Exploratory Trading

Adam Clark-Joseph*

October 13, 2011

Abstract

Empirical research suggests that high-frequency traders (HFTs) tend to be better informed in some respects than their lower-frequency counterparts, but the precise connection between rapid trading and superior information remains unclear. Anecdotal accounts suggest that at least some HFTs use their own trades to gather information, but such "exploratory trading" is poorly understood. In this paper, I formalize the intuitive concept of exploratory trading in a simple model, and I show that exploratory trading and high-frequency trading bear a natural relationship to one another. My model sheds light on the broad question of how HFTs could parlay their speed into valuable information, and it provides explanations for a variety of empirical findings about high-frequency trading. In addition, the exploratory trading model generates a number of new testable predictions about HFT activity.

*Harvard University, E-mail: adjoseph@fas.harvard.edu. I thank Andrei Kirilenko and other seminar participants at the Commodity Futures Trading Commission, as well as seminar participants at Harvard University for their useful feedback, and I thank John Campbell, Andrei Shleifer, Alp Simsek and Jeremy Stein for their invaluable advice and comments. I gratefully acknowledge the support of an NSF Graduate Research Fellowship.

1 Introduction

1.1 High Frequency Trading

Over the past few decades, information technology has permeated and reshaped financial markets. Several recent papers, such as Hendershott and Riordan (2009) [20] and Hendershott *et al.* (2011) [19] document the prevalence algorithmic trading in modern (electronic) financial markets. The research of Hendershott *et al.* provides compelling evidence that general algorithmic trading tends to improve liquidity and aid the price-discovery process. However, algorithmic trading can potentially be used myriad ways, some of which are socially desirable, and some of which may be deleterious to the public good.

Although the research of Hendershott *et al.* suggests that the positive effects of algorithmic trading outweigh the negative effects in aggregate, this leaves open the possibility that some subset of algorithmic traders is doing something harmful. The particular subset of algorithmic traders that has attracted the greatest scrutiny in this regard are the so-called "high-frequency traders" ("HFTs").

1.1.1 Defining High-Frequency Trading

The term "high-frequency trader" lacks a precise definition, but the SEC [10] provides a fairly standard characterization. Following the SEC's terminology, HFTs are proprietary trading entities[1] that generate a large number of trades on a daily basis and are typically attributed the following characteristics:

> "(1) the use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders;
>
> (2) use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies;
>
> (3) very short time-frames for establishing and liquidating positions;
>
> (4) the submission of numerous orders that are cancelled shortly after submission; and
>
> (5) ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions over-night [when markets are closed])" [10, pp. 45-46]

Hasbrouck and Saar (2011) characterize HFTs as a subset of proprietary traders who implement "low-latency strategies"—strategies that respond to market events on a millisecond timescale. Although not all low-latency strategies generate large numbers of trades[2], the specific notion of speed that defines low-latency strategies is a quintessential element high-frequency trading.

[1] I use the term "entities," because HFTs may be organized in a variety of ways. Although some HFTs are organized as firms, others may be proprietary trading desks at some larger organization, and still others might be organized as hedge funds. Cf. the SEC's "Concept release on equity market structure," Concept Release No. 34-61358[10], page 45.

[2] There are many ways to respond to a market event, such as revising a resting limit order, that do not necessarily entail trading.

1.1.2 Empirical Results

High-frequency trading is notoriously difficult to study empirically, because data with sufficient temporal resolution are typically anonymous. Although we can use 13-F forms to track the behavior of institutional investors at a quarterly frequency, there is no general, simple way to track the behavior of a HFT at second- or millisecond-frequency. Nevertheless, there are at least three datasets[3] in which HFT-activity can be distinguished from non-HFT-activity, and analyses of these datasets have established some foundational empirical results about HFTs.

Perhaps the most striking and consistent finding in this empirical literature is that an enormous fraction of trading volume can be attributed to HFTs. Estimates vary, but even the most conservative estimates suggest that HFTs account for more than 30% of total trading volume in U.S. equities, and most estimates fall in the $50\% - 70\%$ range (see [10, 3]). These trading volume figures suggest that HFT activity constitutes an important facet of modern financial markets, which naturally raises the question of exactly how HFTs affect markets.

Three recent studies—Hasbrouck and Saar (2011), Brogaard (2010), and Kirilenko *et al.* (2010)—all examine the same general question of how HFT activity affects markets, but each study takes a very different approach to answering this question. Hasbrouck and Saar study order-level NASDAQ data, and they develop statistical techniques to identify "strategic runs" of orders that they attribute to HFTs. In a more direct approach, Brogaard uses novel dataset of order-level data for 120 stocks (60 listed on NASDAQ, 60 listed on the NYSE); this dataset distinguishes messages from 26 firms that had been identified by NASDAQ as engaging primarily in high frequency trading. Finally, Kirilenko *et al.* analyze audit-trail, transaction-level data for the E-mini S&P 500 stock index futures market, from May 3 to May 6, 2010. Kirilenko *et al.* use this data to sort over 15,000 trading accounts into six categories on the basis of realized trading behavior (one category consisted of HFTs, and the remaining categories consisted of different types of non-HFTs[4]).

Both Brogaard (2010) and Hasbrouck and Saar (2011) reach similar conclusions about the effects of HFT activity. Hasbrouck and Saar conclude that increased HFT activity tends to improve traditional measures of market quality such as short-term volatility and spreads. Brogaard likewise finds evidence that HFTs may dampen intra-day volatility, and that HFTs frequently provide inside quotes. Although Brogaard's results suggest that HFTs supply less additional book-depth than their inside-quote provision would typically imply, he finds no evidence that HFTs flee the market in volatile times. Furthermore, Brogaard finds that HFTs contribute significantly to the price-discovery process. Both

[3]Specifically, the datasets are 1) the computerized trade reconstruction data from the Commodity Futures Trading Commission (CFTC) that Kirilenko *et al.* (2010) [24] use to analyze the Flash Crash, 2) Order-level data for 120 stocks (60 listed on NASDAQ, 60 listed on the NYSE) which flags the orders from 26 firms known by NASDAQ to primarily engage in high-frequency strategies, and 3) Transaction data from the Deutsche Boerse that identifies whether or not each trade's buyer and seller were participants in the "Automated Trading Program"—i.e., whether the orders were generated by an algorithm. See [24, 11] for details on (1), [3] for details on (2), and [14, 20] for details on (3).

[4]Kirilenko *et al.* define "high-frequency trader" much more precisely than do either Hasbrouck and Saar or Brogaard. The accounts that Kirilenko *et al.* classify as HFTs are archetypes of the SEC characterization, but they are not necessarily close analogues of the "HFTs" that the other papers consider.

studies suggest that in aggregate, HFT activity tends to have a positive influence on markets.

Whereas Brogaard (2010) and Hasbrouck and Saar (2011) examine HFT activity across a large number of assets, over relatively long time spans, Kirilenko *et al.* focus on a single asset during a short time span, because their primary interest is the role (if any) that HFTs played in the Flash Crash of May 6, 2010. Kirilenko *et al.* conclude that HFT activity did not actually trigger the Flash Crash, but HFTs *did* exacerbate market volatility through their response to the unusually large selling pressure that day. These conclusions are not incompatible with the broader findings of the other two studies, and some of the more detailed results agree extremely well. For example, Brogaard's analysis suggests that rather than fleeing the market in volatile times, HFTs actually increase their participation somewhat; Kirilenko *et al.* find that HFT trading volume increased dramatically in both absolute and relative terms during short interval on May 6 when the largest price changes occurred. Nevertheless, the discrepancy between the rosy conclusions of the two more general studies and the darker conclusions of the more specific study highlight the need for more thorough and detailed understanding HFT activities.

1.1.3 What HFTs are Actually Doing

Crucial to understanding high-frequency trading is the general question of how speed relates to information. An important point of agreement across empirical studies is that HFTs appear to possess some sort of informational advantage over other market participants. For example, Brogaard finds strong evidence that HFTs avoid providing liquidity to informed traders, while Kirilenko *et al.* find that HFTs tend to trade in the same direction as contemporaneous price changes—i.e., they can anticipate future prices. The popular notion that "HFTs are better informed because they are fast" begs the question of why speed should lead to superior information.

The standard explanation of the relationship between speed and information is that HFTs are able to react to new public information faster than other market participants, so that new public information briefly serves as private information for the HFTs [9, 21, 22, 2, 7]. However, as Hasbrouck and Saar (2011) establish in considerable detail, "at horizons of extreme brevity, however, there is simply not sufficient time for an agent to be reacting to anything except very local market information" [18, pp. 8]. Conceptually, the mechanism underlying the standard explanation amounts to picking off stale quotes. Mcinish and Upson (2011) estimate that HFTs' direct gains from exploiting this kind of differential in U.S. markets constitutes around 10% of their total annual profits; this figure is not negligible, but it highlights the inadequacy of the standard explanation of the relationship between speed and information.

Empirically, Brogaard finds that the aggregate trading of the 26 HFTs in his sample, viewed as the behavior of a single "representative HFT," is consistent with a strategy based on order-imbalance-driven price-reversals. Similarly, Kirilenko *et al.* conclude that the HFTs in their sample exhibit trading patterns that are consistent with some notion of market-making. However, this general resemblance to traditional market-making does not resolve the mystery of exactly what HFTs are doing. First, various HFTs are believed to employ a variety of different strategies[5], so these aggregate results potentially occlude important heterogeneity. More importantly, although traditional, formally-registered market-makers typically have unique information about order-flow, almost no HFTs are formally registered as market-makers [10]. Even if each HFT behaves exactly like a traditional market-maker, the puzzle of how they use speed as a replacement for privileged order-flow information would remain.

Many of the general techniques/strategies disclosed by and imputed[6] to HFTs are standard elements of the non-high-frequency realm. These include arbitrage, statistical arbitrage, directional bets, and market-making. However, the high-frequency activities commonly classified as "liquidity detection" lack low-frequency analogues. This classification covers a number of slightly different techniques— "pinging," "sniffing," etc.—but all of these techniques fundamentally entail the use of orders for the express purpose of gathering information about the market. Obviously, such techniques point to a mechanism by which HFTs might obtain superior information. Less obviously, but much more importantly, this mechanism illuminates the connection between information and trading speed.

1.2 Trade-Revealed Information

I ultimately seek to address how trading speed could translate to superior information. As a first step, I will examine the idea of "exploratory trading"—how an agent might use his own trades to gather valuable information. A central objective of my analysis is to establish that exploratory trading relates naturally and intimately to high-frequency trading. Although exploratory trading may arise in a number of contexts, I frame my model in a potentially high-frequency setting, and focus my analysis on the implications of exploratory trading for understanding high-frequency trading.

The idea that the trading process can reveal new information about an asset dates back at least to Romer (1993). Romer proposed this idea to explain why efficient asset prices could change dramatically even in the absence of external news. Subsequent research extended Romer's basic idea to address a variety of other aspects of asset-price dynamics. At a very general level, this literature explores how the information revealed by some set of initial trades affects the subsequent trades of rational agents. In this paper, I consider a converse issue, namely how optimal initial trades depend on the amount of information that those trades are expected to reveal.

An agent who trades an asset ultimately cares about the prices at which she actually executes trades. Conceptually, we can decompose the price at which an agent executes a trade into a component that depends on the agent's trade, and a component that does not depend on her trade (I will refer to this latter component as the "asset-intrinsic" component). More concretely, we might decompose a realized price into the price impact of the agent's actual trade, and the price that would have prevailed if the agent hadn't

[5]Brogaard notes that some of the 26 HFTs in his sample primarily take liquidity, while others primarily provide liquidity. Although Brogaard cannot observe this heterogeneity directly in his data, the contacts at NASDAQ who constructed the dataset conveyed this important fact. Publications by both the United States Securities Exchange Commission and the Australian Securities and Investments Commission also indicate considerable diversity in HFT strategies. [10, 30]

[6]Cf. publications of various market regulatory bodies, [10, 30]

traded.

In principle, the trading process could reveal information about either the future asset-intrinsic component of price, or the future price-impact of trades. However, essentially all of research in the spirit of Romer (1993) focuses on the question of how past trading can reveal information about the asset-intrinsic component of future prices. Although the revelation of information about price impact is a topic of both practical and theoretical importance, the mechanism by which past trading reveals information about future price impact is relatively transparent. The sole paper on this topic, Hong and Rady (2002), covers this basic mechanism fairly comprehensively.

The theoretical question that I address in the present paper—how optimal initial trades depend on the amount of information that they are expected to reveal—is strictly more complicated than the analogous, converse, Romer (1993)-style questions. In the simplest, two-period setting, the Romer (1993)-style question must be addressed for period 2, but then to answer my style of question, we must determine how optimal period-2 trading profits depend on the amount of information that period-1 trade reveals. Next, we must role back to period 1, solve a highly non-standard inference problem to find the relationship between the period-1 trade and the amount of information the trade is expected to reveal, use this to express the expected conditionally optimal period-2 trading profits in terms of the period-1 trade, and then combine this with the expected direct trading profits from the period-1 trade, and finally maximize the expected total profit with respect to the period-1 trade. The very simplicity that makes learning about price-impact unattractive for studying Romer (1993)-style issues makes "learning about price impact" an ideal setting to develop a baseline model of exploratory trading.

Hong and Rady (2002) briefly discuss the topic that I address:

> "In [a modification of our] set-up, the first-period action impacts on second-period profit because it affects the informed investors' second-period belief about liquidity ... Informed traders may have an incentive to experiment and sacrifice first-period trading profit for more precise information about liquidity, which may improve the profitability of their second-period trades. Unfortunately, even this simple set-up cannot be solved in closed form." (pp. 427)

By using a slight variation on the Hong and Rady (2002) model (which is in turn a slightly modified version of the Kyle (1985) model), I derive a closed-form solution for optimal exploratory trading, I formalize Hong and Rady's conjecture that informed traders may have an incentive to engage in costly experiments in period 1 to increase their expected profits in period 2, and I establish conditions under which the conjecture holds.

With the inner workings of exploratory trading laid bare, the connections between exploratory and high-frequency trading follow easily. These connections shed light on the general question of how trading speed relates to superior information, and they also suggest explanations for a variety of existing empirical findings, as well as generating a number of testable predictions.

2 A Model of Exploratory Trading

For purposes of tractability and expositional clarity, I initially consider exploratory trading in the context of learning about price-impact. Although this is not necessarily the most interesting application, the formal and conceptual results that I derive about exploratory trading in this context extend to more general settings.

2.1 The Baseline Model

Let time be discrete, consisting of periods $t = 1, 2$. Consider the market for a single asset. At the end of period 2, the asset pays a fixed terminal liquidation dividend of μ and the world ends.

The net supply of the asset in period 1 is given by the random variable $-s_1$, which has zero mean and finite variance $\sigma_s^2 > 0$. For simplicity, assume that the net supply of the asset in period 2 is exactly zero.

The respective market demand curves for the asset in periods 1 and 2 are

$$y_1 = \alpha - \beta^{-1} p_1 \tag{1}$$
$$y_2 = \alpha - \beta^{-1} p_2 \tag{2}$$

where p_t denotes the price in period t[7]. The parameter β is drawn from a distribution with strictly positive support, bounded away from zero, with mean b and finite variance $\sigma_\beta^2 > 0$; intuitively, β represents marginal price-impact of a market order. The parameter α depends on β and follows a distribution such that

$$\alpha \equiv \frac{\mu + \xi}{\beta} \tag{3}$$

for some zero-mean random variable $\xi \in \mathcal{L}^1$ that is independent of β and has finite variance $\sigma_\xi^2 > 0$. The product $\alpha\beta$, which I will denote by $p_0 \equiv \alpha\beta$, corresponds to the price at which $y_t = \mathbb{E}[s_t]$. Equation (3) implies that $p_0 \equiv \mu + \xi$, and the distributional assumptions on ξ imply that $\mathbb{E}[p_0|\beta] = \mathbb{E}[p_0] = \mu$.

Consider a single agent—call him the "high-frequency trader," or "HFT"—who submits market orders[8] in periods 1 and 2 with the objective of maximizing his expected aggregate net profits. Denote by x_t the number shares that the HFT purchases in period t. Assume

[7]From a theoretical standpoint, it might be nicer to have a second-period demand curve of the form $y_2 = \alpha - \beta^{-1} p_2 - y_1$, but this would make the algebra much messier. To the extent that the interesting features of this model revolve around the estimation of β, the algebraically simpler case still delivers the same intuitive conclusions.

[8]Nothing fundamental would change if the HFT used limit orders instead of market orders. If we assume both that there is no supply noise in period 2, *and* that the HFT knows $\mu - p_0$ perfectly, then allowing the HFT to use limit orders would lead to a trivial solution wherein the HFT offers to purchase/sell an unlimited quantity at the price $p_0 + \frac{\mu - p_0}{2}$. However, as long as the HFT is uncertain about either the net supply in period 2, or the true value of $\mu - p_0$, allowing the HFT to submit limit orders would not produce any pathologies. The HFT's optimization problem would be less intuitive and far less tractable, but the same basic concepts would apply.

that the HFT knows μ and the product $p_0 \equiv \alpha\beta$[9], but not α or β individually. The HFT observes his own trades, as well as the market-clearing price in each period, but he observes neither the market demand curves, nor the quantity s_1. When the HFT observes p_1 he can use his knowledge of x_1 and p_0 to estimate β, but his uncertainty about s_1 prevents him from learning β exactly.

2.2 How Chronological Time Enters the Model

Although the model is set in abstract discrete time, true chronological time enters the model in an important, albeit implicit, way. The HFT's inference problem basically amounts to determining what portion of a price change was caused by his own trade, and what portion was driven by something else. Random supply noise is a simple "something else," but many other factors could plausibly confound the relationship between the HFT's trade during some time interval and the price change in that interval. However, the prevalence and comparative importance of such confounding factors must tend to decrease as time intervals shorten[10].

As a concrete example, suppose that I sell 10,000 shares of AAPL sometime in October. If I compare Apple's opening price on the first trading day in October with its opening price on the last trading day on October, this would not tell me very much about market-depth around the time of my trade. By contrast, if I compared the price of AAPL 10 milliseconds before and after I sold 10,000 shares, I would probably learn something meaningful about market-depth around the time of my trade.

The role of true, chronological time in the model will ultimately illuminate the relationship between exploratory and high-frequency trading. I address this deeper, more intrinsically interesting issue after solving the baseline model. The formal framework that emerges from solving the baseline model facilitates more precise discussion this central issue.

[9] The HFT's trade at date 1 affects the quality of his information about the slope of the market demand curve depends on his period-1 trades, the quality of his information about the intercept does not. Uncertainty about the intercept does not raise any fundamentally new issues, so streamline my analysis, I will essentially assume that the intercept is known. Rather than literally assuming that the HFT knows α, it turns out to be more natural to suppose that he knows $p_0 \equiv \alpha\beta$. Whereas the lone parameter α does not have an obvious economic meaning, the product $\alpha\beta$ corresponds to the market-clearing price that would prevail in the absence of supply noise if the HFT did not participate in the market. The distributional assumptions about β and α imply that β is independent of p_0, so the HFT's knowledge of p_0 removes the extraneous uncertainty from the model without introducing additional information about β.

[10] This claim can be made rigorous. Suppose that the price change due to confounding factors during some bounded interval $\mathcal{I}$ can be represented as $\int_{\mathcal{I}} f(t)\, dt$, where $f : \mathbb{R} \mapsto \mathbb{R}$ is a measurable function that is integrable on all subsets of $\mathbb{R}$ with finite measure. Let $\mathcal{I}_1, ..., \mathcal{I}_n$ be non-overlapping sub-intervals of $\mathcal{I}$ such that $\cup_{j=1}^{n} \mathcal{I}_j = \mathcal{I}$ (i.e. let the collection of intervals $\{\mathcal{I}_j\}_{j=1}^{n}$ be a partition of $\mathcal{I}$). The average magnitude of the price-change (due to confounding factors) on each sub-interval in this collection is $\frac{1}{n}\sum_{j=1}^{n}\left|\int_{\mathcal{I}_j} f(t)\, dt\right| \leq \frac{1}{n}\sum_{j=1}^{n}\int_{\mathcal{I}_j}|f(t)|\, dt = \frac{1}{n}\int_{\mathcal{I}}|f(t)|\, dt$. By assumption, $\int_{\mathcal{I}}|f(t)|\, dt < \infty$, so $\frac{1}{n}\int_{\mathcal{I}}|f(t)|\, dt \to 0$ as $n \to \infty$. Note that if $\sup_j\left\{\int_{\mathcal{I}_j} 1dt\right\} \leq \varepsilon \int_{\mathcal{I}} 1dt$, it follows that $n \geq \frac{1}{\varepsilon}$. Thus the average magnitude of the price-change on each element of a partition $\mathcal{P}$ of the interval $\mathcal{I}$ approaches zero as the measure of the largest element of $\mathcal{P}$ approaches zero.

3 Solving the Model

Intuitively, the HFT's basic strategy entails buying (selling) the asset when the price is below (above) the terminal value μ. However, the HFT faces downward-sloping market demand curves, so the market-clearing price p_t depends on the HFT's purchase x_t. Consequently, the HFT faces a trade-off between the number of shares that he buys, and the spread $\mu - p_t$ that he earns on each share. If the HFT knows the shape of the market demand curve, then his optimization problem is isomorphic to that of a profit-maximizing monopolist facing a downward-sloping demand curve.

If the HFT does *not* know the shape of the market demand curve, his task is more complex than the standard monopolist's problem. Although the HFT still faces the quantity/spread trade-off, his ability to optimally balance this trade-off depends on the quality of his information about the market demand curve. The price impact of the HFT's trade in the first period provides information about the slope of market demand curve that the HFT can use to better choose his trade in period 2, but this information comes at the expense of trading costs in the first period.

In the remainder of this section, I derive the HFT's optimal trading strategy through backward induction.

3.1 Date $t = 2$

The HFT pays $x_1 p_1$ at date 1, and $x_2 p_2$ at date 2, then he receives a payoff of $\mu(x_1 + x_2)$ at the end of period 2, so the HFT's total realized profit is

$$\begin{aligned} \Pi^{total} &= -x_1 p_1 - x_2 p_2 + \mu(x_1 + x_2) \\ &= x_1(\mu - p_1) + x_2(\mu - p_2) \end{aligned}$$

At date 2, the HFT chooses x_2 to maximize the conditional expectation of his total profit, $\mathbb{E}_1[x_1(\mu - p_1) + x_2(\mu - p_2)]$. Since x_1 and p_1 are determined before the second period, the HFT's choice of x_2 depends only on his conditional expectation of his period-2 trading profits, $\mathbb{E}_1[x_2(\mu - p_2)]$. Thus the HFT solves

$$\begin{aligned} &\max_{x_2} \mathbb{E}_1[x_2(\mu - p_2)] \\ &\quad s.t.\ p_2 = p_0 + \beta x_2 \end{aligned}$$

The HFT's feasible optimal trade at date 2, call it $\hat{x_2}$, is given by

$$\hat{x_2} = \frac{\mu - p_0}{2\mathbb{E}_1[\beta]} \tag{4}$$

Equation (4) confirms our basic intuitions about the HFT's trading strategy; he trades against perceived price dislocations, but he accounts for the anticipated price impact of his trade.

Next, we wish to determine how the profit that the HFT earns from trading $\hat{x_2}$ depends on the quality of his information about the demand curve. Let $x_2^* \equiv \frac{\mu - p_0}{2\beta}$ denote the infeasible optimal trade that the HFT would select if he knew the true value of β, and let

$\pi_2^* \equiv \beta (x_2^*)^2$ denote the associated infeasible maximized profit. Intuitively, as $\hat{x_2}$ deviates from the infeasible optimum x_2^* , we should expect the HFT's realized profits to decline relative to π_2^*. Indeed, in the appendix I show that the true profit that the HFT would earn from the optimal feasible trade $\hat{x}_2$ can be expressed as

$$\hat{\pi_2} = \pi_2^* - \beta (\hat{x}_2 - x_2^*)^2 \tag{5}$$

The HFT's expected feasible profit at a given value of β can be expressed naturally in terms of the mean-square error of $\frac{1}{\mathbb{E}_1[\beta]}$ relative to $\frac{1}{\beta}$, and approximated in terms of the conditional variance of $\mathbb{E}_1[\beta]$:

$$\begin{aligned}
\mathbb{E}[\hat{\pi}_2|\beta, p_0] &= \pi_2^* - \frac{\beta (\mu - p_0)^2}{4}\mathbb{E}\left[\left(\frac{1}{\mathbb{E}_1[\beta]} - \frac{1}{\beta}\right)^2 |\beta\right] && (6)\\
&\approx \pi_2^* - \frac{(\mu - p_0)^2}{4\beta^3} Var(\mathbb{E}_1[\beta]|\beta) && (7)
\end{aligned}$$

Equation (6) illustrates how variability in the HFT's estimate of the market demand curve reduces the profits that the HFT earns from his optimal feasible trading strategy in period 2.

3.2 Date $t = 1$

In section (3.1), I related the HFT's optimal period-2 trading strategy and associated profits to the estimator $\mathbb{E}_1[\beta]$. In particular, I showed that the HFT's expected profits from trading in period 2 increased as $\mathbb{E}_1[\beta]$ became a better estimator of β. As I will show below, the quality of $\mathbb{E}_1[\beta]$ as an estimator of β depends on x_1, the HFT's order in period 1. Consequently, the HFT's optimal trading strategy in period 1 will depend not only on the direct revenues associated with the trade, but also on the extent to which the trade is expected to improve the HFT's information about the market demand curve in the next period.

3.2.1 Exploratory Trading at Date 1

I now consider the case in which the HFT can use the information from his period-1 trade to update his beliefs about β. The key idea is that the HFT's purchase in period 1 induces predictable variation in the market-clearing price that the HFT can exploit to better estimate the slope of the market demand curve.

The market-clearing price in the first period, p_1, can be expressed as

$$\Delta p = \beta (x_1 + s_1) \tag{8}$$

where I define $\Delta p \equiv p_1 - p_0$.

Since the single observation $(\Delta p, x_1)$ constitutes the entirety of the HFT's empirical data, β is underidentifed from the HFT's perspective, regardless of the value of x_1. However, the underidentification of β means simply that the HFT cannot *perfectly* (i.e., consistently) estimate β from a single, noisy observation. Although the traditional binary identified/underidentified classification is well-suited for asymptotic analyses, the present setting calls for finer distinctions.

No finite choice of x_1 will allow the HFT to *completely* disentangle the effects of β from those of supply noise on the basis of a single observation, but the value of x_1 determines how well the HFT can separate the effects of β from those of s_1. Intuitively, we might think of x_1 determining "how well" β is identified. We can make this intuitive notion precise by using equation (6), and considering how x_1 affects the HFT's expected period-2 profits.

The exact effects of x_1 will depend both on the distributions of β and s_1, and on the HFT's knowledge about these distributions, but the HFT will generally tend to learn more about β the larger is the magnitude of x_1. Although I cannot invoke a central limit theorem to sidestep these distributional details in the usual manner, I accomplish something similar by considering the case in which $|x_1|$ becomes large, and applying integrability/moment conditions to characterize tail behavior. We can always bound the variance of $\mathbb{E}_1[\beta]$ by $\mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2\right] \leq \frac{(\sigma_\beta^2+b^2)\sigma_s^2}{x_1^2}$, and under mild regularity conditions, this bound becomes tight as $|x_1|$ becomes large (see mathematical appendix for details). To avoid a morass of unenlightening algebra, I will appeal to this tight bound and make the simplifying assumption that $Var(\mathbb{E}_1[\beta]|\beta) \equiv \mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2|\beta\right]$ is given by

$$Var(\mathbb{E}_1[\beta]|\beta) = K^2\frac{\beta^2\sigma_s^2}{x_1^2} \tag{9}$$

for $x_1^2 \geq K^2\sigma_s^2$, where $1 \geq K > 0$ is some positive constant that depends on the unconditional distributions of β and of s_1.

We can combine (9) with the approximation (6) to characterize the relationship between x_1 and the HFT's expected period-2 profits. Taking (6) to hold exactly[11], we obtain

$$\mathbb{E}[\hat{\pi}_2|p_0] = \frac{(\mu-p_0)^2\,\mathbb{E}\left[\beta^{-1}\right]}{4}\left(1-\frac{K^2\sigma_s^2}{x_1^2}\right) \tag{10}$$

Since the HFT could always choose not to trade at all in period 2, his expected period-2 profits must be non-negative. The right-hand side of equation 10 is negative for $x_1^2 < K^2\sigma_s^2$, so the model is not applicable in that region (this is why I only assume that (9) holds for $x_1^2 \geq K^2\sigma_s^2$).

The expected direct trading profit from trading x_1 is

$$\mathbb{E}[\pi_1|p_0] = (\mu-p_0)\,x_1 - bx_1^2 \tag{11}$$

so the HFT's total expected profit over both periods is

$$\mathbb{E}[\pi_1+\hat{\pi}_2|p_0] = (\mu-p_0)\,x_1 - bx_1^2 + \frac{(\mu-p_0)^2\,\mathbb{E}\left[\beta^{-1}\right]}{4}\left(1-\frac{K^2\sigma_s^2}{x_1^2}\right) \tag{12}$$

The optimal value of x_1 depends on three factors. First, trading costs in the form price impact tend to push the optimal value of x_1 towards zero; in equation (12), trading costs

[11] Since we could adjust the positive constant K to make the approximation good, ignoring the approximation error is more innocuous in this case than it is in general.

enter through the "$-bx_1^2$" term. Second, as the HFT trades more at date 1, he obtains better information about β that he can use to trade more profitably (in expectation) at date 2; in equation (12), the "$-\frac{K^2\sigma_s^2}{x_1^2}$" term reflects this informational benefit to trading. This informational benefit pushes the optimal value of $|x_1|$ away from zero. Finally, the direct gains from trading in period 1, reflected by the "$(\mu - p_0)\,x_1$" term in equation (12), tend to push the optimal value of x_1 away from zero.

Unlike the informational gains from trading, the direct gains from trading in period 1 are completely standard and are thoroughly understood. We can isolate the informational motive for trading from the "direct gain" motive by supposing that the HFT only observes $\mu - p_0$ after he has selected x_1. Recall that I assume $\mathbb{E}\left[\mu - p_0\right] = 0$, so the HFT's initial expectation of his total profit is

$$\mathbb{E}\left[\pi_1 + \hat{\pi}_2\right] = -bx_1^2 + \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4}\left(1 - \frac{K^2\sigma_s^2}{x_1^2}\right) \tag{13}$$

Intuitively, (13) represents the profit that the HFT would expect to obtain if he had to select the value of x_1 before learning p_0. Since $\mathbb{E}\left[\mu - p_0\right] = 0$, the "$(\mu - p_0)\,x_1$" term from (12) vanishes—in expectation, there is no direct gain from a "blind trade".

In the mathematical appendix, I show that the x_1^* that maximizes the HFT's unconditional expected profit, (13), is characterized by

$$(x_1^*)^2 = \begin{cases} \frac{1}{2}\sigma_\xi\sigma_s K\sqrt{\frac{\mathbb{E}\left[\beta^{-1}\right]}{b}} & if\ \sigma_s^2 < \frac{\mathbb{E}\left[\beta^{-1}\right]}{16K^2 b}\sigma_\xi^2 \\ 0 & if\ \sigma_s^2 \geq \frac{\mathbb{E}\left[\beta^{-1}\right]}{16K^2 b}\sigma_\xi^2 \end{cases} \tag{14}$$

The condition $\sigma_s^2 < \frac{\mathbb{E}\left[\beta^{-1}\right]}{16K^2 b}\sigma_\xi^2$ ensures that the maximized expected profit is non-negative (if the maximized expected profit is negative, HFT would simply not participate in the market). Also, the condition "$\sigma_s^2 < \frac{\mathbb{E}\left[\beta^{-1}\right]}{16K^2 b}\sigma_\xi^2$" implies $(x_1^*)^2 > 2K^2\sigma_s^2 > K^2\sigma_s^2$, and therefore guarantees that we are in a valid region of our model. Since I have removed the "direct gain" motive for trading in period 1, the optimal trade characterized in equation (14) is driven entirely by information-seeking concerns.

To rephrase this more bluntly, equation (14) validates the concept of exploratory trading by explicitly characterizing the phenomenon in a simple model. However, the concept of exploratory trading is merely a tool to better unravel the mysteries of high-frequency trading. Although equation (14) marks the end of my analysis of exploratory trading in the abstract, it also marks the beginning of my analysis of the connection between exploratory and high-frequency trading.

4 Exploratory and High-Frequency Trading

In section 2.2, I showed how chronological time implicitly appeared in the baseline model through the distribution of supply noise, s_1. The discussion in 2.2 introduced the idea that the HFT wants to determine the causal effect of his trade on prices, but that confounding factors such as supply noise hinder his inference. The main point of section 2.2 is that the "importance" (in some sense) of these confounding factors depends on the clock-time duration of the interval in which trade occurs. This point remains valid, but we can now state it more precisely. As equation (10) reveals, the HFT's expected period-2 profit depends on σ_s^2, the variance of period-1 supply noise. The relevant measure of the "importance" of period-1 supply noise is simply the variance σ_s^2.

The link between the parameter σ_s^2 and the implicit duration of period 1 makes it possible to investigate chronological-time considerations elsewhere in the model. In particular, we can use the results from section 3.2 to examine the relationship between exploratory trading and chronological time in detail, which will in turn reveal the connections between exploratory trading and high-frequency trading.

4.1 Speed is Necessary for Exploratory Trading

Equation (14) (reprinted below for convenience) suggests a simple but natural connection between exploratory and high-frequency trading.

$$(x_1^*)^2 = \begin{cases} \frac{1}{2}\sigma_\xi \sigma_s K \sqrt{\frac{\mathbb{E}[\beta^{-1}]}{b}} & if\ \sigma_s^2 < \frac{\mathbb{E}[\beta^{-1}]}{16K^2 b}\sigma_\xi^2 \\ 0 & if\ \sigma_s^2 \geq \frac{\mathbb{E}[\beta^{-1}]}{16K^2 b}\sigma_\xi^2 \end{cases}$$

When the variance of supply noise exceeds some threshold ($\sigma_s^2 \geq \frac{\mathbb{E}[\beta^{-1}]}{16K^2 b}\sigma_\xi^2$), the optimal level of exploratory trading drops to zero ($(x_1^*)^2 = 0$). Exploratory trading only arises in the model when the variance of supply noise is sufficiently small, or equivalently when the duration of the first trading period is sufficiently short.

Although a trade reveals some amount of valuable information, trading is also costly (due to price impact). Both the informational gain and the price-impact cost of a trade depend on the trade's magnitude, but only the informational gain depends on the variance of supply noise. As the variance of supply noise increases, the informational gain from a given magnitude of trade decreases. Up to a point, a trader can partially offset this reduction in the informational gain by increasing the magnitude of his trade, but this also increases his trading costs. Eventually, when σ_s^2 becomes sufficiently large, the informational gains become too small to justify the cost of any non-zero level of exploratory trade.

The relationship between exploratory trading and high-frequency trading can also be understood in terms of equation (13). As noted earlier the "$-\frac{K^2\sigma_s^2}{x_1^2}$" term reflects the informational gain from trading in period 1, expressed in terms of the HFT's expected total profit. There are two ways to make $\frac{\sigma_s^2}{x_1^2}$ small: make x_1^2 large, or make σ_s^2 small. Up to this point, we have treated σ_s^2 as fixed, and considered the optimal choice of x_1^2. However, to the extent that σ_s^2 depends of trading speed, there is another dimension along which optimization is possible.

This raises two important points. First, the costs associated with reducing σ_s^2 by increasing trading speed are largely fixed; these include direct data feeds, colocation services, some proprietary software development, and so on. These fixed costs of increasing speed are considerable, and the cost reduction per trade would likely be small, increased trading speed would be most valuable (from the standpoint of exploratory trading) for a trader who intended to engage in a large number of trades. Hence the clear connection between exploratory trade and low-latency trading also suggests a similar connection between exploratory trade and high-frequency trading *per se*.

The second important point that arises from the two possible approaches to making $\frac{\sigma_s^2}{x_1^2}$ small is that in spite of the potential value associated with superior trading speed, this does not necessarily imply that we should observe a Bertrand-competition-style latency arms race (or at least not one driven by exploratory trading). An HFT could potentially overcome some minute latency disadvantage by slightly increasing the magnitude of his exploratory trades.

4.2 Speed is Basically Sufficient for Exploratory Trading

If the HFT trades the quantity $x_1 = \varpi$, the expected informational gain from this trade depends on $|\varpi|$, but the reasons why the HFT selected the quantity $x_1 = \varpi$ are completely irrelevant. Earlier, in order to establish and clarify the theoretical notion of exploratory trading, I isolated the informational motive for trading from the standard "direct gain" motive. The optimal exploratory trade that I characterize in equation (14) and analyze in section 4.1 is the theoretically pure variety, which excludes any "direct gain" component. In this subsection, I consider the potential informational value of trading in a more general context.

An important preliminary result is that in the baseline model, the HFT's information-seeking motives for trading will never conflict with his direct-gain motives. Since trading costs are directly proportional to $-x_1^2$, and the informational benefits of trading are inversely proportional to $-x_1^2$, these two factors can uniquely determine the optimal *magnitude* of x_1^*, but they do not determine the optimal sign of x_1^*—i.e., whether it is optimal to sell $|x_1^*|$ shares, or to buy $|x_1^*|$ shares. If the HFT knew the sign of $\mu - p_0$, he could obtain direct gains from his first-period trade by choosing x_1^* to have the same sign. Given this choice of sign, both the informational benefit factor and the direct gain factor push the magnitude of x_1^* in the same direction. If the HFT considers both direct gains and informational gains when he selects x_1, all of our earlier results remain qualitatively unchanged[12].

Even if the HFT doesn't consider informational gains when he selects x_1, he can still extract information from the result of his trade. The trade generates some information, regardless of whether the HFT chose x_1 optimally or arbitrarily. Furthermore, the informational gain associated with x_1 is still proportional to $\frac{\sigma_s^2}{x_1^2}$, so increased trading speed still translates to increased informational gains, exactly as discussed in section 4.1. The key implication of the preceding results that novel information is a natural by-product of high-frequency trades. Regardless of whether a HFT actually chooses his trades with the intent of gathering information, those trades still generate non-public, non-negligible information.

[12]However, if we introduce direct gains from period-1 trading, the HFT's optimization problem becomes much less tractable, because the quartic equation that arises in the first-order condition no longer has a simple biquadratic form.

5 Discussion

5.1 Relation to Empirical Findings

The preceding analysis of exploratory trading sheds light on a variety of empirical results concerning HFT activity.

5.1.1 Cross-sectional Variation of HFT Participation in Stocks

Brogaard (2010) finds that fraction of total trading activity for which HFTs are responsible varies systematically across stocks. In particular, Brogaard finds that HFTs' relative fraction of market activity tends to be greatest in large market-cap stocks, and stocks with greater market depth. To the extent that we can characterize such stocks as those for which the price-impact of trading is small, the exploratory trading model helps to explain Brogaard's finding.

As equation (14) reveals, the optimal magnitude of exploratory trading increases as the expected price-impact parameter ($b \equiv \mathbb{E}[\beta]$) decreases. In markets where the expected price-impact is small, the expected cost of exploratory trading is also small, so the optimal magnitude of exploratory trading is large, and HFTs obtain a greater amount of information. Thus in addition to the direct increase in HFT activity associated with exploratory trading, the improved information might also encourage HFT participation. Consistent with such "improved information," Brogaard finds that HFTs are best able to avoid providing liquidity to informed traders in large-cap stocks.

5.1.2 Avoiding Informed Traders and Anticipating Price Changes

In absolute terms, Brogaard finds that HFTs avoid providing liquidity to informed traders more successfully than do non-HFTs. As noted earlier, this finding suggests that HFTs tend to know some things that non-HFTs do not. The baseline exploratory trading model is not directly applicable to this setting, but conceptually simple extension of the baseline model offers a natural explanation.

In section I of his (1993) paper, Romer presents a model in which (essentially) agents know the precision of their own information about an asset, but they are unsure whether their is more precise or less precise than that of other agents. Following some unexpected but observable exogenous supply shock, agents observe the slope of the market demand curve from which they can infer the relative precision of their signals. The problem of determining whether other agents' information is more or less precise than your own is basically isomorphic to the problem of detecting informed traders, and Romer establishes a framework in which the slope of the demand curve provides the information necessary to solve this problem. The baseline exploratory trading model directly addresses the issue of estimating the slope of the market demand curve. A natural explanation of how trading speed would help HFTs detect informed traders requires little more[13] than appending Romer's model to the baseline exploratory trading model.

[13] There are a few uninteresting but potentially tedious technical details that arise in a strictly rigorous concatenation of the two models, and I am actively working out these minutiae, but the intuition is straightforward.

The "exploratory trading + Romer" extension also helps to explain the finding of Kirilenko *et al.* that HFTs appear to profitably trade in the same direction as contemporaneous price changes (i.e., anticipate price changes). When agents deduce the relative precision of their signal, they can use this information to infer whether the current market-price properly reflects aggregate information that is appropriately weighted by its precision, and thereby potentially uncover temporary mispricings.

5.1.3 Fleeting Orders

In their 2009 paper, "Technology and liquidity provision: The blurring of traditional definitions," Hasbrouck and Saar (henceforth "HS") investigate what they term "fleeting limit orders," that is, limit orders that are cancelled within two seconds of being placed. Among the explanations of the existence of fleeting limit orders that HS consider is the hypothesis that "fleeting orders are a byproduct of a strategy meant to 'search' for latent liquidity...the search hypothesis implies that fleeting orders are intended to demand, rather than supply, liquidity" (pp.154). HS find a variety of empirical evidence that supports their "search hypothesis."

Notwithstanding the limit order vs. market order difference, rapidly submitting liquidity-demanding demanding orders for the purpose of learning about latent (i.e., hidden) liquidity bears an obvious resemblance to engaging in exploratory trading for the purpose of estimating the current slope of the market demand curve. While HS's results do not definitively establish the empirical importance of exploratory trading of precisely the form I analyze in the present paper, their results clearly support the empirical validity of the general exploratory trading concept.

In the opposite direction, the theory of exploratory trading that I develop in this paper may prove useful for developing structural models of exactly how fleeting orders arise from various trading strategies.

5.1.4 Competition Among HFTs

An interesting stylized fact about the HFT industry is that a relatively small number of HFTs account for a vast majority of HFT activity. For example, Kirilenko *et al.* characterize only 16 of the 15,000+ accounts in their sample as HFTs. Likewise, only 26 HFTs are identified in Brogaard's sample. Although Brogaard notes that these group of 26 excludes some entities that are often classified as HFTs, the 26 identified HFTs are involved in 68.5% of the total dollar volume in Brogaard's sample. All remotely plausible estimates of total the total U.S. dollar volume in which HFTs are involved are below 80%, so omissions from the 26 HFTs identified in Brogaard's data are unlikely to be important in the particular sample that Brogaard studies[14].

While there are some fixed costs to starting a HFT firm, (colocation, direct data-feeds, etc.), these costs are on the order of a few hundred thousand dollars annually. Furthermore, HFTs require relatively little operating capital—Brogaard estimates that the 26 HFTs in his sample, together, require around $117 million to conduct their trades on the 120 stocks.

[14]This does not necessarily mean that HFT entities excluded from Brogaard's 26 are unimportant for other stocks/assets, or even for the stocks that Brogaard considers at times not covered in his sample.

By comparison, Brogaard estimates the annual profit of the 26 HFTs from trading the 120 stocks is around \$74 million[15].

Once again, the baseline exploratory trading model provides some insight. First, the result that all high-frequency trades generate some form of potentially valuable information suggests that there are natural economies of scale for HFTs. Roughly speaking, a HFT who trades more gets better information.

We can take this analysis still further by noting that if there were more than one HFT in the exploratory trading model, then one HFT's trades would look like supply noise to the other HFTs (and vice versa). If there is one incumbent HFT, then his trades would look like noise to a potential entrant, and at least reduce the incumbent's prospective profits. As the number of HFTs in a given market increases, the apparent supply noise from the perspective of a potential entrant also increases. In other words, high-frequency trading inherently imposes barriers to entry. An interesting possibility that arises from this conclusion is that HFTs might engage in excessive trading for purely anti-competitive purposes[16].

5.2 Testable Predictions

The baseline exploratory trading model generates a number of testable predictions.

Although the exploratory trading model provides a possible explanation for the systematic cross-sectional variation in HFT participation, it also makes much crisper predictions about how HFT profits should depend on model parameters. In particular, in the cross-section, HFT profits should decrease in σ_s^2, decrease in $\mathbb{E}[\beta]$, and increase in σ_ξ^2. HFT profits should also tend to be greatest in stocks with higher relative levels of undisplayed liquidity. All of these cross-sectional predictions above have direct time-series analogues. A particularly interesting issues is how changes in σ_s^2 over time affect HFT profits. While distinguishing σ_β^2 from σ_s^2 may require some subtlety, HFT behavior in response to the two types of change would help to isolate the importance/value of the information that HFTs may derive from their own trades.

The audit-level CTR data that Kirilenko *et al.* analyze would be particularly useful for empirical analyses of these predictions.

5.3 Discussion of Modeling Assumptions

Three assumptions of the baseline exploratory trading model merit some discussion.

5.3.1 Exogenous Private Information

For purposes of tractability, the baseline exploratory trading model in this paper analyzes the optimal manner in which an agent can use his own trades to learn about the price-impact of his trades. To motivate the value of such price-impact information, I consider

[15] Note that these figures (\$117 million and \$74 million) correspond to trading in the specific stocks in Brogaard's sample. Brogaard extrapolates that the required capital and profits for the entire U.S. equities market are \$4.68 billion and \$2.8 billion, respectively.

[16] I thank John Campbell for pointing out this intriguing issue.

an agent who has some private knowledge about future prices. In terms of the model, I examine how the HFT can best make use of his knowledge of $\mu - p_0$ during periods 1 and 2, but I take the HFT's knowledge of $\mu - p_0$ to be exogenously given. *I assume that the HFT exogenously knows $\mu - p_0$ purely to simplify the exposition, and to focus on the novel aspect of my model.* In other words, *this assumption is not crucial*!

In earlier drafts of this paper, the HFT inferred both the intercept and the slope, but inference about the intercept turns out to be a standard type of problem, and it complicates the more interesting and unusual inference about the slope. Another, potentially more attractive way to remove the "exogenous private information" assumption is to tack a slight variation of the model in section 1 of Romer (1993) onto the basic exploratory trading model. I am currently working on the details of this extension.

Alternatively, the baseline exploratory trading model can be viewed as an extension of the Kyle (1985) model to a setting in which the informed trader is uncertain about market depth, and the market-maker doesn't act quickly enough to alter his initial strategy. This is not necessarily an attractive option if our ultimate goal is to use exploratory trading to illuminate the inner workings of high-frequency trading, but it is is reasonable if we simple want to think about exploratory trading for its own sake.

5.3.2 Limited Information

Limitations on what the HFT can directly observe are an indispensable component of the exploratory trading model. Clearly, if the HFT could observe the market demand curve directly, he would have no reason to engage in exploratory trading. Similarly, if the HFT could directly observe s_1, the net supply in period 1, he could perfectly infer the value of β without recourse to exploratory trading.

While the rationale for imposing these limitations on what the HFT can observe is obvious from a modeling standpoint, such limitations are slightly harder to motivate from an empirical perspective. In many markets, traders can observe the limit-order book, which seems rather similar to observing the market demand curve. However, while the order book certainly contains some information about the market demand curve, it does not perfectly reveal the true market demand curve. In the context of a real market, it might be more precise to think of exploratory trading revealing information about "the residual component of market demand not explained by the order book" rather than "the market demand curve" *per se,* but ideas are isomorphic. In fact, if we assume in the model that the HFT conditions all of his beliefs β on some not-perfectly-informative order book, our earlier analyses still hold exactly, without any modification[17].

Limiting the HFT's ability to observe s_1 might also be implausible in some markets. However, "supply noise" was just a convenient way to represent the problem of distinguishing between the component of price impact that depended on the HFT's order, and the component of price impact that did not. Alternatively, we could have assumed that the HFT was uncertain about the size of the transitory and permanent components of the price change between dates -1 and 0, so that he could not be sure how much of the price

[17] The HFT's prior for β would be different, but since we never specified the particular form of this prior to begin with, this difference would not change any of our results.

change between dates 0 and 1 could be attributed to the price impact of his trade, and how much could be attributed to some persistent component of the preceding price change.

5.3.3 Slowly-Varying Demand Parameters

Stable demand curves constitute the second vital component of the exploratory trading model. When the HFT trades in period 1, he learns something about the shape of the demand curve in that period. Since the HFT does not obtain this information until after period 1 ends, the information is only valuable to him if it reveals something about the demand curve that he will face in period 2. The demand parameters need not actually remain constant over both periods[18], but the period 2 parameters cannot be independent of the period-1 parameters.

In addition to this temporal stability, the demand curves must be sufficiently well-behaved that their local shape provides meaningful information about their global shape. Since an exploratory order and the subsequent order that it informs will not generally be identical, the HFT must extrapolate. If demand curves are linear, then their local shape is perfectly informative about their global shape, but under more general assumptions, extrapolation error will limit the scale (in terms of order size) at which exploratory trading could be profitable.

Both of the stability requirements above become easier to satisfy in a high-frequency context, where we replace large, infrequent trades with smaller, more frequent ones.

6 Conclusion

In this paper, I address a central puzzle about high-frequency trading, namely how trading speed could translate to superior information. To this end, I analyze the idea of "exploratory trading"—how an agent might use his own trades to gather valuable information. A central result of my analysis is that exploratory trading bears a natural connection to high-frequency trading, and conversely, that high-frequency trading inherently raises issues analogous to those of exploratory trading. This connection between exploratory and high-frequency trading ultimately illuminates the issue of how trading speed could translate to superior information. Beyond these general results, the model of exploratory trading also helps to explain a variety of specific empirical findings and generates a number of testable implications.

References

[1] Terrence Hendershott Alex Boulatov and Dmitry Livdan. Informed trading and portfolio returns. June 2011.

[2] Bruno Biais, Thierry Foucault, and Sophie Moinas. Equilibrium algorithmic trading. Working Paper, October 2010.

[18] Although I abstract away from any sort of strategic or adaptive behavior by the agents who (in aggregate) submit the market demand curve, we could relax this simplifying assumption somewhat without dramatically altering the qualitative behavior of the model.

[3] Jonathan A. Brogaard. High frequency trading and its impact on market quality. Northwestern University Kellogg SOM Working Paper, November 2010.

[4] Miguel Sousa Lobo Bruce Ian Carlin and S. VIiswanathan. Episodic liquidity crises: Cooperative and predatory trading. *The Journal of Finance*, LXII(5), OCTOBER 2007.

[5] Álvaro Cartea and José Penalva. Where is the value in high frequency trading? Banco De España, Documentos de Trabajo N.Âº 1111, June 2011.

[6] Jeff Castura, Robert Litzenberger, Richard Gorelick, and Yogesh Dwivedi. Market efficiency and microstructure evolution in u.s. equity markets: A high-frequency perspective. Working Paper from RGM Advisors, LLC, October 2010.

[7] Giovanni Cespa and Thierry Foucault. Insiders-outsiders, transparency and the value of the ticker. Queen Mary University Dept. of Economics Working Paper No. 628, April 2008.

[8] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Recent trends in trading activity. UCLA Working Paper, January 2010.

[9] Adam Clark-Joseph and Brock Mendel. A model and analysis of high-frequency trading. Harvard University Working Paper, February 2011.

[10] Securities Exchange Commission. Concept release on equity market structure, concept release no. 34-61358. FileNo. 17 CFR Part 242 [Release No. 34-61358; File No. S7-02-10] RIN 3235-AK47, January 2010.

[11] U.S. Securities Exchange Commission. Findings regarding the market events of may 6, 2010, September 2010.

[12] Jaksa Cvitanic and Andrei Kirilenko. High frequency traders and asset prices. CalTech/CFTC Working Paper, March 2011.

[13] David Easley, Marcos M. López de Prado, and Maureen O'Hara. The microstructure of the 'flash crash' flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Portfolio Management*, Forthcoming.

[14] Peter Gomber and Markus Gsell. Algorithmic trading engines versus human traders – do they behave different [sic] in securities markets? Published by the Center for Financial Studies at Goethe-University Frankfurt,, April 2009.

[15] Sanford J. Grossman and Merton H. Miller. Liquidity and market structure. *The Journal of Finance*, 43(3), July 1988.

[16] L.E. Harris. Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments*, 7:1–76, 1998.

[17] Joel Hasbrouck and Gideon Saar. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12:143–172, 2009.

[18] Joel Hasbrouck and Gideon Saar. Low-latency trading. NYU Stern/ Cornell GSM Working Paper, May 2011.

[19] Terrence Hendershott, Charles M. Jones, and Albert J. Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1), February 2011.

[20] Terrence Hendershott and Ryan Riordan. Algorithmic trading and information. NET Institute Working Paper No. 09-08, September 2009.

[21] Robert Jarrow and Philip Protter. A dysfunctional role of high frequency trading in electronic markets. Johnson School Research Paper Series No. 08-2011, March 2011.

[22] Boyan Jovanovic and Albert J. Menkveld. Middlemen in limit-order markets. Working Paper, June 2010.

[23] Michael Kearns, Alex Kulesza, and Yuriy Nevmyvaka. Empirical limitations on high frequency trading profitability. University of Pennsylvania CIS Working Paper, 2010.

[24] Andrei Kirilenko, Mehrdad Samadi, Albert S. Kyle, and Tugkan Tuzun. The flash crash: The impact of high frequency trading on an electronic market. October 2010.

[25] Thomas McInish and James Upson. Strategic liquidity supply in a market with fast and slow traders. September 2011.

[26] Ciamac C. Moallemi, Beomsoo Park, and Benjamin Van Roy. Strategic execution in the presence of an uninformed arbitrageur. Columbia GSB/ Stanford EE Working Paper, December 2010.

[27] Ciamac C. Moallemi and Mehmet Saglam. The cost of latency. Columbia GSB Working Paper, June 2010.

[28] Marco Lutat Tim Uhle Peter Gomber, Bjorn Arndt. High-frequency trading. Pubished by Deutsche Borse AG Market Policy and European Public Affairs, March 2011.

[29] David Romer. Rational asset-price movements without news. *The American Economic Review*, 83(5):1112–1130, December 1993.

[30] Australian Securities and Investments Commission. Australian equity market structure. REPORT 215, November 2010.

[31] X. Frank Zhang. High-frequency trading, stock volatility, and price discovery. Yale SOM Working Paper, December 2010.

A Mathematical Appendix

A.1 Calculations Related to x_2 and π_2

Solving for x_2^*:

$$
\begin{aligned}
& \frac{\partial}{\partial x_2}\left(-\beta\left(\alpha+x_2\right)x_2+\mu x_2\right) \\
&= -\beta\alpha-2\beta x_2+\mu \\
0 &\equiv -\beta\alpha-2\beta x_2^*+\mu \\
x_2^* &= \frac{\mu-\beta\alpha}{2\beta} \\
&= \frac{\mu-p_0}{2\beta}
\end{aligned}
$$

Solving for π_2^*:

$$
\begin{aligned}
\pi_2^* &= -\beta\left(\alpha+\frac{\mu-\beta\alpha}{2\beta}\right)\frac{\mu-\beta\alpha}{2\beta}+\mu\frac{\mu-\beta\alpha}{2\beta} \\
&= -\beta\left(\frac{\beta\alpha+\mu}{2\beta}\right)\frac{\mu-\beta\alpha}{2\beta}+\mu\frac{\mu-\beta\alpha}{2\beta} \\
&= \frac{(\beta\alpha)^2-\mu^2}{4\beta}+\frac{2\mu^2-2\mu\beta\alpha}{4\beta} \\
&= \frac{(\beta\alpha)^2+\mu^2-2\mu\beta\alpha}{4\beta} \\
&= \beta\left(\frac{\mu-\beta\alpha}{2\beta}\right)^2 \\
&= \beta\left(x_2^*\right)^2
\end{aligned}
$$

Solving for $\hat{\pi}_2$:

$$
\begin{aligned}
\hat{\pi}_2 &= -\beta\left(\alpha+\hat{x}\right)\hat{x}+\mu\hat{x} \\
&= \left(-\beta\alpha+\mu\right)\hat{x}-\beta\left(\hat{x}\right)^2 \\
&= \left(\mu-\beta\alpha\right)\left(\hat{x}-x^*+x^*\right)-\beta\left(\hat{x}-x^*+x^*\right)^2 \\
&= \left(\mu-\beta\alpha\right)x^*+\left(\mu-\beta\alpha\right)\left(\hat{x}-x^*\right) \\
&\quad -\beta\left(\hat{x}-x^*\right)^2-2\beta\left(\hat{x}-x^*\right)x^*-\beta\left(x^*\right)^2 \\
&= \left(\mu-\beta\alpha\right)x^*-\beta\left(x^*\right)^2-\beta\left(\hat{x}-x^*\right)^2 \\
&\quad +\left(\mu-\beta\alpha\right)\left(\hat{x}-x^*\right)-2\beta\left(\hat{x}-x^*\right)x^* \\
&= \pi_2^*-\beta\left(\hat{x}-x^*\right)^2+\left(\mu-\beta\alpha-2\beta x^*\right)\left(\hat{x}-x^*\right) \\
&= \pi_2^*-\beta\left(\hat{x}-x^*\right)^2
\end{aligned}
$$

The final equality uses the fact that

$$
\begin{aligned}
\mu-\beta\alpha-2\beta x^* &\equiv \mu-\beta\alpha-2\beta\left(\frac{\mu-\beta\alpha}{2\beta}\right) \\
&\equiv 0
\end{aligned}
$$

Expected Feasible Period-2 Profit Taylor Approximation:

$$\begin{aligned}
\left(\frac{1}{z}-\frac{1}{\beta}\right)^2 &= \left(\frac{1}{\beta}-\frac{1}{\beta}\right)^2 + 2\left(\frac{1}{z}-\frac{1}{\beta}\right)\left(\frac{-1}{z^2}\right)|_{z=\beta}(z-\beta) \\
&\quad +\frac{1}{2}2\left(\frac{-2}{\beta z^3}+\frac{3}{z^4}\right)|_{z=\beta}(z-\beta)^2 + \frac{2}{6}\left(\frac{6}{\beta\zeta^4}-\frac{12}{\zeta^5}\right)(z-\beta)^3 \\
&= \frac{1}{\beta^4}(z-\beta)^2 + 2\zeta^{-5}\left(\frac{\zeta-2\beta}{\beta}\right)(z-\beta)^3 \\
&\leq \frac{1}{\beta^4}(z-\beta)^2 + \frac{32}{3125\beta^5}(z-\beta)^3 \\
&= \frac{z^2-2z\beta+\beta^2}{\beta^4} + \frac{z^3-3z^2\beta+3z\beta^2-\beta^3}{\frac{3125}{32}\beta^5} \\
\left(\frac{1}{z}-\frac{1}{\beta}\right)^2 &= \frac{1}{\beta^4}(z-\beta)^2 + 2\beta^{-5}\left(\frac{\beta-2\beta}{\beta}\right)(z-\beta)^3 + \frac{2}{24}\left(\frac{10}{\eta^6}-\frac{4}{\beta\eta^5}\right)(z-\beta)^4 \\
&= \frac{1}{\beta^4}(z-\beta)^2 - \frac{2}{\beta^5}(z-\beta)^3 + \frac{\eta^{-6}}{6}\left(\frac{5\beta-2\eta}{\beta}\right)(z-\beta)^4 \\
&\geq \frac{1}{\beta^4}(z-\beta)^2 - \frac{2}{\beta^5}(z-\beta)^3 - \frac{1}{4374\beta^6}(z-\beta)^4
\end{aligned}$$

A.2 Calculations Related to the HFT's Inference Problem

Bounding the variance of $\mathbb{E}_1[\beta]$

$$\begin{aligned}
\left(\frac{\Delta p}{x_1}-\beta\right)^2 &= \beta^2\left(\frac{s_1}{x_1}\right)^2 \\
\mathbb{E}\left[\left(\frac{\Delta p}{x_1}-\beta\right)^2\right] &= \frac{\left(\sigma_\beta^2+b^2\right)\sigma_s^2}{x_1^2} \\
&\geq \mathbb{E}\left[\left(\mathbb{E}_1[\beta]-\beta\right)^2\right]
\end{aligned}$$

The last line uses the fact that the conditional expectation of β will be at least as good an estimator (in the $\mathcal{L}^2$-sense) as the expectation of β conditioned upon the value of $\frac{\Delta p}{x_1}$, which in turn will be at least as good an estimator of β (in the $\mathcal{L}^2$-sense) as $\frac{\Delta p}{x_1}$ itself.

Next, since x_1 is known (and we shall assume $x_1 \neq 0$) we can normalize equation (8) by dividing through by x_1:

$$\frac{\Delta p}{x_1} = \beta + \beta\epsilon_x$$

where we define $\epsilon_x \equiv \frac{s_1}{x_1}$. Suppose that for a given fixed value of x_1, the conditional expectation $\mathbb{E}_1[\beta]$ is a smooth function of $\frac{\Delta p}{x_1}$, say $\mathbb{E}_1[\beta] = g\left(\frac{\Delta p}{x_1}\right)$, with continuous derivative $g'(\cdot)$ that is not identically zero. By standard delta-method-type Taylor approximation arguments, it is easy to show that

$$\begin{aligned}(x_1)^2\,\mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2\right] &\rightarrow \mathbb{E}\left[g'(\beta)^2\right] Var(\Delta p)\\ &= \mathbb{E}\left[g'(\beta)^2\right]\left(\sigma_\beta^2+b^2\right)\sigma_s^2\end{aligned}$$

as $|x_1|\rightarrow\infty$. Since we assume that the derivative $g'(\cdot)$ is not identically zero, the term on the right-hand side of the equation above, $\mathbb{E}\left[g'(\beta)^2\right]\left(\sigma_\beta^2+b^2\right)\sigma_s^2$, is strictly positive. Therefore we can choose $J>0$ such that for all x_1 satisfying $(x_1)^2>J$, we have

$$(x_1)^2\,\mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2\right] > \frac{1}{2}\mathbb{E}\left[g'(\beta)^2\right]\left(\sigma_\beta^2+b^2\right)\sigma_s^2$$

Hence as $|x_1|\rightarrow\infty$, we can bound $\mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2\right]$ below by a function of the form $\frac{A\sigma_s^2}{x_1^2}$, where A is some strictly positive constant. □

The restrictive assumption in the argument above was taking the conditional expectation $\mathbb{E}_1[\beta]$ to be a smooth function of $\frac{\Delta p}{x_1}$. Although this condition is easy to satisfy if we restrict attention to $\frac{\Delta p}{x_1}>0$, the condition becomes more restrictive if we permit $\frac{\Delta p}{x_1}\leq 0$. Because we assume that the support of β is strictly positive, a negative value of $\frac{\Delta p}{x_1}$ implies that $s_1<-x_1$, whereas a positive value of $\frac{\Delta p}{x_1}$ does not imply this. (More precisely, the break-point for $\frac{\Delta p}{x_1}$ is the infimum of the support of β, which is strictly positive, but the same ideas apply.) Although $\mathbb{E}_1[\beta]$ will typically be a continuous function of $\frac{\Delta p}{x_1}$ in a neighborhood of zero, it is not obvious that it is likely to be everywhere-differentiable in that neighborhood—see the expressions below.

$$\begin{aligned}f(\theta|\Delta p=c;\ x_1) &= \frac{p_s\left(\frac{x_1}{\theta}\left(\frac{\Delta p}{x_1}-\theta\right)\right)f(\theta)}{\int p_s\left(\frac{x_1}{t}\left(\frac{\Delta p}{x_1}-t\right)\right)f(t)\,dt}\\ &= \frac{p_{\epsilon_x}\left(\frac{1}{\theta}\left(\frac{\Delta p}{x_1}\right)-1\right)f(\theta)}{\int p_{\epsilon_x}\left(\frac{1}{t}\left(\frac{\Delta p}{x_1}\right)-1\right)f(t)\,dt}\\ \mathbb{E}[\beta|\Delta p=c;\ x_1] &= \frac{\int \theta p_{\epsilon_x}\left(\frac{1}{\theta}\left(\frac{\Delta p}{x_1}\right)-1\right)f(\theta)\,d\theta}{\int p_{\epsilon_x}\left(\frac{1}{t}\left(\frac{\Delta p}{x_1}-t\right)\right)f(t)\,dt}\end{aligned}$$

Of course, the probability of encountering this break-point decreases as $|x_1|$ increases, and Chebychev's inequality implies that this probability is bounded above by a function of the form $\frac{H}{x_1^2}$ for some strictly positive constant H. Since we are only interested in deriving a lower bound for $\mathbb{E}\left[(\mathbb{E}_1[\beta]-\beta)^2\right]$, the basic result above should probably hold under more general conditions, but the proof would be more delicate and more involved.

A.3 Maximizing Total Expected Profits

Expected Period-2 Profits as a Function of x_1: Using equation (9), we can replace $Var\left(\mathbb{E}_1[\beta]|\beta\right)$ with $K^2\frac{\beta^2\sigma_s^2}{x_1^2}$ in equation(6). Then, taking the approximation to hold exactly for simplicity, we get

$$\mathbb{E}\left[\hat{\pi}_2|\beta,p_0\right] = \pi_2^* - K^2\frac{(\mu-p_0)^2}{4\beta}\frac{\sigma_s^2}{x_1^2}$$

By taking expectations of both sides of the above equation with respect to β, we obtain an expression for the unconditional expectation of $\hat{\pi}_2$ as a function of x_1:

$$\begin{aligned}\mathbb{E}\left[\hat{\pi}_2|p_0\right] &= \mathbb{E}\left[\pi_2^*\right] - \frac{K^2\mathbb{E}\left[\beta^{-1}\right]\sigma_s^2\xi^2}{4x_1^2} \\ &= \frac{\xi^2\mathbb{E}\left[\beta^{-1}\right]}{4} - \frac{K^2\mathbb{E}\left[\beta^{-1}\right]\sigma_s^2\xi^2}{4x_1^2} \\ &= \frac{\xi^2\mathbb{E}\left[\beta^{-1}\right]}{4}\left(1-\frac{K^2\sigma_s^2}{x_1^2}\right)\end{aligned}$$

Recall that we have defined $\xi \equiv \beta\alpha - \mu$ and $p_0 \equiv \beta\alpha$, so $\xi \equiv p_0 - \mu$.

The direct trading profit from trading x_1

$$\begin{aligned}\pi_1 &= x_1\left(\mu - p_0 - \beta x_1\right) \\ &= \left(\mu - p_0\right)x_1 - \beta x_1^2 \\ &= \xi x_1 - \beta x_1^2\end{aligned}$$

The optimal choice of x_1 in the absence of direct trading gains:

$$\mathbb{E}\left[\pi_1+\hat{\pi}_2\right] = -bx_1^2 + \frac{\sigma_\xi^2\mathbb{E}\left[\beta^{-1}\right]}{4}\left(1-\frac{K^2\sigma_s^2}{x_1^2}\right)$$

$$\max_{x_1}\left\{-bx_1^2 + \frac{\sigma_\xi^2\mathbb{E}\left[\beta^{-1}\right]}{4}\left(1-\frac{K^2\sigma_s^2}{x_1^2}\right)\right\}$$

$$\begin{aligned}FOC:\ -2bx_1^* + 2\frac{\sigma_\xi^2\mathbb{E}\left[\beta^{-1}\right]}{4}\frac{K^2\sigma_s^2}{(x_1^*)^3} &\equiv 0 \\ \Rightarrow \frac{\sigma_\xi^2\mathbb{E}\left[\beta^{-1}\right]K^2\sigma_s^2}{4} &= b\left(x_1^*\right)^4 \\ \Rightarrow \sigma_\xi\sigma_s K\sqrt{\frac{\mathbb{E}\left[\beta^{-1}\right]}{4b}} &= \left(x_1^*\right)^2\end{aligned}$$

The maximized value of the objective function is therefore given by:

$$
\begin{aligned}
\max_{x_1} \mathbb{E}\left[\pi_1 + \hat{\pi}_2\right] &= -b\left(x_1^*\right)^2 + \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4}\left(1 - \frac{K^2\sigma_s^2}{\left(x_1^*\right)^2}\right) \\
&= -b\left(x_1^*\right)^2 + \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4} - \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4}\frac{K^2\sigma_s^2}{\left(x_1^*\right)^2} \\
&= \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4} - b\sigma_\xi\sigma_s K\sqrt{\frac{\mathbb{E}\left[\beta^{-1}\right]}{4b}} \\
&\quad - \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right] K^2\sigma_s^2}{4}\frac{2\sqrt{b}}{\sigma_\xi\sigma_s K\sqrt{\mathbb{E}\left[\beta^{-1}\right]}} \\
&= \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4} - \frac{\sigma_\xi\sigma_s K\sqrt{\mathbb{E}\left[\beta^{-1}\right] b}}{2} - \frac{\sigma_\xi\sigma_s K\sqrt{\mathbb{E}\left[\beta^{-1}\right] b}}{2} \\
&= \frac{\sigma_\xi^2 \mathbb{E}\left[\beta^{-1}\right]}{4} - \sigma_\xi\sigma_s K\sqrt{\mathbb{E}\left[\beta^{-1}\right] b} \\
&= \frac{\sigma_\xi\sqrt{\mathbb{E}\left[\beta^{-1}\right]}}{4}\left(\sigma_\xi\sqrt{\mathbb{E}\left[\beta^{-1}\right]} - 4\sigma_s K\sqrt{b}\right)
\end{aligned}
$$

Hence the maximized value of the expected total profit (in the absence of direct trading gains) will be positive when σ_ξ^2 is sufficiently larger than σ_s^2, specifically, when

$$
\sigma_\xi^2 > \sigma_s^2 \frac{16K^2 b}{\mathbb{E}\left[\beta^{-1}\right]}
$$

If σ_ξ^2 does not satisfy the condition above, the HFT would obtain a higher expected profit (zero) by not participating in the market than he would by participating in an otherwise optimal manner.