We also provide new insights on the impact of HFT competition: we show that it lowers the short-horizon volatility of most stocks, and that it is instrumental in making the market less concentrated by helping the viability of smaller trading venues.

1. Literature Review

Our paper joins a rapidly expanding body of literature on HFT in financial markets. For recent surveys on the topic of HFT, see Jones (2013) and Goldstein, Kumar, and Graves (2014). Among the theoretical contributions are those of Han, Khapko, and Kyle (2014), Hoffmann (2014), Biais, Foucault, and Moinas (2015), Ait-Sahalia and Saglam (2016), Foucault, Hombert, and Rosu (2016), Jovanovic and Menkveld (2016), and Rosu (2016). In particular, Jarrow and Protter (2012) show that, when HFT firms respond to common signals, their correlated activity affects market prices, thereby increasing market volatility and generating abnormal profits for these firms. While we indeed show that multiple HFT firms pursue correlated strategies, we find that the short-horizon volatility of most stocks loads negatively, not positively, on a measure of cross-sectional correlation between HFT strategies. Budish, Cramton, and Shim (2015) and Menkveld and Zoican (2016) model HFT firms that pursue heterogeneous strategies in the market, which we also document empirically.

Many empirical contributions focus on intraday analysis of aggregate HFT behavior (e.g., Carrion 2013; Hasbrouck and Saar 2013; Brogaard, Hendershott, and Riordan 2014; Jarnecic and Snape 2014; Hirschey 2016; Kirilenko et al. 2016). Several papers use data on trading by individual HFT firms, rather than aggregate behavior, to investigate HFT strategies (Baron, Brogaard, and Kirilenko 2012; Hagstromer and Norden 2013; Hagstromer, Norden, and Zhang 2014; Benos and Sagade 2016), although they focus predominantly on whether HFT firms are demanding or supplying liquidity. Clark-Joseph (2014) uses index futures data from the CMA to examine exploratory trading by HFT firms.

More closely related to our analysis, Hagstromer and Norden (2013) find that HFT firms that predominantly supply liquidity appear to mitigate intraday volatility, which complements our finding pertaining to the relation between HFT competition and volatility, although they do not address the competition dimension.[2] Breckenfelder (2013) and Brogaard and Garriott (2016) examine one aspect of competition: the entry and exit of HFT firms. Specifically, Brogaard and Garriott (2016) analyze data from one alternative trading system in Canada and show that new entrants take volume away from

[2] Brogaard, Carrion, Moyaert, Riordan, Shkilko, and Sokolov (2016) identify short intervals with large price movements and show that NASDAQ HFT firms in the aggregate supply rather than demand liquidity during these intervals, hence possibly dampening volatility. Hasbrouck and Saar (2013) also find that aggregate HFT activity appears to lower the intraday volatility of NASDAQ stocks.

incumbents even as they increase the overall market share of HFT firms, and these effects decline with each successive entry. They also find that market liquidity improves after the entry of an HFT firm and deteriorates after an exit (especially when there are only one or two HFT firms trading in a given stock). Breckenfelder (2013) uses data from the Stockholm Stock Exchange and finds the opposite result: deterioration of liquidity for entries of HFT firms and improvement for exits. Menkveld (2013) examines the strategy of one HFT firm and makes the case that this particular firm enhances the viability of a new trading venue, which is related to our result that competition between HFT firms is positively related to the viability of smaller trading venues and negatively related to the viability of the dominant trading venue.

Chaboud et al. (2014) investigate algorithmic trading on the foreign-exchange EBS platform. While they observe only the aggregate trading of algorithmic traders, they create a measure of correlated algorithmic trading by comparing the frequency with which algorithmic traders are on both sides of a transaction with a benchmark model that assumes an independent matching of algorithmic traders and humans. The higher their measure, the fewer triangular arbitrage opportunities are observed in the market (and hence the more efficient are prices), but they find no relation between their measure and the autocorrelation of returns (or excess volatility). Benos et al. (2015) use transactions data from the London Stock Exchange to study how ten HFT firms that are regulated by the U.K. Financial Conduct Authority interact with each other. They use vector autoregressions to show that there is a positive dynamic relationship between HFT firms: aggressive buying by one firm tends to follow aggressive buying by another (and similarly for aggressive selling), and firms tend to trade in response to the past trading activity of other firms. Benos et al. construct a measure of concurrent directional trading on the part of their ten HFT firms that is meant to capture correlated behavior and show that it has positive contemporaneous and lagged relationships with returns.[3]

Lastly, our paper joins several other papers that use Canadian order-level data to investigate HFT-related issues. Malinova, Park, and Riordan (2014) study the cost of trading around changes in market

[3] Anand and Venkataraman (2016) investigate the correlated activity of market makers on the Toronto Stock Exchange, some of which are HFT firms that operate as Electronic Liquidity Providers (ELPs). They find that market makers scale back their activity when market conditions are unfavorable, which can be one of the drivers behind commonality in liquidity.

structure that mainly affected HFT firms and other algorithmic traders. Comerton-Forde, Malinova, and Park (2016) study the nature of liquidity provision around changes in dark trading regulation, while Korajczyk and Murphy (2016) examine HFT liquidity provision to large institutional trades. Brogaard, Hendershott, and Riordan (2016) study how both limit orders and trades of high-frequency traders contribute to price discovery.

2. The Canadian Market: Sample, Data, and HFT Firms

Our data come from the Investment Industry Regulatory Organization of Canada (IIROC). All trading venues in Canada are required to provide data feeds to IIROC, which performs both real-time and post-trade market surveillance of trading activities. Traders need to obtain an IIROC membership to directly connect to trading venues in Canada, and IIROC admits only security dealers as members. Other financial firms, such as asset managers, banks, insurance companies, and proprietary trading firms, can trade through dealers' brokerage arms or via direct market access (DMA) arrangements provided by dealers. DMA arrangements allow non-dealer trading firms to directly access trading venues without having to hand over their orders to brokers for execution.

During our sample period (June 2010 through March 2011), Canada has five trading venues organized as electronic limit order books that trade stocks listed on the Toronto Stock Exchange: Alpha ATS Limited Partnership (ALF), Chi-X Canada ATS (CHX), Omega ATS (OMG), Pure Trading (PTX), and the Toronto Stock Exchange (TSX).[4] Trading on crossing networks ("dark trading") in Canada during our sample period is limited to essentially one dark pool (MATCH Now) with no more than a 3% market share.[5]

2.1 Sample

Our empirical work is carried out on 30 trading days that are selected to capture variation across market conditions. We rank the daily returns of the S&P/TSX Composite Index from June 2010 through March

[4] Alpha became a stock exchange on April 1, 2012. In July 2012, Alpha was acquired by the TMX Group, which also owns TSX. During our sample period, however, Alpha and TSX were independent trading venues.

[5] MATCH Now also provides a real-time data feed to IIROC and is included in our data. Liquidnet Canada, another dark pool, executed only a few trades each day during our sample period. As a result, IIROC did not require it to participate in the real-time centralized data feed, instead requiring it to submit trade information manually at the end of the trading day.

2011, and select the 10 worst days (down days), the 10 best days (up days), and the 10 days closest to (and centered on) zero return (flat days). In other words, we take the two extremes in terms of days in which the market went up or down the most, as well as the days with the least return movement. This design allows us to examine whether the correlation structure of HFT strategies depends on market conditions (as summarized by the daily return on a broad index).[6]

Our sample stocks consist of 52 constituents of the S&P/TSX 60 Index, which accounts for approximately 73% of Canada's total equity market capitalization. Eight stocks are excluded from the Index, as they were converted from income trusts to corporate structures (five), were delisted (one), had their symbols changed (one), or were listed for less than one year before the start of our sample period (one). Panel A of Table 1 presents summary statistics for the sample.[7] The mean market capitalization is 19.4 billion Canadian dollars (CAD), with an average stock price of 39.1 CAD, and average daily volume of 78.2 million CAD. Panel A clearly shows that our sample period encompasses three distinct market conditions: the average daily returns of stocks on down, flat, and up days are –1.72%, –0.08%, and 1.66%, respectively.

2.2 Data

The order-level data we obtain from IIROC cover all Canadian trading venues, and contain information about order submissions, cancellations, modifications, and executions with 10-millisecond time stamps. The time stamps from all trading venues are synchronized with the regulator's time stamps and reflect the local time at which a message (a general term used for submissions, cancellations, and executions of orders) is processed. The record of each message contains the following information: ticker symbol, order side (buy or sell), trading venue, price, total quantity, non-displayed quantity, broker ID, trader ID, order type (e.g., client orders, inventory trading), various flags (e.g., short sell, market on close, opening trade),

[6] While the U.S. enforces price priority only for the top of the book (the best bid and ask prices), the Canadian Securities Administrators (CSA) introduced an Order Protection Rule, effective on February 1, 2011, whereby the full depth of the book is protected across all marketplaces, requiring all visible, immediately accessible, better-priced limit orders to be filled before other limit orders at inferior prices. Only one of our 30 days (a down day) is post February 1, 2011, and hence in a market environment that is subject to the new order protection rule. In Section 2.4, we show that the correlations between HFT firms are similar in the down, flat, and up days. Therefore, we do not believe that the inclusion of this day affects our results.

[7] Data on stock characteristics are obtained from the Summary Information Database of the Canadian Financial Markets Research Centre (CFMRC).

and order/trade ID. Trade messages are identified as buyer-initiated or seller-initiated. Events in an order's life, including modification, partial fill, full fill, and cancellation have the same order ID.

One advantage of our data is that the same trader IDs are used on all trading venues in Canada. Most HFT firms in Canada do not operate as licensed dealers but rather gain access through DMA arrangements with one or multiple dealers. We obtain tables that identify the trader IDs of all trading firms that use DMA arrangements. Hence, we can accurately detect the activity of each HFT firm on all trading venues irrespective of whether it trades via multiple DMA arrangements with dealers or uses multiple trader IDs.

2.3 Identifying HFT Firms

We use our own procedure to identify HFT firms rather than adopting a classification provided by an exchange. Since we have both trader IDs and a mapping of the trader IDs to the firms, we aggregate all trader IDs that belong to the same firm. We operate at the firm level because there are no rules (to the best of our knowledge) that guide how firms use trader IDs. One possible concern is that firms assign multiple trader IDs to the order flow of an algorithm and send orders via DMA arrangements with multiple dealers to make it more difficult for outsiders to ascertain their activity. Routing via multiple dealers could also be driven by the desire to limit dependence on one dealer. We choose to work at the HFT-firm level to make our analysis robust to whatever gaming could be going on in terms of the firm's discretionary assignment of trader IDs to its orders. If an HFT firm uses multiple algorithms and actually designates a separate trader ID to each algorithm, our procedure will lump them together, although the PCA we implement in Section 3 could potentially differentiate these separate strategies. It is important to note that for dealers who may have brokerage arms in addition to proprietary trading operations, we exclude orders and trades made in the capacity of an agent, and include in our measurement of their HFT activity only those orders and trades identified in the order type field as proprietary activity.

We use an out-of-sample procedure to identify HFT firms to ensure that the identification is exogenous to our empirical analysis. We use data from the 22 trading days in September 2010 to identify the HFT firms, while we carry out the empirical work on the 30 down, flat, and up days as described in

Section 2.1.[8] To reduce the number of trading firms that we scrutinize more closely, we rank firms on several criteria and look at those that rank highly on at least one criterion (e.g., the number of times per day that the firm's inventory position crosses zero).[9] Our use of multiple criteria is motivated by our desire not to limit our sample to firms that pursue a certain strategy that requires a particular profile (e.g., a high order-to-trade ratio), but rather to allow for HFT firms that implement a variety of algorithms. We emphasize that some firms subsequently identified as engaged in HFT rank highly on only one criterion.

To help us differentiate the HFT firms from among the trading firms on our short list, we use two qualitative criteria. First, trading firms that participate in the Toronto Stock Exchange's Electronic Liquidity Providers Program are categorized as HFT firms. This is a program that offers fee incentives to firms that use proprietary capital and high-frequency electronic trading algorithms to provide liquidity on the exchange.[10] Second, we search newspapers and the web for information about the firms. Some firms have websites on which they state that they engage in proprietary trading or explicitly state that they pursue high-frequency strategies. Other firms are mentioned in newspaper articles as engaging in HFT. We use this information to ensure that we identify HFT firms even if they do not rank highly on some of the quantitative data criteria. Our procedure results in 31 firms identified as HFT firms. Some of these firms have a DMA arrangement with dealers while others are dealers engaged in proprietary trading. Our goal in this identification procedure is to ensure that we study as complete a set of HFT firms as possible.

Panel B of Table 1 provides summary statistics for the 31 HFT firms as well as for categories formed on market share of volume. Specifically, MS1 consists of four firms with a market share greater than 4%, MS2 consists of six firms with market share between 1% and 4%, and MS3 consists of the rest of the firms. Overall, these 31 HFT firms have 46.4% of the market share in terms of volume. The average number of daily trades of a firm in our sample is 19,445, but it ranges from 102,035 for firms in MS1 to 2,489 for firms in MS3. Similarly, the number of daily messages a firm sends to the market

[8] To ensure that the firms we identify operate during our sample period, we require HFT firms to be active (i.e., to trade) in at least 10 out of the 30 sample days.

[9] We compute several measures for each trading firm: (1) the order-to-trade ratio (defined as submissions and cancellations of limit and marketable orders divided by trades), (2) the number of times a firm's intraday inventory positions cross the end-of-day positions (or zero), (3) cross-trading-venue activity in the same time-stamp or a neighboring time-stamp, (4) median time-to-cancellation of non-marketable limit orders, and (5) the number of daily trades.

[10] ELPs are either independent proprietary trading firms or proprietary trading desks within large banks or financial firms, and the program requires that they passively trade at least 65% of their volume.

(submissions and cancellations of non-marketable limit orders, as well as the number of marketable orders) is 1,063,974, ranging from 5,496,423 for firms in MS1 to 239,157 on average for firms in MS3. The mean (median) number of times the intraday inventory position of a firm in our sample crosses its end-of-day inventory increases with market share: 4.3 (1.8) for MS3, 13.4 (4.3) for MS2, and 73.0 (70.1) for MS1.

2.4 Measures of HFT Activity

The main measure of HFT firm activity that we employ emphasizes actions initiated by the HFT firm. Our measure, MSG, is defined as the sum of three components: the number of submissions of non-marketable limit orders, the number of cancellations of non-marketable limit orders, and the number of marketable order executions. Hence, MSG for an interval (say one second) describes the total number of messages that the HFT firm sends to the market during the interval to initiate a change in its position (either in terms of presence in the limit order book or to transact immediately).[11] We investigate two additional measures to ensure the robustness of our conclusions. The first measure, TRD, is the number of trades made by an HFT firm in an interval. These trades could occur as the result of submitting marketable orders or the execution of previously submitted limit orders that rested in the book. The second measure, LMT, comprises all submissions and cancellations of non-marketable limit orders, and hence describes the actions the firm takes in the interval solely to change its presence in the limit order book.

We conducted all the tests using the three measures. In Section 3, to economize on the presentation of the results of the PCA, we only report the results for MSG. In Sections 4 and 5, where the regression and partial correlations analyses enable an efficient presentation of the results for the three measures, we report the results side-by-side. Our preference for MSG as the main measure is partially based on the fact that the vast majority of HFT activity is done in terms of submissions and cancellations of orders, not actual trading, and hence a measure that includes these submissions and cancellations

[11] An AMEND order type is considered as two messages, a cancellation and a resubmission, for the purpose of our measures. Our measures include both non-displayed and displayed orders. A refresh of an iceberg order (when the displayed part is executed and shares from the non-displayed part become displayed) does not lead to a change in price or quantity, hence only the initial order is counted. This standardizes the treatment for the various order types according to their economic functions, and let us summarize all activity in terms of submissions and cancellations of non-marketable orders and executions of marketable orders.

would offer a more complete portrayal of their activity. This idea is also highlighted in Brogaard, Hendershott, and Riordan (2015) and Subrahmanyam and Zheng (2016) who stress the importance of HFT limit order behavior.

Before proceeding to the main analysis, however, we use simple correlations to establish two stylized facts that are directly relevant to the economics of HFT strategies and hence impact the design of our tests in the rest of the paper. First, we look at whether the predominant correlation in HFT strategies involves directional activity or total activity. Second, we examine whether correlations in HFT strategies differ on days in which the market experiences large positive or negative returns.

The first stylized fact, concerning directional activity, is motivated by empirical studies on herding in financial markets (e.g., Wermers 1999; Khandani and Lo 2007, 2011; Choi and Sias 2009; Pedersen 2009; Brown, Wei, and Wermers 2014) that recognize the destabilizing influence that simultaneous actions in one direction (buying or selling) by institutional investors can have on asset prices. In Figure 1, we compare the magnitudes of correlations in total HFT activity (buy plus sell orders) and directional HFT activity (buy minus sell orders) for the three activity measures (MSG, TRD, and LMT).[12] Our cross-sectional (or market-wide) correlation measure provides information on whether the strategies of HFT firms are correlated across stocks at a given time. For each one-second time interval, we compute the correlation coefficient between the activities of pairs of HFT firms across the stocks in the sample, and average the correlations for all pairs of firms in a certain group (e.g., our market share subgroups MS1, MS2, and MS3). The time-series (or stock-specific) correlation provides information on whether the strategies of HFT firms are correlated over time for a given stock. For each stock, we compute the correlation coefficient between the activities of pairs of HFT firms over all time intervals, and average across all pairs of firms in a certain group.

We observe a striking result: the correlations involving total activity are four to twelve times the magnitude of those involving directional activity. For example, Panel A of Figure 1 shows that the one-

[12] Since limit order submissions and cancellations have opposing economic implications, we construct the directional activity measure such that we count the cancellation of a limit buy (sell) order by adding to the number of limit sell (buy) orders. For robustness, we also compute a directional activity measure such that submissions and cancellations of buy (sell) orders are both counted as buy (sell) orders, and the resulting correlations are very close. In particular, the result we present in Figure 1 that the correlations of total activity are much larger in magnitude than the correlations of directional activity holds for both measures of directional activity.

second cross-sectional (or market-wide) correlation of the total activity measure (MSG) of the HFT firms with the highest market share is 0.359 versus 0.073 for the directional activity measure (NetMSG). For the total trading measure (TRD), the correlation of all HFT firms is 0.313 versus 0.025 for net trading (buys minus sells). A similar pattern is observed for the time-series (or stock-specific) correlations in Panel B. The picture that emerges is consistent with what one would expect for HFT firms that operate simultaneously on both sides of the market, rather than pursuing very strong directional trading for long periods of time. In other words, it appear as if much of the HFT firms' activity involves either placing buy and sell orders simultaneously or buying and selling very rapidly within the same one-second interval. Many HFT firms design strategies to interact with uninformed order flow. Market microstructure theory often specifies uninformed order flow is non-directional (in contrast to informed order flow), which could potentially explain our finding of low directional correlations of HFT strategies. Given these very low directional correlations, we focus in the rest of the paper on analyzing correlations in total activity.

The second stylized fact, on correlated HFT activity in different market conditions, is motivated by the concern that HFT firms react to adverse market conditions (in terms of declining prices) by changing their strategies and hence cause greater fragility. Our study is designed to enable us to analyze this issue in greater detail. Specifically, we carry out the empirical work on the ten days with the largest negative index returns from June 2010 through March 2011 (with average daily returns of −1.72%), the ten best days (with average daily returns of 1.66%), and the ten days during which the index moved very little (with average daily returns of −0.08%).

Figure 2 presents the correlations of the three HFT activity measures (MSG, TRD, and LMT) for these down, flat, and up days, alongside the correlations for the entire 30-day period. We observe that the market-wide correlations for the MSG measure of the largest HFT firms (MS1) are almost identical on down, flat, and up days (0.354, 0.361, and 0.362, respectively). Similarly, the stock-specific correlations of the trading measure (TRD) for all HFT firms exhibit similar patterns on down, flat, and up days (0.039, 0.043, and 0.041, respectively). We do not observe that these three distinct market environments result in correspondingly distinct correlations.[13] This stylized fact, like the previous one concerning lack of

[13] Hasbrouck and Saar (2013) analyze the impact of a low-latency activity measure, which they view as a proxy for the activity of HFT firms, in two periods: one in which the NASDAQ Composite Index went up 4.34% and another

correlation in directional trading, is reassuring in terms of market fragility, although it does not preclude the possibility that the correlations would increase during times of extreme stress, such as the Flash Crash in the U.S. The absence of such extreme episodes in Canadian markets, however, prevents us from examining this possibility empirically. We verified that the results of all our tests do not differ materially on down, flat, or up days. We also classified all 10-minute intervals into three categories based on volatility and three categories based on volume, and looked at whether correlations in HFT strategies differ when we focus on intraday periods (10-minute intervals) distinguished by higher volume or volatility. We could not discern any clear patterns across the categories. Given these findings, results in the rest of the paper are for the entire sample period rather than breaking the results down by market conditions.

3. HFT Competition and Product Differentiation

3.1 Correlated Activity and Competition

The HFT industry is shrouded in secrecy. Most HFT firms are private and hence reveal no financial or operating information, and details about the profitability of proprietary trading desks of larger, publicly-listed firms are not disclosed to the public. Firms use restrictive clauses in employment contracts as well as litigation to deter or prevent exiting employees from taking computer codes used for trading algorithms. In general, HFT firms do not reveal information about the operation or the objectives of their algorithms beyond speaking generally about concepts such as “liquidity provision” and “arbitrage.”

How can one go about investigating the extent of competition in such an industry? Examining the economic profits of firms would be ideal, but it is practically impossible to examine the profitability of most HFT firms or trading desks as the costs associated with hiring and retaining the individuals who develop the algorithms as well as other operating costs are simply unavailable. Researchers can only ascertain firms’ net trading revenues (or what is left when shares are bought and sold by the firm) from datasets like ours that describe trading activity. While in Section 3.4 we look at the net trading revenues of the HFT firms, revenues without costs cannot be used to make a correct inference about economic

in which it declined 7.99%. Like us, they find that the impact of their proxy on market quality was similar in both periods.

profits or the competitiveness of an industry.

We therefore gather evidence on the extent of competition in the industry by looking at the number of competitors. From a theoretical perspective, the importance of the number of competitors goes beyond the traditional (static) equilibrium concepts. Since firms must share collusive profits, a higher number of competitors results in each of them gaining less. As a result, the gain from deviating increases and the long-term benefit of maintaining collusion is reduced. Therefore, even in dynamic collusion models coordination is more difficult with a larger number of firms (e.g., Ivaldi et al. 2003). Because the number of direct competitors in an industry is crucial to the absence of collusion, defining industry boundaries (or product categories) is important. Hoberg and Phillips (2016), for example, design a new classification scheme using text-based analysis of product descriptions to define industries.

We pursue a similar objective: classifying HFT firms that directly compete with each other by identifying firms that follow similar strategies. This enables us to determine how many firms compete in certain “products” (or strategies), and to establish whether the major players are monopolists that pursue markedly distinct strategies or whether multiple firms compete with each other. We identify competing firms by considering how manifestations of strategies (e.g., the messages HFT firms send to the market) correlate between HFT firms. The more highly correlated are the strategies of two firms, the more likely it is that they pursue the same profit opportunities (i.e., respond to the same trading signals) and follow a similar business model.

There are two conceptual issues we need to clarify. First, the correlated behavior of firms could also potentially characterize collusive behavior: Green and Porter (1984) describe a situation in which collusion results in recurrent episodes of patterns in product prices and firm profits. The assumptions in their model, however, do not fit the HFT industry. For example, the industry in their model is assumed to be stable over time. The HFT industry, on the other hand, keeps changing, with low barriers to entry enticing new start-ups to enter while others exit. Second, firms in their model are not able to engage in product differentiation, while developing a better algorithm is nearly essential in the HFT industry. Green and Porter assume in their model that information about the industry and its environment, like competitors’ cost functions, is public, which is the opposite of what is found in the HFT industry. The above considerations suggest that correlated behavior in the HFT industry is unlikely to be a manifestation of collusive behavior.

The second conceptual issue we want to clarify is that we use the terms “strategy” and “activity” somewhat interchangeably. While a strategy is typically a plan of action, we do not know the specifics of the algorithms employed by each HFT firm. We can measure only the outcomes of a strategy (e.g., submissions and cancellations of orders), although it is natural to recognize that the actions we observe are the manifestations of a strategy, and hence to treat measures of these actions as representing the strategy. In other words, while we measure the correlated activity of HFT firms, we are really interested in the similarity in their strategies. Are these necessarily the same? One can presumably construct examples in which the two could diverge, like when two different strategies produce the same action in one particular state of the world. For example, market making and aggressive news trading could both predict high activity when there is an outburst of news (Foucault, Hombert, and Rosu (2016)). However, these two strategies will have completely different predictions when there is no news.

While such examples focus on particular states of the world, the strength of our methodology is that we use over 36 million observations to estimate the similarities in strategies. We look at every one-second interval when the market is open, for 30 days (which represent three distinct market conditions), in 52 different stocks. Only if the strategies of various HFT firms produce similar actions across the many days, stocks, and states of the world in our observations, would our methodology identify an underlying common strategy and point to the firms that follow it. Therefore, while we examine correlated activity, our inference is about the similarity in the strategies of HFT firms.

Researchers that utilize datasets of aggregate HFT activity (e.g., Carrion 2013; Brogaard, Hendershott, and Riordan 2014; Jarnecic and Snape 2014) essentially assume that all HFT firms are in the same business. We know from Hagstromer and Norden (2013) that an HFT firm may pursue a strategy that differs from that of another firm along a particular dimension: the percentage of passive trading. We go one-step further by using a data-driven methodology (PCA) to decompose the correlation matrix of HFT activity as a tool with which to define the subsets of HFT firms that compete with each other directly in a particular strategy. This analysis helps us understand HFT strategies in several ways. First, it tells us whether there are underlying common strategies that our HFT firms follow and that represent much of the variability in their activity. Second, it shows us how close the strategy of each firm is to these common strategies, and therefore helps us establish the extent of competition in pursuing each underlying common strategy. Third, we analyze the empirical representations of these underlying common strategies (i.e., the principal component scores) and how they relate to the market environment.

3.2 HFT Firm Loadings on the Principal Components

Principal component analysis is essentially a data reduction technique. In our application, we think about the HFT firms' activities as the "variables" that we seek to summarize. The input for the PCA is a matrix with 31 columns (one for each HFT firm) and the number of rows is equal to the number of stocks times the number of intervals in the sample period. In other words, the stocks are stacked one on top of another and we analyze both the time-series and cross-sectional sources of variation in HFT activity together.

We are using multiple measures throughout the paper to describe the activity of HFT firms, though in this section we concentrate on the MSG measure, which comprises all messages an HFT firm sends to the market in an interval. While the choice of which measure to use for the PCA could matter, we find that the HFT firms we identify as following underlying common strategies using one measure are also identified as such using other measures. For example, there are ten firms with large loadings on the principal components when we use the TRD measure, all of which we also identify using the MSG message. Three of the other firms that MSG identifies as following underlying common strategies have loadings just below the cutoff when looking at the TRD measure, thus from an economic perspective (given that any such cutoff is arbitrary), we get a similar inference regarding the firms that follow underlying common strategies. The identification of firms that follow the three underlying common strategies is identical whether we use MSG or LMT. When we implement the PCA just on the buys component of MSG (or the sells component), we get as an output the same exact HFT firms following the three underlying common strategies.[14] All these measures that one can use as an input to the PCA are simply different descriptors of the activity generated by the HFT firms' strategies, and we find that they lead to a similar inference regarding which HFT firms follow common strategies. Therefore, to

[14] We also examine a couple of additional measures for robustness. For example, the same exact HFT firms that we find using the MSG measure also follow the principal components when we analyze just the removal of HFT firms' liquidity from the book, which may be of interest to market regulators concerned about market stability. We also looked at depth in the book (up to 10 price levels from the best prices), and found a similar inference except that most of the firms with large loadings on the first principal component when using the MSG measure loaded on the second principal component in the depth analysis, and vice versa.

economize on the exposition in terms of tables and discussion, we present in this section the analysis using the MSG measure of activity.

The PCA decomposes the correlation matrix (rather than the covariance matrix) of the strategies. In other words, the variables that describe the activity (or strategy) of each HFT firm are standardized to have zero mean and unit variance. Standardizing the variables eliminates the possibility that one of them would dominate the procedure because it has a much higher scale or range. From an economic perspective, this means that our procedure gives the very active HFT firms the same weight as any other HFT firm (each contributing one unit of variance to the total variance). This helps us focus on the resemblance of strategies to one another even if some firms are larger than others.

When conducting the PCA, we must choose how many principal components to retain for subsequent analysis.[15] The corresponding economic question is how many separate underlying strategies are common to a significant number of the HFT firms in our sample. As usual with such data-driven methodologies, that determination is made based on patterns we identify in the data. Specifically, we conduct a Scree test by plotting the eigenvalues of the principal components and looking for a natural break. Each of the 31 firms contributes 100/31=3.2% of the variance. However, meaningful principal components would naturally explain more of the variance. We find that the first principal component explains 11.66% of the variance, the second 4.5%, and the third 4.02%, together accounting for over 20% of total variation. Additional principal components account for less than 4% of the variance each, and there seems to be a natural break after three components. Hence, we extract three principal components for subsequent analysis.

The second choice when implementing the methodology involves determining the rotation of the principal components. The rotation is meant to help us interpret the loadings, which are the coefficients of each HFT firm on each of the principal components. We utilize the commonly-used varimax orthogonal rotation. The loading of an HFT firm on a particular principal component using this rotation has a simple interpretation: it is equivalent to the bivariate correlation between the HFT firm's activity and the underlying common strategy represented by the principal component. Our conclusions are robust to using

[15] In a PCA, the first component accounts for the largest portion of total variance and successive principal components account for remaining portions of the variance that were not accounted for by previous principal components.

other rotations.[16]

Table 2 presents the loadings from the PCA of the MSG measure using one-second intervals. Each principal component can be viewed as representing a certain underlying common strategy that multiple HFT firms follow and which results in the correlated behavior of the firms, and therefore we use the terms "underlying common strategy" and "principal component" interchangeably. The larger (i.e., closer to 1) the loading of a particular HFT firm on a principal component, the greater is the similarity of the firm's activity to that underlying common strategy. For each principal component, the loadings are sorted from the most positive to the most negative, and the 31 firms are represented by F01 through F31. The largest loading on the first principal component, for example, belongs to firm F14 and signifies 0.76 correlation between the strategy of this firm and the underlying common strategy represented by the principal component.

While the choice of a cutoff for the magnitude above which a loading is considered economically significant is somewhat arbitrary, it helps to have some cutoff in mind when looking at the results. We consider loadings significant from an economic standpoint if they are greater than or equal to 0.35, and mark them with an asterisk in the table. Eight HFT firms have significant loadings on the first principal component, suggesting that it represents a strategy that is common to these eight firms. Similarly, six firms exhibit significant loadings on the second principal component, ranging from 0.71 to 0.35, while four HFT firms exhibit significant loadings on the third principal component. Four firms appear to follow more than one strategy: F31 and F28 (significant loadings on the first and second principal components) and F17 and F20 (significant loadings on the first and third principal components).

One advantage to using PCA is that the procedure is well understood by researchers, and has straightforward output: the principal components are linear combinations of the 31 HFT firms' activities such that the components explain the maximum variation in that activity. A disadvantage of the procedure is that most of the loadings are non-zero, thus the cutoff question needs to be addressed. There have been recent attempts to propose methodologies that tackle this issue. A new approach called sparse principal

[16] We also conducted the analysis using an oblique rotation (promax) that allows the principal components to be correlated. The results are basically identical to those with the orthogonal rotation: the loadings are similar in magnitude and the same HFT firms load on the same principal components. Furthermore, the regressions on component scores that we present in Section 3.3 yield the same results irrespective of whether an orthogonal or an oblique rotation is used.

component analysis (SPCA) combines a PCA with the Lasso variable selection technique and balances the goals of explaining the variance and achieving a pattern of sparse loadings. The Lasso imposes a constraint on the sum of the absolute values of the coefficients, causing some coefficients to shrink toward zero.

For robustness, we implement the SPCA methodology proposed in Zou, Hastie, and Tibshirani (2006). To facilitate comparison with the results of the standard PCA, we set the parameters of the SPCA in such a way that 8, 6, and 4 HFT firms have non-zero loadings on the first, second, and third principal components, respectively. We find that the set of HFT firms that follow the underlying common strategies identified by the SPCA is identical to the one from the standard PCA methodology. The only difference is that the SPCA is designed to select non-overlapping variables, while the standard PCA methodology allows a variable to have significant loadings on multiple principal components; thus, our PCA methodology identifies four HFT firms as following more than one underlying common strategy. Given the familiarity with the standard procedure and the nearly identical results, we proceed in the Subsection 3.3 to analyze the component scores from the standard PCA methodology.

To summarize so far, the first important takeaway from the principal component analysis is that there are 14 HFT firms competing in three underlying common strategies, with each underlying common strategy followed by multiple firms. On the other hand, 17 HFT firms do not appear to pursue any of these common strategies but rather pursue more unique strategies. How important are the common strategies relative to the unique strategies? The 14 firms that compete in one of the three common strategies represent most of the HFT activity: 96.21% of the messages that HFT firms send to the market and 78.97% of the volume they trade. Therefore, we find that competing (as opposed to monopolist) HFT firms generate most of the HFT activity.

3.3 Regressions on Principal Component Scores

The challenge in interpreting the results of a PCA lies in understanding the economic nature of the underlying common strategies represented by these principal components. To gain additional insights, we use another output of the methodology: the component scores. Each principal component is essentially a linear combination of the observed variables (the 31 HFT firms' MSG measures), and the component scores are computed from these variables using the estimated loadings in Table 2 as weights. Since there is a separate score for each principle component in each stock and time interval, we can use regressions to examine the relations between these component scores (as dependent variables) and the market environment.

The first three explanatory variables in our regression specification provide information about the HFT firms. We are interested in characterizing the extent to which HFT firms operate simultaneously across trading venues. The first variable, *HFTcrossmsg*, counts messages sent only by HFT firms that submit messages to multiple trading venues in the same interval. Similarly, we are interested in the degree to which HFT firms trade passively by supplying liquidity. The second variable, *HFTliqsupply*, count the executed limit orders of all HFT firms in each interval. The third variable, |*HFTinventory*|, is the absolute value of the aggregate inventory position of all HFT firms in Canadian dollars (cumulative from the beginning of the day and assuming that all of them start the day with zero inventory).

The next two explanatory variables represent the degree of integration of the Canadian market: *PriceAlign* is the percentage of time that the three trading venues with the highest market share post the same bid or ask price, while *SpreadAlign* is the percentage of time that these three trading venues have the same bid-ask spread. The next three explanatory variables represent the state of liquidity in the market (aggregated across all trading venues): total depth up to 10 cents from the market-wide best bid or offer (MWBBO), the magnitude of depth imbalance at the MWBBO (defined as the absolute value of the difference between the number of shares at the bid and at the ask), and percentage MWBBO spread.[17] The last three explanatory variables represent market conditions in the interval: return (computed from the last transaction price in each interval), volatility (computed as the absolute value of return), and the average time between trades in the interval. All explanatory variables are standardized to have zero mean and unit standard deviation to make the coefficients comparable across variables.

Table 3 shows the regression coefficients together with *t*-statistics computed from double-

[17] All measures are time-weighted. Since messages are time-stamped to 10 milliseconds, there are orders that are submitted and cancelled (or executed) within the same time stamp. When we consider orders that stay in the book to provide liquidity (e.g., for our measures of depth), we assume that a submitted order stays in the book for 10 milliseconds. The exceptions are the following special order types: immediate or cancel orders (IOC), fill or kill orders (FOK), all or nothing orders (AON), dealer's AG orders (that are generated to fulfill their market making obligations and execute against an incoming order), odd lot orders (OL), and marketable orders, which are assumed to be executed or canceled upon arrival to the market and hence are not added to the limit order book.

clustered standard errors (along both the stocks and intervals dimensions) to focus our attention on the most significant relationships.[18] Looking at the regression on the first principal component scores, we observe that the coefficient on HFT cross-venue activity (*HFTcrossmsg*) is positive and highly statistically significant. HFT firms that load on this principal component are also more active when prices are more volatile (a positive and significant coefficient on |Ret|), and there is less depth in the book (*Depth10*) and greater imbalance at the top of the book (|*TopDepthImb*|). These are times at which some trading venues post better spreads than others (a negative and significant coefficient on *SpreadAlign*) and there appears to be a need to move liquidity across trading venues, which the HFT firms indeed do with their more intense cross-venue message activity. This suggests that the first principal component could represent a cross-venue arbitrage strategy.

We want to be clear that, as always is the case with principal component analysis, the principal components do not possess an inherent economic interpretation. In every application of this methodology, however, researchers are tasked with giving them one based on the output. We evaluate the regression coefficients of the principal component scores on attributes of HFT activity and market conditions, and our interpretation simply reflects our reading of these coefficients. Whether these labels (e.g., cross-venue arbitrage) are accurate or not does not detract from our use of principal components as data-driven representations of underlying common strategies in a way that can help us understand the competitive landscape of HFT firms. Our use of the principal components in the rest of the paper does not depend on these labels.

Turning to the second principal component, we find that the coefficient on trading passively by providing liquidity (*HFTliqsupply*) is large and highly statistically significant. This underlying common strategy is more active when prices and spreads are more aligned across the trading venues (positive coefficients on *PriceAlign* and *SpreadAlign*), and may represent times when market-making activity is most profitable (i.e., when market-making firms earn the spread rather than lose to changing prices).

[18] We are primarily interested in contemporaneous relations, which is why Table 3 presents the results of regressions in which we line up the component scores with observations of the market environment over the same interval. We also run regressions in which the component scores are regressed on lagged and lead values of the market variables. We find that the signs of the coefficients on almost all variables are the same in the contemporaneous, lagged, and lead regressions: 31 out of 33 possible combinations (11 variables times 3 principal components) had identical signs. These results are available from the authors.

Similarly, this strategy is more active when there is greater depth in the book and a smaller imbalance in the top prices, creating ample opportunities for market making without excessive risk.[19] This combination of coefficients, and especially the large coefficient on passive trading that is traditionally associated with market makers, could indicate that this underlying common strategy is related to market making.

The underlying common strategy represented by the third principal component has a combination of statistically significant coefficients that suggests a distinct strategy, although it shares some attributes with the first principal component. HFT firms that load on the third principal component are also more active when prices are more volatile, when depth in the book is lower, and there is greater imbalance in the top prices. Two of the economic relationships that make the third principal component different are a significant negative coefficient on *PriceAlign*, which means that the best prices are not the same across the trading venues, and a negative and significant coefficient on *HFTliqsupply*, which indicates more active (rather than passive) HFT trading. This pattern of relations with the market environment could represent a short-horizon directional strategy (e.g., momentum) that requires quick trading using marketable orders and is more profitable when there is less depth throughout the book and prices are fragmented across the trading venues.[20]

There could certainly be other labels for the underlying common strategies that would fit such patterns in regression coefficients. The important takeaway from the regressions, however, is not the labels. Rather, it is the recognition that there is heterogeneity in the underlying common strategies, as evidenced by the distinct relations we document between each principal component and the variables that represent various aspects of the market environment. Furthermore, there appears to be a significant

[19] Note that the coefficients on spread and volatility are not statistically different from zero. When spreads are larger, market makers' profit potential may be higher, and as a result they increase their activity. By supplying more liquidity, however, they work to narrow the spread. The contemporaneous relation we estimate could incorporate both of these contrasting effects, resulting in a coefficient that is indistinguishable from zero. As for volatility, it is well known that market makers profit from transitory volatility (e.g., they buy when prices are pushed down too much and sell when they reverse). At the same time, market makers lose to informed traders (or simply faster traders) when the source of volatility is informational. Our volatility measure combines both transitory and permanent volatility, which could be why we do not see a strong effect as these two would have opposite implications for the intensity of market makers' activity.

[20] Out of the 52 stocks in our sample, 37 are cross-listed in the U.S. We ran the regressions separately for these 37 stocks and added two variables that describe liquidity in the U.S. market: percentage NBBO spread and volume. We find positive coefficients on these two variables in the regressions for the first and third principal components and negative coefficients in the regressions for the second principal component. The inclusion of these variables, however, does not make any difference for the other 11 variables in the regressions: we basically find the same signs on the other variables; hence, our conclusions as to the nature of the strategies are unchanged.

number of HFT firms that pursue unique strategies, although they represent a small portion of HFT activity. The heterogeneity in HFT strategies is important insofar as the insights generated by thinking about HFT as a single "entity" could be rather limited because aggregating HFT activity hides heterogeneity that is important to understanding how HFT firms interact with markets and ultimately affect them.

3.4 Net Trading Revenues

Our PCA identifies three underlying common strategies that are pursued by multiple HFT firms: eight, six, and four HFT firms load on the first, second, and third principal components, respectively. Evidence that each strategy is followed by several HFT firms could suggest that competition drives down possible rents earned by these firms for their activities in the market.[21] While there are also HFT firms with unique strategies in which we do not observe competition, the aggregate economic rents earned by these firms is likely more limited due to their small market share.

As noted, a direct analysis of the economic profits of HFT firms or trading desks cannot be conducted as their business costs are unavailable. HFT is carried out mostly by private firms that do not divulge any information about their activity or by trading desks of larger financial services firms that do not report separate performance and costs figures.[22] We can, however, look at the firms' net trading revenues (or what is left when shares are bought and sold by the firm). We operationalize the concept of net trading revenues as follows. We sum the positive cash inflows (how much the HFT firm gets from

[21] The impact of competition between HFT firms on rents could depend on the specifics of the particular strategy. If HFT firms are viewed as informed traders, the models of Holden and Subrahmanyam (1992) and Foster and Viswanathan (1993) show that it is enough to have two competing informed traders to almost instantaneously eliminate their informational advantage and have their profits vanish in a continuous market. Li (2013) takes another approach. The informed traders in her model are endowed with identical flows of long-lived private information (as opposed to one-shot long-lived private information in the aforementioned papers). As a result, the aggregate profits of the informed traders decrease in the number of informed traders (they are inversely proportional to the square root of the number of informed traders), but do not vanish.

[22] There are three exceptions. One exception is the HFT firm Virtu Financial, which went public in April 2015. However, Virtu trades a broad range of asset classes including equities, foreign exchange, commodities, options, and fixed income in over 200 market centers in 35 countries. Virtu's financial statements provide little help in trying to assess the relevant cost structure. For example, they do not break down results according to countries and do not provide the cost structure for equities versus other asset classes. KCG, another publicly-traded firm with a market-making HFT arm but with other businesses such as brokerage and trading venues on three continents, does not provide information in its financial statements that would make estimating the costs for our Canadian market feasible. Flow Traders, an HFT firm listed on Euronext Amsterdam, breaks down some financials by continent only. Virtu, KCG, and Flow Traders were not publicly-listed firms during our sample period, although one of the firms that merged to become KCG (Knight) was publicly traded during that time period.

selling shares) and negative cash outflows (how much it pays for buying shares), and add or subtract the trading venue fee or liquidity rebate associated with each trade (depending on whether the HFT firm provides or demands liquidity).[23] We compute the net trading revenue for each HFT firm in each stock and on each day, assuming that shares left (short positions) at the end of the day are "liquidated" (closed) using either the end-of-day midquote or the closing price, and start every day with zero inventory (Brogaard, Hendershott, and Riordan 2014).

Net trading revenues are likely impacted by the trading environment, and hence could differ depending on whether we look at the 10 down days, 10 flat days, or 10 up days in our sample period. Panel A of Table 4 provides evidence consistent with this conjecture. For each 10-day period, we consider all observations of daily net trading revenues for each HFT firm in each stock (~9,000 observations, as not all HFT firms trade all stocks every day). The column NetRevenues shows the median of these observations and hence gives a representative number for the net trading revenues of a typical HFT firm in a typical stock/day. We find that net trading revenues for a down day are very modest: between 35 and 39 CAD (depending on whether we value inventory with the closing price or the midquote). The number for an up day is similar (around 34 CAD), while HFT firms make less money per stock on a flat day when market prices do not vary as much (around 11 CAD). Why does the median value of net trading revenues seem so modest? It is clear that in an industry with low barriers to entry, some of the HFT firms we study could be new entrants that are struggling to be profitable and will not survive. Others might have had successful algorithms at some point in the past, but became less profitable or even loss making when other firms with better algorithms entered the market. We indeed observe that while most HFT firms have positive net trading revenues for the 30-day sample period, there are firms with negative numbers.

Given that our PCA methodology identifies three underlying common strategies, we next examine whether the firms that follow these strategies have higher or lower net trading revenues than firms that have more idiosyncratic algorithms. Panel B of Table 4 presents the median HFT/day/stock net trading revenues separately for the firms that follow each of the three principal components as well as the

[23] We obtain information on the structure of fees and liquidity rebates for each of the trading venues over our sample periods. Liquidity rebates are payments made by trading venues to HFT firms that take the passive side of trades (i.e., whose limit orders are executed by incoming marketable orders). See Malinova and Park (2014) for a discussion of liquidity rebates in the Canadian market. Note that while our net trading revenues computations incorporate the fees and rebates associated with the trading venues, they do not include brokerage or clearance fees.

17 firms that do not follow any of them. We find that firms that follow the first and third principal component have net trading revenues of around 70 and 105 CAD, respectively, both of which are much larger than the approximately 10 CAD in net trading revenues for the firms that do not follow any of the underlying common strategies. Firms that follow the second principal component have lower net trading revenues (16.6 or 22.5 CAD, depending on our treatment of end-of-day inventory), with one of the two estimates not statistically different from the revenues of the 17 firms with unique strategies.

Several comments are in order concerning these results. First, the higher net trading revenues for firms that follow the underling common strategies we identify could explain why we observe multiple firms pursuing each strategy: these are strategies with a larger revenue potential, and they elicit more competition from firms. Second, there is a wide dispersion in revenues and some HFT firms make much more money than others. This skewness implies that the mean per HFT/stock/day is much larger than the median, for example, for the firms that follow the first principal component: 603.68 CAD (mean) vs. 68.54 CAD (median). A similar picture whereby most revenues are concentrated in the hand of a smaller number of HFT firms is also reported in Baron, Brogaard, and Kirilenko (2012) for the E-mini S&P 500 futures market. Third, the strategy that we (loosely) associate with market making has the lowest median revenues among the underlying common strategies, which is also consistent with evidence in Baron et al. (2012), who find that firms that primarily provide liquidity have lower trading revenues. Fourth, these results represent an equal mixture of down, flat, and up days in our sample period, but the distribution of such trading environments may not be the same over any calendar period.

Lastly, the *p*-values in Table 4 are from a Wilcoxon test for which we treat each HFT/day/stock number as an independent observation, which may be somewhat questionable. For example, it could be that the performance of a particular HFT firm is correlated across the stocks it trades on any particular day. In additional tests, we aggregate the revenues of each HFT firm on a given day from all the stocks it trades, leaving us with 240 observations (8 HFT firms, 30 days), 180 observations, and 120 observations for the first, second, and third principal components, respectively, as well as 481 observations for the 17 firms that do not follow an underlying common strategy (some firms do not trade in all 30 days). The tests yield similar results: net trading revenues for the firms that follow the first and third principal components are larger and statistically different from those of the 17 firms that do not follow any principal component, while the net trading revenues for the firms that follow the second principal component are not statistically different.

What can we say from these numbers about the economic profits of the HFT firms that operate in Canada? Not much. Baron et al. (2016) look at the annual reports of three large HFT firms (KCG, Virtu Financial, and Flow Traders) and note that variation in trading revenues should translate reasonably well into variation in profits because these firms have a rather small fixed cost component. Even if variation in trading revenues could be used as a proxy for profit variation, we are interested in a different question: whether the *level* of economic profits is close to zero. This question is difficult to answer just knowing the trading revenues for a heterogeneous set of trading firms, some large and some small, that follow distinct trading strategies. Still, our PCA identifies multiple firms that compete in each of the underlying common strategies. While we find that these strategies have a larger revenue potential (presumably because they provide important services to the market like arbitraging price differences across markets or providing liquidity), our ability to identify multiple competitors that follow each strategy is encouraging as competition could lower the economic rents these HFT firms earn.

4. Correlated HFT Activity and Volatility

We next examine whether HFT competition could exacerbate volatility. If greater HFT competition means that multiple HFT firms engage in strategies that are highly correlated across stocks, their dominance in the trading environment could potentially amplify idiosyncratic occurrences at the microstructure level and spread them across the market. We stress that our focus is not on the question whether the magnitude of HFT activity impacts volatility but rather we study how the short-horizon volatility of stocks loads on $MktCor_{mt}$, which is our time-varying market-wide variable that captures HFT competition across stocks.

For each one-second time interval, we compute the correlation coefficient between the activities of pairs of HFT firms across the stocks in the sample, and average the correlations for all pairs of firms (as in Section 2.4) to obtain $MktCor_{mt}$. In Table 5, we present the mean $MktCor_{mt}$ for the three measures that describe HFT strategies: MSG, TRD, and LMT. We observe that for all measures, the correlations between the largest HFT firms (MS1) are higher than the correlations between other HFT firms. While

MSG is a comprehensive measure in that it incorporates every action a firm initiates over an interval, TRD and LMT represent other aspects of an HFT strategy. A priori, it is unclear which measure is more important in explaining volatility. On the one hand, TRD focuses on trade executions, and as such it could be more tightly related to price changes. On the other hand, most of the activity of HFT firms involves order flow that does not culminate in trades (limit order submission and cancellation). Including such orders in the representation of the strategy may result in a better description of HFT activity and therefore could be more appropriate when analyzing how HFT activity impacts volatility. We use the MSG, TRD, and LMT measures to examine whether the results are sensitive to the specific representation of the strategies.

We investigate whether HFT competition impacts the short-horizon volatility of individual stocks by running the following regression:

$$\left|r_{it}\right| = a_{0i} + a_{1i} MktCor_{mt} + a_{2i} HFT_{mt} + a_{3i} r_{mt} + a_{4i}\left|r_{mt}\right| + a_{5i} Volume_{mt} + error_{it}\,, \tag{1}$$

where $|r_{it}|$ is the absolute value of the interval return for stock i in interval t, which we use as our measure of interval volatility. The right-hand side of equation (1) contains only market-wide variables.[24] First, we are interested in studying the effect of correlation in HFT strategies, not the magnitude of HFT activity, and therefore we add as a control variable (HFT_{mt}) the sum across all HFT firms of the same measure of HFT activity that we use for $MktCor_{mt}$ (MSG, TRD, or LMT). Second, like in a typical market model, r_{mt} is the value-weighted return of all stocks in our sample.[25] Third, we control for market volatility (the absolute value of the market return) and market volume (aggregate volume in all stocks) in the interval.

We do not believe that the model in equation (1) suffers from endogeneity in the form of either reverse causality or omitted variables. With respect to the former, it is generally accepted that there is no reverse causality when one regresses a stock attribute on market attributes. When we run a market model, for example, we normally assume that the stock return does not "cause" the market return. Similarly, the volatility of a single stock in equation (1) does not cause market volatility, nor does it cause the market-wide correlation of HFT strategies that is computed using HFT activity across all stocks in the sample. With respect to the latter, the issue of omitted variables as a source of endogeneity in equation (1) is

[24] The returns are computed from trade prices (using the trade closest to the end of an interval).
[25] The results are similar when we use the equal-weighted return rather than the value-weighted return as a proxy for the market portfolio.

addressed by including control variables. Conventional wisdom suggests that the activity of HFT firms is driven by volume and volatility.[26] It is certainly possible that a market volatility shock could impact both stock volatility (if stocks respond to market conditions) and the correlation of HFT strategies. Once market volatility is a controlled for in the regression, however, $MktCor_{mt}$ no longer serves as a proxy for it.[27] Therefore, a solution to the omitted variables problem is to include important market drivers of stock volatility in the regression, which is why we include both market volatility and volume.

Panel A of Table 6 presents summary statistics for 52 regressions (one for each stock) in which the dependent variable is the stock's one-second interval volatility. We find that the mean loading on the HFT competition variable is negative, reflecting the negative loadings in 43 to 50 (depending on the measure of HFT activity) of the 52 individual stock regressions. It is instructive to contrast this predominantly negative loading with the predominantly positive loadings on both the volatility of the market and the magnitude of HFT activity. As expected, the volatility of individual stocks increases with the volatility of the market. Similarly, higher HFT activity in the market may represent times when there is more fundamental news and hence intensified price changes for individual stocks. The loadings on HFT competition are predominantly negative, however, suggesting that HFT competition does not serve as a proxy for market volatility, but rather represents a separate and distinct effect.[28] Our results, therefore, contrast with Jarrow and Protter's (2012) prediction that correlated HFT activity would increase volatility.

Panel B of Table 6 presents the regression results for all three interval lengths (one-second, 10-second, or 60-second). We show only the $MktCor_{mt}$ coefficients but add three statistical tests: a joint F-test from a seemingly unrelated regression estimation (SURE) of the 52-equation system, a nonparametric

[26] The decrease in HFT activity in 2011 and 2012 is attributed to a decline in volume and volatility in the market. See, for example "High-speed trading no longer hurtling forward" by Nathaniel Popper, *The New York Times*, October 14, 2012, and "High frequency trading loses its luster" by Ivy Schmerken, *Wall Street and Technology*, April 1, 2013.

[27] The right-hand-side variables in equation (1) are not highly correlated with each other (e.g., the correlation of market volatility and market volume is 0.108), thus there is no problem in having all of them in the same regression.

[28] Out of the 52 stocks, 37 are cross-listed in the U.S. A possible concern is that the model in equation (1) is misspecified because it excludes market return, volatility, and volume from the U.S. We therefore used the SPY exchange-traded fund that tracks the S&P 500 Index to construct variables at the one-second frequency for the U.S. market return, volatility, and volume (where we use the SPY volume as a proxy for overall U.S. market volume). The correlations between the U.S. and Canadian variables were modest, which facilitated adding the three U.S. variables to the regressions in equation (1). The addition of the U.S. market variables did not materially affect our results. In particular, most stocks had a negative and statistically significant loading on the HFT competition variable, similar to the results in Table 6.

Sign Test, and a Wilcoxon test.[29] The results confirm that the negative effect of competition between HFT strategies on the short-horizon volatility of individual stocks is observed irrespective of whether we use one-second, 10-second, or 60-second intervals. Furthermore, all statistical tests reject the hypothesis that the loadings on the HFT competition variable are equal to zero. The magnitude of the effect is also economically meaningful. An increase in MSG (TRD) correlation from 0 to 0.5 would result in a decrease of 31% (32%) in one-second volatility.[30]

Why would greater competition between HFT firms' strategies decrease rather than increase stock return volatility? Jovanovic and Menkveld (2016) posit that HFT firms trade on "hard information," such as price changes in same-industry stocks or the market index. Greater competition to turn these cross-stock "private" signals into public information implies lower adverse selection costs and hence a lower price impact of trades. As short-horizon (e.g., one-second) volatility is predominantly driven by the price impact of trades (as opposed to the public release of fundamental news, which is relatively rare for a stock), it follows that greater competition between HFT firms would decrease short-horizon volatility. It is also possible that the negative loadings are driven by the impact of correlated HFT strategies on the transitory, rather than the permanent, price impact of trades if they reflect greater competition in market-making activity. If returns and order flows of various stocks are correlated, efficient market making would necessitate algorithms that consider multiple stocks (e.g., Ho and Stoll 1983). Competition between such market-making algorithms would appear as higher cross-sectional correlations of HFT strategies, and such competition would decrease the transitory price impact of trades, lowering short-horizon volatility.

We next investigate the channel through which competition between HFT firms lowers the short-horizon volatility of stocks by examining competition in each of the underlying common strategies. We run a regression similar to equation (1) but replace the single $MktCor_{mt}$ variable, which is computed as the average of paired correlations of all HFT firms, with three separate market-wide competition variables

[29] Given the nature of the data (small N, large T) and our desire to estimate a separate vector of coefficients for each stock, which can be interpreted as loadings from a multiple-factor model, the natural approach is seemingly unrelated regression estimation (SURE). Note, however, that because we have identical regressors in all regressions (the common factors), SURE is essentially the same as regression-by-regression OLS (i.e., it does not provide efficiency gains), except we can use it to construct a joint coefficient test.

[30] While an increase in correlation from 0 to 0.5 is a rather large change, even a one standard deviation increase in correlation would decrease the one-second interval volatility between 6.3% (MSG) and 16.6% (TRD).

each computed as the average of paired correlations only for the firms that load significantly on one of the principal components (*PC1Cor*$_{mt}$, *PC2Cor*$_{mt}$, and *PC3Cor*$_{mt}$). In this regression, we include three control variables for the magnitude of the HFT activity, each aggregating the messages (across all stocks in the market) of the HFT firms that load on one of the principal components.[31] As in Table 6, we use MSG, TRD, and LMT to measure HFT activity, for robustness.[32]

Table 7 presents the results of the regressions. For both the MSG and LMT measures of HFT activity, the second principal component seems to drive a significant portion of the relation: its coefficient is the most negative, the *t*-statistic computed from the cross-sectional distribution of the coefficient points to a highly significant relation, and there are 39 stocks (out of 52) for which this variable has negative and statistically significant loadings in the individual stock regressions (compared with 10 positive and significant loadings). This principal component represents the underlying common strategy that we (loosely) associate with market making based on the component scores regressions in Section 3. If we are willing to accept this characterization of the second principal components, this result suggests that the reduction in volatility is driven by lower transitory volatility due to competition between market makers (Ho and Stoll 1983). The regressions with the TRD measure show statistically significant results for all three principal components, but the strongest relations appear to be with the second and third principal components (43 and 41 negative and statistically significant loadings in the individual regressions, respectively).[33] If we believe that the third principal component represents an underlying common strategy that is related to short-horizon directional speculation, this result could suggest that the volatility reduction is due at least in part to competition between HFT firms in activity that impounds hard information into prices, as in Jovanovic and Menkveld (2016).

To summarize, we find that the short-horizon volatility of most stocks loads negatively on a market-wide variable that describes correlated HFT activity across stocks, which is contrary to concerns that correlated HFT activity may increase market fragility. Therefore, competition between HFT firms

[31] The variables are not highly correlated with one another (the highest correlation is 0.3), hence there is no difficulty in having all three of them in the same regression specification.
[32] Identifying the HFT firms that follow the three underlying common strategies, though, is done based on the principal component analysis of MSG from Section 3.
[33] As in Table 6, joint F-tests from the SURE specification reject the hypothesis that the coefficients on HFT competition for these principal components are equal to zero.

could benefit the market in more the one way: not just the possible reduction in economic rents but also a reduction in short-horizon volatility. It is important to stress, though, that this result is observed during "normal" times, and may not hold during a market breakdown episode like the U.S. Flash crash in May of 2010.

5. Correlated HFT Activity and Competition between Trading Venues

Equity trading markets in the U.S., Canada, and many other countries are characterized by multiple trading venues on which stocks can be traded. Such trading fragmentation may create negative externalities in the form of lower price integrity and higher costs as liquidity is scattered across trading venues. Against these negative externalities, proponents of this structure argue that the competition it induces between trading venues results in lower fees and greater innovation in terms of trading technology and services.[34]

HFT is pertinent to the debate on market fragmentation: it alleviates the negative externalities of fragmentation and facilitates competition between trading venues. First, HFT firms can act as market consolidators, transforming the environment into a virtual central electronic limit order book by moving liquidity from one venue to another and ensuring that prices are the same across the venues. Second, a new venue must secure enough liquidity provision to attract trading, thus the willingness of HFT firms to provide liquidity on new trading venues can be essential in fostering competition between venues.[35] What is less well understood, however, is how competition between HFT firms, which is reflected in the pursuit of similar strategies by multiple firms, affects the competitiveness of trading venues and the concentration of trading in the market.

5.1 Correlated HFT Strategies and Market Concentration

To examine market concentration of trading volume in stock *i*, we compute the Herfindahl-Hirschman Index (HHI_i) of market share for the five trading venues. The lower the HHI_i, the less concentrated the

[34] O'Hara and Ye (2011) examine the consequences of market fragmentation in the U.S. Note that the focus of much of the recent literature has been on fragmentation caused by crossing networks (e.g., Buti, Rindi, and Werner 2016), while the type of fragmentation that exists during our sample period in Canada consists almost exclusively of trading venues that are structured as electronic limit order books.

[35] Menkveld (2013) notes that a new trading venue in Europe became viable only when a large HFT began trading on it.

market.[36] The average level of concentration for a stock in our sample is 0.879 when we use one-second intervals, and it declines to 0.778 and 0.629 in 10-second and 60-second intervals, respectively. The stock-specific variable that captures HFT competition, $StockCor_i$, describes how the strategies of HFT firms are correlated over time for each of our measures of HFT activity (MSG, TRD, and LMT). As in Section 2.4, we construct $StockCor_i$ by computing the correlation coefficient between the activities of pairs of HFT firms over all time intervals, and averaging across all pairs of firms. Patterns in $StockCor_i$ are similar to those observed in Table 5 for the market-wide correlations. In particular, the correlations are higher when we look at the activity of the largest HFT firms.

It is important to note that the relation between these two variables—market concentration and the correlation of HFT strategies—is best viewed as being jointly determined in equilibrium rather than as one causing the other. While competition between HFT firms could enhance the viability of smaller trading venues, the multiplicity of trading venues could also increase the profit opportunities from arbitrage and market making across trading venues and lead to intensified competition between HFT firms. Therefore, we examine the cross-sectional association between HHI_i and $StockCor_i$ without assuming causality, but still control for several variables that could influence the level of concentration in a stock. Specifically, we use three variables—market capitalization, price level, and the standard deviation of 30-minute returns over the sample period—to control for heterogeneity in fundamental attributes across stocks. We employ two additional variables, time-weighted average spread and bid-and-offer (BBO) depth, to control for the liquidity environment of the stock. Lastly, since our goal is to study the effect of correlation in HFT strategies rather than the magnitude of HFT activity per se, we add the magnitude of HFT activity in stock *i* as a control variable.[37]

Table 8 presents the partial correlations between HHI_i and $StockCor_i$ when we control for market capitalization, price level, volatility, spread, depth, and the magnitude of HFT activity. The results are very strong: the partial correlation is negative and highly statistically significant in eight of the nine

[36] The HHI is computed as the sum of squared market shares of the trading venues, and with five trading venues it is always between 0.2 (if volume is equally divided among the five trading venues) and 1 (if volume concentrates on one trading venue).
[37] In other words, if $StockCor_i$ is computed using the MSG measure, we add as a control variable the number of all messages by HFT firms in that stock during each interval. Similarly, when we compute our competition variable using the TRD (LMT) measures, we use the number of all trades (submissions and cancellations of limit orders) as controls.

specifications (three measures times three lengths of time intervals). This finding is consistent with the hypothesis that more highly correlated HFT strategies, which we view as a manifestation of competition between HFT firms, are associated with a less concentrated market structure. Competition between HFT firms and competition between trading venues are therefore tightly linked in a "competition begets competition" manner.

As in Section 4, we investigate the channel through which competition between HFT firms interacts with market concentration by examining competition in each of the underlying common strategies. If one is willing to accept the suggestive evidence that the first principal component is related to cross-venue arbitrage, then we would expect this underlying common strategy to be a driver of market concentration because it improves the quality of prices displayed on the smaller trading venues. Menkveld (2013) finds that HFT market making is important for the success of a new trading venue. Therefore, if one is willing to entertain the possibility that the second principal component is related to market making, then we would expect a strong negative relation with market concentration.

We therefore carry out a similar analysis with separate competition variables computed only for the firms that load significantly on one of the principal components ($StockCorPC1_i$, $StockCorPC2_i$, or $StockCorPC3_i$).[38] The results in Table 8 show that the partial correlations between HHI_i and $StockCorPC1_i$ are negative and statistically significant for all interval lengths and strategy measures. There is a similar strong negative relation between HHI_i and $StockCorPC2_i$. However, none of the partial correlations between HHI_i and $StockCorPC3_i$ are statistically different from zero, irrespective of the HFT activity measure or the interval length. This suggests that the negative relation between HFT competition and market concentration could be driven by competition in two common strategies that we label as cross-venue arbitrage (first principal component) and market making (second principal component). Note that, although the principal components cannot be interpreted with a high degree of confidence, our results are intuitively consistent with the labels we discuss in Section 3.

[38] To be consistent with the manner in which we construct the competition variable, the magnitude of HFT activity in stock *i* that we use as a control variable is computed only for the firms that load significantly on each principal component.

5.2 Correlated HFT Strategies and the Competitiveness of Trading Venues

In this subsection, we would like to further our understanding of the result that HFT competition is associated with a less concentrated market by examining whether competition between HFT firms enhances the viability (or competitiveness) of specific trading venues. For each of the five trading venues that are organized as electronic limit order books, we examine whether a higher correlation between HFT strategies on a specific venue increases the percentage of time it displays the best prices or narrowest spreads.

Panel A of Table 9 presents market share summary statistics for the five trading venues, as well as two measures of trading venue competitiveness (or viability): (i) the percentage of time that the trading venue posts either the best bid or the best ask in the market (where the market is defined as the aggregation of all five trading venues), and (ii) the percentage of time that the bid–ask spread on the trading venue is the narrowest spread in the market.[39] We find that there is a dominant trading venue in Canada with a market share in terms of volume of 69.26%. This trading venue displays the best price 92% of the time (averaged across all stocks in our sample) with the lowest standard deviation (6.1%) and the highest minimum (71.1%). The other trading venues also frequently display the best prices, although the variability in the cross-section is greater. Similarly, the largest trading venue has the narrowest spread 76.8% of the time (compared with 54.2%, 36.1%, 26.9%, and 9.4% for the other four trading venues).

We compute the HFT competition variable for each trading venue separately using only activity on that trading venue. In other words, $StockCor_{iv}$ indicates whether the strategies of HFT firms are correlated over time for stock i on a particular trading venue v.[40] For each trading venue v, we run the following cross-sectional regression:

$$\begin{aligned} C_{iv} = a_0 + a_1 StockCor_{iv} + a_2 HFT_{iv} + a_3 Spread_{iv} + a_4 BBOdepth_{iv} + \\ a_5 MktCap_i + a_6 Price_i + a_7 Volatility_i + error_{iv} \end{aligned}, \qquad (2)$$

where C_{iv} stands for one of the two competitiveness measures for stock i on trading venue v, and we use

[39] We note that two or more trading venues could potentially be at the best bid or ask at the same time (or have the same narrowest spread). We also recognize that activity on one trading venue could affect activity on another trading venue as in the multi-market inventory model of Lescourret and Moinas (2015).

[40] Not all HFT firms in our sample are very active on multiple trading venues. We therefore focus on firms undertaking substantial activity on all five trading venues, which we define as sending at least 10,000 messages during our sample period to each of the trading venues. Our results are not sensitive to this particular cutoff.

MSG (total messages sent by the HFT firm) as our measure of HFT strategies.[41] As in Section 5.1, including the magnitude of HFT activity on the specific trading venue as a control variable (HFT_{iv}) ensures that we detect the effect of correlation in HFT strategies, not the magnitude of their activity per se. Spreads and depth (now computed from the best prices on a single trading venue) control for the liquidity of the stock, while the last three variables control for heterogeneity in fundamental attributes across stocks.

To present results for the five trading venues and the three interval lengths side-by-side in Panel B of Table 9, we only report the coefficient on $StockCor_{iv}$ from each regression (with heteroscedasticity-consistent *t*-statistics). We identify an interesting pattern. For both dependent variables, the coefficients on the smaller trading venues are positive, although not all of them are statistically significant. For the largest trading venue in terms of market share, however, we find the opposite result: the coefficient is negative and statistically significant in almost all regressions. Therefore, the impact of correlated HFT strategies on the viability or competitiveness of a trading venue depends on the nature of that venue: it benefits smaller venues that introduce competition into the market, thus detracting from the dominant position of the largest trading venue.

The disparity in results between the smaller trading venues and the dominant exchange suggests that reverse causality is less likely to be a concern in these regressions. We also believe that the economics of trading on these venues is such that the percentage of time a trading venue posts the best prices is determined by the strategies of the HFT firms, rather than vice versa, because HFT firms are the dominant players on these trading venues in terms of liquidity provision. Our results therefore suggest that at least part of the negative relation between market concentration and competition between HFT firms that we document in Section 5.1 is driven by HFT competition enhancing the viability of smaller trading venues (in terms of displaying better prices and smaller spreads), which increases their market share.

6. Conclusions

In this paper, we examine product market competition in a particularly interesting subset of the financial

[41] Results using the other measures of HFT strategies are similar in nature and are omitted from the table to economize on the presentation.

services industry: firms that carry out high-frequency trading. The difficulty in pursing such a study has to do with the idiosyncrasies of an industry in which the product is often not publicly defined or easily observed by market participants. While regulators can have data on the actions of HFT firms, attempting to understand the product attributes (i.e., the essence of the strategy behind the actions) is often difficult if not outright impossible.

Against this backdrop, the idea behind our study is to define the main product categories in this industry by looking at whether multiple HFT firms follow related strategies. The empirical tool we use, principal component analysis, enables us to employ detailed data on the actions of individual HFT firms to identify representations of underlying strategies that are common to multiple firms. In and of itself, our finding that there are at least three main product categories that are attractive to multiple HFT firms is important because it shows that considering HFT as a single entity in terms of its impact on the market may afford limited insights. We show that these three underlying common strategies differ from one another not just in terms of the types of HFT activity (e.g., cross-venue messages or passive liquidity provision), but also in terms of the market and limit order book conditions associated with them. These strategies serve different functions and should be regarded separately by both regulators and market participants.

While much of the literature in economics regards competition as a good attribute that lowers the economic rents of firms and facilitates innovation, there are cases in which such competition can have negative byproducts.[42] Since competition between HFT firms manifests itself in more correlated actions across stocks, it is quite natural to be concerned that it will result in higher short-horizon volatility. We find the opposite result: individual stock volatility loads negatively on market-wide HFT competition, which suggests that competition in this industry could be especially important not just for rent reduction but also in terms of reduction in transitory price movements that exacerbate short-horizon volatility.

Lastly, the market structure for equity trading is currently fragmented among many trading venues. HFT firms are in a unique position to alleviate some of the negative byproducts of this fragmentation by effecting rapid price adjustments across trading venues that enhance the price integrity

[42] For example, product market competition can make it more difficult to infer a manager's action given the output of the firm (Golan, Parlour, and Rajan 2015), and in general could exacerbate the problem of incentivizing managers (e.g., Scharfstein 1988).

of the market. Given their dominance in the market, HFT firms can also help smaller trading venues attract order flow by supplying liquidity (Menkveld 2013). We find that greater HFT competition on smaller trading venues is helpful for the competitive position of these venues, but the opposite result is observed for the dominant trading venue in Canada. This is likely one of the drivers behind the negative relation we document between market concentration and HFT competition.

While our findings contribute several important insights to the study of the industrial organization of the HFT industry, much is still unknown. In particular, there is no evidence on the process of product development and testing in this industry, which has implications both for the barriers to entry (or the amount of funding required to develop profitable algorithms) and for the stability of the markets (runaway algorithms can wreak havoc and cause market disruptions). The importance of specific strategies to the well-functioning of our markets is also not completely understood. While we suggest that there is ample competition in the main strategies that HFT firms follow, we could not unequivocally show that the three product categories are competitive to the extent that economic rents have vanished. We view this as an open question, and hope that future studies will be able to provide a more definitive answer to this question, as well as deepen our knowledge of the other attributes of this fascinating industry.

References

Ait-Sahalia, Yacine and Mehmet Saglam, 2016, High Frequency Market Making, Working paper, Princeton University.

Anand, Amber, and Kumar Venkataraman, 2016, Market conditions, fragility, and the economics of market making, *Journal of Financial Economics* 121, 327-349.

Baron, Matthew, Jonathan Brogaard, Bjorn Hagstromer, and Andrei Kirilenko, 2016, Risk and return in high frequency trading, Working paper, Cornell University.

Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko, 2012, The trading profits of high frequency traders, working paper, Princeton University.

Benos, Evangelos, James Brugler, Erik Hjalmarsson, and Filip Zikes, 2015, Interactions among high-frequency traders, Working Paper No. 523, Bank of England.

Benos, Evangelos, and Satchit Sagade, 2016, Price discovery and the cross-section of high-frequency trading, forthcoming in the *Journal of Financial Markets,*.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2015, Equilibrium fast trading, *Journal of Financial Economics* 116, 292-313.

Breckenfelder, H. Johannes, 2013, Competition between high-frequency traders, and market quality, Working paper, Stockholm School of Economics, SHoF & SIFR.

Brogaard, Jonathan A., and Corey Garriott, 2016, High-frequency trading competition, Working paper, University of Washington and Bank of Canada.

Brogaard, Jonathan A., Terrence Hendershott, and Ryan Riordan, 2014, High frequency trading and price discovery, *Review of Financial Studies* 27, 2267-2306.

Brogaard, Jonathan A., Terrence Hendershott, and Ryan Riordan, 2016, Price discovery without trading: Evidence from limit orders, Working paper, University of Washington.

Brogaard, Jonathan A., Allen Carrion, Tjobait Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov, 2016, High-frequency trading and extreme price movements, Working paper, University of Washington.

Brown, Nerissa C., Kelsey D. Wei, Russ Wermers, 2014, Analyst recommendations, mutual fund herding, and overreaction in stock prices, *Management Science* 60, 1-20.

Budish, Eric, Peter Cramton, and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics* 130, 1547-1621.

Buti, Sabrina, Barbara Rindi, and Ingrid M. Werner, 2016, Dark pool trading strategies, market quality, and welfare, forthcoming in the *Journal of Financial Economics*.

Carrion, Allen, 2013, Very fast money: High-frequency trading on the NASDAQ, *Journal of Financial Markets* 16, 680-711.

Chaboud, Alain P., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2014, Rise of the machines: Algorithmic trading in the foreign exchange market, *Journal of Finance*, LXIX, 5, October, 2045-2084.

Choi, Nicole, and Richard W. Sias, 2009, Institutional industry herding, *Journal of Financial Economics* 94, 469-491.

Clark-Joseph, Adam D., 2014, Exploratory trading, Working paper, University of Illinois.

Comerton-Forde, Carole, Katya Malinova, and Andreas Park, 2016, Regulating dark trading: order flow segmentation and market quality, Working paper, University of Toronto and FIRN.

Foster, F. Douglas, and S. Viswanathan, 1993, The effect of public information and competition on trading volume and price volatility, *Review of Financial Studies* 6, 23-56.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu, 2016, News trading and speed, *Journal of Finance* 71, 335-382.

Golan, Limor, Christine A. Parlour, and Uday Rajan, 2015, Competition, Managerial slack, and corporate governance, *Review of Corporate Finance Studies* 4, 43-68.

Goldstein, Michael A., Pavitra Kumar, and Frank C. Graves, 2014, Computerized and high-frequency trading, forthcoming, *The Financial Review*, 49, 177-202.

Green, Edward J., and Robert H. Porter, Noncooperative collusion under imperfect price information, *Econometrica* 52, 87-100.

Hagstromer, Bjorn, and Lars Norden, 2013, The diversity of high-frequency traders, *Journal of Financial Markets* 16, 741-770.

Hagstromer, Bjorn, Lars Norden, and Dong Zhang, 2014, How aggressive are high-frequency traders? *The Financial Review*, 49, 395–419.

Han, Jungsuk, Mariana Khapko, and Albert S. Kyle, 2014, Liquidity with high-frequency market making, Working paper, Swedish House of Finance.

Hasbrouck, Joel, and Gideon Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646-679.

Hirschey, Nicholas H., 2016, Do high-frequency traders anticipate buying and selling pressure? Working paper, London Business School.

Ho, Thomas S.Y., and Hans R. Stoll, 1983, The dynamics of dealer markets under competition, *Journal of Finance* 38, 4, September, 1053-1074.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, forthcoming in the *Journal of Political Economy*.

Hoffmann, Peter, 2014, A dynamic limit order market with fast and slow traders, *Journal of Financial Economics* 113, 156-169.

Holden, Craig W., and Avanidhar Subrahmanyam, 1992, Long-lived private information and imperfect competition, Journal of Finance 47, 247-270.

Ivaldi Marc, Bruno Jullien, Patrick Rey, Paul Seabright, and Jean Tirole, 2003, The economics of tacit collusion, Final Report for DG Competition, European Commission.

Jarnecic, Elvis, and Mark Snape, 2014, The provision of liquidity by high-frequency participants, *The Financial Review*, 49, 371–394.

Jarrow, Robert A., and Philip Protter, 2012, A dysfunctional role of high frequency trading in electronic markets, *International Journal of Theoretical and Applied Finance*, 15, 3, 1-15.

Jones, Charles M., 2013, What do we know about high-frequency trading? Working paper, Columbia Business School.

Jovanovic, Boyan, and Albert J. Menkveld, 2016, Middlemen in Limit Order Markets, Working paper, New York University.

Khandani, Amir E., and Andrew W. Lo, 2007, What happened to the quants in August 2007? *Journal of Investment Management* 5, 29-78.

Khandani, Amir E., and Andrew W. Lo, 2011, What happened to the quants in August 2007? Evidence from factors and transactions data, *Journal of Financial Markets*, 14, 1-46.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2016, The flash crash: The impact of high frequency trading on an electronic market, forthcoming in the *Journal of Finance*.

Korajczyk, Robert A., and Dermot Murphy, 2016, High frequency market making to large institutional trades, Working paper, Northwestern University.

Lescourret, Laurence, and Sophie Moinas, 2015, Liquidity supply across multiple trading venues. Working paper. ESSEC Business School.

Li, Su, 2013, Multiple insiders with long lived flow of private information, and high frequency competition, Working paper, U.S. Securities and Exchange Commission.

Malinova, Katya, and Andreas Park, 2015, Subsidizing liquidity: the impact of make/take fees on market quality, *Journal of Finance* 70, 509-536.

Malinova, Katya, Andreas Park, and Ryan Riordan, 2014, Do retail traders suffer from high frequency traders? Working paper, University of Toronto.

Menkveld, Albert J., 2013, High frequency trading and the new-market makers, *Journal of Financial Markets*, 16, 712–740.

Menkveld, Albert J., and Marius A. Zoican, 2016, Need for speed? Exchange latency and liquidity, Working paper, Duisenberg School of Finance - Tinbergen Institute Discussion Paper.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality? *Journal of Financial Economics*, 100, 3, 459-474.

Pedersen, Lasse Heje, 2009, When everyone runs for the exit, *International Journal of Central Banking* 5, 177-199.

Rosu, Ioanid, 2016, Fast and slow informed trading, Working paper, HEC.

Scharfstein, David, 1988, Product market competition and managerial slack, *Rand Journal of Economic*s 19, 147–55.

Subrahmanyam, Avanidhar, and Hui Zheng, 2016, Limit order placement by high-frequency traders, Working paper, UCLA.

Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance*, LIV, 2, April, 581-622.

Zou, Hui, Trevor Hastie, and Robert Tibshirani, Sparse principal component analysis, 2006, *Journal of Computational and Graphical Statistics* 15, 265-286.

Table 1
Summary Statistics

Our sample consists of 52 stocks from the S&P/TSX60 Index. We rank the daily returns of the S&P/TSX Composite Index from June 2010 through March 2011, and select the 10 worst days (down days), the 10 best days (up days), and the 10 days closest to and centered on zero return (flat days) for a 30-day sample period. Panel A presents summary statistics for the sample stocks: market capitalization, price, standard deviation of 30-minute returns, daily volume (in Canadian dollars), and daily return. Panel B presents summary statistics for the 31 high-frequency trading (HFT) firms that we identify using data from the Investment Industry Regulatory Organization of Canada (IIROC). These 31 firms are further categorized into three subgroups according to market share of volume: MS1 (market share of at least 4%; 4 firms), MS2 (market share of between 1% and 4%; 6 firms), and MS3 (the rest of the HFT firms). Market share (in terms of volume or trades) is computed by dividing the trading undertaken by each HFT firm by total trading in the market. Trades consist of executions of both marketable orders and non-marketable orders. Sub/Canc is the number of (non-marketable) limit order submissions and cancellations (i.e., all non-execution messages) that the firm sends to the market, where a modification of an order counts as a cancellation and a resubmission. Messages are all the orders (both marketable and non-marketable) and cancellations that a firm sends to the market. Orders/Trades Ratio is defined for each HFT firm as Messages divided by Trades. The mean order-to-trade ratio is the average of the order-to-trade ratios of the individual HFT firms, and therefore due to its non-linear nature and the heterogeneity of the firms is not equal to the cross-sectional mean number of messages divided by the cross-sectional mean number of trades. CrossEndInventory is the number of times per day that an HFT firm's intraday inventory position crosses its end-of-day inventory position. We compute the measures for each HFT firm using all days in our sample period, and then provide in the table cross-firm means and medians for all HFT firms, as well as for subgroups by market share.

Panel A: Sample Stocks

Days	Stocks		MktCap (Million CAD)	Price (CAD)	StdRet (30min Ret)	CADVolume (1000; daily)	Return (daily)
All	S&P60	Mean	19,412	39.1	0.40%	78,156	-0.04%
		Median	11,401	36.9	0.40%	51,120	-0.04%
Down Days	S&P60	Mean	19,203	38.3	0.43%	81,195	-1.72%
		Median	11,299	36.4	0.43%	55,892	-1.56%
Flat Days	S&P60	Mean	19,499	39.7	0.34%	70,614	-0.08%
		Median	11,840	37.4	0.34%	50,078	-0.07%
Up Days	S&P60	Mean	19,535	39.5	0.39%	82,903	1.66%
		Median	11,574	37.0	0.37%	54,097	1.46%

Panel B: HFT Firms

HFT Firms		MktShare (Volume)	MktShare (Trades)	Trades (daily)	Sub/Canc (daily)	Messages (daily)	Orders/Trades Ratio	CrossEnd Inventory
All	Mean	1.50%	1.70%	19,445	1,056,838	1,063,974	92.7	14.9
	Median	0.34%	0.44%	5,035	97,146	97,288	40.4	2.4
MS1	Mean	6.58%	8.95%	102,035	5,466,547	5,496,423	49.5	73.0
	Median	7.23%	8.60%	98,048	3,272,863	3,312,989	42.2	70.1
MS2	Mean	2.67%	2.08%	23,730	982,137	995,868	38.3	13.4
	Median	2.79%	2.11%	24,003	147,090	162,729	6.9	4.3
MS3	Mean	0.19%	0.22%	2,489	238,236	239,157	116.5	4.3
	Median	0.05%	0.04%	714	40,129	40,268	71.4	1.8

Table 2

Principal Components Analysis: Loadings

This table presents the loadings from a principal component analysis of HFT strategies. The measure of HFT activity that we use to characterize the strategies is MSG, which comprises all messages an HFT firm actively sends to the market in an interval (submission of non-marketable limit orders, cancellation of non-marketable limit orders, and marketable limit orders that result in trade executions). We conduct the analysis using one-second intervals. For the purpose of the analysis, we view the 31 HFT firms as the "variables," while the observations are all intervals during the 30-day sample period for all sample stocks. The principal component analysis uses the varimax orthogonal rotation, and the first three principal components are retained for further analysis. The loading of an HFT firm on each of the principal components signifies the extent to which the firm's activity corresponds to the underlying common strategy represented by that principal component; it is equivalent to the bivariate correlation between the firm's measure and the principal component, and is therefore between –1 and 1. For each principal component, the loadings are sorted from the most positive to the most negative. An asterisk indicates all principal components that are ≥ 0.35.

HFT	PC1 Loading	HFT	PC2 Loading	HFT	PC3 Loading
F14	0.76 *	F27	0.71 *	F17	0.51 *
F16	0.67 *	F08	0.52 *	F23	0.48 *
F04	0.57 *	F31	0.41 *	F19	0.47 *
F24	0.54 *	F05	0.40 *	F20	0.43 *
F31	0.48 *	F02	0.39 *	F26	0.33
F28	0.46 *	F28	0.35 *	F12	0.22
F20	0.40 *	F30	0.34	F08	0.20
F17	0.38 *	F06	0.27	F18	0.18
F29	0.33	F12	0.24	F14	0.15
F01	0.32	F26	0.22	F29	0.13
F21	0.26	F10	0.17	F21	0.12
F27	0.26	F01	0.16	F03	0.12
F06	0.26	F20	0.13	F10	0.08
F07	0.23	F04	0.12	F30	0.05
F26	0.23	F24	0.11	F05	0.03
F08	0.22	F09	0.11	F02	0.03
F11	0.15	F23	0.10	F06	0.03
F12	0.08	F11	0.07	F16	0.02
F19	0.07	F18	0.06	F25	0.01
F05	0.04	F14	0.05	F13	0.01
F22	0.03	F03	0.04	F27	0.00
F10	0.01	F15	0.03	F09	0.00
F15	0.01	F25	0.02	F22	-0.01
F09	0.00	F22	0.01	F15	-0.01
F25	0.00	F13	0.01	F04	-0.02
F03	0.00	F29	0.00	F28	-0.07
F13	0.00	F16	-0.03	F31	-0.09
F02	-0.04	F21	-0.07	F24	-0.13
F30	-0.06	F07	-0.12	F07	-0.13
F23	-0.07	F19	-0.12	F11	-0.13
F18	-0.08	F17	-0.16	F01	-0.30

Table 3

Regressions using Principal Component Scores

This table presents regressions of principal component scores on variables that represent the market environment. In the PCA of the MSG measure, the 31 high-frequency trading (HFT) firms are the "variables," while the observations are all 1-second intervals during the 30-day sample period for all sample stocks. The PCA uses the varimax orthogonal rotation, and the first three principal components are retained for further analysis. There are 11 right-hand-side variables in each regression that describe attributes of the HFT strategies, as well as the market environment. We identify only the HFT firms that simultaneously (i.e., in the same interval) submit messages to multiple trading venues, and HFTcrossmsg is the aggregate number of these messages for the 31 HFT firms. HFTliqsupply is the aggregate number of passive (i.e., liquidity supplying) trades for the HFT firms, and |HFTinventory| is the absolute value of the aggregate inventory position of all HFT firms in Canadian dollars (cumulative from the beginning of the day and assuming that all of them start the day with zero inventory). The next two variables represent the degree of integration of the Canadian market. Specifically, PriceAlign is the percentage of time that the three trading venues with the highest market share post the same bid or ask prices, while SpreadAlign is the percentage of time that the three trading venues have the same bid–ask spread. The next three variables represent the state of liquidity in the market (aggregated across all trading venues): total depth up to 10 cents from the market-wide best bid or offer (MWBBO), depth imbalance at the MWBBO (defined as the absolute value of the difference between the number of shares at the bid and at the ask), and percentage MWBBO spread. The last three variables represent market conditions in the interval: return (computed from the last transaction price in each interval), volatility (computed as the absolute value of return), and the average time between trades in the interval. All right-hand-side variables are standardized to have zero mean and unit standard deviation to make the coefficients comparable across variables. We report the regression coefficients together with *t*-statistics computed using double-clustered (interval and stock) standard errors.

	PC1		PC2		PC3	
Variable	Coef.	*t*-stat.	Coef.	*t*-stat.	Coef.	*t*-stat.
HFTcrossmsg	0.6717	8.25	0.1548	1.55	0.3330	3.74
HFTliqsupply	-0.0285	-1.07	0.3556	10.61	-0.0973	-4.12
|HFTinventory|	0.0051	1.68	-0.0047	-1.40	0.0013	0.90
PriceAlign	-0.0050	-0.41	0.0299	2.60	-0.0386	-4.46
SpreadAlign	-0.0261	-5.00	0.0159	2.53	0.0013	0.33
Depth10	-0.0428	-4.84	0.0737	4.76	-0.0278	-2.43
|TopDepthImb|	0.0057	2.55	-0.0193	-2.74	0.0078	2.10
%Spread	-0.0077	-2.07	0.0030	1.43	0.0068	2.01
Return	0.0015	2.04	0.0007	0.91	0.0007	0.98
|Return|	0.0262	2.54	-0.0023	-0.39	0.0256	3.46
Time-bet-Trades	-0.0850	-5.54	-0.1051	-6.53	0.0150	1.73
Intercept	-2.91E-07	0.00	4.47E-08	0.00	-3.14E-07	0.00
R^2	49.61%		27.40%		9.08%	

Table 4
Net Trading Revenues

This table presents net trading revenues for the HFT firms. We use data from the Investment Industry Regulatory Organization of Canada (IIROC) to identify 31 high-frequency trading (HFT) firms. Our sample consists of 52 stocks from the S&P/TSX 60 (S&P). For each HFT firm, we compute net trading revenues per day per stock by summing all the proceeds from selling shares (in Canadian dollars) and subtracting the cost of purchasing shares. We obtain information about fees and rebates from each trading venue and adjust the costs and revenues from buying and selling shares on each trading venue for these fees and rebates. If the numbers of shares bought on a given day for a given stock is not the same as the number of shares sold, we assume that the inventory of shares (the short position) is liquidated (covered) at the end of the day at either the closing price or the closing midquote, resulting in two estimates of NetRevenues (denoted CloseP and MidQ, respectively). In Panel A, we present net trading revenues computed separately for each market environment in our 30-day sample period: 10 worst days (down days), the 10 best days (up days), and the 10 days closest to and centered on zero return (flat days). We present the median net trading revenues (NetRevenues) across all the HFT/stock/day observations, as well as *p*-values from a Wilcoxon test of one market environment versus another and a chi-squared test for equality across the three market environments. In Panel B, we present the net trading revenues for the firms that follow the three underlying common strategies we identify using the principal component analysis: PC1 (8 firms), PC2 (6 firms), and PC3 (4 firms), as well as the median for the HFT firms that do not follow any underlying common strategy (NonPC, 17 firms). We also provide *p*-values from a Wilcoxon test of one group versus another (e.g., PC1 vs. NonPC), as well as from a chi-squared test of PC1 vs. PC2 vs. PC3. All *p*-values are for a two-sided test.

Panel A: Net Trading Revenues by Market Environment

		NetRevenues	vs. DF *p*-value	vs. DU *p*-value	vs. DF vs. DU Chi-squared *p*-value
Down Days	MidQ	35.90	<.001	0.366	<.001
	CloseP	38.56	<.001	0.382	<.001
Flat Days	MidQ	11.01		<.001	
	CloseP	11.42		<.001	
Up Days	MidQ	34.06			
	CloseP	34.59			

Panel B: Net Trading Revenues of HFT Firms by Underlying Common Strategy

		NetRevenues	vs. NonPC *p*-value	vs. PC2 *p*-value	vs. PC3 *p*-value	vs. PC2 vs. PC3 Chi-squared *p*-value
PC1	MidQ	68.54	<.001	<.001	<.001	<.001
	CloseP	70.39	<.001	<.001	<.001	<.001
PC2	MidQ	16.64	0.101		<.001	
	CloseP	22.54	0.037		<.001	
PC3	MidQ	105.00	<.001			
	CloseP	105.10	<.001			
NonPC	MidQ	10.30				
	CloseP	10.95				

Table 5

Market-Wide Correlations

This table presents cross-sectional or market-wide correlations of HFT strategies. We use data from the Investment Industry Regulatory Organization of Canada (IIROC) to identify 31 high-frequency trading (HFT) firms. These 31 firms are further categorized into three subgroups according to market share of volume: MS1 (market share of at least 4%; 4 firms), MS2 (market share of between 1% and 4%; 6 firms), and MS3 (the rest of the HFT firms). The market-wide correlations indicate whether the strategies of HFT firms are correlated across stocks at a given time. In each time interval, we compute the correlation coefficient between the activities of two HFT firms across the stocks in the sample, and average the correlations for all pairs of firms in a certain group (where ALL consists of the 31 HFT firms). We examine three measures of HFT activity: (i) the number of "messages" (MSG) HFT firms send to the market, where messages are defined as submissions and cancellations of nonmarketable limit orders as well as executions of marketable limit orders, (ii) trades (TRD), and (iii) submissions/cancellations of nonmarketable limit orders (LMT). The measures representing HFT strategies as well as the correlations are computed separately for three interval lengths: I1 (1-second intervals), I2 (10-second intervals), and I3 (60-second intervals). Our sample consists of 52 stocks from the S&P/TSX60 (S&P).

Strategy Measure	HFT Group	I1	I2	I3
MSG	MS1	0.359	0.332	0.317
	MS2	0.131	0.120	0.122
	MS3	0.174	0.127	0.108
	ALL	0.198	0.147	0.130
TRD	MS1	0.417	0.436	0.518
	MS2	0.219	0.122	0.155
	MS3	0.189	0.067	0.052
	ALL	0.313	0.152	0.131
LMT	MS1	0.355	0.328	0.313
	MS2	0.127	0.110	0.107
	MS3	0.175	0.129	0.108
	ALL	0.197	0.147	0.127

Table 6

Regressions of Individual Stock Volatility on Market-wide HFT Competition

This table presents the results of regressions of individual stock interval return volatility on the market-wide correlations of high-frequency trading (HFT) strategies. Our sample consists of 52 stocks from the S&P/TSX 60. For each stock, we estimate the following regression over all intervals in the sample period:

$$\left|r_{it}\right| = a_{0i} + a_{1i}MktCor_{mt} + a_{2i}HFT_{mt} + a_{3i}r_{mt} + a_{4i}\left|r_{mt}\right| + a_{5i}Volume_{mt} + error_{it},$$

where $|r_{it}|$ is the absolute value of the return for stock i in interval t (computed from the trade price closest to the end of the interval), which is our measure of interval return volatility, r_{mt} is the value-weighted return of all stocks in our sample, $|r_{mt}|$ is the absolute value of the market return, and $Volume_{mt}$ is the aggregate volume in all stocks. We are chiefly interested in the impact of the cross-sectional correlation between HFT firms as measured by $MktCor_{mt}$, which is defined as the average over all pairs of HFT firms of their correlation across stocks (for a particular activity measure). $MktCor_{mt}$ is a market-wide attribute of the extent of competition (or similarity in strategies) between HFT firms in the market, and is computed for each interval in the 30-day period. We use three measures of HFT strategies: MSG (all messages HFT firms send to the market), TRD (all their trades), and LMT (only submissions and cancellations of non-marketable limit orders). HFT_{mt} is the magnitude of HFT activity in the market (which is the sum across all HFT firms of the same measure that we use for $MktCor_{mt}$). For each variable, we present the average coefficient across the 52 stocks, t-statistics computed from the cross-sectional distribution of the regression coefficients, the number of negative coefficients, the number of negative coefficients that are significant at the 5% level (from a two-sided test), the number of positive coefficients, the number of significant positive coefficients, and the average R^2. In Panel A, we present the coefficients on all variables from the regressions using the 1-second intervals. In Panel B, we present just the coefficient on $MktCor_{mt}$ side-by-side for the 1-second, 10-second, and 60-second intervals. We add three statistical tests to help determine whether the coefficients of the individual regressions are different from zero: a joint F-test from a seemingly unrelated regression estimation (SURE) of the 52-equation system, the nonparametric sign test, and the Wilcoxon test.

Panel A: Summary Statistics for 52 Regressions of Individual Stock Volatility on Market-wide HFT Competition

		Avg. Coef.	CS t-stat.	# Coef. < 0	# t-stat. < -1.96	# Coef. > 0	# t-stat. > 1.96	Avg. R^2
MSG	*Intercept*	8.00E-06	3.67	6	5	46	45	5.47%
	$MktCor_{mt}$	-1.72E-05	-4.18	43	39	9	6	
	HFT_{mt}	5.76E-09	6.53	5	5	47	47	
	r_{mt}	-6.29E-03	-1.67	36	10	16	1	
	$\|r_{mt}\|$	8.22E-01	8.72	0	0	52	52	
	$Volume_{mt}$	1.35E-12	9.68	5	3	47	45	
TRD	*Intercept*	1.39E-05	7.06	1	1	51	48	5.81%
	$MktCor_{mt}$	-1.78E-05	-9.58	50	49	2	1	
	HFT_{mt}	2.67E-07	8.31	5	4	47	47	
	r_{mt}	-6.01E-03	-1.59	36	10	16	2	
	$\|r_{mt}\|$	7.27E-01	7.13	0	0	52	52	
	$Volume_{mt}$	-3.45E-13	-3.68	43	22	9	3	
LMT	*Intercept*	7.99E-06	3.68	7	5	45	45	5.46%
	$MktCor_{mt}$	-1.71E-05	-4.14	43	40	9	6	
	HFT_{mt}	5.74E-09	6.50	5	5	47	47	
	r_{mt}	-6.30E-03	-1.67	36	10	16	1	
	$\|r_{mt}\|$	8.24E-01	8.76	0	0	52	52	
	$Volume_{mt}$	1.38E-12	9.70	5	3	47	46	

Panel B: Coefficient on $MktCor_{mt}$ from 52 Regressions of Individual Stock Volatility on Market-wide HFT Competition

Variable	**Interval**	Avg. Coef.	CS *t*-stat.	Joint Test (*p*-value)	Sign Test (*p*-value)	Wilcoxon (*p*-value)	# Coeff. < 0	# *t*-stat. < -1.96	# Coeff. > 0	# *t*-stat. > 1.96	Avg. R^2
MSG	**I1**	-1.72E-05	-4.18	<0.001	<0.001	<0.001	43	39	9	6	5.47%
	I2	-2.43E-04	-6.55	<0.001	<0.001	<0.001	46	43	6	5	9.65%
	I3	-6.51E-04	-6.83	<0.001	<0.001	<0.001	45	36	7	4	14.93%
TRD	**I1**	-1.78E-05	-9.58	<0.001	<0.001	<0.001	50	49	2	1	5.81%
	I2	-1.71E-04	-7.96	<0.001	<0.001	<0.001	45	42	7	2	9.84%
	I3	-2.13E-04	-3.48	<0.001	0.018	0.001	35	28	17	6	14.39%
LMT	**I1**	-1.71E-05	-4.14	<0.001	<0.001	<0.001	43	40	9	6	5.46%
	I2	-2.50E-04	-6.61	<0.001	<0.001	<0.001	46	42	6	5	9.65%
	I3	-7.14E-04	-7.27	<0.001	<0.001	<0.001	46	37	6	4	14.95%

Table 7

Individual Stock Volatility and Market-wide HFT Competition by Principal Component

This table presents the results of regressions of individual stock interval volatility on three variables that describe market-wide correlation of HFT strategies: one for each set of firms that follow an underlying common strategies. For each stock, we estimate the following regression over all one-second intervals in the sample period:

$$\begin{aligned}|r_{it}| = a_{0i} + a_{1i}PC1Cor_{mt} + a_{2i}PC2Cor_{mt} + a_{3i}PC3Cor_{mt} + \\ a_{4i}PC1HFT_{mt} + a_{5i}PC2HFT_{mt} + a_{6i}PC3HFT_{mt} + a_{7i}r_{mt} + a_{8i}|r_{mt}| + a_{9i}Volume_{mt} + error_{it}\end{aligned},$$

where $|r_{it}|$ is the absolute value of the return for stock *i* in interval *t*, which is our measure of interval return volatility, r_{mt} is the value-weighted return of all stocks in our sample, $|r_{mt}|$ is the absolute value of the market return, and $Volume_{mt}$ is the aggregate volume in all stocks. We compute the HFT competition variable $PC1Cor_{mt}$ as the average over the market-wide (cross-sectional) correlations of all pairs of HFT firms from among the eight firms that load significantly on the first principal component, and similarly we compute $PC2Cor_{mt}$ ($PC3Cor_{mt}$) for the firms that load significantly on the second (third) principal component. We add the magnitude of HFT activity of the firms that load on the three principal components as control variables: $PC1HFT_{mt}$, $PC2HFT_{mt}$, and $PC3HFT_{mt}$. We conduct the regression analysis for three measures of HFT strategies: MSG (all messages HFT firms send to the market), TRD (all their trades), and LMT (submissions and cancellations of non-marketable limit orders). The table presents for each variable the average coefficient across the 52 stocks, *t*-statistics computed from the cross-sectional distribution of the regression coefficients, the number of negative coefficients, the number of negative coefficients that are significant at the 5% level (from a two-sided test), the number of positive coefficients, the number of significant positive coefficients, and the average R^2.

		Avg. Coef	Avg. t-stat	# Coef < 0	# t-stat < -1.96	# Coef > 0	# t-stat > 1.96	Avg. R^2
MSG	*Intercept*	9.39E-06	3.99	8	7	44	44	6.13%
	$PC1Cor_{mt}$	-2.52E-06	-0.81	33	30	19	13	
	$PC2Cor_{mt}$	-8.90E-06	-4.29	41	39	11	10	
	$PC3Cor_{mt}$	-6.84E-07	-0.68	37	29	15	9	
	$PC1HFT_{mt}$	-1.20E-09	-0.63	34	33	18	14	
	$PC2HFT_{mt}$	7.96E-09	3.87	10	7	42	39	
	$PC3HFT_{mt}$	9.25E-09	5.76	8	7	44	42	
	r_{mt}	-6.66E-03	-1.58	33	8	19	3	
	$\|r_{mt}\|$	8.56E-01	10.81	0	0	52	52	
	$Volume_{mt}$	2.02E-12	10.90	4	2	48	45	
TRD	*Intercept*	4.37E-05	7.27	1	0	51	50	12.56%
	$PC1Cor_{mt}$	-1.01E-05	-4.84	42	27	10	3	
	$PC2Cor_{mt}$	-1.91E-05	-9.78	48	43	4	0	
	$PC3Cor_{mt}$	-2.96E-05	-3.80	48	41	4	3	
	$PC1HFT_{mt}$	-2.35E-08	-0.71	34	15	18	9	
	$PC2HFT_{mt}$	8.13E-07	0.27	25	11	27	12	
	$PC3HFT_{mt}$	4.03E-07	4.95	11	2	41	28	
	r_{mt}	1.51E-03	0.15	23	4	29	4	
	$\|r_{mt}\|$	9.87E-01	20.81	0	0	52	52	
	$Volume_{mt}$	-2.20E-13	-0.41	31	10	21	11	
LMT	*Intercept*	9.42E-06	4.00	8	7	44	44	6.12%
	$PC1Cor_{mt}$	-2.54E-06	-0.82	33	31	19	13	
	$PC2Cor_{mt}$	-8.65E-06	-4.24	41	39	11	10	
	$PC3Cor_{mt}$	-7.39E-07	-0.73	37	29	15	9	
	$PC1HFT_{mt}$	-7.74E-10	-0.42	34	34	18	15	
	$PC2HFT_{mt}$	6.98E-09	3.52	12	7	40	39	
	$PC3HFT_{mt}$	8.96E-09	5.76	7	7	45	42	
	r_{mt}	-6.64E-03	-1.58	33	8	19	3	
	$\|r_{mt}\|$	8.58E-01	10.88	0	0	52	52	
	$Volume_{mt}$	2.05E-12	10.86	4	2	48	45	

Table 8

Market Concentration and the Correlation of HFT Strategies

This table presents the results of an analysis that relates concentration of trading across the trading venues to competition between high-frequency trading (HFT) firms. The five trading venues we investigate are organized as electronic limit order books and together execute approximately 97.7% of the trading volume in Canada during our sample period. Our sample consists of 52 stocks from the S&P/TSX 60 (S&P). To examine market concentration, we compute the Herfindahl-Hirschman Index (HHI_i) of market share in terms of volume for the five trading venues for each stock. The HHI is computed as the sum of squared market shares of the trading venues, and the lower the HHI the less concentrated the market. We use data from the Investment Industry Regulatory Organization of Canada (IIROC) to identify 31 high-frequency trading (HFT) firms. To represent HFT competition we construct the variable $StockCor_i$, which describes how the strategies of HFT firms are correlated over time for a given stock. To construct the variable for each stock, we compute the correlation coefficient between the activities of pairs of HFT firms over all time intervals, and average across all pairs of firms. We examine three measures of HFT activity: (i) the number of "messages" (MSG) HFT firms send to the market, where messages are defined as submissions and cancellations of nonmarketable limit orders as well as executions of marketable limit orders, (ii) trades (TRD), and (iii) submissions/cancellations of nonmarketable limit orders (LMT). The measures representing HFT strategies as well as the correlations are computed separately for three interval lengths: I1 (1-second intervals), I2 (10-second intervals), and I3 (60-second intervals). The first line in the table presents the partial correlations between *HHI* and *StockCor* when we control for market capitalization, price level, volatility, spread, depth, and the magnitude of HFT activity. In the next three lines, we replace the *StockCor* variable with HFT competition variables computed only from pairs of HFT firms that load significantly on one of the three principal components (e.g., *StockCorPC1* for the eight firms that load significantly on the first principal component). We also replace the control variable for the magnitude of all HFT activity with the magnitude of HFT activity only for the HFT firms that load significantly on that principal component. We report in parentheses *p*-values for the (two-sided) hypothesis that the partial correlation is equal to zero.

	MSG			TRD			LMT		
	I1	I2	I3	I1	I2	I3	I1	I2	I3
StockCor	-0.603	-0.611	-0.635	-0.333	-0.380	-0.177	-0.608	-0.622	-0.656
	(<.001)	(<.001)	(<.001)	(0.024)	(0.009)	(0.239)	(<.001)	(<.001)	(<.001)
StockCorPC1	-0.539	-0.587	-0.689	-0.375	-0.424	-0.535	-0.539	-0.588	-0.686
	(<.001)	(<.001)	(<.001)	(0.010)	(0.003)	(<.001)	(<.001)	(<.001)	(<.001)
StockCorPC2	-0.564	-0.483	-0.490	-0.502	-0.422	-0.500	-0.554	-0.500	-0.536
	(<.001)	(0.001)	(0.001)	(<.001)	(0.004)	(<.001)	(<.001)	(<.001)	(<.001)
StockCorPC3	0.031	-0.086	-0.150	-0.075	-0.111	-0.038	0.022	-0.083	-0.152
	(0.840)	(0.571)	(0.319)	(0.622)	(0.462)	(0.802)	(0.885)	(0.581)	(0.313)

Table 9

Trading Venue Competitiveness and the Correlation of HFT Strategies

This table presents summary statistics of measures of the competitiveness (or viability) of trading venues and regression results that relate them to competition between high-frequency trading (HFT) firms. The five trading venues we investigate are all organized as electronic limit order books and together execute 97.7% of the volume in Canada during our sample period. We denote the five trading venues in the table by the letters A through E. Panel A provides cross-sectional summary statistics for market share in terms of volume, as well as for the two measures of trading venue viability or competitiveness: (i) %TimeBestPrices, defined as the percentage of time that the trading venue posts either the best bid or the best ask in the market (where the market is defined as the aggregation of all five trading venues), and (ii) %TimeSmallSpreads, defined as the percentage of time that the bid–ask spread on the trading venue is the narrowest spread in the market. In Panel B we examine whether correlated activity of HFT firms on a particular trading venue is helpful for the competitive position of the trading venue by manifesting in better prices and spreads. Not all HFT firms in our sample are very active on multiple trading venues. We therefore focus on the firms with substantial activity on all five trading venues, which we define as sending at least 10,000 messages to each of the five trading venues. There are eight HFT firms that satisfy this criterion, and we compute trading-venue-specific time-series correlations that provide information regarding whether the strategies of these HFT firms are correlated over time for a given stock on a particular trading venue. These correlations are similar in nature to $StockCor_i$, except they are computed for each trading venue separately using only activity on that trading venue. For each trading venue v, we run the following cross-sectional regression:

$$C_{iv} = a_0 + a_1 StockCor_{iv} + a_2 MSG_{iv} + a_3 Spread_{iv} + a_4 BBOdepth_{iv} + a_5 MktCap_i + a_6 Price_i + a_7 Volatility_i + error_{iv}$$

where C_{iv} is one of the two viability measures, and we use MSG (total messages sent by the HFT firm) as our measure of HFT strategies for computing $StockCor_{iv}$. The next two variables, average spread and bid-and-offer (BBO) depth, are computed separately for each trading venue and are meant to control for the liquidity environment of the stock on that trading venue. The last three control variables—market capitalization, price level, and the standard deviation of 30-minute returns over the sample period—are meant to control for heterogeneity in fundamental attributes across stocks. To economize on the size of the table, and because we want to present results for the five trading venues and the three time intervals (I1, I2, and I3), we report in Panel B only the coefficients on $StockCor_{iv}$ from each regression (with heteroscedasticity-consistent *t*-statistics in parentheses).

Panel A: Summary Statistics for Market Share and Competitiveness Measures

		A	B	C	D	E
%Market Share of Volume	Mean	14.3%	11.8%	0.4%	1.9%	69.3%
	Std. Dev.	6.7%	3.4%	0.6%	1.3%	8.0%
	Min	4.8%	3.7%	0.0%	0.4%	45.1%
	25^{th} Perc	9.5%	9.8%	0.1%	1.2%	64.3%
	Median	13.5%	12.2%	0.1%	1.6%	70.4%
	75^{th} Perc	16.7%	14.0%	0.5%	2.3%	75.9%
	Max	34.8%	17.6%	2.7%	7.2%	82.8%
	N	52	52	52	52	52
%TimeBestPrices	Mean	65.2%	83.2%	32.7%	66.0%	92.0%
	Std. Dev.	22.6%	13.7%	17.4%	33.0%	6.1%
	Min	15.0%	39.6%	8.1%	0.1%	71.1%
	25^{th} Perc	50.9%	80.6%	20.8%	37.0%	89.8%
	Median	71.9%	87.0%	28.0%	79.6%	92.5%
	75^{th} Perc	84.9%	91.0%	40.7%	96.3%	96.7%
	Max	92.2%	98.0%	82.8%	99.3%	98.9%
	N	52	52	52	52	52
%TimeSmallSpread	Mean	36.1%	54.2%	9.4%	26.9%	76.8%
	Std. Dev.	24.0%	15.8%	14.3%	19.9%	8.0%
	Min	1.8%	13.2%	0.1%	0.0%	50.4%
	25^{th} Perc	13.4%	44.2%	1.9%	6.4%	72.8%
	Median	37.0%	58.1%	3.3%	27.3%	78.5%
	75^{th} Perc	57.1%	65.3%	10.7%	44.9%	82.5%
	Max	83.5%	83.8%	69.3%	67.6%	89.8%
	N	52	52	52	52	52

Panel B: Coefficients on $StockCor_{iv}$ from Regressions of Competitiveness Measures

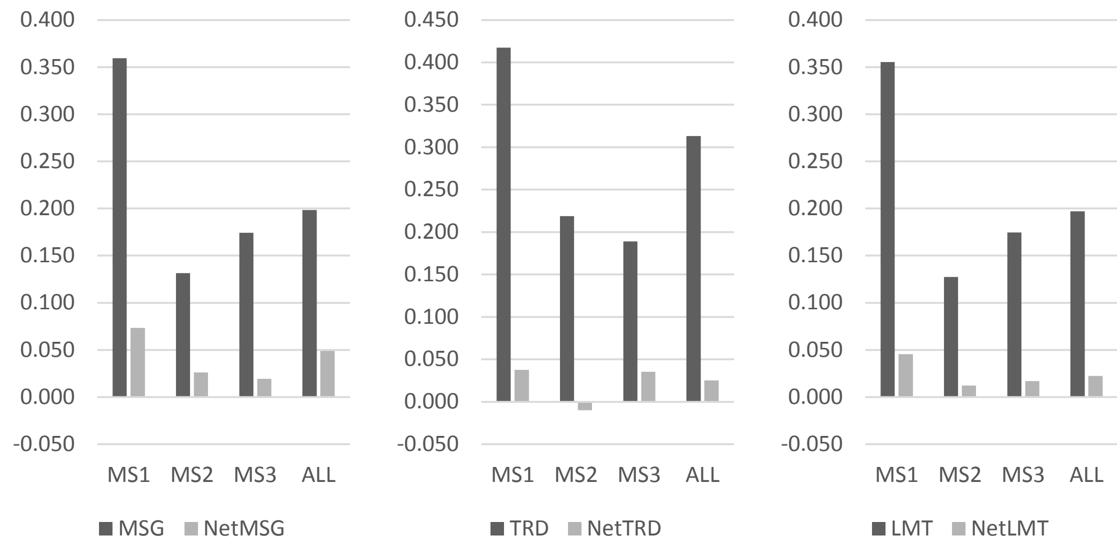
	Interval	Venue A	Venue B	Venue C	Venue D	Venue E
%TimeBestPrices	I1	0.658	0.220	0.970	4.281	-0.340
		(2.15)	(0.92)	(1.27)	(3.20)	(-3.64)
	I2	0.704	0.325	1.474	2.091	-0.199
		(2.85)	(1.46)	(2.17)	(1.11)	(-2.15)
	I3	0.661	0.348	1.427	0.781	-0.096
		(2.70)	(1.76)	(3.31)	(0.53)	(-1.26)
%TimeSmallSpread	I1	1.218	0.891	0.735	2.472	-0.875
		(2.64)	(2.61)	(1.17)	(2.77)	(-4.63)
	I2	1.188	0.891	0.966	1.375	-0.660
		(2.77)	(2.73)	(1.79)	(1.39)	(-3.37)
	I3	1.042	0.794	0.818	0.539	-0.438
		(2.60)	(2.68)	(2.48)	(0.72)	(-2.72)

Figure 1

Correlations of HFT Strategies: Total versus Directional

In this figure, we compare the correlations of HFT strategies for total versus directional measures. We use data from the Investment Industry Regulatory Organization of Canada (IIROC) to identify 31 HFT firms. These firms are further categorized into three subgroups according to market share—MS1 (market share > 4%), MS2 (market share of between 1% and 4%), and MS3 (the rest). We compare the magnitudes of the cross-sectional or market-wide correlations in Panel A, and the time-series or stock-specific correlations in Panel B for three measures of total HFT activity (defined as buy plus sell orders) and directional ("Net") HFT activity (defined as buy minus sell orders): MSG (all messages HFT firms send to the market), TRD (all their trades), and LMT (submissions and cancellations of non-marketable limit orders). The market-wide correlations indicate whether the strategies of HFT firms are correlated across stocks in a particular time interval. For each one-second time interval, we compute the correlation coefficient between the activities of two HFT firms across the stocks in the sample, and average the correlations for all pairs of firms. The stock-specific correlations indicate whether the strategies of HFT firms are correlated over time for a particular stock. For each stock, we compute the correlation coefficient between HFT activities of any two HFT firms, and average these across all pairs of firms.

Panel A: Market-wide Correlations



Panel B: Stock-specific Correlations

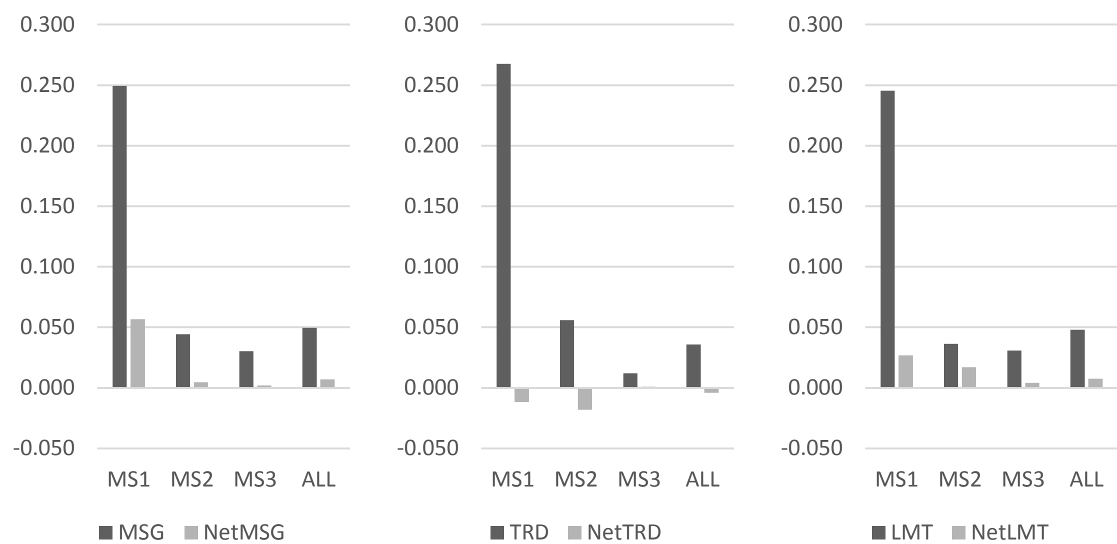
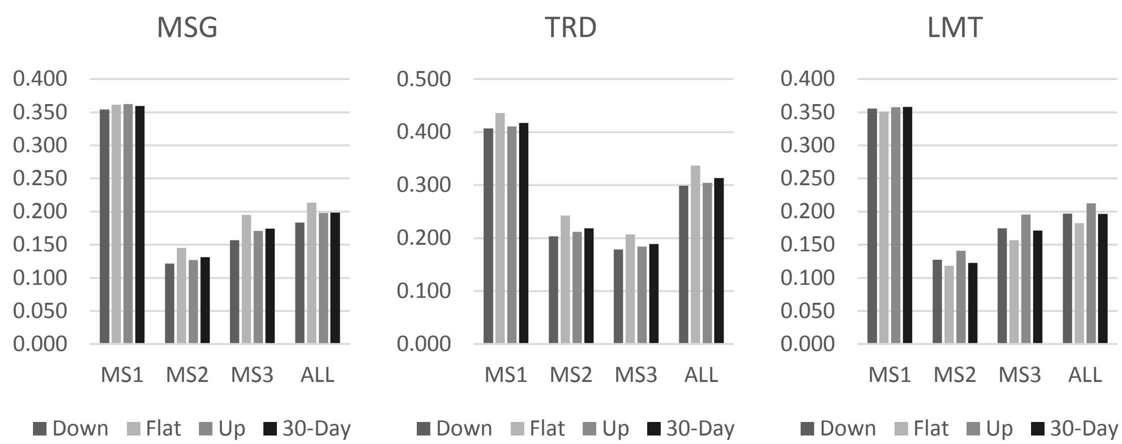


Figure 2
Correlations of HFT Strategies and Market Conditions

This figure compares the correlations of HFT strategies for varying market conditions. We use data from the Investment Industry Regulatory Organization of Canada (IIROC) to identify 31 HFT firms. These firms are further categorized into three subgroups according to market share—MS1 (market share > 4%), MS2 (market share of between 1% and 4%), and MS3 (the rest). We rank the daily returns of the S&P/TSX Composite Index from June 2010 through March 2011, and select the 10 worst days (down days), the 10 best days (up days), and the 10 days closest to and centered on zero return (flat days), for a total of 30 days (sample period). We compare the magnitudes of the cross-sectional or market-wide correlations in Panel A, and the time-series or stock-specific correlations in Panel B for down days, flat days, and up days. For each period we examine the correlations of HFT strategies in terms of the number of "messages" (MSG) they send to the market, where messages are defined as submissions and cancellations of nonmarketable limit orders as well as execution of marketable limit orders. The market-wide correlations indicate whether the strategies of HFT firms are correlated across stocks in a particular time interval. In each one-second time interval, we compute the correlation coefficient between the activities of two HFT firms across the stocks in the sample, and average the correlations for all pairs of firms. The stock-specific correlations indicate whether the strategies of HFT firms are correlated over time for a particular stock. For each stock, we compute the correlation coefficient between HFT activities of any two HFT firms, and average these across all pairs of firms.

Panel A: Market-wide Correlations



Panel B: Stock-specific Correlations