Fast and Slow Informed Trading*

Ioanid Roşu†

June 3, 2015

Abstract

This paper develops a model in which traders receive a stream of private signals, and differ in their information processing speed. In equilibrium, the fast traders (FTs) quickly reveal a large fraction of their information, and generate most of the volume, volatility and profits in the market. If a FT is averse to holding inventory, his optimal strategy changes considerably as his aversion crosses a threshold. He no longer takes long-term bets on the asset value, gets most of his profits in cash, and generates a "hot potato" effect: after trading on information, the FT quickly unloads part of his inventory to slower traders. The results match evidence about high frequency traders.

Keywords: Trading volume, inventory, volatility, high frequency trading, price impact, mean reversion.

*Earlier versions of this paper circulated under the title "High Frequency Traders, News and Volatility." The author thanks Kerry Back, Laurent Calvet, Thierry Foucault, Johan Hombert, Pete Kyle, Stefano Lovo, Victor Martinez, Daniel Schmidt, Dimitri Vayanos, Jiang Wang; finance seminar participants at Copenhagen Business School, HEC Paris, Univ. of Durham, Univ. of Leicester, Univ. Paris Dauphine, Univ. Madrid Carlos III, ESSEC, KU Leuven; and conference participants at the European Finance Association 2014 meetings, American Finance Association 2013 meetings, Society for Advancement of Economic Theory in Portugal, Central Bank Microstructure Conference in Norway, Market Microstructure Many Viewpoints Conference in Paris for valuable comments.

†HEC Paris, Email: rosu@hec.fr.

1 Introduction

Today's markets are increasingly characterized by the continuous arrival of vast amounts of information. A media article about high frequency trading reports on the hedge fund firm Citadel: *"Its market data system, for example, contains roughly 100 times the amount of information in the Library of Congress. [...] The signals, or alphas, that prove to have predictive power are then translated into computer algorithms, which are integrated into Citadel's master source code and electronic trading program."* ("Man vs. Machine," CNBC.com, September 13th 2010). The sources of information from which traders obtain these signals usually include company-specific news and reports, economic indicators, stock indexes, prices of other securities, prices on various other trading platforms, limit order book changes, as well as various "machine readable news" and even "sentiment" indicators.[1]

At the same time, financial markets have seen in recent years the spectacular rise of algorithmic trading, and in particular of high frequency trading.[2] This coincidental arrival raises the question whether or not at least some of the HFTs do process information and trade very quickly in order to take advantage of their speed and superior computing power. Recent empirical evidence suggests that this is indeed the case.[3] But, despite the large role played by high frequency traders (HFTs) in the current financial landscape, there has been relatively little progress in explaining their strategies in connection with information processing.

We consider the following questions regarding HFTs: What are the optimal trading strategies of HFTs who process information? Why do HFTs account for such a large share of the trading volume? What explains the race for speed among HFTs? What are the effects of HFTs on measures of market quality, such as liquidity and price volatility?

[1] *"Math-loving traders are using powerful computers to speed-read news reports, editorials, company Web sites, blog posts and even Twitter messages—and then letting the machines decide what it all means for the markets."* ("Computers That Trade on the News," New York Times, December 22nd 2010).

[2] Hendershott, Jones, and Menkveld (2011) report that from a starting point near zero in the mid-1990s, high frequency trading rose to as much as 73% of trading volume in the United States in 2009. Chaboud, Chiquoine, Hjalmarsson, and Vega (2014) consider various foreign exchange markets and find that starting from essentially zero in 2003, algorithmic trading rose by the end of 2007 to approximately 60% of the trading volume for the euro-dollar and dollar-yen markets, and 80% for the euro-yen market.

[3] See Brogaard, Hendershott, and Riordan (2014), Baron, Brogaard, and Kirilenko (2014), Kirilenko, Kyle, Samadi, and Tuzun (2014), Hirschey (2013), Benos and Sagade (2013), Brogaard, Hagströmer, Nordén, and Riordan (2013).

How can HFT order flow anticipate future order flow and returns? What explains the "intermediation chains" or "hot potato" effects found among HFTs? Why do some HFTs have low inventories? Regarding the last question, some recent literature identifies HFTs as traders with both high trading volume *and* low inventories (see Kirilenko *et al.* 2014, SEC 2010). But then, a natural question arises: why would having low inventories be part of the definition of HFTs?

In this paper, we provide a theoretical model of informed trading which parsimoniously addresses these questions. Because we want to study speed differences among informed traders, we start with the standard framework of Kyle (1985), and modify it along several dimensions.[4] First, the asset's fundamental value is not constant but follows a random walk process, and each risk-neutral informed trader, or speculator, gradually receives signals about the asset value increments. Second, there are multiple speculators who differ in their speed, in the sense that some speculators receive their signal with a lag. Third, each speculator can trade only on lagged signals with a lag of at most m, where m is an exogenously given number.

It is the last assumption that sets our model apart from previous models of informed trading. A key effect of this assumption is to prevent the "rat race" phenomenon discovered by Holden and Subrahmanyam (1992), by which traders with identical information reveal their information so quickly, that the equilibrium breaks down at the "high frequency" limit, when the number of trading rounds approaches infinity. In our model, the speculators reveal only a fraction of their total private information, and this has a stabilizing effect on the equilibrium. Economically, we think of this assumption as equivalent to having a positive information processing cost per signal (and per trading round).[5] Indeed, since one of our results is that the value of information decays fast, even a tiny information processing cost would make speculators optimally ignore their signals after a sufficiently large number of lags m.

[4]As in Kyle (1985), we assume that informed traders submit only market orders; this is a plausible assumption for informed HFTs (see Brogaard, Hendershott, and Riordan 2014). Also, we set the model in continuous time, which makes it easier to solve for the equilibrium.

[5]Intuitively, information processing is costly because speculators need to avoid trading on stale information, and this involves (i) constantly monitoring public information to verify that their signal has not been incorporated into the price, and (ii) extracting the predictable part of their signal from past order flow, so that speculators trade only on the unpredictable (non-stale) part.

To simplify the analysis, we restrict our attention to the particular case when $m = 1$, which we call the *benchmark model*. In this model, speculators can trade using only their current signal and its lagged value. Thus, there are two types of speculators: *fast traders* (or FTs), who observe the signal instantly; and *slow traders* (or STs), who observe the signal after one lag. The benchmark model has the advantage that the equilibrium can be described in closed form. In the Internet Appendix we verify numerically that the main results in the benchmark model carry through to the general case ($m > 1$).

Our first main result in the benchmark model is that the FTs generate most of the trading volume, volatility, and profits. To understand why, consider the decision of N fast traders about what weight to use on the last signal they have received. Because the dealer sets a price function which is linear in the aggregate order size, each FT faces a Cournot-type problem and trades such that the price impact of his order is on average $1/(N + 1)$ of his signal. That brings the expected aggregate price impact to $N/(N+1)$ of the signal, and leaves on average only $1/(N+1)$ of the signal unknown to the dealer. Thus, once the STs observe the lagged signal, they now have much less private information to exploit. Moreover, the ST profits are further diminished by competition with FTs, who also trade on the lagged signal. Empirically, Baron, Brogaard, and Kirilenko (2014) find out that the profits of HFTs are concentrated among a small number of incumbents, and the profits appear to be correlated with speed.

Our second main result is that volume, volatility and liquidity are increasing with the number of FTs. First, more competition from FTs makes the prices more informative overall, and thus increases liquidity (measured, as in Kyle 1985, by the inverse price impact coefficient). As the market is more liquid, FTs face a lower price impact, and therefore trade even more aggressively. This creates an amplification mechanism that allows the aggregate FT trading volume to be increasing roughly linearly with the number of FTs. The effect of FTs on volatility is more muted but still positive; this is because in our model price volatility is bounded above by the fundamental volatility of the asset. Empirically, in line with our theoretical results, Hendershott, Jones, and Menkveld (2011), Boehmer, Fong, and Wu (2014), and Zhang (2010) document a positive effect of HFTs on liquidity. Moreover, the last two papers find a positive effect of HFTs on volatility. We should point out, however, that our model is more likely to

apply only to the subcategory of informed, market taking HFTs, and not to all HFTs. Thus, our results should be interpreted with caution.

Our third main result in the benchmark model is the existence of anticipatory trading: the order flow of fast traders predicts the order flow of slow traders in the next period. This comes from the fact that the fast traders' signal does not fully get incorporated into the price, hence the slow traders have an incentive to use the signal in the next period, after they remove the stale (predictable) part. Anticipatory trading is therefore related to speculator order flow autocorrelation. Our model predicts that the speculator order flow autocorrelation is positive, although it is small if the number of fast traders is large. Empirically, Brogaard (2011) finds that the autocorrelation of aggregate HFT order flow is indeed small and positive. Also, using Nasdaq data on high-frequency traders, Hirschey (2013) finds that HFT order flow anticipates future order flow.

Despite being able to match several stylized facts about HFTs in our benchmark model, a few questions remain. Why do many HFTs have low inventories, both intraday and at the day close?[6] Why do HFTs engage in "hot potato" trading (or "intermediation chains"), in which HFT pass their inventories to other traders?[7] What is the role of speed in explaining these phenomena?

To provide some theoretical guidance on these issues, we extend our benchmark model to include one trader with inventory costs. These costs can arise from risk aversion or from capital constraints, but we take a reduced form approach and assume the costs are quadratic in inventory, with a coefficient called *inventory aversion* (see Madhavan and Smidt 1993). We call this additional trader the *Inventory-averse Fast Trader*, or IFT.[8] We call this extension the *model with inventory management*. In addition to choosing the weight on his current signal, the IFT also chooses the rate at which he

[6]SEC (2010) characterizes HFTs by their "*very short time-frames for establishing and liquidating positions*" and argues that HFTs end "*the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions over-night).*" See also Kirilenko *et al.* (2014), Brogaard, Hagströmer, Nordén, and Riordan (2013), or Menkveld (2013).

[7]Weller (2014) analyzes both theoretically and empirically "intermediation chains" in which uninformed HFTs unwind inventories to slower, fundamental traders. Kirilenko *et al.* (2014) mention "hot potato effect" during the Flash Crash episode of May 6, 2010, when some HFTs would churn out their inventories very quickly to trade with other HFTs.

[8]We do not introduce more than one IFT since the model would be much more difficult to solve. The IFT is assumed fast because without slower traders it is not profitable to manage inventory.

mean reverts his inventory to zero each period. Without discussing yet optimality, suppose the IFT does inventory management, i.e., chooses a positive rate of inventory mean reversion. What are the effects of this choice?

The first effect of inventory management is that the IFT keeps essentially all his profits in cash. To see this, suppose the IFT chooses a coefficient of mean reversion of 10%. This translates into the inventory being reduced by a fraction of 10% in each trading round. Therefore, the IFT's inventory tends to become small over many rounds, and because our model is set in the high frequency limit (in continuous time), the inventory becomes in fact negligible.[9] We call this result the *low inventory effect*.

The second effect is that the IFT no longer makes profits by betting on the fundamental value of the asset. This stands in sharp contrast to the behavior of a risk-neutral speculator, such as the fast trader in the benchmark model. Indeed, the FT accumulates inventory in the direction of his information, since he knows his signals are correlated with the asset's liquidation value. By contrast, although the IFT initially trades on his current signal, he subsequently fully reverses the bet on that signal by removing a fraction of his inventory each trading round. Thus, the IFT's direct revenue from each signal eventually decays to zero. We call this result the *information decay effect*.

The third effect of inventory management is that, in order to make a profit, the IFT must (i) anticipate the slow trading, and (ii) trade in the opposite direction to slow trading. By *slow trading* here we simply mean the part of order flow that involves the speculators' lagged signals.[10] To understand this effect, consider how the IFT uses a given signal. The information decay effect means that the IFT's final revenues from betting on his signal are zero. Therefore, the IFT must benefit from inventory reversal. Since any trade has price impact, inventory reversal makes a profit only if gets pooled with order flow in the opposition direction, so that the IFT's price impact is negative. But in order to be *expected* profit, the opposite order flow must come from speculators who use lagged signals, i.e., from slow trading. We call this result the *hot potato effect*, or the *intermediation chain effect*.[11]

[9]Formally, the inventory follows an autoregressive process, hence its variance has the same order as the variance of the signal, which at high frequencies is negligible.

[10]A subtle point is that slow *trading* does not need to come from actual slow *traders*. Slow trading can also arise from fast traders who use their lagged signals as part of their optimal trading strategy.

[11]In our simplified framework, the intermediation chain only has one link, between the IFT and the

The reason behind this terminology is that the IFT's current signal (the "potato") produces undesirable inventory (is "hot") and must be passed on to slower traders in order to produce a profit. Thus, speed is important to the IFT. Without slower trading, there is no hot potato effect, and the IFT makes a negative expected profit from any trading strategy that mean reverts his inventory to zero. Note also that the hot potato generates a complementarity between the IFT and slow traders: Stronger inventory mean reversion by the IFT reduces the price impact of the STs, who can trade more aggressively. But more aggressive trading by the STs allows stronger mean reversion from the IFT.

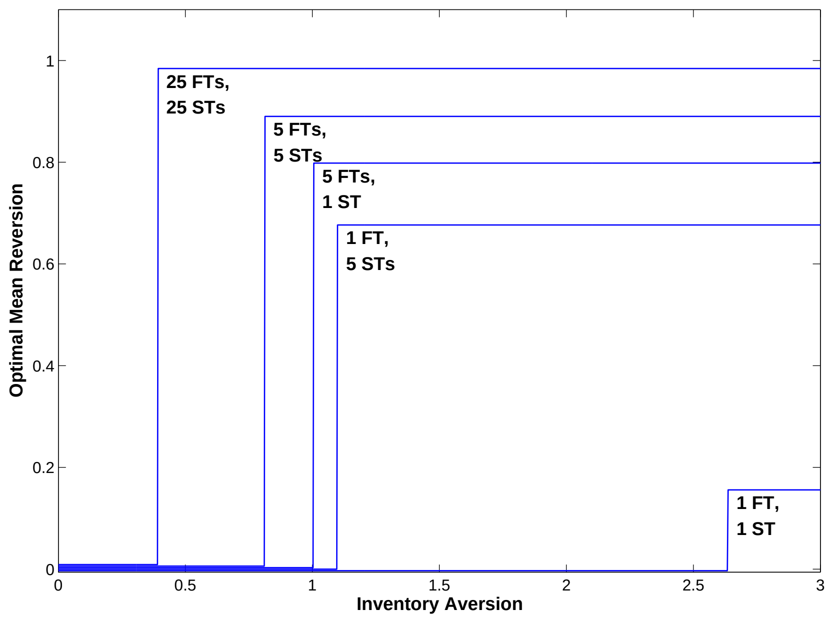
The optimal strategy of the IFT produces two contrasting types of behavior, depending on how his inventory aversion compares to a threshold. Below the threshold, the IFT behaves like a risk-neutral speculator, and lowers his inventory costs simply by reducing the weight on his signals. He does not manage inventory at all, because the information decay effect ensures that even a small but positive inventory mean reversion eventually destroys all revenues from the fundamental bets. With inventory aversion above the threshold, the IFT manages inventory and has all his profits in cash. The IFT benefits not from fundamental bets on his signals, but from the hot potato effect.

Figure 1 illustrates the optimal mean reversion for the IFT as a function of his inventory aversion coefficient. We see that, as his inventory aversion rises, the IFT changes discontinuously from the regime with no inventory mean reversion to the regime with positive inventory mean reversion. The threshold at which this discontinuity occurs depends on the number of fast traders (FTs) and slow traders (STs) in the model. This threshold is decreasing in both parameters, because the amount of slow trading is increasing in both parameters. Slow trading is clearly increasing in the number of slow traders. But it is also increasing in the number of fast traders because (i) the fast traders also use their lagged signals, and (ii) more fast traders make the market more liquid, which allows slow trading to be more aggressive.

Our results speak to the literature on high-frequency trading. One may think that in practice HFTs have very low inventories because either (i) HFTs have very high

slow traders. But we conjecture that in a model where speculators use more than one lag for their signals, the intermediation chains become longer, depending on the number of lags.

Figure 1: Optimal Inventory Mean Reversion. This figure plots the optimal mean reversion coefficient of an inventory-averse fast trader (IFT), when he competes with N_F fast traders (FTs) and N_S slow traders (STs), with $N_F, N_S \in \{1, 5, 25\}$. On the horizontal axis is the IFT's inventory aversion coefficient. The optimal mean reversion coefficient is computed using the results of Section 5, in the inventory management model with parameters N_F and $N_L = N_F + N_S$. The other parameter values are: $\sigma_w = 1$, $\sigma_u = 1$.



risk aversion, or (ii) HFTs do not have superior information and wish to maintain zero inventory to avoid averse selection on their positions in the risky asset. Our results suggest that this is not necessarily the case. Indeed, Figure 1 suggests (and we rigorously prove in Proposition 6) that in the limit when the number of speculators is large, the threshold inventory aversion converges to zero, and the optimal mean reversion is close to one. In other words, even with low inventory aversion, the IFT chooses very large mean reversion. Yet, even at these high rates of mean reversion the IFT does not loses more than about 50% of his average profits from inventory management (the advantage being that he has all his profits in cash).

We predict that in practice the fast speculators are sharply divided into two categories. In both categories speculators trade with a large volume. But in one category speculators accumulate inventory by taking fundamental bets. In the other category speculators have very low inventories; they initially trade on their signals but then

quickly pass on part of their inventory to slower traders. These covariance patters produce testable implications of our model.

The division of fast speculators in two categories appears consistent with the empirical findings of Kirilenko *et al.* (2014), who study trading activity in the E-mini S&P 500 futures during several days around the Flash Crash of May 6, 2010. The "opportunistic traders" described in their paper resembles our risk-neutral fast traders: opportunistic traders have large volume, appear to be fast, and accumulate relatively large inventories. By contrast the "high frequency traders" in their paper, while they are also fast and trade in large volume, keep very low inventories. Indeed, HFTs in their sample liquidate 0.5% of their aggregate inventories on average each second.

Related Literature

Our paper contributes to the literature on trading with asymmetric information. We show that competition among informed traders, combined with noisy trading strategies, produces a large informed trading volume and a quick information decay.[12] The market is very efficient because competition among informed traders makes them trade aggressively on their common information. This intuition is present in Holden and Subrahmanyam (1992) and Foster and Viswanathan (1996). The former paper finds that the competition among informed traders is so strong, that in the continuous time limit there is no equilibrium in smooth strategies. Our contribution to this literature is to show that that there exists an equilibrium in *noisy* strategies. This rests on two key assumptions: (i) *noisy information*, i.e., speculators learn over time by observing a stream of signals, and (ii) *finite lags*, i.e., speculators only use a signal for a fixed number of lags—which is plausible if there is a positive information processing cost per signal.

Without the finite lags assumption, noisy information by itself does not generate noisy strategies, as Back and Pedersen (1998) show. Chau and Vayanos (2008), Caldentey and Stacchetti (2010), and Li (2012) find that noisy information coupled with either model stationarity or a random liquidation deadline produces strategies that are still smooth as in Kyle (1985), but towards the high frequency limit they have almost

[12] A speculator's strategy is *smooth* if the volatility generated by that speculator's trades is of a lower magnitude compared to the volatility from noise trading; and *noisy* if the magnitudes are the same.

infinite weight. Thus, the market in these papers is nearly strong-form efficient, which makes speculators' strategies appear noisy (there is no actual equilibrium in the limit). By contrast, in our model the market is not strong-form efficient even in the limit, yet strategies are noisy. Foucault, Hombert, and Roşu (2015) propose a model in which a single speculator receives a signal one instant before public news. The speculator's strategy is noisy, but for a different reason than in our model: the speculator optimally trades with a large weight on his forecast of the news. Yet a different mechanism occurs in Cao, Ma, and Ye (2013). In their model, informed traders must disclose their trades immediately after trading, and therefore traders optimally obfuscate their signal by adding a large noise component to their trades.

Our paper also contributes to the rapidly growing literature on High Frequency Trading.[13] In much of this literature, it is the speed *difference* that has a large effect in equilibrium. The usual model setup has certain traders who are faster in taking advantage of an opportunity that disappears quickly. As a result, traders enter into a winner-takes-all contest, in which even the smallest difference in speed has a large effect on profits. (See for instance the model with speed differences of Biais, Foucault, and Moinas (2014), or the model of news anticipation of Foucault, Hombert, and Roşu (2015).) By contrast, our results regarding volume and volatility remain true even if all informed traders have the same speed. This is because in our model the need for speed arises endogenously, from competition among informed traders. In our model, being "slow" simply means trading on lagged signals. Since in equilibrium speculators also use lagged signals (the unanticipated part, to be precise), in some sense all traders are slow as well. Yet, it is true in our model that a genuinely slower trader makes less money, since he can only trade on older information that has already lost much of its value.

Our results regarding the optimal inventory of informed traders are, to our knowledge, new. Theoretical models of inventory usually attribute inventory mean reversion to passive market makers, who do not possess superior information, but are concerned

[13]See Biais, Foucault, and Moinas (2014), Aït-Sahalia and Saglam (2014), Budish, Cramton, and Shim (2014), Foucault, Hombert, and Roşu (2015), Du and Zhu (2014), Li (2014), Hoffmann (2014), Pagnotta and Philippon (2013), Weller (2014), Cartea and Penalva (2012), Jovanovic and Menkveld (2012), Cvitanić and Kirilenko (2010).

with absorbing order flow.[14] Our paper shows that an informed investor with inventory costs (the "IFT") can display behavior that makes him appear like a market maker, even though he only submits market orders (as in Kyle 1985). Indeed, in our model the IFT does not take fundamental bets, passes his risky inventory to slower traders (the hot potato effect), and keeps essentially all his money in cash. To obtain these results, even a small inventory aversion of the IFT suffices, but only if enough slow trading exists.

A related paper is Hirshleifer, Subrahmanyam, and Titman (1994). In their 2-period model, risk averse speculators with a speed advantage first trade to exploit their information, after which they revert their position because of risk aversion; while the slower speculators trade in the same direction as the initial trade of the faster speculators. The focus of Hirshleifer, Subrahmanyam, and Titman (1994) is different, as they are interested in information acquisition and explaining behavior such as "herding" and "profit taking." Our goal is to analyze the inventory problem of fast informed traders in a fully dynamic context, and to study the properties of the resulting optimal strategies.

The paper is organized as follows. Section 2 describes the model setup. Section 3 solves for the equilibrium in the particular case with two categories of traders: fast and slow. Section 4 discusses the effect of fast and slow traders on various measures of market quality. Section 5 introduces and extension of the baseline model in which a new trader (the IFT) has inventory costs. Then, it analyzes the IFT's optimal strategy and its effect on equilibrium. Section 6 concludes. All proofs are in the Appendix or the Internet Appendix. The Internet Appendix solves for the equilibrium in the general case, and analyzes several modifications and extensions of our baseline model.

2 Model

Trading for a risky asset takes place continuously over the time interval $[0, T]$, where we use the normalization:[15]

$$T = 1. \tag{1}$$

[14] See Ho and Stoll (1981), Madhavan and Smidt (1993), Hendershott and Menkveld (2014), as well as many references therein.

[15] To eliminate confusion with later notation, we often use T instead of 1. This way, we can denote below $t - \mathrm{d}t$ by $t - 1$ without much confusion.

Trading occurs at intervals of length $\mathrm{d}t$ apart. Throughout the text, we refer to $\mathrm{d}t$ as representing one period, or one trading round. The liquidation value of the asset is

$$v_T \;=\; \int_0^T \mathrm{d}v_t, \quad \text{with} \quad \mathrm{d}v_t \;=\; \sigma_v \mathrm{d}B_t^v, \tag{2}$$

where B_t^v is a Brownian motion, and $\sigma_v > 0$ is a constant called the *fundamental volatility.* We interpret v_T as the "long-run" value of the asset; in the high frequency world, this can be taken to be the asset value at the end of the trading day. The increments $\mathrm{d}v_t$ are then the short term changes in value due to the arrival of new information. The risk-free rate is assumed to be zero.

There are three types of market participants: (a) $N \geq 1$ risk neutral speculators, who observe the flow of information at different speeds, as described below; (b) noise traders; and (c) one competitive risk neutral dealer, who sets the price at which trading takes place.

Information and Speed. At $t = 0$, there is no information asymmetry between the speculators and the dealer. Subsequently, each speculator receives the following flow of signals:

$$\mathrm{d}s_t \;=\; \mathrm{d}v_t + \mathrm{d}\eta_t, \quad \text{with} \quad \mathrm{d}\eta_t \;=\; \sigma_\eta \mathrm{d}B_t^\eta, \tag{3}$$

where $t \in (0, T]$ and B_t^η is a Brownian motion independent from all other variables. Denote by

$$w_t \;=\; \mathsf{E}\big(v_T \;\big|\; \{s_\tau\}_{\tau \leq t}\big) \tag{4}$$

the expected value conditional on the information flow until t. We call w_t the *value forecast*, or simply *forecast*. Because there is no initial information asymmetry, $w_0 = 0$. Denote by σ_w the instantaneous volatility of w_t, or the *forecast volatility.* The increment of the forecast w_t, and the forecast variance are given, respectively, by

$$\mathrm{d}w_t \;=\; \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\,\mathrm{d}s_t, \qquad \sigma_w^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}w_t)}{\mathrm{d}t} \;=\; \frac{\sigma_v^4}{\sigma_v^2 + \sigma_\eta^2}. \tag{5}$$

When deriving empirical implications, we call σ_w the *signal precision*, as a precise signal (small σ_η) corresponds to a large σ_w.

Speculators obtain their signal with a lag $\ell \in \{0, 1, 2, \ldots\}$. A *$\ell$-speculator* is a trader who at $t \in (0, T]$ observes the signal from ℓ periods before, $\mathrm{d}s_{t-\ell\,\mathrm{d}t}$. To simplify notation, we use the following convention:

$$\text{Notation for trading times:} \quad t - \ell \quad \text{instead of} \quad t - \ell\,\mathrm{d}t. \tag{6}$$

For instance, instead of $\mathrm{d}s_{t-\ell\,\mathrm{d}t}$ we write $\mathrm{d}s_{t-\ell}$.

Trading and Prices. At each $t \in (0, T]$, denote by $\mathrm{d}x_t^i$ the market order submitted by speculator $i = 1, \ldots, N$ at t, and by $\mathrm{d}u_t$ the market order submitted by the noise traders, which is of the form $\mathrm{d}u_t = \sigma_u \mathrm{d}B_t^u$, where B_t^u is a Brownian motion independent from all other variables. Then, the aggregate order flow executed by the dealer at t is

$$\mathrm{d}y_t \;=\; \sum_{i=1}^{N} \mathrm{d}x_t^i + \mathrm{d}u_t. \tag{7}$$

The dealer is risk neutral and competitive, hence she executes the order flow at a price equal to her expectation of the liquidation value conditional on her information. Let $\mathcal{I}_t = \{y_\tau\}_{\tau < t}$ be the dealer's information set just before trading at t. The order flow at date t, $\mathrm{d}y_t$, executes at

$$p_t \;=\; \mathsf{E}\big(v_T \mid \mathcal{I}_t \cup \mathrm{d}y_t\big). \tag{8}$$

Together with the price, another important quantity is the dealer's expectation at t of the k-lagged signal $\mathrm{d}w_{t-k}$:

$$z_{t-k,t} \;=\; \mathsf{E}\big(\mathrm{d}w_{t-k} \mid \mathcal{I}_t\big). \tag{9}$$

Equilibrium Definition. In general, a trading strategy for a ℓ-speculator is a process followed by his risky assset position, x_t, which is measurable with respect to his information set $\mathcal{J}_t^{(\ell)} = \{y_\tau\}_{\tau<t} \cup \{s_\tau\}_{\tau \le t-\ell}$. For a given trading strategy, the speculator's expected profit π_τ, from date τ onwards, is

$$\pi_\tau \;=\; \mathsf{E}\left(\int_\tau^T (v_T - p_t)\mathrm{d}x_t \mid \mathcal{J}_\tau^{(\ell)} \right). \tag{10}$$

As in Kyle (1985), we focus on linear equilibria. Specifically, we consider strategies

which are linear in the unpredictable part of their signals,[16]

$$\mathrm{d}w_{t-k} - z_{t-k,t}, \quad k \;=\; \ell,\; \ell+1,\; \ldots \tag{11}$$

We restrict strategies to exclude signals older than a fixed number of lags m (which is allowed to depend on the speculator's speed parameter ℓ). This assumption can be justified by costly information processing, as explained at the end of this section. Formally, the ℓ-speculator's strategy is of the form:

$$\mathrm{d}x_t \;=\; \gamma_{\ell,t}(\mathrm{d}w_{t-\ell} - z_{t-\ell,t}) + \gamma_{\ell+1,t}(\mathrm{d}w_{t-\ell-1} - z_{t-\ell-1,t}) + \cdots + \gamma_{m,t}(\mathrm{d}w_{t-m} - z_{t-m,t}). \tag{12}$$

A linear equilibrium is such that: (i) at every date t, each speculator's trading strategy (12) maximizes his expected trading profit (10) given the dealer's pricing policy, and (ii) the dealer's pricing policy given by (8) and (9) is consistent with the equilibrium speculators' trading strategies.

Finally, the speculators consider the covariance structure of $z_{t-k,t}$ to be independent of their strategy. More precisely, for all $j, k \geq 0$, the speculators consider the numbers

$$Z_{j,k,t} \;=\; \mathsf{Cov}\big(\mathrm{d}w_{t-j}, z_{t-k,t}\big) \tag{13}$$

to depend only on j, k, and t. Thus, the covariance terms $Z_{j,k,t}$ are interpreted as being computed by the dealer, as part of her (publicly known) pricing rules.[17]

Model Notation. If all speculators in the model have a strategy of the form (12) with the same $m \geq 0$, we call it the *model with m lags*, and write $\mathcal{M}_m$. In the paper, we focus on the particular case with $m = 1$ lags. In this setup, the 0-speculators are called the *fast traders*, and the 1-speculators are called the *slow traders*; thus, we call $\mathcal{M}_1$ the

[16]Intuitively, if the strategy had a predictable component, the dealer's price would adjust and reduce the speculator's profit. We formalize this intuition in a discrete version of our model in Internet Appendix M. In the paper, however, we work in continuous time since it is easier to obtain analytical solutions. Similarly, Kyle (1985) directly assumes that the speculator's strategy in continuous time is linear in the unpredictable part of the fundamental value, $v - p_t$.

[17]For instance, the coefficient ρ_t in the dealer's pricing rule $z_{t-1,t} = \rho_t \mathrm{d}y_t$ is computed using the covariance term $\mathsf{Cov}(\mathrm{d}w_t, \mathrm{d}y_t)$ (see equation (A11)). Hence, even though a speculator affects $\mathrm{d}y_t$ by his strategy, he can consider the covariance term $\mathsf{Cov}(\mathrm{d}w_t, \mathrm{d}y_t)$ to be independent of his strategy. In Internet Appendix M.3, we explore an alternative specification in which the speculator takes into account his effect on $\mathrm{d}y_t$. We find, however, that the overall effect on the equilibrium coefficients is very small.

model with fast and slow traders.

If some ℓ-speculators have strategies of the form (12) with different m_ℓ, we call this the *mixed* model with m lags, where m is the maximum of all m_ℓ. We are particularly interested in the mixed model with $m = 1$ lags in which 0-speculators (fast traders) only trade on their current signal ($m_0 = 0$) and the 1-speculators (slow traders) only use their lagged signal ($m_1 = 1$). We call this the *bechmark model* with fast and slow traders, and denote it by $\mathcal{B}_1$. In Section 3, we solve for the equilibrium in both $\mathcal{M}_1$ and $\mathcal{B}_1$, and show that $\mathcal{M}_1$ can be regarded as a particular case of $\mathcal{B}_1$.

Information Processing. The assumption that speculators cannot use lagged signals beyond a given bound can be justified by introducing an information processing cost $\delta > 0$ per individual signal and per unit of time. More precisely, we consider an alternative model in which a ℓ-speculator can use all past signals, but must pay a fixed cost $\delta_\ell \, \mathrm{d}t$ each time he trades with a nonzero weight ($\gamma_{k,t}$) on his k-lagged signal (see equation 12). Then, intuitively, because the value of information decays with the lag, and the speculator does not want to accumulate too large a cost, he must stop using lagged signals beyond an upper bound. In Result 1 we show that for a particular value of δ the alternative model is equivalent to $\mathcal{M}_1$.

In this paper, we do not model the exact nature of the speculators' signals and their processing costs. But, intuitively, an information processing cost per signal (and per trading round) is plausible, because in practice speculators must constantly monitor each signal in order to avoid trading on stale (predictable) information. In our model, this can be done by simply removing the predictable part ($z_{t,t-k}$) from the lagged signal ($\mathrm{d}w_{t-k}$). In practice, however, speculators must monitor various sources of public information (such as news reports, economic data, or trading information in various related securities), to extract the part of the signal has not yet been incorporated into the price.

Note that an individual processing cost implicitly means that speculators cannot simply rely on free public signals, such as the price, to shortcut the learning process. This is because in reality prices may contain other relevant information about the fundamental value, along which the speculators are adversely selected. We formalize this intuition in Internet Appendix L, where we introduce an orthogonal dimension of the fundamental value, and show that trading strategies that rely on prices make an average loss.

3 Equilibrium with Fast and Slow Traders

In this section, we analyze the important case in which speculators use signals with a maximum lag of one. There are two types of speculators: (i) the *Fast Traders*, or FTs, who observe the signal with no delay (called 0-speculators in Section 2); and (ii) the *Slow Traders*, or STs, who observe the signal with a delay of one lag (called 1-speculators). The trading strategy of FTs and STs is of the form (see (12)):

$$\mathrm{d}x_t \;=\; \gamma_t(\mathrm{d}w_t - z_{t,t}) + \mu_t(\mathrm{d}w_{t-1} - z_{t-1,t}), \quad t \;\in\; (0,T], \tag{14}$$

where the weight γ_t must be zero for a ST. There are two possibilities: either the FT can trade on both the current and the lagged signals, or the FT can trade only on the current signal, i.e., the FT's weight γ_t must be zero.[18] The former case is the model denoted by $\mathcal{M}_1$, the *model with fast and slow traders*. The latter case is the model denoted by $\mathcal{B}_1$, the *benchmark model*.

In Section 3.1, we solve for the equilibrium of the model $\mathcal{M}_1$ in closed form. One important implication is that the FTs and STs trade identically on their lagged signal (μ_t is the same for all). Therefore, if we require the FTs to use only their current signal (as in $\mathcal{B}_1$) and introduce an equal number of additional STs, then the aggregate behavior remains essentially the same. Hence, the model $\mathcal{M}_1$ can be regarded as a particular case of $\mathcal{B}_1$, and we are justified in calling $\mathcal{B}_1$ the *benchmark* model with fast and slow traders. This more general model can also be solved in closed form, by using essentially the same formulas as in Section 3.1. We discuss the benchmark model in Section 3.2.

3.1 The Model with Fast and Slow Traders

In this section, we solve for the equilibrium of the model $\mathcal{M}_1$ with fast and slow traders. From (14), the FTs have a strategy of the form $\mathrm{d}x_t = \gamma_t(\mathrm{d}w_t - z_{t,t}) + \mu_t(\mathrm{d}w_{t-1} - z_{t-1,t})$, while the STs have a strategy of the same form, except that μ_t must be zero. The current signal ($\mathrm{d}w_t$) is not predictable from the past order flow, hence the dealer sets $z_{t,t} = 0$. The lagged signal ($\mathrm{d}w_{t-1}$) has already been used by the FTs in the previous

[18] Intuitively, this can occur if the FT must pay a higher processing cost per signal than the ST; see the discussion at the beginning of Section 3.2.

trading round, hence the dealer can use the past order flow to compute the predictable part $z_{t-1,t}$.[19] To simplify notation, let $\widetilde{\mathrm{d}w}_{t-1}$ be the unanticipated part at t of the lagged signal:

$$\widetilde{\mathrm{d}w}_{t-1} \;=\; \mathrm{d}w_{t-1} - z_{t-1,t}. \tag{15}$$

In Theorem 1, we show that there exists a closed-form linear equilibrium of the model. The equilibrium is *symmetric*, in the sense that the FTs have identical trading strategies, and so do the STs. We also provide asymptotic results when the number N_F of fast traders is large. We say that X_∞ is the asymptotic value of a number X which depends on N_F, if the ratio X/X_∞ converges to 1 as N_F approaches infinity, and we write:

$$X \approx X_\infty \quad \iff \quad \lim_{N_F \to \infty} \frac{X}{X_\infty} \;=\; 1. \tag{16}$$

Let "F" refer to the fast traders, and "S" to the slow traders. Denote by N_F the number of fast traders, and by N_S the number of slow traders. We denote the total number of speculators by

$$N_L \;=\; N_F + N_S. \tag{17}$$

This is the same as the number of speculators who use their lagged signals, hence the "L" notation. We also call N_L the *number of lag traders.*

Theorem 1. *Consider the model $\mathcal{M}_1$ with $N_F > 0$ fast traders and $N_S \geq 0$ slow traders; let $N_L = N_F + N_S$. Then, there exists a symmetric linear equilibrium with constant coefficients, of the form ($t \in (0,T]$):*

$$\begin{aligned} \mathrm{d}x_t^F &= \gamma \mathrm{d}w_t + \mu \widetilde{\mathrm{d}w}_{t-1}, & \mathrm{d}x_t^S &= \mu \widetilde{\mathrm{d}w}_{t-1}, \\ \widetilde{\mathrm{d}w}_{t-1} &= \mathrm{d}w_{t-1} - \rho \mathrm{d}y_{t-1}, & \mathrm{d}p_t &= \lambda \mathrm{d}y_t, \end{aligned} \tag{18}$$

where the coefficients γ, μ, ρ, λ are given by:

$$\begin{aligned} \gamma &= \frac{1}{\lambda}\frac{1}{N_F+1}, & \mu &= \frac{1}{\lambda}\frac{1}{N_L+1}\frac{1}{1+b}, \\ \rho &= \frac{\sigma_w}{\sigma_u}\sqrt{(1-a)(a-b^2)}, & \lambda &= \rho\,\frac{N_F}{N_F-b}, \end{aligned} \tag{19}$$

[19] In Theorem 1, we show that that the dealer sets $z_{t-1,t} = \rho \mathrm{d}y_{t-1}$ for some constant coefficient ρ.

with

$$\omega \;=\; 1+\frac{1}{N_F}\,\frac{N_L}{N_L+1}, \qquad b \;=\; \frac{\sqrt{\omega^2+4\frac{N_L}{N_L+1}}\;-\;\omega}{2}, \qquad a \;=\; \frac{N_F-b}{N_F+1}. \tag{20}$$

We have the following asymptotic limits when N_F is large:

$$\omega_\infty \;=\; a_\infty \;=\; 1 \qquad b_\infty \;=\; \frac{\sqrt{5}-1}{2}, \qquad \lambda_\infty \;=\; \rho_\infty \;=\; \frac{\sigma_w}{\sigma_u}\,\frac{1}{\sqrt{N_F}}. \tag{21}$$

The number b is increasing in both N_F and N_S. Moreover, $\omega \in [1,2)$, $a \in (0,1)$, $b \in [0, b_\infty)$.

One consequence of the Theorem is that FTs and STs trade with the same intensity (μ) on their lagged signals. This is true because the current signal $\mathrm{d}w_t$ is uncorrelated with the lagged signal $\widetilde{\mathrm{d}w}_{t-1}$, which implies that the FTs and the STs get the same expression for the expected profit that comes from the lagged signal.[20]

We now discuss some comparative statics regarding the optimal weights γ and μ (for brevity, we omit the proofs). The fast traders' optimal weight γ is decreasing in the number of fast traders, yet it is increasing in the number of slow traders. The first statement simply reflects that, when the number of fast traders is larger, these traders must divide the pie into smaller slices. The same logic applies to the coefficient on the lagged signal: μ is decreasing in both N_F and N_S, as the fast and slow traders compete in trading on their common lagged signal. This last intuition also shows that the fast traders' weight γ is increasing in the number of slow traders. Indeed, when there is more competition from slow traders, the fast traders have an incentive to trade more aggressively on their current signal, as the slow traders have not yet observed this signal.

The next Corollary helps to get more intuition for the equilibrium.

[20]This result does not generalize to the case when there are more lags ($M > 1$). In Internet Appendix I, we see that there is a positive autocorrelation between the signals of higher lags, which reflects a more complicated covariance structure. Mathematically, this translates into the covariance matrix A having non zero entries $A_{i,j}$ when $i > j \geq 1$.

Corollary 1. *In the context of Theorem 1, we have the following formulas:*

$$\lambda\,\bar{\gamma} \;=\; \frac{N_F}{N_F+1}, \qquad \lambda\,\bar{\mu} \;=\; \frac{1}{1+b}\,\frac{N_L}{N_L+1},$$
$$\frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)}{\mathrm{d}t} \;=\; (1-a)\,\sigma_w^2 \;=\; \frac{1+b}{N_F+1}\,\sigma_w^2, \qquad \frac{\mathsf{Cov}\big(\widetilde{\mathrm{d}w_t}, w_t\big)}{\mathrm{d}t} \;=\; \frac{1-a}{1+b}\,\sigma_w^2 \;=\; \frac{\sigma_w^2}{N_F+1}. \tag{22}$$

The first equation in (22) implies that $\lambda\bar{\gamma}\mathrm{d}w_t = \frac{N_F}{N_F+1}\mathrm{d}w_t$, which shows that most of the current signal ($\mathrm{d}w_t$) is incorporated into the price by the fast traders. The intuition comes from the Cournot nature of the equilibrium. Indeed, when trading on the current signal, the benefit of each of each FT increases linearly with the intensity of trading γ on his signal; while the price at which he eventually trades increases linearly with the aggregate quantity demanded. Given that the price impact of the other $N_F - 1$ fast traders aggregates to $\frac{N_F-1}{N_F+1}\mathrm{d}w_t$, the FT is a monopsonist against the residual supply curve, and trades such that his price impact is half of $\frac{2}{N_F+1}\mathrm{d}w_t$, i.e., his price impact equals $\frac{1}{N_F+1}\mathrm{d}w_t$.

After incorporating $\frac{N_F}{N_F+1}\mathrm{d}w_t$ in trading round t, the fast traders must compete with the slow traders for the remaining $\frac{1}{N_F+1}\mathrm{d}w_t$ in the next trading round. As explained before, the speculators must trade a multiple of the unanticipated part of the lagged signal, $\widetilde{\mathrm{d}w_t} = \mathrm{d}w_t - \rho\mathrm{d}y_t$. Thus, when trading on the lagged signal, the benefit of each speculator—fast or slow—increases linearly with the intensity of trading μ, and is proportional to the covariance $\mathsf{Cov}\big(\widetilde{\mathrm{d}w_t}, w_t\big)$. At the same time, each speculator faces a price that increases linearly with the aggregate quantity demanded, and which is proportional to the lagged signal variance $\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)$. The argument is now similar to the Cournot one above, except that everything gets multiplied by the ratio $\mathsf{Cov}(\widetilde{\mathrm{d}w_t}, w_t)/\,\mathsf{Var}(\widetilde{\mathrm{d}w_t})$, which according to (22) is equal to $1/(1+b)$. This justifies the second equation in (22). It also implies that in the case of the lagged signal only a fraction $1/(1+b)$ of it is incorporated by the speculators into the price.

We use the results in Theorem 1 to compute the expected profits of the fast traders and the slow traders.

Proposition 1. *In the context of Theorem 1, the expected profit of the FTs and STs at*

$t = 0$ *from their equilibrium strategies are given, respectively, by:*

$$\begin{aligned} \frac{\pi^F}{\sigma_w^2} &= \frac{\gamma}{N_F+1} + \frac{1}{N_F+1}\,\frac{\mu}{N_L+1}, \\ \frac{\pi^S}{\sigma_w^2} &= \frac{1}{N_F+1}\,\frac{\mu}{N_L+1}. \end{aligned} \tag{23}$$

The ratio of slow profits to fast profits is therefore

$$\frac{\pi^S}{\pi^F} = \frac{1}{1+\frac{(N_L+1)^2(1+b)}{N_F+1}} \implies \frac{\pi^S}{\pi^F} \approx \frac{N_F}{(N_F+N_S)^2}\,\frac{1}{1+b_\infty}. \tag{24}$$

Thus, even if there is only one ST ($N_S = 1$), the ST profits are small compared to the FT profits. The reason is that FTs trade also on their lagged signals, and thus compete with the STs. Indeed, FTs compete for trading on $\mathrm{d}w_t$ only among themselves, while they also compete with the STs for trading on the lagged signal $\widetilde{\mathrm{d}w}_{t-1}$.

Finally, Proposition 1 gives an estimate for the information processing cost δ that would be sufficient to discourage speculators from trading on lagged signals beyond one, if that were not imposed by the model. We state the following numerical result.

Result 1. *Consider the alternative model setup with N_F fast speculators and N_S slow speculators, in which each speculator can use past signals at any lag, but must pay for each signal (used with nonzero weight) an information processing cost of*

$$\delta = \frac{1}{N_F+1}\,\frac{\mu}{N_L+1}\,\sigma_w^2. \tag{25}$$

Then, the alternative model is equivalent to the model with fast and slow traders ($\mathcal{M}_1$).

3.2 The Benchmark Model

We now consider the *benchmark model* $\mathcal{B}_1$, in which the fast traders use only the current signal, while the slow traders use only the lagged signal.[21] The strategies of the FTs

[21] As in Result 1, $\mathcal{M}_{0,1}$ is equivalent to an alternative setup with information processing costs, in which (i) the STs pay the cost δ from (25), while (ii) the FTs pay a cost slightly higher than δ. Indeed, if a FT paid δ, he would be indifferent between using his lagged signal and not using it; while with a slightly higher cost, he would be strictly worse off and would ignore his lagged signal.

and STs are, respectively, of the form

$$\mathrm{d}x_t^F \;=\; \gamma_t \mathrm{d}w_t, \qquad \mathrm{d}x_t^S \;=\; \mu_t \widetilde{\mathrm{d}w}_{t-1}, \tag{26}$$

where $\widetilde{\mathrm{d}w}_{t-1} = \mathrm{d}w_{t-1} - \rho_t \mathrm{d}y_{t-1}$. The dealer sets the price using the rule $\mathrm{d}p_t = \lambda_t \mathrm{d}y_t$. Let $N_F \geq 1$ be the number of FTs and $N_L \geq 0$ the number of STs.

The next result shows that the model $\mathcal{M}_1$ with N_F fast traders and N_S slow traders produces essentially the same outcome as the benchmark model $\mathcal{B}_1$ with N_F fast traders and $N_L = N_F + N_S$ slow traders.

Corollary 2. *Consider (a) the model $\mathcal{M}_1$ with $N_F \geq 1$ fast traders and $N_S \geq 0$ slow traders; and (b) the benchmark model $\mathcal{B}_1$ with N_F fast traders and $N_L = N_F + N_S$ slow traders. Then, the equilibrium coefficients γ, μ, λ, ρ in the two models are identical.*

This Corollary is obtained by simply following the proof of Theorem 1 to solve for the equilibrium in the $\mathcal{B}_1$ model. The key step is to observe that in Theorem 1 the fast trader's choice of μ is the same as the slow trader's choice of μ, and therefore it does not matter who does the optimization, as long as the total number of speculators using their lagged signal is the same.

We finally note that the benchmark model $\mathcal{B}_1$ with $N_F > 0$ fast traders and N_L slow traders has two important particular cases:

- If $N_L \geq N_F$, $\mathcal{B}_1$ is equivalent to the model $\mathcal{M}_1$ with N_F fast traders and $N_S = N_L - N_F$ slow traders;
- If $N_L = 0$, $\mathcal{B}_1$ is the model $\mathcal{M}_0$ (with 0 lags).

4 Market Quality with Fast and Slow Traders

In this section, we study the effect of fast and slow trading on various measures of market quality. The setup is the benchmark model $\mathcal{B}_1$ with $N_F \geq 1$ fast traders and $N_L \geq 0$ slow traders. In this context, "fast trading" is the speculators' aggregate trading on their current signal, and "slow trading" is the speculators' aggregate trading on their lagged signal. The measures of market quality analyzed are illiquidity (measured by the price

impact coefficient), trading volume, price volatility, price informativeness, the speculator participation rate, and the speculator's order flow autocorrelation. The main conclusion of this section is that fast trading has the strongest effect on most of our measures of market quality, while slow trading has a relatively smaller effect. The only measure that depends crucially on slow trading is the speculators' order flow autocorrelation, which becomes positive only in the presence of slow trading. This is shown to be related to anticipatory trading: the order flow coming from fast traders anticipates the order flow coming from the slow traders in the next period.

4.1 Measures of Market Quality

We first decompose the aggregate speculator order flow into fast trading and slow trading. Denote by $\mathrm{d}\bar{x}_t$ be the aggregate speculator order flow. Let $\bar{\gamma}$ be the aggregate weight on the current signal ($\mathrm{d}w_t$), and $\bar{\mu}$ the aggregate weight on the lagged signal ($\widetilde{\mathrm{d}w}_{t-1}$). We decompose the aggregate speculator order flow $\mathrm{d}\bar{x}_t$ into two components: *fast trading*, which represents the aggregate trading on the current signal; and *slow trading*, which represents the aggregate trading on the lagged signal:

$$\mathrm{d}\bar{x}_t \;=\; \underbrace{\bar{\gamma}\,\mathrm{d}w_t}_{\text{Fast Trading}} \;+\; \underbrace{\bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1}}_{\text{Slow Trading}}, \quad \text{with} \quad \bar{\gamma} \;=\; N_F\gamma, \quad \bar{\mu} \;=\; N_L\mu. \tag{27}$$

As in Theorem 1, we define $b = \rho\bar{\mu}$. We call b the *slow trading coefficient*. Then, slow trading exists (is nonzero) only if the number of traders who use their lagged signal is positive, or equivalently if $b > 0$:

$$\text{Slow Trading exists} \qquad \Longleftrightarrow \qquad N_L \;>\; 0 \qquad \Longleftrightarrow \qquad b \;>\; 0. \tag{28}$$

Note that the case when there is no slow trading coincides with the model $\mathcal{M}_0$ with 0 lags from Section 2. In that model, N_F fast traders use only their current signal.

We now define the measures of market quality. Recall that the dealer sets a price that changes in proportion to the total order flow $\mathrm{d}y = \mathrm{d}\bar{x}_t + \mathrm{d}u_t$:

$$\mathrm{d}p_t \;=\; \lambda\,\mathrm{d}y_t \;=\; \lambda\left(\bar{\gamma}\,\mathrm{d}w_t + \bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1} + \mathrm{d}u_t\right), \tag{29}$$

First, as it is standard in the literature, we define *illiquidity* to be the price impact coefficient λ. Thus, the market is considered illiquid if the price impact of a unit of trade is large, i.e., if the coefficient λ is large.

Second, we define *trading volume* as the infinitesimal variance of the aggregate order flow $\mathrm{d}y_t$:

$$TV \;=\; \sigma_y^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}y_t)}{\mathrm{d}t}. \tag{30}$$

We argue that this is a measure of trading volume. Indeed, in each trading round the actual aggregate order flow is given by $\mathrm{d}y_t$. Thus, one can interpret trading volume as the absolute value of the order flow: $|\mathrm{d}y_t|$. From the theory of normal variables, the average trading volume is given by $\mathsf{E}\big(|\mathrm{d}y_t|\big) = \sqrt{\frac{2}{\pi}}\,\sigma_y$. With our definition $TV = \sigma_y^2$, we observe that TV is monotonic in $\mathsf{E}\big(|\mathrm{d}y_t|\big)$, and thus TV can be used a measure of trading volume. Using (29), we compute the trading volume in our model by the formula

$$TV \;=\; \bar{\gamma}^2\,\sigma_w^2 + \bar{\mu}^2\,\sigma_{\widetilde{w}}^2 + \sigma_u^2, \quad \text{with} \quad \sigma_{\widetilde{w}}^2 \;=\; \frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)}{\mathrm{d}t}. \tag{31}$$

The trading volume measure TV can be decomposed into the speculator trading volume and the noise trading volume:

$$TV \;=\; TV^s + TV^n, \quad \text{with} \quad TV^s \;=\; \bar{\gamma}^2\,\sigma_w^2 + \bar{\mu}^2\,\sigma_{\widetilde{w}}^2, \quad TV^n \;=\; \sigma_u^2. \tag{32}$$

Third, we define *price volatility* σ_p to be the square root of the instantaneous price variance:

$$\sigma_p \;=\; \left(\frac{\mathsf{Var}(\mathrm{d}p_t)}{\mathrm{d}t}\right)^{1/2}. \tag{33}$$

From (29), it follows that the instantaneous price variance can be computed simply as the product of the illiquidity measure λ and the trading volume $TV = \sigma_y^2$. Thus,

$$\sigma_p^2 \;=\; \lambda^2\; TV \;=\; \lambda^2\,\Big(\bar{\gamma}^2\,\sigma_w^2 + \bar{\mu}^2\,\sigma_{\widetilde{w}}^2 + \sigma_u^2\Big). \tag{34}$$

Fourth, we define *price informativeness* as a measure inversely related to the forecast error variance $\Sigma_t = \mathsf{Var}\big((w_t - p_{t-1})^2\big)$. Thus, if prices are informative, they stay close to the forecast w_t, i.e., the variance Σ_t is small. In Internet Appendix I, in the general

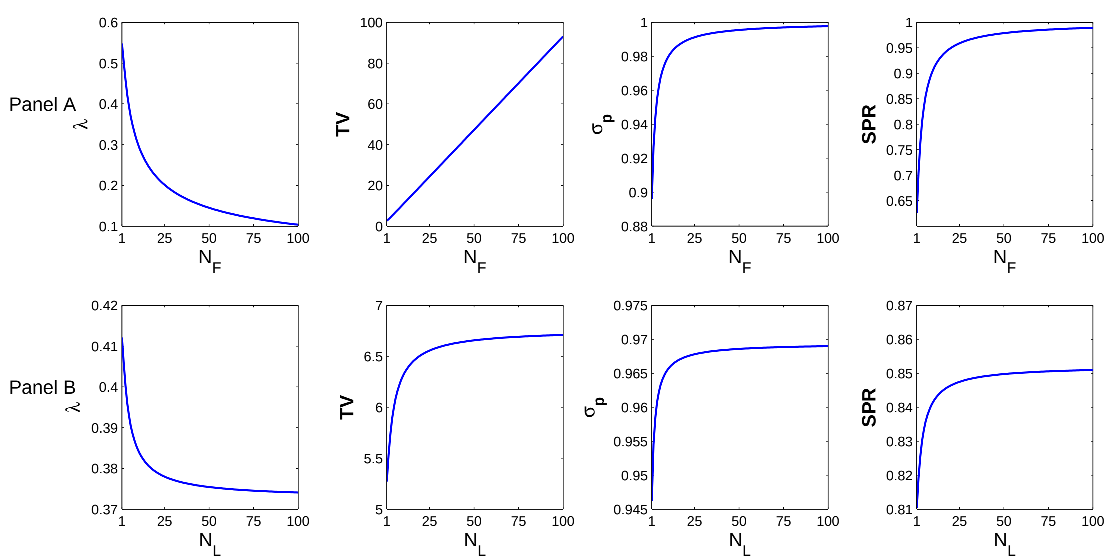
model with at most m lagged signals ($\mathcal{M}_m$) we show that Σ_t evolves according to $\Sigma_t' = \sigma_w^2 - \sigma_p^2$, where σ_p^2 is the price variance (Proposition I.1). Therefore, since Σ_t' is inversely monotonic in the price variance, we do not use it as a separate measure of market quality.

Fifth, the *speculator participation rate* is defined as the ratio of speculator trading volume over total trading volume:

$$SPR \;=\; \frac{TV^s}{TV} \;=\; \frac{\bar{\gamma}^2\,\sigma_w^2 + \bar{\mu}^2\,\sigma_{\tilde{w}}^2}{\bar{\gamma}^2\,\sigma_w^2 + \bar{\mu}^2\,\sigma_{\tilde{w}}^2 + \sigma_u^2}. \tag{35}$$

SPR can also be interpreted as the fraction of price variance due to the speculators.

Figure 2: Market Quality with Fast and Slow Traders. This figure plots the following measures of market quality: (i) illiquidity λ; (ii) trading volume TV; (iii) price volatility σ_p; and (iv) speculator participation rate SPR. Panel A plots the dependence of the four market quality measures on the number of fast traders N_F, while taking the number of slow traders $N_L = 5$. Panel B plots the dependence of the four market quality measures on N_L, while taking $N_F = 5$. The other parameters are $\sigma_w = 1$, $\sigma_u = 1$.



4.2 Comparative Statics on Market Quality

We now give explicit formulas for our measures of market quality. As before, we use asymptotic notation when N_F is large: $X \approx Y$ stands for $\lim\limits_{N_F \to \infty} \frac{X}{Y} = 1$.

Proposition 2. *Consider the benchmark model with $N_F \geq 1$ fast traders and $N_L \geq 0$ slow traders. Then, the price impact coefficient, trading volume, price volatility, and speculator participation rate satisfy:*

$$\begin{aligned} \lambda &= \frac{\sigma_w}{\sigma_u} \frac{\sqrt{(1+b)(a-b^2)}}{\sqrt{N_F+1}} \frac{N_F}{N_F-b}, \qquad & TV &= \sigma_u^2 (N_F+1) \frac{a}{(1+b)(a-b^2)}, \\ \sigma_p^2 &= \sigma_w^2 \frac{N_F^2}{(N_F+1)(N_F-b)}, \qquad & SPR &= a + \frac{b^2(1+b)}{N_F-b}, \end{aligned} \tag{36}$$

where $b^2 + b\big(1 + \frac{1}{N_F} \frac{N_L}{N_L+1}\big) = \frac{N_L}{N_L+1}$, *and* $a = \frac{N_F-b}{N_F+1}$.

Panel A of Figure 2 shows how the four measures of market quality vary with the number of fast traders N_F, while holding the number of slow traders N_L constant. Panel B of Figure 2 shows how the four measures of market quality vary with N_L, while holding N_F constant. We find that all four market quality measures vary in the same direction with respect to N_F and N_L. Nevertheless, the number of fast traders has a much stronger effect on these measures than the number of slow traders.

To get more intuition about the effect of fast trading on market quality, we consider the simplest case, when $N_L = 0$. Since all speculators trade only on their current signal, this case coincides with the model $\mathcal{M}_0$ as defined in Section 2. In this model there is no slow trading ($\bar{\mu} = 0$), hence the slow trading coefficient b is zero. Moreover, $a = \frac{N_F-b}{N_F+1} = \frac{N_F}{N_F+1}$. Thus, we can solve the model $\mathcal{M}_0$ by simply using Proposition 2. Nevertheless, it is instructive to solve for the equilibrium of $\mathcal{M}_0$ independently.

Proposition 3. *Consider the model $\mathcal{M}_0$, with N_F fast traders whose trading strategy is of the form $\mathrm{d}x_t = \gamma_t \mathrm{d}w_t$. Then, the optimal coefficient γ is constant and equal to $\gamma = \frac{1}{\lambda} \frac{1}{N_F+1} = \frac{\sigma_u}{\sigma_w} \frac{1}{\sqrt{N_F}}$. The price impact coefficient, trading volume, price volatility, and speculator participation rate satisfy, respectively,*

$$\lambda = \frac{\sigma_w}{\sigma_u} \frac{\sqrt{N_F}}{N_F+1}, \qquad TV = \sigma_u^2 (N_F+1), \qquad \sigma_p^2 = \sigma_w^2 \frac{N_F}{N_F+1}, \qquad SPR = \frac{N_F}{N_F+1}. \tag{37}$$

Using Proposition 3, we now discuss in more detail the effect of the number N_F of fast traders on the measures of market quality. First, we note by quickly inspecting the formulas in Proposition 3, that we obtain the same qualitative results as those displayed

in Figure 2. Namely, illiquidity is decreasing in N_F, while the other three measures are increasing in N_F.

An important consequence of Proposition 3 is that in our model the speculator participation rate can be made arbitrarily close to 1 if the number of fast traders is large. Thus, noise trading volatility is only a small part of the total volatility. This stands in sharp contrast for instance with the models of Kyle (1985) or Back, Cao, and Willard (2000), in which virtually *all* instantaneous price volatility is generated by the noise traders at the high frequency limit (in continuous time).

The market is more efficient when the number of fast traders is large. Indeed, in the proof of Proposition 3 we show that the rate of change of the forecast error variance Σ' is constant and equal to $\frac{\sigma_w^2}{N_F+1}$. Since by assumption there is no initial informational asymmetry ($\Sigma_0 = 0$), it follows that $\Sigma_t \leq \frac{\sigma_w^2}{N_F+1}$ for all t. In other words, the price stays close to the fundamental value at all times. Thus, a larger number N_F of fast traders, rather than destabilizing the market, makes the market more efficient.

The trading volume TV strongly increases with the number of fast traders. This occurs because of the competition among FTs make them trade more aggressively. By trading more aggressively, FTs reveal more information, which as we see later lowers the traders' price impact. This has an amplifier effect on the trading aggressiveness, such that the trading volume grows essentially linearly in the number of speculators (see equation (37)). Moreover, the speculator participation rate SPR also increases in N_F, since SPR is the fraction of trading volume caused by the speculators.

Surprisingly, a larger number of fast traders make the market more liquid, as more information is revealed when there are more competing speculators. This appears to be in contradiction with the fact that more informed trading should increase the amount of adverse selection. To understand the source of this apparent contradiction, note that illiquidity is measured by the price impact λ of one unit of volume. But, while the trading volume TV strongly increases in N_F in an unbounded way, its price impact is bounded by magnitude of the signal $\mathrm{d}w_t$.[22] Thus, the price impact *per unit of volume* actually decreases, indicating that prices are more informative. This makes the market

[22] In Internet Appendix I, we make this intuition rigorous in the general case; see the discussion surrounding Proposition I.4.

overall more liquid. This result is consistent with the empirical studies of Zhang (2010), Hendershott, Jones, and Menkveld (2011), and Boehmer, Fong, and Wu (2014).

To understand the effect of fast traders on the price volatility σ_p, consider the pricing formula $\mathrm{d}y_t = \lambda \mathrm{d}y_t$, which implies $\sigma_p^2 = \lambda^2 TV$. There are two effects of N_F on the price volatility σ_P. First, the trading volume TV strongly increases in N_F, which has a positive effect on σ_P. Second, price impact λ decreases in N_F, which has a negative effect on σ_P. The first effect is slightly stronger than the second, hence the net effect is that price volatility σ_P increases in N_F. This result is consistent with the empirical studies of Boehmer, Fong, and Wu (2014) and Zhang (2010).

A few caveats are in order. First, all these studies analyze the effects of HFT activity, where activity is proxied either by turnover or by intensity of order-related message traffic, and *not* by the number of HFTs present in the market. An answer to this concern is that, as we have seen, trading volume does increase in N_F. Second, in our paper we do not model *passive* HFTs, that is, HFTs that offer liquidity via limit orders. Therefore, it is possible that an increase in the number of passive HFTs decreases price volatility, which would cancel the opposite effect of the number of *active* HFTs. For instance, Hasbrouck and Saar (2012) document a negative effect of HFTs on volatility, possibly because they also consider passive HFTs, which by providing liquidity have a stabilizing effect on price volatility. Moreover, Chaboud, Chiquoine, Hjalmarsson, and Vega (2014) find essentially no relation. In our model, the dependence of price volatility on N_F is weak, which may explain the mixed results in the empirical literature.

Next, we discuss how the various measures of market quality depend on the speculators' signal precision σ_w. Note that, according to equation (5), the signal precision is related to the fundamental volatility σ_v by a monotonic relation: $\sigma_w = \frac{\sigma_v}{(1+\sigma_\eta^2/\sigma_v^2)^{1/2}}$. Therefore, we only analyze the dependence of market quality on signal precision, while keeping in mind that these results apply equally to the fundamental volatility.

The price volatility σ_p increases in signal precision, indicating that speculators trade more aggressively when they have a more precise signal. Indeed, σ_p is the volatility of $\mathrm{d}p_t$, which is the price impact of the aggregate order flow. In particular, the order flow coming from the FTs has an aggregate price impact which is proportional to $\mathrm{d}w_t$.[23]

[23]From Proposition 3, the FTs' order flow equals $\lambda N_F \gamma \mathrm{d}w_t = \frac{N_F}{N_F+1}\mathrm{d}w_t$.

Thus, price volatility increases in the signal precision.

A larger signal precision σ_w generates more adverse selection between fast traders and the dealer, hence the illiquidity λ is increasing in the signal precision. However, the trading volume TV is independent of σ_w. To get some intuition for this result, note that $TV = \frac{\sigma_p^2}{\lambda^2}$. Since both the numerator and denominator increase with signal precision, the net effect is ambiguous. Proposition 3 shows that the two effects exactly offset each other.

4.3 Order Flow Autocorrelation and Anticipatory Trading

We start by analyzing the autocorrelation of the components of the order flow. Since the dealer is competitive and risk neutral, the total order flow $\mathrm{d}y_t$ has zero autocorrelation. But because the dealer cannot identify the part of the order flow that comes from speculators, the *speculator* order flow can in principle be autocorrelated.

As in Section 4.1, in the benchmark model with fast and slow traders, the aggregate speculator order flow decomposes into its fast trading and slow trading components:

$$\mathrm{d}\bar{x}_t \;=\; \underbrace{\mathrm{d}\bar{x}_t^F}_{\text{Fast Trading}} \;+\; \underbrace{\mathrm{d}\bar{x}_t^S}_{\text{Slow Trading}}, \quad \text{with} \quad \mathrm{d}\bar{x}_t^F \;=\; \bar{\gamma}\,\mathrm{d}w_t, \quad \mathrm{d}\bar{x}_t^S \;=\; \bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1}, \tag{38}$$

with $\bar{\gamma} = N_F \gamma$ and $\bar{\mu} = N_L \mu$. As before, we say that slow trading exists if $b = \rho\bar{\mu} > 0$, or equivalently $N_L > 0$.

We define *speculator order flow autocorrelation* by $\mathsf{Corr}\big(\mathrm{d}\bar{x}_t, \mathrm{d}\bar{x}_{t+1}\big)$. Because $\mathrm{d}\bar{x}_{t+1}^F$ is orthogonal to both components of $\mathrm{d}\bar{x}_t^F$, we obtain the decomposition:

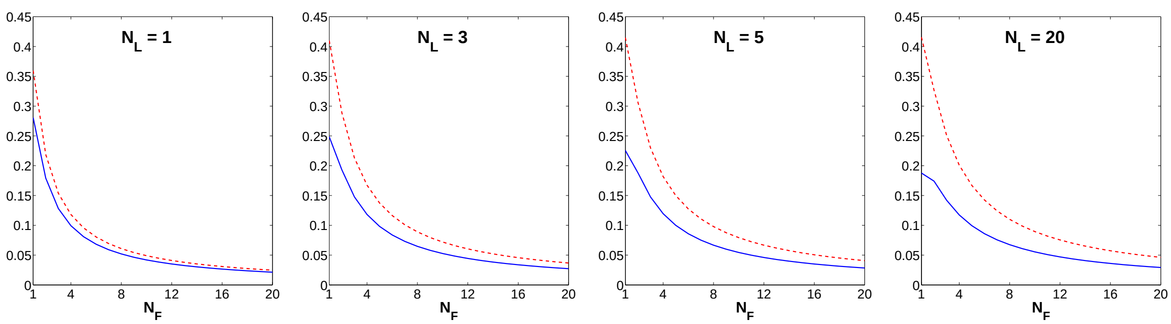
$$\rho_{\bar{x}} \;=\; \mathsf{Corr}\big(\mathrm{d}\bar{x}_t, \mathrm{d}\bar{x}_{t+1}\big) \;=\; \underbrace{\frac{\mathsf{Cov}\big(\mathrm{d}\bar{x}_t^F, \mathrm{d}\bar{x}_{t+1}^S\big)}{\mathsf{Var}(\mathrm{d}\bar{x}_t)}}_{\text{Anticipatory Trading}} \;+\; \underbrace{\frac{\mathsf{Cov}\big(\mathrm{d}\bar{x}_t^S, \mathrm{d}\bar{x}_{t+1}^S\big)}{\mathsf{Var}(\mathrm{d}\bar{x}_t)}}_{\text{Expectation Adjustment}}. \tag{39}$$

We denote the *anticipatory trading* part by ρ_{AT} and the *expectation adjustment* part by ρ_{EA}. The first component arises because fast trading at t anticipates slow trading at $t+1$. Indeed, there is a positive correlation between fast trading at t and slow trading at $t+1$ ($\bar{\mu}\widetilde{\mathrm{d}w}_t$). The second component arises because slow trading at $t+1$ is based on lagged signals, adjusted by subtracting the dealer's expectation which incorporates past

lagged signals. Because of this expectation adjustment, we see below that the slow order flow is negatively autocorrelated. Formally, slow trading at $t+1$ ($\bar{\mu}\widetilde{\mathrm{d}w_t}$) is proportional to the lagged signal minus dealer's expectation, $\widetilde{\mathrm{d}w_t} = \mathrm{d}w_t - \rho \mathrm{d}y_t$. But the dealer's expectation is proportional on the total order flow at t, which includes the previous slow trading ($\mathrm{d}y_t = \bar{\gamma}\, \mathrm{d}w_t + \bar{\mu}\, \widetilde{\mathrm{d}w_{t-1}} + \mathrm{d}u_t$). We compute:

$$\rho_{\bar{x}} = \rho_{AT} + \rho_{EA}, \quad \text{with} \quad \rho_{AT} = \bar{\mu}\bar{\gamma}\, \frac{\mathsf{Var}\big(\mathrm{d}w_t\big)}{\mathsf{Var}\big(\mathrm{d}\bar{x}_t\big)}, \quad \rho_{EA} = -\rho\bar{\mu}^3\, \frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_{t-1}}\big)}{\mathsf{Var}\big(\mathrm{d}\bar{x}_t\big)}. \tag{40}$$

Figure 3: Speculator Order Flow Autocorrelation. This figure plots the speculator order flow autocorrelation $\rho_{\bar{x}}$ (solid line) and the anticipatory trading component ρ_{AT} (dashed line) as a function of the number N_F of fast traders. The four graphs correspond to four values of the number N_L of speculators using their lagged signal: $N_L = 1, 3, 5, 20$.



Proposition 4. *Consider the benchmark model with $N_F \geq 1$ fast traders and $N_L \geq 0$ slow traders. Then, the speculator order flow autocorrelation and its components satisfy*

$$\rho_{\bar{x}} = \frac{b(b+1)(a-b^2)}{a^2+b^2(1-a)}\, \frac{1}{N_F+1}, \qquad \frac{\rho_{AT}}{\rho_{\bar{x}}} = \frac{a}{a-b^2}, \qquad \frac{\rho_{EA}}{\rho_{\bar{x}}} = -\frac{b^2}{a-b^2}, \tag{41}$$

where a and b are as in Proposition 2. Moreover, $\rho_{\bar{x}}$ is strictly positive if and only if there exists slow trading, i.e., $N_L > 0$.

One implication of Proposition 4 is that, as long as there exists slow trading, the speculator order flow autocorrelation $\rho_{\bar{x}}$ is nonzero. To understand why, note that both the anticipatory trading component and the expectation adjustment component depend on the existence of slow trading. Formally, if there is no slow trading, $\bar{\mu} = 0$ implies that both components of the speculator order flow autocorrelation are zero.

Figure 3 shows how the speculator order flow autocorrelation ($\rho_{\bar{x}}$) and its anticipatory trading component (ρ_{AT}) depend on the number of fast traders (N_F) for four different values of the number of slow traders ($N_L = 1, 3, 5, 20$). We see that both $\rho_{\bar{x}}$ and ρ_{AT} are decreasing in N_F. Indeed, when the number of fast traders is large, there is only $\frac{1}{N_F+1}$ of the signal left in the next period for the slow traders. Hence, one should expect the autocorrelation to decrease by the order of $\frac{1}{N_F+1}$, which is indeed the case. For instance, when $N_L = 5$, we see that the speculator order flow autocorrelation is 22.56% when there is one FT, but decreases to 2.84% when there are 20 FTs. Our results are consistent with the empirical literature on HFTs. For instance, Brogaard (2011) finds that the autocorrelation of aggregate HFT order flow is small but positive.

The anticipatory trading component ρ_{AT} is increasing in the number of slow traders N_L (to see this, fix for instance $N_F = 10$ in each of the four graphs in Figure 3). The intuition is simple: when the number of slow traders is larger, fast trading in each period can better predict the slow trading the next period, hence the correlation ρ_{AT} is larger. Using Nasdaq data on high-frequency traders, Hirschey (2013) finds that HFT order flow anticipates non-HFT order flow. But Nasdaq defines HFTs along several criteria including the use of large trading volume and low inventories. In our model, these are indeed the characteristics of fast traders, but not those of slow traders (see the next section for a discussion about traders' inventories). Thus, if in our model we classified fast traders as HFTs and slow traders as non-HFTs, our previous results would imply that HFT order flow anticipates non-HFT order flow.

5 Inventory Management

In this section, we analyze the inventory problem of fast traders. Because the benchmark model cannot address this problem (when speculators are risk-neutral, their inventory follows a random walk), we modify the model by introducing an additional trader with inventory costs.[24] We call this new trader the *Inventory-averse Fast Trader*, or IFT, and the resulting setup the *model with inventory management*, or the *model with an IFT*.

[24] Introducing more than one inventory-averse trader makes the problem considerably more complicated, as the number of state variables increases.

To get intuition for the model with inventory management, we first solve for the optimal strategy of the IFT in a partial equilibrium framework, taking as fixed the behavior of the other speculators and the dealer. The solution of this problem is provided in closed form. Then, we continue with a general equilibrium analysis. We show that the equilibrium reduces to a non-linear equation in one variable, which can be solved numerically. We then study the properties of the general equilibrium, and the effect of the inventory management on market quality.

5.1 Model

To define the model with inventory management, we consider a setup similar to the benchmark model, but we replace one risk-neutral fast trader with an inventory-averse fast trader (IFT). Specifically, the IFT maximizes an expected utility U of the form (recall that $T = 1$):

$$U \;=\; \mathsf{E}\left(\int_0^T (v_1 - p_t)\mathrm{d}x_t\right) \;-\; C_I\,\mathsf{E}\left(\int_0^T x_t^2 \mathrm{d}t\right), \tag{42}$$

where x_t is his inventory in the risky asset, and $C_I > 0$ is a constant. We call C_I the *inventory aversion coefficient*. We do not identify the exact source of inventory costs for the IFT, but these can be thought to arise either from capital constraints or from risk aversion.

In this model, there are N_F fast traders, N_L slow traders, and one IFT. The equilibrium concept is similar to the linear equilibrium from Section 2. But, because the inventory problem is very difficult in a more general formulation, we assume directly that the speculators' strategies have constant coefficients, and that the dealer has pricing rules as in the benchmark model. Thus, the fast trader $i = 1, \ldots, N_F$ and the slow trader $j = 1, \ldots, N_L$ have strategies, respectively, of the form:

$$\mathrm{d}x_{i,t}^F \;=\; \gamma_i \mathrm{d}w_t, \qquad \mathrm{d}x_{j,t}^S \;=\; \mu_j \widetilde{\mathrm{d}w}_{t-1}. \tag{43}$$

The dealer has pricing rules of the form:

$$\mathrm{d}p_t \;=\; \lambda \mathrm{d}y_t, \qquad z_{t-1,t} \;=\; \rho \mathrm{d}y_{t-1}, \tag{44}$$

where $\mathrm{d}y_t$ is the aggregate order flow at t, and $z_{t-1,t} = \mathsf{E}_t(\mathrm{d}w_{t-1})$ is the dealer's expectation of the current signal given the past order flow. The coefficient λ is chosen so that the dealer breaks even, meaning that her expected profit is zero.[25]

Since the IFT has quadratic inventory costs, it is plausible to expect that his optimal trading strategy is linear in the inventory.[26] Therefore, we assume that the IFT's strategy is of the following type:

$$\mathrm{d}x_t \;=\; -\Theta\, x_{t-1} + G\, \mathrm{d}w_t, \tag{45}$$

with constant coefficients $\Theta \in [0,2)$ and $G \in \mathbb{R}$. Equivalently, the IFT's inventory x_t follows an $AR(1)$ process

$$x_t \;=\; \phi\, x_{t-1} + G\, \mathrm{d}w_t, \qquad \phi \;=\; 1 - \Theta, \tag{46}$$

with autoregressive coefficient $\phi \in (-1, 1]$.[27]

If $\Theta > 0$, in each trading round the IFT removes a fraction Θ of his current inventory, with the goal of bringing his inventory eventually to zero. One measure of how quickly the inventory mean reverts to zero is the *inventory half life*. This is defined as the average number of periods (of length $\mathrm{d}t$) that the process needs to halve the distance from its mean, i.e.,

$$\text{Inventory Half Life} \;=\; \frac{\ln(1/2)}{\ln(\phi)}\,\mathrm{d}t \;=\; \frac{\ln(1/2)}{\ln(1-\Theta)}\,\mathrm{d}t. \tag{47}$$

[25] Note that because of inventory management, the aggregate order flow is no longer completely unpredictable by the dealer. Nevertheless, the only source of predictability is the IFT's inventory, and, as we prove later, this inventory in equilibrium is very small because of fast mean reversion. Moreover, not being able to properly compute the expectation of IFT's inventory does not mean that the dealer loses money. Indeed, we have assumed that the dealer chooses λ so that her expected profit is zero.

[26] This is standard in the literature. See for instance Madhavan and Smidt (1993), but also Hendershott and Menkveld (2014), or Ho and Stoll (1981).

[27] A standard result is that the $AR(1)$ process becomes explosive (with infinite mean and variance) if ϕ is outside $[-1, 1]$, or equivalently if Θ is outside $[0, 2]$.

Hence, the inventory half life is of the order of $\mathrm{d}t$. This in practice can be short (minutes, seconds, milliseconds), which means that when $\Theta > 0$ the IFT does very quick, "real-time" inventory management.

We end this section with a brief discussion of the different types of inventory management. In Section 5.2 we will see that there is a discontinuity between the cases $\Theta = 0$ and $\Theta > 0$. To explain this discontinuity, we introduce a new case, $\Theta = 0_+$, in which the IFT mean reverts his inventory, but much more smoothly (formal details are below). It turns out that this intermediate inventory management regime indeed connects continuously the other two. Thus, there are three different cases (regimes):

- $\Theta = 0$, the *neutral regime*: the IFT's strategy is of the form $\mathrm{d}x_t = G\mathrm{d}w_t$, similar to the strategy of a (risk-neutral) fast trader.

- $\Theta > 0$, the *fast regime*: the IFT's strategy is of the form $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$. In this regime, the inventory half life is of the order of $\mathrm{d}t$.

- $\Theta = \theta\mathrm{d}t$, the *smooth regime*: the IFT's strategy is of the form $\mathrm{d}x_t = -\theta x_{t-1}\mathrm{d}t + G\mathrm{d}w_t$, with $\theta \in (0, \infty)$.[28] In this regime, the inventory half life $\frac{\ln(1/2)}{\ln(1-\theta\mathrm{d}t)}\,\mathrm{d}t = \frac{\ln(2)}{\theta}$, which is much larger than the inventory half life in the fast regime.

The smooth regime is discussed in detail in Internet Appendix K. We find that indeed the smooth regime connects continuously the cases $\Theta = 0$ (neutral regime) with the case $\Theta > 0$ (fast regime).[29] However, we show that the smooth regime is not optimal for the IFT when there is enough slow trading (this is true for instance if the $N_L \geq 2$ and $N_F \geq 1$). Therefore, in the rest of the paper we assume that there is enough slow trading, and ignore the smooth regime.

5.2 Optimal Inventory Management

In this section, we do a partial equilibrium analysis, and solve for the optimal strategy of the IFT while fixing the behavior of the other players. This allows us to get insight

[28]This is called an Ornstein-Uhlenbeck process.

[29]More precisely, $\theta = 0$ in the smooth regime coincides with $\Theta = 0$; while the limit when $\theta \nearrow \infty$ in the smooth regime coincides with the limit when $\Theta \searrow 0$ in the fast regime.

about the IFT's behavior, without having to do a full equilibrium analysis. We leave this more general analysis to Section 5.3.

Consider the inventory management model with one IFT, N_F fast traders and N_L slow traders. Let γ, μ be the coefficients arising from the strategies of the FTs and STs (not necessarily optimal), and λ, ρ the coefficients from the dealer's pricing rules. Define additional *model coefficients* by:

$$R = \frac{\lambda}{\rho}, \qquad \gamma^- = N_F \gamma, \qquad \bar{\mu} = N_L \mu, \qquad a^- = \rho \gamma^-, \qquad b = \rho \bar{\mu}. \tag{48}$$

Proposition 5 analyzes the inventory management regime, where by definition the IFT has a trading strategy with positive mean reversion ($\Theta > 0$). For this result, the strategy need not be optimal.

Proposition 5. *In the inventory management model, let* $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$ *be the IFT's strategy (not necessarily optimal), with* $\Theta > 0$*. Suppose* $b \in (-1, 1)$*. Then, the IFT has zero inventory costs, and all his expected profits are in cash. His expected profit* π *satisfies:*

$$\pi = \lambda \left(\bar{\mu} G \frac{1 - a^-}{1 + \phi b} - G^2 \frac{b + \frac{1}{1+\phi}}{1 + \phi b} \right) \sigma_w^2. \tag{49}$$

Because the IFT reduces his inventory by a fraction $\Theta > 0$ in each trading round, his inventory decays exponentially on average. As our model is set at the high frequency limit (in continuous time), the decrease in IFT inventory is very quick, and the inventory remains infinitesimal at all times.[30]

In general, the expected profit π of any speculator satisfies:

$$\pi = \mathsf{E} \int_0^T (v_T - p_t) \mathrm{d}x_t = \underbrace{\mathsf{E}\big(v_T \, x_T\big)}_{\text{Risky Component}} + \underbrace{\mathsf{E} \int_0^T (-p_t) \, \mathrm{d}x_t}_{\text{Cash Component}}. \tag{50}$$

The *risky component* is the expected profit due to the accumulation of inventory in the risky asset. This does not translate into cash profits until the liquidation date, $T = 1$. The *cash component* is the expected profit that comes from changes in the cash account due to trading. Because the IFT has an infinitesimal inventory, his risky component of

[30]Mathematically, the average squared inventory $\mathsf{E}(x_t^2)$ is of the order of $\mathrm{d}t$; see equation (A32).

profits is negligible. Hence, all IFT profits come from the cash component, as stated in Proposition 5.

As a result of keeping all his profits in cash, the behavior of the IFT is very different than the behavior of a risk-neutral speculator. Indeed, while the risk-neutral speculator trades directly on his private information, the IFT benefits only indirectly, from timing his trades and unloading his inventory to slower traders.

To understand why the IFT behaves differently, suppose he observes a new signal $\mathrm{d}w_t$. Initially, the IFT trades on his signal ($G\mathrm{d}w_t$), but subsequently he fully reverses his trade by unloading a positive fraction of his inventory each period. Therefore, the only way for the IFT to make money is to ensure that the inventory reversal is done at a profit. This can occur for instance if the IFT expects that when he sells, other traders buy even more, and as a result his overall price impact is negative. But this is only possible if there exist slow traders, whose lagged signals can be predicted by the IFT.

In general, the expected profit of a speculator who manages inventory satisfies the following formula:[31]

$$\pi \;=\; \mathsf{E}\int_0^T x_{t-1}\,\mathrm{d}p_t. \tag{51}$$

Thus, inventory management is profitable only when the speculator can use his past inventory (x_{t-1}) to forecast the current price change ($\mathrm{d}p_t$). In particular, this formula explains why the IFT trades at $t-1$ an amount $G\mathrm{d}w_{t-1}$ even though he knows that subsequently he will fully reverse his trade. He trades like this because his signal $\mathrm{d}w_{t-1}$ anticipates the slow trading at t, which in turn affects $\mathrm{d}p_t$. To make this intuition more precise, the next result specializes the formula (51) to our model.

Corollary 3. *In the context of Proposition 5, the IFT's expected profit satisfies:*

$$\pi \;=\; \mathsf{E}\int_0^T x_{t-1}\left(\lambda\,\bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1}\right) - \lambda\Theta\;\mathsf{E}\int_0^T x_{t-1}^2. \tag{52}$$

If there is no slow trading ($\bar{\mu}=0$), the IFT's makes negative expected profits.

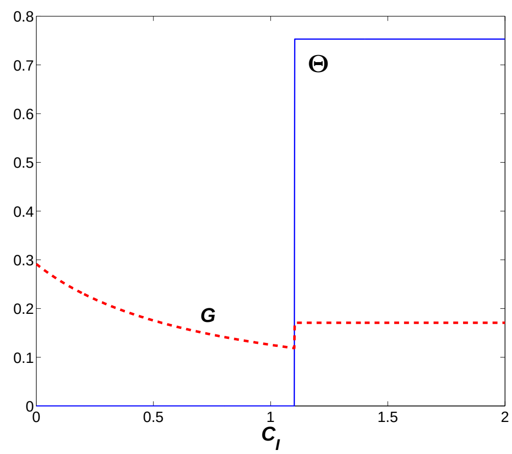
Using Corollary 3, we see that the IFT's speculative trade $G\mathrm{d}w_{t-1}$ is part of the

[31] For the IFT, see equation (A35). The result is true in general when the speculator has infinitesimal inventory. Indeed, if we integrate $\mathrm{d}(x_t p_t) = p_t\mathrm{d}x_t + x_{t-1}\mathrm{d}p_t$, we get $x_T p_T$, which is zero in expectation since x_T is infinitesimal. Hence, $\pi = \mathsf{E}\int_0^T(-p_t)\mathrm{d}x_t = \mathsf{E}\int_0^T x_{t-1}\mathrm{d}p_t$.

IFT's inventory x_{t-1}, and is also correlated with the price change $\mathrm{d}p_t$ via the slow trading component of the order flow $\bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1}$. Without slow trading ($\bar{\mu} = 0$), there is no correlation, hence no revenue source for the IFT. Therefore, the IFT makes negative profits on average, as he loses from the price impact of his trades.

The main result of this section describes the IFT's optimal strategy when there is enough slow trading, i.e., the slow trading coefficient b is above a threshold.

Figure 4: Optimal IFT Inventory Management. This figure plots the coefficients of the IFT's optimal trading strategy ($\mathrm{d}x_t = -\Theta\, x_{t-1} + G\,\mathrm{d}w_t$) in the inventory management model with $N_F = 5$ fast traders and $N_L = 5$ slow traders. On the horizontal axis is the IFT's inventory aversion, C_I. The parameter values are $\sigma_w = 1$, $\sigma_u = 1$. For the model coefficients, we use the equilibrium values from Section 5.3: $a^- = 0.7088$, $b = 0.5424$, $\lambda = 0.3782$, $\rho = 0.3439$. The formulas for G, Θ, and $\bar{C}_I$ are from Theorem 2.



Theorem 2. *In the inventory management model, suppose the model coefficients satisfy $0 \leq a^-, b < 1$ and $\lambda, \rho > 0$. In addition, suppose $b > \frac{\sqrt{17}-1}{8} = 0.3904$.*[32] *Let $\bar{C}_I = 2\lambda\left(\frac{(1-Ra^-)^2(1+\sqrt{1-b})^2}{R^2 b(1-a^-)^2} - 1\right)$. Then, if $C_I < \bar{C}_I$, the optimal strategy of the IFT is to set*

$$\Theta \;=\; 0, \qquad G \;=\; \frac{1-Ra^-}{2\lambda + C_I}. \tag{53}$$

If $C_I > \bar{C}_I$, the optimal strategy of the IFT is to set

$$\Theta \;=\; 2 - \frac{\sqrt{1-b}}{b} \;\in\; (0,2), \qquad G \;=\; \frac{1-a^-}{2\rho\left(1+\frac{1}{\sqrt{1-b}}\right)}. \tag{54}$$

[32] In equilibrium (section 5.3) we have the following numerical results: the condition $b < 1$ is always satisfied, and the condition $b > \frac{\sqrt{17}-1}{8}$ is equivalent to having (i) $N_L \geq 2$ and (ii) $N_L \geq 6$ if $N_F = 0$.

Thus, there are two different types of behavior (regimes) for the IFT, depending on how his inventory aversion compares to a threshold value ($\bar{C}_I$).

- (*Neutral regime*) If the inventory aversion coefficient is small (below $\bar{C}_I$), the IFT sets $\Theta = 0$ and controls his inventory by choosing his weight G. As his inventory aversion gets larger, the IFT reduces his inventory costs by decreasing G. The tradeoff is that a smaller G also reduces expected profits. The behavior of the IFT when $\Theta = 0$ is essentially the same as the behavior of a FT.

- (*Fast regime*) If the inventory aversion is large (above $\bar{C}_I$), the IFT manages his inventory by choosing a positive mean reversion coefficient ($\Theta > 0$). There is no longer a tradeoff between expected profit and inventory costs, as the IFT has zero inventory costs. Hence, the IFT chooses the weight G and the mean reversion Θ to maximize expected profit (more details below).

Theorem 2 implies that a small change in IFT's inventory aversion can have a large effect on the IFT's behavior. Figure 4 plots the coefficients of the optimal strategy when $N_F = 5$, $N_S = 5$. We see that when the IFT's inventory aversion rises above the threshold $\bar{C}_I = 0.1021$, his optimal mean reversion coefficient jumps from $\Theta = 0$ to $\Theta = 0.7530$. Also, his optimal weight jumps from $G = 0.1186$ (the left limit of G at the threshold) to $G = 0.1708$ (the constant value of G above the threshold).

The sharp discontinuity between the two regimes arises because the IFT has zero inventory costs in the fast regime ($\Theta > 0$). Let $U^*_{\Theta>0}$ and $U^*_{\Theta=0}$ be the maximum expected utility of the IFT respectively when he manages inventory versus when he does not. Because the IFT has zero inventory costs in the fast regime, $U^*_{\Theta>0}$ does not depend on C_I; while in the neutral regime $U^*_{\Theta=0}$ is decreasing in C_I and is precisely equal to $U^*_{\Theta>0}$ at the threshold value $C_I = \bar{C}_I$.[33] This implies that the neutral regime is optimal when C_I is below the threshold, while the fast regime is optimal when C_I is above the threshold.

Note that a necessary condition for Theorem 2 is the existence of enough slow trading. Formally, the slow trading coefficient must be larger than the threshold $b = 0.3904$, which numerically is true for instance if there are $N_F \geq 1$ fast traders and $N_L \geq 2$ slow traders. If slow trading is below the threshold, we show that a similar analysis holds,

[33] Formally, these statements follow from equations (A43) and (A39) in the Appendix.

but with the fast regime replaced by the *smooth regime*, in which the IFT still manages inventory but with a strategy of the form: $\mathrm{d}x_t = -\theta x_t \mathrm{d}t + G\mathrm{d}w_t$. (See Section J.1 in the Internet Appendix.) To simplify presentation, we assume that there is enough slow trading, and ignore the smooth regime in the rest of the paper.

Using our results, we predict that in practice fast speculators are sharply divided into two categories. In both categories speculators generate large trading volume. But in one category the speculators make fundamental bets and accumulate inventories, while in the other category speculators mean revert their inventories very quickly, and keep their profits in cash. Our results appear consistent with the "opportunistic traders" and the "high frequency traders" described in Kirilenko *et al.* (2014). Both opportunistic traders and HFTs have large volume and appear to be fast. But while opportunistic traders have relatively large inventories, the HFTs in their sample (during several days around the Flash Crash of May 6, 2010) liquidate 0.5% of their aggregate inventories on average each second. This implies that HFT inventories have an $AR(1)$ half life of a little over 2 minutes.

We finish this section with a brief discussion of how the IFT's optimal strategy is correlated with slow trading. Corollary 3 shows that if there is no slow trading, the IFT cannot make positive profits. Theorem 2 shows that with enough slow trading, the IFT can manage inventory and make positive profits (see equation (A43) in the Appendix). In the previous discussion, we have argued that this is possible only if the IFT trades in the opposite direction to the slow trading. We now prove this is indeed the case.

Corollary 4. *In the context of Theorem 2, suppose the IFT is sufficiently averse ($C_I > \bar{C}_I$). Denote by $\mathrm{d}\bar{x}_t^S = \bar{\mu}\widetilde{\mathrm{d}w}_{t-1}$ the slow trading component of the speculator order flow. Then, the IFT's optimal strategy is negatively correlated with slow trading:*

$$\mathsf{Cov}\big(\mathrm{d}x_t, \mathrm{d}\bar{x}_t^S\big) \;=\; -\Theta\,\mathsf{Cov}\big(x_{t-1}, \mathrm{d}\bar{x}_t^S\big) \;<\; 0. \tag{55}$$

We call this phenomenon the *hot potato effect*, or the *intermediation chain effect*. The intuition is that the IFT's current signal generates undesirable inventory and must be passed on to slower traders in order to produce a profit. The passing of inventory can be thought as the beginning of an intermediation chain. Kirilenko *et al.* (2014) and

Weller (2014) document such hot potato effects among high frequency traders.

5.3 Equilibrium Results

In this section, we solve for the full equilibrium of the inventory management model. For simplicity, we assume that the IFT is *sufficiently averse*, meaning that his inventory aversion is above a certain threshold (formally, above the threshold value $\bar{C}_I$ from Theorem 2). Then, the solution can be expressed almost in closed form, except for the slow trading coefficient b, which satisfies a non-linear equation in one variable.

Theorem 3. *Consider the inventory management model with one sufficiently averse IFT, N_F fast traders, and N_L slow traders. Suppose there is an equilibrium in which the speculators's strategies are: $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$ (the IFT), $\mathrm{d}x_t^F = \gamma \mathrm{d}w_t$ (the FTs), $\mathrm{d}x_t^S = \mu \widetilde{\mathrm{d}w}_{t-1}$ (the STs); and the dealer's pricing rules are: $\mathrm{d}p_t = \lambda \mathrm{d}y_t$, $\widetilde{\mathrm{d}w}_t = \mathrm{d}w_t - \rho \mathrm{d}y_t$. Denote the model coefficients R, a^-, b as in (48). Suppose $\frac{\sqrt{17}-1}{8} < b < 1$. Then, the equilibrium coefficients satisfy equations (A44)–(A46) from the Appendix.*

Conversely, suppose the equations (A44)–(A46) have a real solution such that $\frac{\sqrt{17}-1}{8} < b < 1$, $a < 1$, $\lambda > 0$. Then, the speculators' strategies and the dealer's pricing rules with these coefficients provide an equilibrium of the model.

Rather than relying on numerical results to study the equilibrium, we start by providing asymptotical results when the number of FTs and STs is large. The advantage is that the asymptotic results can be expressed in closed form, and thus help provide a clearer intuition for the equilibrium. Let $\bar{C}_I$ be the threshold aversion from Theorem 2. Let π be the expected profit of a sufficiently averse IFT ($C_I \geq \bar{C}_I$), and $\pi^{C_I=0}$ the maximum expected profit of a risk-neutral IFT ($C_I = 0$), where the behavior of the other speculators and the dealers is taken to be the same. Let γ_0 the benchmark FT weight, and $\pi_0^F = \frac{\gamma_0}{N_F+2}\,\sigma_w^2$ the benchmark profit of a FT, as in Proposition (1). We use the asymptotic notation: $X \approx X_\infty$ stands for $\lim\limits_{N_F, N_L \to \infty} \frac{X}{X_\infty} = 1$.

Proposition 6. *Consider (i) the inventory management model with one sufficiently averse IFT, N_F fast traders, and N_L slow traders, and (ii) the benchmark model with $N_F + 1$ fast traders and N_L slow traders. Then, the equilibrium coefficients γ, μ, λ, ρ*

are asymptotically equal across the two models when N_F and N_L are large. Also, $a \approx 1$, $b \approx b_\infty = 0.6180$, *and we have the following asymptotic formulas:*

$$\begin{aligned} &\Theta \;\approx\; 1, \qquad \frac{G}{\gamma_0} \;\approx\; 1 - b_\infty \;=\; 0.3820, \qquad \frac{\pi}{\pi_0^F} \;\approx\; 2b_\infty - 1 \;=\; 0.2361, \\ &\frac{\pi}{\pi^{C_I=0}} \;\approx\; \tfrac{4}{5}\, b_\infty \;=\; 49.44\%, \qquad \bar{C}_I \;\approx\; \tfrac{1+5b_\infty}{2}\,\lambda_\infty \;\approx\; 2.0451\,\frac{\sigma_w}{\sigma_u}\,\frac{1}{\sqrt{N_F+1}}. \end{aligned} \tag{56}$$

The first implication of Proposition 6 is that model with inventory management is asymptotically the same as the benchmark model when both N_F and N_L are large. This is not surprising, since when there are many other speculators, the IFT has a relatively smaller and smaller role in the limit.

The behavior of the IFT is more surprising. First, when there are many other speculators, the IFT's inventory mean reversion becomes extreme (Θ approaches 1). This means that the IFT's inventory half life becomes essentially zero, as the IFT removes most of his inventory each period. This extreme mean reversion is possible because the existence of a sufficient amount of slow trading allows the hot potato effect to generate positive profits for the IFT. Furthermore, the equation $\pi \approx 49.44\% \times \pi^{C_I=0}$ implies that even under extreme inventory mean reversion ($\Theta = 1$) the IFT can trade so that he only loses on average only about 50% of his maximum expected profits when he has zero inventory aversion.[34]

The equation $\bar{C}_I \approx 2.0451\,\frac{\sigma_w}{\sigma_u}\,\frac{1}{\sqrt{N_F+1}}$ implies that the threshold inventory aversion above which the IFT chooses to mean revert his inventory becomes very small when the number of competing fast traders is large. This is perhaps counterintuitive, since one may think that the IFT chooses fast inventory mean reversion because he has very high inventory aversion. This is not the case, however. Indeed, even when the IFT has small inventory aversion, a sufficient amount of slow trading is enough to convince the IFT to engage in very fast inventory mean reversion. This is because inventory management is a zero/one proposition. Once the IFT engages in inventory management ($\Theta > 0$), any profits from fundamental bets become zero, and the hot potato effect is the sole source of profits.

[34]This recalls the saying attributed to Joseph Kennedy (the founder of the Kennedy dynasty) that "*I would gladly give up half my fortune if I could be sure the other half would be safe.*"

We now compare the IFT with the other speculators. For the IFT, we consider the following variables: (i) IFT's trading volume, measured by the his order flow variance $TV_x = \mathsf{Var}(\mathrm{d}x_t)/\mathrm{d}t$, as in Section 4.1, (ii) IFT's order flow autocorrelation, $\rho_x = \mathsf{Corr}(\mathrm{d}x_t, \mathrm{d}x_{t+1})$; and (iii) $\beta_{x,\bar{x}^S} = \mathsf{Cov}(\mathrm{d}x_t, \mathrm{d}\bar{x}_t^S)/\mathsf{Var}(\mathrm{d}\bar{x}_t^S)$, which is the regression coefficient of the IFT's strategy $(\mathrm{d}x_t)$ on the slow trading component $(\mathrm{d}\bar{x}_t^S)$. We are also interested in the individual FT volume, TV_{x^F}; the aggregate FT volume, $TV_{\bar{x}^F}$; and the aggregate ST volume, $TV_{\bar{x}^S}$; the aggregate FT order flow autocorrelation, $\rho_{\bar{x}^F}$; and the aggregate ST order flow autocorrelation, $\rho_{\bar{x}^S}$.

The next result computes all these quantities, and provides asymptotic results when both N_F and N_L are large. Some of these results provide new testable implications, regarding the relationship between trading volume, order flow covariance, and inventory.

Proposition 7. *In the context of Theorem 3, consider a sufficiently averse IFT. Then, the variables defined above satisfy the following formulas:*

$$
\begin{aligned}
&\frac{TV_x}{TV_{x^F}} = \frac{2G^2}{(1+\phi)\gamma^2} \approx 4 - 6b_\infty = 0.2918, \quad \frac{TV_{\bar{x}^S}}{TV_{\bar{x}^F}} = \frac{b^2(1-a)}{(a^-)^2} \approx \frac{b_\infty}{N_F+1}, \\
&\rho_x = -\frac{\Theta}{2} \approx -\frac{1}{2}, \qquad \rho_{\bar{x}^F} = 0, \qquad \rho_{\bar{x}^S} \approx -b_\infty = -0.6180, \\
&\beta_{x,\bar{x}^S} = -\frac{\Theta(1-a^-)}{2b(1+2\sqrt{1-b})} \approx -\frac{3+b_\infty}{5(N_F+1)} = -\frac{0.7236}{N_F+1}.
\end{aligned}
\tag{57}
$$

The last result illustrates the hot potato effect. The IFT's order flow has a negative beta on the STs' aggregate order flow, which means that the IFT and the STs trade in opposite directions. As the number of FTs becomes larger, there is more information released to the public by the trades of the fast traders, hence there is less room for slow trading. As a result, the hot potato effect is less intense when there is a large number of FTs.

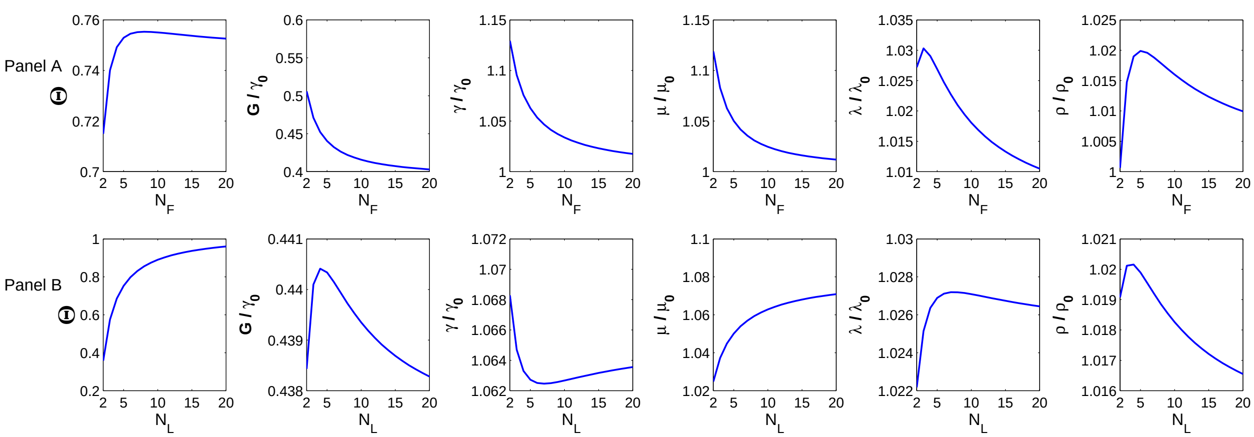
Proposition 7 implies that in the limit when N_F and N_L are large, the IFT's trading volume is about 30% of the individual FT trading volume. This implies that the IFT's trading volume is comparable to that of a regular FT. By contrast, just as in the benchmark model, the volume coming from STs is much smaller than the volume coming from FTs. This confirms our intuition that in an empirical analysis which selects traders based on volume, the IFT and the FTs are in the category with large trading volume,

while the STs are in the category with small trading volume.

If we compare order flow autocorrelations, we see that the IFT is similar to the STs, but not to the FTs. Indeed, the IFT and the STs have negative and large order flow autocorrelation. By contrast, the FTs have zero order flow autocorrelation.[35] Finally, if we compare inventories, the IFT has infinitesimal inventory, while the variance of the other speculators' inventory increases over time.[36] Nevertheless, the STs' inventories are smaller relative to the FTs' inventories, since the STs have smaller volume.

We now present some numerical results for the equilibrium coefficients. Figure 5 plots the equilibrium coefficients (Θ, G, γ, μ, λ, ρ). We normalize some variables X in the inventory management model by the corresponding variable X_0 in the benchmark model. Panel A of Figure 5 plots the variables against N_F, while holding N_L constant. Panel B plots the same variables against N_L, while holding N_F constant.[37]

Figure 5: Equilibrium Coefficients with Inventory Management. This figure plots the equilibrium coefficients that arise in the inventory management model. Some variables X are normalized by the corresponding variable X_0 in the benchmark model. The coefficients are Θ, G, γ, μ, λ, ρ. Panel A plots the dependence of the six variables on the number of fast traders N_F, while taking the number of slow traders $N_S = 5$. Panel B plots the dependence of the six variables on N_S, while taking $N_F = 5$. The other parameters are $\sigma_w = 1$, $\sigma_u = 1$.



As expected, we find that the mean reversion coefficient Θ is increasing in the number

[35]Even if we allowed FTs to trade on lagged signals, one can see that the FTs would still have very small order flow autocorrelation (of the order of $\frac{1}{N_F+1}$) because of their large trading volume.

[36]For the IFT, $\mathsf{Var}(x_t) = \frac{G^2}{1-\phi^2}\,\sigma_w^2\,\mathrm{d}t$ (see equation (A25) in the Appendix); while for the FT, $\mathsf{Var}(x_t^F) = t\sigma_w^2$, as the FT's inventory follows a random walk.

[37]We consider $N_F, N_L \geq 2$. The reason is that in order to apply Theorem 2, we need to have $b > \frac{\sqrt{17}-1}{8}$. This is true in equilibrium if $N_F, N_L \geq 2$.

of slow traders N_L. This is because the IFT needs slow traders in order to make profits. The IFT's weight G is less than half the benchmark weight γ_0, indicating that the IFT shifts towards inventory management in order to make profits. This leaves more room for fundamental profits, which explains why both the FTs and the STs are better off with inventory management than in the benchmark model (γ/γ_0 and μ/μ_0 are both above one), despite the price impact λ being larger than in the benchmark (we see that $\lambda/\lambda_0 > 1$). The reason why the market is more illiquid in the inventory management model is that the IFT trades much less intensely on his signal (G is less than half of γ_0), and therefore the informational efficiency is lower. To see directly that the market is less informationally efficient in the inventory management model, we use the fact that in our model price volatility is a proxy for informational efficiency (see the discussion in Section 4). Then, we verify numerically that indeed $\sigma_p/\sigma_{p,0} < 1$, which implies that with inventory management the market is less informationally efficient.

6 Conclusion

We have presented a theoretical model in which traders continuously receive signals over time about the value of an asset, but only use each signal for a finite number of lags (which can be justified by an information processing cost per signal). We have found that competition among speculators reveals much private information to the public, and the value of information decays fast. Therefore, a trader who is just one instant slower than the other traders loses the majority of the profits by being slow. Another consequence is that the market is very efficient and liquid. As a feedback effect, because of the small price impact (high market liquidity), the informed traders are capable of trading even more aggressively. In equilibrium, the fast speculator trading volume is very large and dominates the overall trading volume. We have also considered an extension of the model in which a fast speculator, called the inventory-averse fast trader (IFT), has quadratic inventory costs. We find that a sufficiently averse IFT has a very different behavior compared to a risk-neutral fast trader. The IFT keeps his profits in cash, makes no fundamental bets on the value of the risky asset, and quickly passes his inventory to "slow traders," who use their lagged signals. This hot potato effect is

possible because the existence of slower traders more than reverses the price impact of the IFT.

Appendix A. Proofs

Notation Preliminaries

Recall that $t+1$ is notation for $t + \mathrm{d}t$, and

$$T \;=\; 1. \tag{A1}$$

In general, a tilde above a symbol denotes normalization by σ_w. For instance, if σ_u is the instantaneous volatility of the noise trader order flow, and σ_y the instantaneous volatility of the total order flow, we denote by

$$\tilde{\sigma}_u \;=\; \frac{\sigma_u}{\sigma_w}, \qquad \tilde{\sigma}_y \;=\; \frac{\sigma_y}{\sigma_w}, \quad \text{with} \quad \sigma_u^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}u_t)}{\mathrm{d}t}, \qquad \sigma_y^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}y_t)}{\mathrm{d}t}. \tag{A2}$$

Consider a trading strategy $\mathrm{d}x_t$ for $t \in (0, T]$, where $T = 1$. Denote by $\tilde{\pi}$ the normalized expected profit at $t = 0$ from the strategy $\mathrm{d}x_t$:

$$\tilde{\pi} \;=\; \frac{1}{\sigma_w^2}\,\mathsf{E}\left(\int_0^T (w_t - p_t)\mathrm{d}x_t\right). \tag{A3}$$

For variances and covariances, a tilde the symbol means normalization by both σ_w^2 and $\mathrm{d}t$. For instance, denote by

$$\widetilde{\mathsf{Var}}\big(\widetilde{\mathrm{d}w_t}\big) \;=\; \frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2 \mathrm{d}t} \;=\; A_t, \qquad \widetilde{\mathsf{Cov}}\big(w_t, \widetilde{\mathrm{d}w_t}\big) \;=\; \frac{\mathsf{Cov}\big(w_t, \widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2 \mathrm{d}t} \;=\; B_t. \tag{A4}$$

Proof of Theorem 1. We look for an equilibrium with the following properties: (i) the equilibrium is symmetric, in the sense that the FTs have identical trading strategies, and the same for the STs; (ii) the equilibrium coefficients are constant with respect to time.

To solve for the equilibrium, in the first step we take the dealer's pricing functions as given, and solve for the optimal trading strategies for the FTs and STs. In the second

step, we take the speculators' trading strategies as given, and we compute the dealer's pricing functions. In Section 2, we have assumed that the speculators take the signal covariance structure as given (see equation (13)). In the current context, this means that the speculators take the following covariances as given and constant:

$$A_t \;=\; \widetilde{\mathsf{Var}}\big(\widetilde{\mathrm{d}w_t}\big) \;=\; \frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\mathrm{d}t}, \qquad B_t \;=\; \widetilde{\mathsf{Cov}}\big(w_t, \widetilde{\mathrm{d}w_t}\big) \;=\; \frac{\mathsf{Cov}\big(w_t, \widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\mathrm{d}t} \tag{A5}$$

Thus, in the rest of the Appendix we consider that the dealer also sets A and B, in addition to setting λ and ρ.

Speculators' Optimal Strategy (γ, μ)

Since we search for an equilibrium with constant coefficients, we assume that the speculators take as given the dealer's pricing rules $\mathrm{d}p_t = \lambda \mathrm{d}y_t$ and $z_{t-1,t} = \rho \mathrm{d}y_{t-1}$, and also the covariances $A = \widetilde{\mathsf{Var}}(\widetilde{\mathrm{d}w_t})$ and $B = \widetilde{\mathsf{Cov}}(w_t, \widetilde{\mathrm{d}w_t})$.

Consider a FT, indexed by $i = 1, \ldots, N_F$. He chooses $\mathrm{d}x_t^i = \gamma_t^i \mathrm{d}w_t + \mu_t^i \widetilde{\mathrm{d}w_{t-1}}$, and assumes that at each $t \in (0, T]$, the price satisfies:

$$\mathrm{d}p_t \;=\; \lambda\, \mathrm{d}y_t, \quad \text{with} \quad \mathrm{d}y_t \;=\; \big(\gamma_t^i + \gamma_t^{-i}\big)\mathrm{d}w_t + \big(\mu_t^i + \mu_t^{-i}\big)\widetilde{\mathrm{d}w_{t-1}} + \mathrm{d}u_t, \tag{A6}$$

where the superscript "$-i$" indicates the aggregate quantity from the other speculators. Since $\mathrm{d}w_t$ and $\widetilde{\mathrm{d}w_{t-1}}$ are both orthogonal on the public information set $\mathcal{I}_t$, and $p_{t-1} \in \mathcal{I}_t$, it follows that $\mathrm{d}x_t^i$ is orthogonal to p_{t-1} as well. The normalized expected profit of FT i at $t = 0$ satisfies:

$$\begin{aligned} \tilde{\pi}^F \;&=\; \frac{1}{\sigma_w^2}\,\mathsf{E}\int_0^T \Big(w_t - p_{t-1} - \lambda\Big((\gamma_t^i + \gamma_t^{-i})\mathrm{d}w_t + (\mu_t^i + \mu_t^{-i})\widetilde{\mathrm{d}w_{t-1}} + \mathrm{d}u_t \Big)\Big)\mathrm{d}x_t^i \\ &=\; \gamma_t^i - \lambda\gamma_t^i\big(\gamma_t^i + \gamma_t^{-i}\big) + \mu_t^i B - \lambda\mu_t^i\big(\mu_t^i + \mu_t^{-i}\big)A. \end{aligned} \tag{A7}$$

This is a pointwise optimization problem, hence it is enough to consider the profit at $t = 0$, and maximize the expression over γ_t^i and μ_t^i. The solution of this problem is $\lambda\gamma_t^i = \frac{1-\lambda\gamma_t^{-i}}{2}$, and $\lambda\mu_t^i = \frac{B/A-\lambda\mu_t^{-i}}{2}$. The ST $j = 1, \ldots, N_S$ solves the same problem, only that his coefficient on $\mathrm{d}w_t$ is $\gamma_t^j = 0$. Thus, all γ's are equal for the FTs, and all μ's are equal for the FTs and STs. We also find that they are constant, and since

$N_L = N_F + N_S$, we have

$$\gamma \;=\; \frac{1}{\lambda}\,\frac{1}{1+N_F}, \qquad \mu \;=\; \frac{B/A}{\lambda}\,\frac{1}{1+N_L}. \tag{A8}$$

Dealer's Pricing Rules (λ, ρ, A, B)

The dealer takes the speculators' strategies as given, and assumes that the aggregate order flow is of the form:

$$\mathrm{d}y_t \;=\; \mathrm{d}u_t + \bar{\gamma}\,\mathrm{d}w_t + \bar{\mu}\,\widetilde{\mathrm{d}w}_{t-1}, \quad \text{with} \quad \bar{\gamma} \;=\; N_F\,\gamma, \quad \bar{\mu} \;=\; N_L\,\mu. \tag{A9}$$

Moreover, the dealer assumes that, in their trading strategy, the speculators set:

$$\widetilde{\mathrm{d}w}_{t-1} \;=\; \mathrm{d}w_{t-1} - \rho^*\,\mathrm{d}y_{t-1}. \tag{A10}$$

Naturally, later we require that in equilibrium the dealer's pricing coefficient ρ coincides with the coefficient ρ^* used by the speculators.

Since the order flow $\mathrm{d}y_t$ is orthogonal to the dealer's information set $\mathcal{I}_t$, the dealer sets $\lambda_t, \rho_t, A_t, B_t$ such that the following equations are satisfied:

$$\begin{aligned}
\lambda_t \;&=\; \frac{\widetilde{\mathsf{Cov}}(w_t,\mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\bar{\gamma}+\bar{\mu}\,B_{t-1}}{\sigma_{y,t}^2}, \qquad \mathrm{d}p_t \;=\; \lambda_t \mathrm{d}y_t,\\
\rho_t \;&=\; \frac{\widetilde{\mathsf{Cov}}(\mathrm{d}w_t,\mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\bar{\gamma}}{\sigma_{y,t}^2}, \qquad \widetilde{\mathrm{d}w}_t \;=\; \mathrm{d}w_t - \rho_t \mathrm{d}y_t,\\
\sigma_{y,t}^2 \;&=\; \widetilde{\mathsf{Var}}(\mathrm{d}y_t^2) \;=\; \tilde{\sigma}_u^2 + \bar{\gamma}^2 + \bar{\mu}^2\,A_{t-1},\\
B_t \;&=\; \widetilde{\mathsf{Cov}}\big(w_t, \mathrm{d}w_t - \rho^*\mathrm{d}y_t\big) \;=\; (1-\rho^*\bar{\gamma}) - \rho^*\bar{\mu}B_{t-1},\\
A_t \;&=\; \widetilde{\mathsf{Var}}\big(\mathrm{d}w_t - \rho^*\mathrm{d}y_t\big) \;=\; 1 - 2\rho^*\bar{\gamma} + (\rho^*)^2\sigma_{y,t}^2\\
&=\; 1 - 2\rho^*\bar{\gamma} + (\rho^*)^2(\tilde{\sigma}_u^2+\bar{\gamma}^2) + (\rho^*)^2\bar{\mu}^2\,A_{t-1}.
\end{aligned} \tag{A11}$$

Consider the last equation in (A11), $A_t = 1 - 2\rho^*\bar{\gamma} + (\rho^*)^2(\tilde{\sigma}_u^2+\bar{\gamma}^2) + (\rho^*)^2\bar{\mu}^2\,A_{t-1}$, which is a recursive equation in A_t. Then, Lemma A.1 (below) implies that A does not depend on t, as long as $|\rho^*\bar{\mu}| < 1$. But, since the dealer takes the speculators' strategies as given, we can use the equilibrium condition $\rho^*\bar{\mu} = b \in (0,1)$. The same method shows that

B does not depend on t. Moreover, Lemma A.1 can be used to compute the constant values of A and B:

$$A \;=\; \frac{(1-\rho^*\bar{\gamma})^2 + (\rho^*)^2\tilde{\sigma}_u^2}{1-(\rho^*\bar{\mu})^2}, \qquad B \;=\; \frac{1-\rho^*\bar{\gamma}}{1+\rho^*\bar{\mu}}. \tag{A12}$$

Then, equation (A11) shows that λ, ρ, $\tilde{\sigma}_y$ are independent on t as well.

Equilibrium Conditions

We now use the equations derived above to solve for the equilibrium values of γ, μ, λ, $\rho = \rho^*$, A, B, $\tilde{\sigma}_y$. Denote by

$$a \;=\; \rho\,\bar{\gamma}, \qquad b \;=\; \rho\,\bar{\mu}, \qquad R \;=\; \frac{\lambda}{\rho}. \tag{A13}$$

From (A12) we have $A = \frac{(1-a)^2+\rho^2\tilde{\sigma}_u^2}{1-b^2}$. Then, substitute A in $\tilde{\sigma}_y^2 = \tilde{\sigma}_u^2 + \bar{\gamma}^2 + \bar{\mu}^2 A$ from (A11), to obtain $\rho^2\tilde{\sigma}_y^2 = \frac{\rho^2\tilde{\sigma}_u^2+(a^2+b^2-2ab^2)}{1-b^2}$. To summarize,

$$B \;=\; \frac{1-a}{1+b}, \quad A \;=\; \frac{(1-a)^2+\rho^2\tilde{\sigma}_u^2}{1-b^2}, \quad \rho^2\tilde{\sigma}_y^2 \;=\; \frac{\rho^2\tilde{\sigma}_u^2+(a^2+b^2-2ab^2)}{1-b^2}. \tag{A14}$$

Using (A11), we get $R = \frac{\lambda}{\rho} = \frac{\bar{\gamma}+\bar{\mu}B}{\bar{\gamma}} = \frac{a+b\frac{1-a}{1+b}}{a} = \frac{a+b}{a(1+b)}$. Also, the equation for ρ implies $\rho = \frac{\bar{\gamma}}{\tilde{\sigma}_y^2} = \frac{\rho a}{\rho^2\tilde{\sigma}_y^2}$. Using the formula for $\rho^2\tilde{\sigma}_y^2$ in (A14), we compute $\rho^2\tilde{\sigma}_u^2 = (1-a)(a-b^2)$. Using this formula, we obtain $\rho^2\tilde{\sigma}_y^2 = a$ and $A = 1-a$. To summarize,

$$R \;=\; \frac{\lambda}{\rho} \;=\; \frac{a+b}{a(1+b)}, \qquad \rho^2\tilde{\sigma}_u^2 \;=\; (1-a)(a-b^2), \qquad \rho^2\tilde{\sigma}_y^2 \;=\; a, \qquad A = 1-a. \tag{A15}$$

From (A8), we have $\frac{N_F}{N_F+1} = \lambda\bar{\gamma} = \frac{\lambda}{\rho}\,a = \frac{a+b}{1+b}$. From this, $a = \frac{N_F-b}{N_F+1}$, and $B = \frac{1-a}{1+b} = \frac{\frac{1+b}{N_F+1}}{1+b} = \frac{1}{N_F+1}$. Also, $\frac{B}{A}\,\frac{N_L}{N_L+1} = \lambda\bar{\mu} = \frac{\lambda}{\rho}\,b = \frac{b(a+b)}{a(1+b)}$. Since $\frac{B}{A} = \frac{1}{1+b}$, we have $\frac{N_L}{N_L+1} = \frac{b(a+b)}{a}$, or $\frac{a}{b(1+b)}\,\frac{N_L}{N_L+1} = \frac{a+b}{1+b}$. The two formulas for $\frac{a+b}{1+b}$ imply $b(1+b)\frac{N_F}{N_F+1} = a\frac{N_L}{N_L+1}$. To summarize,

$$a \;=\; \frac{N_F-b}{N_F+1}, \qquad B \;=\; \frac{1}{N_F+1}, \qquad b(1+b)\frac{N_F}{N_F+1} \;=\; \frac{N_F-b}{N_F+1}\,\frac{N_L}{N_L+1}. \tag{A16}$$

From $\frac{\lambda}{\rho}\,a = \frac{N_F}{N_F+1}$ and $a = \frac{N_F-b}{N_F+1}$, we get $\frac{\lambda}{\rho} = \frac{N_F}{N_F-b}$, as stated.

From (A16), we obtain the quadratic equation $b^2 + b\omega = \frac{N_L}{N_L+1}$, with $\omega = 1 + \frac{1}{N_F}\frac{N_L}{N_L+1}$. One solution of this quadratic equation is $b = \frac{\omega + \left(\omega + 4\frac{N_L}{N_L+1}\right)^{1/2}}{2} \geq 1$, which leads to a negative $\tilde{\sigma}_y^2$ (see (A14)). Thus, we must choose the other solution, $b = \frac{-\omega + \left(\omega + 4\frac{N_L}{N_L+1}\right)^{1/2}}{2} \geq 0$. Let $b_\infty = \frac{\sqrt{5}-1}{2}$. Since $b_\infty^2 + b_\infty = 1$ and $\omega \geq 1$, we have $b_\infty^2 + b_\infty\omega \geq 1$. Moreover, since $b^2 + b\omega = \frac{N_L}{N_L+1} < 1$, we get $b^2 + b\omega < b_\infty^2 + b_\infty\omega$. But the function $b^2 + b\omega$ is strictly increasing in b when $b \geq 0$, hence we obtain $b < b_\infty$. Thus, $b \in [0, b_\infty)$, as stated in the Theorem. We also obtain $a = \frac{N_F - b}{N_F+1} \in (0,1)$. The proof of the exact formulas in (19) is now complete.

We now derive the asymptotic formulas in (19). When N_F is large, note that $a = \frac{N_F}{N_F - b} \approx a_\infty = 1$, $\omega = 1 + \frac{1}{N_F}\frac{N_L}{N_L+1} \approx \omega_\infty = 1$. Therefore, we also get $b \approx b_\infty = \frac{\sqrt{5}-1}{2}$. One can now verify that the formulas for γ_∞, μ_∞, λ_∞, and ρ_∞ are as stated in (19).

We now show how b depends on N_F and N_L (the dependence on N_S is the same as the dependence on $N_L = N_F + N_S$). Consider the function $F(\beta, \omega) = \sqrt{\omega^2 + 4\beta} - \omega$, and note that $b = F(\beta, \omega)/2$, with $\beta = \frac{N_L}{N_L+1}$ and $\omega = 1 + \frac{\beta}{N_F}$. We compute $\frac{\partial \beta}{\partial N_F} = \frac{\partial \beta}{\partial N_L} = \frac{1}{(N_L+1)^2}$, $\frac{\partial \omega}{\partial N_F} = -\frac{N_L(N_L+1) - N_F}{N_F^2(N_L+1)^2} < 0$, $\frac{\partial \omega}{\partial N_L} = \frac{1}{N_F(N_L+1)^2} > 0$. Also, $\frac{\partial F}{\partial \beta} = \frac{2}{\sqrt{\omega^2 + 4\beta}} > 0$, and $\frac{\partial F}{\partial \omega} = \frac{\beta}{\sqrt{\omega^2+4\beta}} - 1 = -\frac{b}{\sqrt{\omega^2+4\beta}} < 0$. Then, $\frac{\partial(2b)}{\partial N_F} = \frac{\partial F}{\partial \beta} \cdot \frac{\partial \beta}{\partial N_F} + \frac{\partial F}{\partial \omega} \cdot \frac{\partial \omega}{\partial N_F} > 0$, and $\frac{\partial(2b)}{\partial N_L} = \frac{\partial F}{\partial \beta} \cdot \frac{\partial \beta}{\partial N_L} + \frac{\partial F}{\partial \omega} \cdot \frac{\partial \omega}{\partial N_L} = \frac{1}{(N_L+1)^2\sqrt{\omega^2+4\beta}}\left(2 - \frac{b}{N_F}\right) > 0$, where the last inequality follows from $b \in (0,1)$.

We end the analysis of the equilibrium conditions, by proving several more useful inequalities for a and b. Denote by $\beta_F = \frac{N_F}{N_F+1}$ and recall that $\beta = \frac{N_L}{N_L+1}$. Then, b satisfies the quadratic equation $b^2 + b\omega = \beta$, with $\omega = 1 + \frac{\beta}{N_F}$. Now start with the straightforward inequality $\beta < \beta_F + 1$, and multiply it by β_F. We get $\beta\beta_F < \beta_F^2 + \beta_F$. Since $\beta_F = 1 - \frac{\beta_F}{N_F}$, we get $\beta(1 - \frac{\beta_F}{N_F}) < \beta_F^2 + \beta_F$, or equivalently $\beta < \beta_F^2 + \beta_F(1 + \frac{\beta}{N_F})$. Since $b^2 + b\omega = \beta$ and $\omega = 1 + \frac{\beta}{N_F}$, we get $b^2 + b\omega < \beta_F^2 + \beta_F\omega$. Because the function $f(x) = x^2 + x\omega$ is increasing in $x \in (0,1)$, we have $b < \beta_F = \frac{N_F}{N_F+1}$. This inequality is equivalent to $N_F - b > N_F b$. Dividing by $N_F + 1$, we get $a = \frac{N_F - b}{N_F+1} > \frac{N_F b}{N_F+1} = b\beta_F$. But we have already seen that $\beta_F > b$, hence $a > b\beta_F > b^2$. To summarize,

$$b \; < \; \frac{N_F}{N_F+1}, \qquad a \; > \; b^2. \tag{A17}$$

Lemma A.1 can now be used to show that the coefficients A and B are constant. Indeed, in the proof of the Theorem, we have seen that both A_t and B_t satisfy recursive equations of the form $X_t = \alpha + \beta X_{t-1}$, with $\beta \in (-1, 1)$. Then, Lemma A.1 implies that X_t converges to a fixed number $\frac{\alpha}{1-\beta}$, regardless of the starting point. But, since we work in continuous time, and $t+1$ actually stands for $t + \mathrm{d}t$, the convergence occurs in an infinitesimal amount of time. Thus, X_t is constant for all t, and that constant is equal to $\frac{\alpha}{1-\beta}$. □

We now state the Lemma that is used in the proof of Theorem 1.

Lemma A.1. *Let $X_1 \in \mathbb{R}$, and consider a sequence $X_t \in \mathbb{R}$ which satisfies the following recursive equation:*

$$X_t - \beta X_{t-1} \;=\; \alpha, \quad t \geq 2. \tag{A18}$$

Then the sequence X_t converges to $\bar{X} = \frac{\alpha}{1-\beta}$, regardless of the initial value of X_1, if and only if $\beta \in (-1, 1)$.

Proof. First, note that $\bar{X}$ is well defined as long as $\beta \neq 1$. If we denote by $Y_t = X_t - \bar{X}$, the new sequence Y_t satisfies the recursive equation $Y_t - \beta Y_{t-1} = 0$. We now show that Y_t converges to 0 (and $\bar{X}$ is well defined) if and only if $\beta \in (-1, 1)$. Then, the difference equation $Y_t - \beta Y_{t-1} = 0$ has the following general solution:

$$Y_t \;=\; C\beta^t, \quad t \geq 1, \quad \text{with} \quad C \;\in\; \mathbb{R}. \tag{A19}$$

Then, Y_t is convergent for any values of C if and only if all $\beta \in (-1, 1]$. But in the latter case, $1 - \beta = 0$, which makes $\bar{X}$ nondefined. □

Proof of Corollary 1. In the proof of Theorem 1, equation (A8) implies $\lambda\bar{\gamma} = \frac{N_F}{N_F+1}$, $\lambda\bar{\mu} = \frac{B}{A}\frac{N_L}{N_L+1}$. But from (A14) and (A15), we have $\frac{B}{A} = \frac{1}{1+b}$, which proves the first row in (22). The second row in (22) just rewrites the formulas for A and B from equations (A14) and (A15). □

Proof of Proposition 1. From Corollary 1, $\lambda\bar{\gamma} = \frac{N_F}{N_F+1}$ and $\lambda\bar{\mu} = \frac{B}{A}\frac{N_L}{N_L+1}$. From (A7),

the equilibrium normalized expected profit of the FT is

$$\tilde{\pi}^F \;=\; \gamma - \lambda\gamma\bar{\gamma} + \mu B - \lambda\mu\bar{\mu}A \;=\; \gamma\Big(1 - \frac{N_F}{N_F+1}\Big) + B\mu\Big(1 - \frac{N_L}{N_L+1}\Big) \qquad \text{(A20)}$$

From (A16), $B = \frac{1}{N_F+1}$, which proves the desired formula for π^F. The profit of the ST is the same as for the FT, but with $\gamma = 0$. The last statement now follows from the asymptotic results in Theorem 1. □

Justification of Result 1. According to Proposition 1, $\delta\,\mathrm{d}t$ is the expected profit that speculators get per unit of time $\mathrm{d}t$ from trading on their lagged signal $(\widetilde{\mathrm{d}w_{t-1}})$. Given that all speculators break even on this lag, they would not trade on any signal with a larger lag, as this would cost them the same (δ), but would bring a lower profit. For this last statement we use the results of Internet Appendix I (Proposition I.3), where we show numerically and asymptotically that the profit generated by lagged signals is decreasing in the number of lags. □

Proof of Proposition 2. Since $1-a = \frac{1+b}{N_F+1}$, equation (19) implies that $\lambda = \rho\frac{N_F}{N_F-b} = \frac{\sigma_w}{\sigma_u}\sqrt{(1-a)(a-b^2)}\frac{N_F}{N_F-b} = \frac{\sigma_w}{\sigma_u}\frac{\sqrt{(1+b)(a-b^2)}}{\sqrt{N_F+1}}\frac{N_F}{N_F-b}$, which proves the first equation in (36).

By definition, the trading volume is $TV = \sigma_y^2$. From (A15), $TV = \sigma_y^2 = \tilde{\sigma}_y^2\sigma_w^2 = \frac{a\sigma_w^2}{\rho^2}$. From (19), $\rho^2 = \frac{\sigma_w^2}{\sigma_u^2}(1-a)(a-b^2)$, hence $TV = \sigma_u^2\frac{a}{(1-a)(a-b^2)}$. Substituting $1-a = \frac{1+b}{N_F+1}$, we get $TV = \sigma_u^2(N_F+1)\frac{a}{(1+b)(a-b^2)}$, which proves the second equation in (36).

The price volatility is $\sigma_p^2 = \lambda^2\,TV = \big(\frac{\lambda}{\rho}\big)^2\rho^2\,TV = \big(\frac{\lambda}{\rho}\big)^2 a\sigma_w^2$. From (19), $\frac{\lambda}{\rho} = \frac{N_F}{N_F-b}$, hence $\sigma_p^2 = \big(\frac{N_F}{N_F-b}\big)^2\frac{N_F-b}{N_F+1}\,\sigma_w^2 = \frac{N_F^2}{(N_F+1)(N_F-b)}\,\sigma_w^2$, which proves the third equation in (36).

The speculator participation rate is $SPR = \frac{\bar{\gamma}^2\,\sigma_w^2+\bar{\mu}^2\,\sigma_{\tilde{w}}^2}{TV} = \frac{\rho^2(\bar{\gamma}^2\,\sigma_w^2+\bar{\mu}^2\,\sigma_{\tilde{w}}^2)}{a\sigma_w^2}$. Since $\rho\bar{\gamma} = a$, $\rho\bar{\mu} = b$, and $\sigma_{\tilde{w}}^2 = (1-a)\sigma_w^2$, we get $SPR = \frac{a^2+b^2(1-a)}{a}$. This proves the last equation in (36), since $\frac{1-a}{a} = \frac{1+b}{N_F-b}$. □

Proof of Proposition 3. As in Theorem 1, we start with the FT's choice of optimal trading strategy. Each FT $i = 1,\ldots,N_F$ observes $\mathrm{d}w_t$, and chooses $\mathrm{d}x_t^i = \gamma_t^i\mathrm{d}w_t$ to maximize the expected profit:

$$\pi_0 \;=\; \mathsf{E}\left(\int_0^T\Big(w_t - p_{t-1} - \lambda_t(\mathrm{d}x_t^i + \mathrm{d}x_t^{-i} + \mathrm{d}u_t)\Big)\mathrm{d}x_t^i\right) \;=\; \int_0^T \gamma_t^i\sigma_w^2\mathrm{d}t - \lambda_t\gamma_t^i\big(\gamma_t^i + \gamma_t^{-i}\big)\sigma_w^2\mathrm{d}t, \qquad \text{(A21)}$$

where the superscript "$-i$" indicates the aggregate quantity from the other FTs. This is a pointwise quadratic optimization problem, with solution $\lambda_t \gamma_t^i = \frac{1-\lambda_t \gamma_t^{-i}}{2}$. Since this is true for all $i = 1, \dots, N_F$, the equilibrium is symmetric and we compute $\gamma_t = \frac{1}{\lambda_t} \frac{1}{1+N_F}$.

The dealer takes the FTs' strategies as given, thus assumes that the aggregate order flow is of the form $\mathrm{d}y_t = \mathrm{d}u_t + N_F \gamma_t \mathrm{d}w_t$. To set λ_t, the dealer sets p_t such that $\mathrm{d}p_t = \lambda_t \mathrm{d}y_t$, with $\lambda_t = \frac{\mathsf{Cov}(w_t, \mathrm{d}y_t)}{\mathsf{Var}(\mathrm{d}y_t)} = \frac{N_F \gamma_t \sigma_w^2}{\sigma_u^2 + N_F^2 \gamma_t^2 \sigma_w^2}$. This implies $\lambda_t^2 \sigma_u^2 + (N_F \gamma_t \lambda_t)^2 \sigma_w^2 = N_F \gamma_t \lambda_t \sigma_w^2$. But $N_F \lambda_t \gamma_t = \frac{N_F}{N_F+1}$. Hence, $\lambda_t^2 \sigma_u^2 + \left(\frac{N_F}{N_F+1}\right)^2 \sigma_w^2 = \frac{N_F}{N_F+1} \sigma_w^2$, or $\lambda_t^2 \sigma_u^2 = \frac{N_F}{(N_F+1)^2} \sigma_w^2$, which implies the formula $\lambda = \frac{\sigma_w}{\sigma_u} \frac{\sqrt{N_F}}{N_F+1}$. We then compute $\gamma_t = \frac{1}{\lambda_t} \frac{N_F}{1+N_F} = \frac{\sigma_u}{\sigma_w} \frac{1}{\sqrt{N_F}}$.

We have $TV = \sigma_y^2 = N_F^2 \gamma^2 \sigma_w^2 + \sigma_u^2$. But $N_F \gamma = \frac{\sigma_u}{\sigma_w} \sqrt{N_F}$, hence $TV = \sigma_u^2(1 + N_F)$. Next, $\sigma_p^2 = \lambda^2 TV = \frac{\sigma_w^2}{\sigma_u^2} \frac{N_F}{(N_F+1)^2} \sigma_u^2 (N_F + 1) = \sigma_w^2 \frac{N_F}{N_F+1}$. Also, $SPR = \frac{TV - \sigma_u^2}{TV} = \frac{\sigma_u^2 (N_F+1) - \sigma_u^2}{\sigma_u^2 (N_F+1)} = \frac{N_F}{N_F+1}$.

Finally, we compute Σ'. From the formula above for λ, we get $\mathsf{Var}(\mathrm{d}p_t) = \lambda^2 \, \mathsf{Var}(\mathrm{d}y_t) = \lambda \, \mathsf{Cov}(w_t, \mathrm{d}y_t) = \mathsf{Cov}(w_t, \mathrm{d}p_t)$. Since $\Sigma_t = \mathsf{Var}(w_t - p_{t-1}) = \mathsf{E}\big((w_t - p_{t-1})^2\big)$, we compute $\Sigma_t' = \frac{1}{\mathrm{d}t} \mathsf{E}\big(2(\mathrm{d}w_{t+1} - \mathrm{d}p_t)(w_t - p_{t-1}) + (\mathrm{d}w_{t+1} - \mathrm{d}p_t)^2\big) = -2 \frac{\mathsf{Cov}(w_t, \mathrm{d}p_t)}{\mathrm{d}t} + \sigma_w^2 + \frac{\mathsf{Var}(\mathrm{d}p_t)}{\mathrm{d}t} = \sigma_w^2 - \sigma_p^2 = \frac{\sigma_w^2}{N_F+1}$. □

Proof of Proposition 4. We use the formulas from the Proof of Theorem 1. Since $\widetilde{\mathrm{d}w}_t$ is orthogonal on $\mathrm{d}y_t$, we have $\widetilde{\mathsf{Cov}}\big(\widetilde{\mathrm{d}w}_t, \mathrm{d}w_t\big) = \widetilde{\mathsf{Cov}}\big(\widetilde{\mathrm{d}w}_t, \widetilde{\mathrm{d}w}_t\big) = A = 1 - a = \frac{1+b}{N_F+1}$. Then, $\widetilde{\mathsf{Cov}}\big(\widetilde{\mathrm{d}w}_t, \widetilde{\mathrm{d}w}_{t-1}\big) = \widetilde{\mathsf{Cov}}\big(\mathrm{d}w_t - \rho\bar{\gamma}\mathrm{d}w_t - \rho\bar{\mu}\widetilde{\mathrm{d}w}_{t-1}, \widetilde{\mathrm{d}w}_{t-1}\big) = -\rho\bar{\mu}A$. Therefore,

$$
\begin{aligned}
\widetilde{\mathsf{Cov}}\big(\mathrm{d}\bar{x}_{t+1}, \mathrm{d}\bar{x}_t\big) &= \widetilde{\mathsf{Cov}}\big(\bar{\gamma}\mathrm{d}w_{t+1} + \bar{\mu}\widetilde{\mathrm{d}w}_t, \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1}\big) = \bar{\mu}\bar{\gamma}A + \bar{\mu}^2\big(-bA\big) \\
\widetilde{\mathsf{Var}}\big(\mathrm{d}\bar{x}_t\big) &= \widetilde{\mathsf{Var}}\big(\bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1}\big) = \bar{\gamma}^2 + \bar{\mu}^2 A.
\end{aligned} \tag{A22}
$$

By multiplying both the numerator and denominator by ρ^2, we compute

$$
\rho_{\bar{x}} = \frac{\bar{\mu}\bar{\gamma}A}{\bar{\gamma}^2 + \bar{\mu}^2 A} - \frac{b\bar{\mu}^2 A}{\bar{\gamma}^2 + \bar{\mu}^2 A} = \frac{ab(1-a)}{a^2 + b^2(1-a)} - \frac{b^3(1-a)}{a^2 + b^2(1-a)} = \rho_{AT} + \rho_{EA}. \tag{A23}
$$

Then, $\rho_{\bar{x}} = \frac{ab - b^3}{a^2 + b^2(1-a)} (1-a) = \frac{(a-b^2)b}{a^2 + b^2(1-a)} \frac{1+b}{N_F+1}$, which implies the desired formulas.

We now prove that $\rho_{\bar{x}} > 0$ if and only if there exists slow trading. When there is no slow trading, $b = \rho\bar{\mu} = 0$, hence $\rho_{\bar{x}} = 0$. When there is slow trading, we show that $\rho_{\bar{x}} = \frac{b(b+1)(a-b^2)}{a^2 + b^2(1-a)} \frac{1}{N_F+1} > 0$. Indeed, we have $b > 0$, $a < 1$, and from equation (A17), $a - b^2 > 0$. □

Proof of Proposition 5. If x_t is the IFT's inventory in the risky asset, denote by

$$\begin{aligned}
\Omega_t^{xx} &= \frac{\mathsf{E}\big(x_t^2\big)}{\sigma_w^2 \mathrm{d}t}, & \Omega_t^{xe} &= \frac{\mathsf{E}\big(x_t(w_t - p_t)\big)}{\sigma_w^2 \mathrm{d}t}, & X_t &= \frac{\mathsf{E}\big(x_t \widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2 \mathrm{d}t} \\
\Omega_t^{xw} &= \frac{\mathsf{E}\big(x_t w_t\big)}{\sigma_w^2 \mathrm{d}t}, & \Omega_t^{xp} &= \frac{\mathsf{E}\big(x_t p_t\big)}{\sigma_w^2 \mathrm{d}t}, & Z_t &= \frac{\mathsf{E}\big(x_{t-1} \mathrm{d}y_t\big)}{\sigma_w^2 \mathrm{d}t}.
\end{aligned} \tag{A24}$$

Since $\Theta > 0$, we have $\Theta \in (0,2)$, or $\phi = 1 - \Theta \in (-1,1)$. From (46), x_t satisfies the recursive equation $x_t = \phi x_{t-1} + G\mathrm{d}w_t$. We compute $\Omega_t^{xx} = \frac{\mathsf{E}((x_t)^2)}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}((\phi x_{t-1} + G\mathrm{d}w_t)^2)}{\sigma_w^2 \mathrm{d}t} = \phi^2 \Omega_{t-1}^{xx} + G^2$. Since $\phi^2 \in (-1,1)$, we apply Lemma A.1 to the recursive formula $\Omega_t^{xx} = \phi^2 \Omega_{t-1}^{xx} + G^2$. Then, Ω_t^{xx} is constant and equal to:

$$\Omega^{xx} \;=\; \frac{G^2}{1-\phi^2} \;=\; \frac{G^2}{\Theta(1+\phi)}, \tag{A25}$$

which is the usual variance formula for the $AR(1)$ process. The order flow at t is $\mathrm{d}y_t = -\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1} + \mathrm{d}u_t$, with $\bar{\gamma} = \gamma^- + G$. Then, Z_t is a function of X_{t-1}:

$$Z_t \;=\; \frac{\mathsf{E}\big(x_{t-1}\mathrm{d}y_t\big)}{\sigma_w^2 \mathrm{d}t} \;=\; -\Theta\Omega_{t-1}^{xx} + \bar{\mu}X_{t-1} \;=\; -\frac{G^2}{1+\phi} + \bar{\mu}X_{t-1}. \tag{A26}$$

The recursive formula for X_t is $X_t = \frac{\mathsf{E}(x_t \widetilde{\mathrm{d}w_t})}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}((\phi x_{t-1} + G\mathrm{d}w_t)(\mathrm{d}w_t - \rho \mathrm{d}y_t))}{\sigma_w^2 \mathrm{d}t} = -\phi\rho Z_t + G - G\rho\bar{\gamma} = -\phi\rho\bar{\mu}X_{t-1} + \phi\frac{\rho G^2}{1+\phi} + G - G\rho\bar{\gamma} = -\phi b X_{t-1} + G(1-a^-) - \frac{\rho G^2}{1+\phi}$. By assumption, $0 \le b < 1$, hence $\phi b \in (-1,1)$. Lemma A.1 implies that X_t is constant and equal to

$$X \;=\; \frac{G(1-a^-) - \frac{\rho G^2}{1+\phi}}{1+\phi b}. \tag{A27}$$

From (A26), Z_t is also constant and satisfies:

$$Z \;=\; \bar{\mu}X - \frac{G^2}{1+\phi} \;=\; \bar{\mu}G\frac{1-a^-}{1+\phi b} - G^2 \frac{b + \frac{1}{1+\phi}}{1+\phi b}. \tag{A28}$$

We are interested in $\Omega_t^{xe} = \Omega_t^{xw} - \Omega_t^{xp}$. The recursive equation for Ω_t^{xw} is $\Omega_t^{xw} = \frac{\mathsf{E}(x_t w_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}\big((\phi x_{t-1} + G\mathrm{d}w_t)(w_{t-1} + \mathrm{d}w_t)\big)}{\sigma_w^2 \mathrm{d}t} = \phi\Omega_{t-1}^{xw} + G$. Since $\phi \in (-1,1)$, Lemma A.1 implies that Ω_t^{xw} is constant and equal to

$$\Omega^{xw} \;=\; \frac{G}{\Theta}. \tag{A29}$$

The recursive formula for Ω_t^{xp} is $\Omega_t^{xp} = \frac{\mathsf{E}(x_t p_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}\Big((\phi x_{t-1} + G\mathrm{d}w_t)(p_{t-1}+\lambda \mathrm{d}y_t)\Big)}{\sigma_w^2 \mathrm{d}t} = \phi\Omega_{t-1}^{xp} + \lambda\phi Z + \lambda G\bar{\gamma}$. Lemma A.1 implies that Ω_t^{xp} is constant and equal to $\Omega^{xp} = \frac{\lambda\phi Z + \lambda G\bar{\gamma}}{\Theta}$. It follows that $\Omega_t^{xe} = \Omega_t^{xw} - \Omega_t^{xp}$ is constant and satisfies:

$$-\Theta\Omega^{xe} \;=\; -\Big(\Theta\Omega^{xw} - \Theta\Omega^{xp}\Big) \;=\; -G + \lambda\phi Z + \lambda G\bar{\gamma}. \tag{A30}$$

The IFT's expected profit satisfies $\pi_{\Theta>0} = \mathsf{E}\int_0^T (w_t - p_t)\mathrm{d}x_t = \mathsf{E}\int_0^T \Big(w_{t-1} - p_{t-1} + \mathrm{d}w_t - \lambda \mathrm{d}y_t\Big)\Big(G\mathrm{d}w_t - \Theta x_{t-1}\Big)$. Hence, the IFT's normalized expected profit is $\tilde{\pi}_{\Theta>0} = \int_0^T \Big(-\Theta\Omega^{xe} + G - \lambda\bar{\gamma}G + \lambda\Theta Z\Big)\mathrm{d}t$. If we use the formula (A30) for Ω^{xe}, we obtain:

$$\tilde{\pi}_{\Theta>0} \;=\; \lambda Z \;=\; \lambda\left(\bar{\mu}G\frac{1-a^-}{1+\phi b} - G^2\frac{b+\frac{1}{1+\phi}}{1+\phi b}\right), \tag{A31}$$

where the second equality comes from (A28). This proves (49).

We now show that the inventory costs are zero, which implies that the IFT's expected utility is the same as his expected profit. According to equation (A25), $\Omega_t^{xx} = \frac{G^2}{\Theta(1+\phi)}$ is constant. Then, by the definition (A24) of Ω_t^{xx}, we have

$$\mathsf{E}\big(x_t^2\big) \;=\; \frac{G^2}{\Theta(1+\phi)}\sigma_w^2 \mathrm{d}t, \tag{A32}$$

which implies that the expected squared inventory of the IFT is infinitesimal, and therefore becomes zero when integrated up over $[0,1]$ ($\mathrm{d}t^2 = 0$). Hence, from the definition (42), the inventory costs of the IFT are $C_I \int_0^T \mathsf{E}\big(x_t^2\big)\mathrm{d}t = 0$.

To show that all IFT's expected profits are in cash, consider the decomposition

$$\pi_{\Theta>0} \;=\; \mathsf{E}\int_0^T (v_T - p_t)\mathrm{d}x_t \;=\; \mathsf{E}\int_0^T w_t \mathrm{d}x_t - \mathsf{E}\int_0^T p_t \mathrm{d}x_t, \tag{A33}$$

which is the same as (50). We need to show that the first term (the risky component) is zero. From (A29), $\Omega_t^{xw} = \frac{G}{\Theta}$. Thus, we compute

$$\frac{\mathsf{E}\int_0^T w_t \mathrm{d}x_t}{\sigma_w^2} \;=\; \frac{\mathsf{E}\int_0^T \big(w_{t-1} + \mathrm{d}w_t\big)\big(-\Theta x_{t-1} + G\mathrm{d}w_t\big)}{\sigma_w^2} \;=\; -\Theta\Omega^{xw} + G \;=\; 0. \tag{A34}$$

which implies that the risky component is indeed zero. This finishes the proof. $\square$

Proof of Corollary 3. From (A24), $Z_t = \frac{\mathsf{E}(x_{t-1}\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t}$, which implies $\mathsf{E}(x_{t-1}\mathrm{d}y_t) = Z_t\sigma_w^2\mathrm{d}t$. From this, $\mathsf{E}\int_0^T x_{t-1}\mathrm{d}p_t = \lambda\int_0^T Z_t\sigma_w^2\mathrm{d}t = \lambda Z\sigma_w^2$, since Z_t is constant. But from (A31), the IFT's expected profit is $\pi_{\Theta>0} = \lambda Z\sigma_w^2$. Therefore,

$$\pi_{\Theta>0} \;=\; \mathsf{E}\int_0^T x_{t-1}\mathrm{d}p_t. \tag{A35}$$

Now, write $\mathrm{d}p_t = \lambda\mathrm{d}y_t = \lambda\big(-\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1} + \mathrm{d}u_t\big)$. Since x_{t-1} is orthogonal to $\mathrm{d}w_t$ and $\mathrm{d}u_t$, we get $\mathrm{d}p_t = \lambda\big(\bar{\mu}\widetilde{\mathrm{d}w}_{t-1} - \Theta x_{t-1}\big)$. If we substitute this formula in (A35), we obtain (52). $\square$

Proof of Theorem 2. Let $\Theta = 0$. Then, the IFT's strategy is of the form $\mathrm{d}x_t = G\mathrm{d}w_t$. We compute the IFT's expected profit $\pi_{\Theta=0} = \mathsf{E}\int_0^T(w_t - p_t)\mathrm{d}x_t = \mathsf{E}\int_0^1\big(w_{t-1} - p_{t-1} + \mathrm{d}w_t - \lambda\mathrm{d}y_t\big)\big(G\mathrm{d}w_t\big) = \mathsf{E}\int_0^1\big(\mathrm{d}w_t - \lambda\mathrm{d}y_t\big)\big(G\mathrm{d}w_t\big) = \mathsf{E}\int_0^1\big(\mathrm{d}w_t - \lambda\bar{\gamma}\mathrm{d}w_t\big)\big(G\mathrm{d}w_t\big) = G(1-\lambda\bar{\gamma})\sigma_w^2$. But $\lambda\bar{\gamma} = \lambda G + \lambda\gamma^- = \lambda G + Ra^-$. The normalized IFT's expected profit is:

$$\tilde{\pi}_{\Theta=0} \;=\; G(1-\lambda\bar{\gamma}) \;=\; G(1-Ra^-) - \lambda G^2. \tag{A36}$$

To compute the IFT's inventory costs, denote by $\Omega_t^{xx} = \frac{\mathsf{E}(x_t^2)}{\sigma_w^2}$. We compute $\frac{\mathrm{d}\Omega_t^{xx}}{\mathrm{d}t} = \frac{1}{\sigma_w^2\mathrm{d}t}\mathsf{E}\big(2x_{t-1}\mathrm{d}x_t + (\mathrm{d}x_t)^2\big) = \frac{1}{\sigma_w^2\mathrm{d}t}\mathsf{E}\big(2Gx_{t-1}\mathrm{d}w_t + G^2(\mathrm{d}w_t)^2\big) = G^2$. Since $\Omega_0^{xx} = 0$, the solution of this first order ODE is $\Omega_t^{xx} = tG^2$, for all $t \in [0,1]$. Hence, the inventory costs are equal to

$$C_I\,\mathsf{E}\int_0^1 x_t^2\mathrm{d}t \;=\; C_I\,G^2\int_0^1 t\mathrm{d}t \;=\; \frac{C_I}{2}G^2, \tag{A37}$$

From (A36) and (A37), the IFT's normalized expected utility when $\Theta = 0$ is:

$$\tilde{U}_{\Theta=0} \;=\; G\big(1-Ra^-\big) - G^2\left(\lambda + \frac{C_I}{2}\right). \tag{A38}$$

The function $\tilde{U}_{\Theta=0}$ attains its maximum at $G = \frac{1-Ra^-}{2\lambda+C_I} = \frac{1-Ra^-}{2\lambda\left(1+\frac{C_I}{2\lambda}\right)}$, as stated in the Theorem. The maximum value is:

$$\tilde{U}^{\max}_{\Theta=0} \;=\; \frac{\big(1-Ra^-\big)^2}{2(2\lambda+C_I)}. \tag{A39}$$

Let $\Theta > 0$, which is equivalent to $\phi = 1 - \Theta \in (-1, 1)$. In the proof of Proposition 5, we have already computed the IFT's expected profit (see (49)) and showed that the IFT's inventory costs are zero. Hence, the IFT's expected utility is the same as his expected profit, and satisfies $\tilde{U}_{\Theta>0} = \tilde{\pi}_{\Theta>0} = \frac{\lambda}{\rho}\Big(bG\frac{1-a^-}{1+\phi b} - \rho G^2\frac{b+\frac{1}{1+\phi}}{1+\phi b}\Big)$. The first order condition with respect to G implies that at the optimum $G = \frac{b(1-a^-)}{2\rho\left(b+\frac{1}{1+\phi}\right)}$, as stated in the Theorem. The second order condition for a maximum is $\lambda\frac{b+\frac{1}{1+\phi}}{1+\phi b} > 0$, which follows from $\lambda > 0$, $b \in [0, 1)$, and $\phi \in (-1, 1)$. For the optimum G, the normalized expected utility (profit) of the IFT is:

$$\tilde{U}_{\Theta>0} \;=\; \frac{\big(Rb(1-a^-)\big)^2}{4\lambda(1+\phi b)\big(b+\frac{1}{1+\phi}\big)}. \tag{A40}$$

We now analyze the function

$$f(\phi) \;=\; (1+\phi b)\left(b+\frac{1}{1+\phi}\right) \quad\Longrightarrow\quad f'(\phi) \;=\; \frac{b^2(1+\phi)^2+b-1}{(1+\phi)^2}. \tag{A41}$$

The polynomial in the numerator has two roots:

$$\phi_1 \;=\; -1+\frac{\sqrt{1-b}}{b} \qquad \phi_2 \;=\; -1-\frac{\sqrt{1-b}}{b}. \tag{A42}$$

By assumption $b < 1$, hence both roots are real. Clearly, we have $\phi_2 < -1$. We show that $\phi_1 \in (-1, 1)$. First, note that ϕ_1 is decreasing in b. For $b = 1$ we have $\phi_1 = -1$; while for $b = \frac{\sqrt{17}-1}{8}$ (which satisfies $4b^2 + b = 1$) we have $\phi_1 = 1$. Since by assumption $\frac{\sqrt{17}-1}{8} < b < 1$, it follows that indeed $\phi_1 \in (-1, 1)$. Thus, $f'(\phi)$ is negative on $(-1, \phi_1)$ and positive on $(\phi_1, 1)$. Hence, $f(\phi)$ attains its minimum at $\phi = \phi_1$, which implies that the normalized expected utility $\tilde{U}_{\Theta>0}$ from (A40) attains its maximum at $\phi = \phi_1$, or $\Theta = 2 - \frac{\sqrt{1-b}}{b}$, as stated in the Theorem. The maximum value (over both G and Θ) is:

$$\tilde{U}^{\max}_{\Theta>0} \;=\; \frac{\big(Rb(1-a^-)\big)^2}{4\lambda b(1+\sqrt{1-b})^2}. \tag{A43}$$

To determine the cutoff value for the inventory aversion coefficient C_I, we set $\tilde{U}^{\max}_{\Theta=0} = \tilde{U}^{\max}_{\Theta>0}$. From (A39) and (A43), algebraic manipulation shows that the cutoff value is

$\bar{C}_I = 2\lambda\Big(\frac{(1-Ra^-)^2(1+\sqrt{1-b})^2}{R^2 b(1-a^-)^2} - 1\Big)$, as stated in the Theorem. $\square$

Proof of Corollary 4. Let $\Theta > 0$. We are in the context of Theorem 2, where $b > \frac{\sqrt{17}-1}{8} > 0$ and $\rho > 0$, hence $\bar{\mu} = \frac{b}{\rho} > 0$. The IFT's strategy is $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$, while the slow trading component is $\mathrm{d}\bar{x}_t^S = \bar{\mu}\widetilde{\mathrm{d}w}_{t-1}$. Since $\mathrm{d}w_t$ is orthogonal to $\widetilde{\mathrm{d}w}_{t-1}$, $\mathsf{Cov}(\mathrm{d}x_t, \mathrm{d}\bar{x}_t^S) = -\Theta\,\mathsf{Cov}(x_{t-1}, \mathrm{d}\bar{x}_t^S) = -\Theta\bar{\mu}\,\mathsf{Cov}(x_{t-1}, \widetilde{\mathrm{d}w}_{t-1})$. This proves the equality in (55). Since $\Theta > 0$ and $\bar{\mu} > 0$, it remains to prove the inequality $\mathsf{Cov}(x_{t-1}, \widetilde{\mathrm{d}w}_{t-1}) > 0$. But $\mathsf{Cov}(x_{t-1}, \widetilde{\mathrm{d}w}_{t-1}) = X\sigma_w^2\mathrm{d}t$ (see (A24)). From (A27), $X = \frac{G(1-a^-) - \frac{\rho G^2}{1+\phi}}{1+\phi b}$. Substituting the optimal G and $\phi = 1 - \Theta$ from Theorem 2, we obtain $X = \frac{(1-a^-)^2}{4(1+\sqrt{1-b})}$. As in Theorem 2, $a^-, b \in [0,1)$, hence $X > 0$ and the proof is complete. $\square$

Proof of Theorem 3. Consider the following implicit equation in b

$$\frac{2b(1+b)(2B+1)}{n_L} = \frac{Q}{B^2(a^-+b)} + \frac{3bB + 2b^2B - 1 - b}{b}(1-a^-) - 2, \tag{A44}$$

where the following substitutions are made:[38]

$$\begin{aligned}
B &= \frac{1}{\sqrt{1-b}}, \qquad q = (B+1)\Big(2(B^2-1) - n_F(3B^2-2)\Big),\\
a^- &= \frac{-q \pm \sqrt{q^2 + n_F B^5\big((4-n_F)B + 2(2-n_F)\big)}}{B^2\big((4-n_F)B + 2(2-n_F)\big)},\\
Q &= B^3(a^-)^2 + 2(3B^3 + 3B^2 - 2B - 1)a^- + (B^3 + 2B^2 - 2).
\end{aligned} \tag{A45}$$

We write the equations for the other coefficients:

$$\begin{aligned}
R &= \frac{4(B+1)B^2\big(a^-+b\big)}{Q}, \qquad a = \frac{(2B+1)a^- + 1}{2(B+1)}\\
\rho^2 &= \left((a-b^2) + \frac{2bB-1}{2B+1}(1-a)\right)(1-a)\,\frac{\sigma_w^2}{\sigma_u^2}, \qquad \lambda = R\,\rho\\
\Theta &= 2 - \frac{\sqrt{1-b}}{b}, \qquad G = \frac{1-a}{\rho(2B+1)}, \qquad \gamma = \frac{a^-}{\rho N_F}, \qquad \mu = \frac{b}{\rho N_L}.
\end{aligned} \tag{A46}$$

The proof is now left to Internet Appendix J (see Sections J.4 and J.5). $\square$

[38] To be rigorous, we have included the case when a^- is negative. However, numerically this case never occurs in equilibrium, because it leads to $\lambda < 0$, which contradicts the FT's second order condition (J56) in Internet Appendix J.

Proof of Proposition 6. See Internet Appendix J (Section J.5). □

Proof of Proposition 7. See Internet Appendix J (Section J.5). □

REFERENCES

[1] Aït-Sahalia, Yacine, and Mehment Saglam (2014): "High Frequency Traders: Taking Advantage of Speed," Working Paper.

[2] Back, Kerry, Henry Cao, and Gregory Willard (2000): "Imperfect Competition among Informed Traders," *Journal of Finance*, 55, 2117–2155.

[3] Back, Kerry, and Hal Pedersen (1998): "Long-Lived Information and Intraday Patterns," *Journal of Financial Markets*, 1, 385–402.

[4] Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko (2014): "Risk and Return in High Frequency Traders," Working Paper.

[5] Benos, Evangelos, and Satchit Sagade (2013): "High-Frequency Trading Behaviour and Its Impact on Market Quality: Evidence from the UK Equity Market," Working Paper.

[6] Biais, Bruno, Thierry Foucault, and Sophie Moinas (2014): "Equilibrium Fast Trading," *Journal of Financial Economics*, forthcoming.

[7] Boehmer, Ekkehart, Kingsley Fong, and Julie Wu (2014): "International Evidence on Algorithmic Trading," Working Paper.

[8] Brogaard, Jonathan (2011): "High Frequency Trading and Its Impact on Market Quality," Working Paper.

[9] Brogaard, Jonathan, Björn Hagströmer, Lars Nordén, and Ryan Riordan (2013): "Trading Fast and Slow: Colocation and Market Quality," Working Paper.

[10] Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014): "High-Frequency Trading and Price Discovery," *Review of Financial Studies*, 27, 2267–2306.

[11] Budish, Eric, Peter Cramton, and John Shim (2014): "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," Working Paper.

[12] Cao, Huining Henry, Yuan Ma, and Dongyan Ye (2013): "Disclosure, Learning, and Coordination," Working paper.

[13] Caldentey, René, and Ennio Stacchetti (2010): "Insider Trading with a Random Deadline," *Econometrica*, 78, 245–283.

[14] Cartea, Álvaro, and José Penalva (2012): "Where is the Value in High Frequency Trading?," *Quarterly Journal of Finance*, 2, 1–46.

[15] Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega (2014): "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," *Journal of Finance*, 69, 2045–2084.

[16] Chau, Minh, and Dimitri Vayanos (2008): "Strong-Form Efficiency with Monopolistic Insiders," *Review of Financial Studies*, 18, 2275–2306.

[17] Cvitanić, Jakša, and Andrei Kirilenko (2010): "High Frequency Traders and Asset Prices," Working Paper.

[18] Du, Songzi, and Haoxiang Zhu (2014): "Welfare and Optimal Trading Frequency in Dynamic Double Auctions," Working Paper.

[19] Foster, Douglas, and S. Viswanathan (1996): "Strategic Trading When Agents Forecast the Forecast of Others," *Journal of Finance*, 51, 1437–1478.

[20] Foucault, Thierry, Johan Hombert, and Ioanid Roşu (2015): "News Trading and Speed," *Journal of Finance*, forthcoming.

[21] Hasbrouck, Joel, and Gideon Saar (2013): "Low-Latency Trading," *Journal of Financial Markets*, 16, 646–679.

[22] Hendershott, Terrence, Charles Jones, and Albert Menkveld (2011): "Does Algorithmic Trading Improve Liquidity?," *Journal of Finance*, 66, 1–33.

[23] Hendershott, Terrence, and Albert Menkveld (2014): "Price Pressures," *Journal of Financial Economics*, 114, 405–423.

[24] Hirschey, Nicholas (2013): "Do High-Frequency Traders Anticipate Buying and Selling Pressure?," Working Paper.

[25] Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman (1994): "Security Analysis and Trading Patterns When Some Investors Receive Information Before Others," *Journal of Finance*, 49, 1665–1698.

[26] Ho, Thomas, and Hans Stoll (1981): "Optimal Dealer Pricing Under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9, 47–73.

[27] Hoffmann, Peter (2014): "A Dynamic Limit Order Market with Fast and Slow Traders," *Journal of Financial Economics*, 113, 156–169.

[28] Holden, Craig, and Avanidhar Subrahmanyam (1992): "Long-Lived Private Information and Imperfect Competition," *Journal of Finance*, 47, 247–270.

[29] Jovanovic, Boyan, and Albert Menkveld (2012): "Middlemen in Limit-Order Markets," Working Paper.

[30] Kirilenko, Andrei, Albert Kyle, Mehrdad Samadi, and Tugkan Tuzun (2014): "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," Working Paper.

[31] Kyle, Albert (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

[32] Li, Su (2012): "Speculative Dynamics I: Imperfect Competition, and the Implications for High Frequency Trading," Working Paper.

[33] Li, Wei (2014): "High Frequency Trading with Speed Hierarchies," Working Paper.

[34] Lyons, Richard (1997): "A Simultaneous Trade Model of the Foreign Exchange Hot Potato," *Journal of International Economics*, 42, 275–298.

[35] Madhavan, Ananth, and Seymour Smidt (1993): "An Analysis of Changes in Specialist Inventories and Quotations," *Journal of Finance*, 48, 1595–1628.

[36] Pagnotta, Emiliano, and Thomas Philippon (2013): "Competing on Speed," Working Paper.

[37] SEC (2010): "Concept Release on Equity Market Structure," Release No. 34-61358; File No. S7-02-10.

[38] Subrahmanyam, Avanidhar, and Sheridan Titman (1999): "The Going-Public Decision and the Development of Financial Markets," *Journal of Finance*, 54, 1045–1082.

[39] Weller, Brian (2014): "Intermediation Chains," Working Paper.

[40] Zhang, X. Frank (2010): "High Frequency Trading, Stock Volatility, and Price Discovery," Working Paper.

Internet Appendix for "Fast and Slow Informed Trading"

Contents

I The General Model with m Lags **2**

I.1 Notation Preliminaries 2

I.2 The General Speed Case 3

I.3 The Equal Speed Case 6

I.4 Proofs of Results 11

J Equilibrium with Inventory Management **27**

J.1 Optimal Inventory Management 27

J.2 Optimal Strategies of Fast and Slow Traders 29

J.3 Dealer's Pricing Rules with Inventory Management 30

J.4 Equilibrium Conditions 31

J.5 Proofs of Results 32

K Smooth Inventory Management **52**

K.1 The Expected Utility of the IFT 53

K.2 Optimal Smooth Inventory Management 54

K.3 Proofs of Results 56

L Robust Trading Strategies **63**

L.1 Motivation 63

L.2 Multidimensional Asset Value 64

L.3 Smooth Trading with Multidimensional Value 66

L.4 Proofs of Results 67

M Discrete Time Models **71**

M.1 General Model 71

M.2 Equilibrium of the General Model 73

M.3 Fast and Slow Traders in Discrete Time 74

M.4 Proofs of Results 77

I The General Model with m Lags

In this section, we analyze the general model with $m \geq 0$ lags. As defined in Section 2 of the paper, this is the model $\mathcal{M}_m$ in which speculators do not use their signals beyond lag m. Recall that a ℓ-speculator observes the signals after ℓ lags ($\ell = 0, 1, \ldots, m$). The strategy of a ℓ-speculator is assumed to be of the form:

$$\mathrm{d}x_t \;=\; \gamma_{\ell,t}(\mathrm{d}w_{t-\ell} - z_{t-\ell,t}) + \gamma_{\ell+1,t}(\mathrm{d}w_{t-\ell-1} - z_{t-\ell-1,t}) + \cdots + \gamma_{m,t}(\mathrm{d}w_{t-m} - z_{t-m,t}), \quad \text{(I1)}$$

where $z_{t-j,m}$ is the dealer's expectation of the j-lagged signal $\mathrm{d}w_{t-j}$.

In the paper, we have already analyzed the particular cases $m = 0$ (Proposition 3 in Section 4) and $m = 1$ (Section 3.1). For these two cases, the equilibrium is described in closed form. For the case $m > 1$, we show in Section I.2 that the equilibrium reduces to system of non-linear equations in the model coefficients. In Section I.3 we discuss the particular case in which all speculators have the same speed: if N_ℓ is the number of ℓ-speculators, then $N_0 > 0$, and $N_1 = \cdots = N_m = 0$. The proofs are provided in Section I.4.

I.1 Notation Preliminaries

To simplify the presentation, we write the system of equations using matrix notation. All vectors are in column format, and we denote by X' the transpose a vector X.

Because some variables, such as the coefficients γ_j in (I1) are not defined for all entries $j = 0, 1, \ldots, m$, we collect them in a vector denoted by $\gamma^{(\ell)}$, which contains only the entries $j = \ell, \ell+1, \ldots, m$. If we need to sum $\gamma^{(\ell)}$ over different ℓ, we make a slight abuse of notation and denote by $\gamma^{(\ell)}$ the vector with the same entries above ℓ, but padded with zeros at the entries $j = 0, 1, \ldots, \ell - 1$. In vector notation,

$$\gamma^{(\ell)} \;=\; \left[\, \gamma_\ell^{(\ell)} \,,\, \ldots \,,\, \gamma_m^{(\ell)} \,\right]' \qquad \text{or} \qquad \gamma^{(\ell)} \;=\; \left[\, 0 \,,\, \ldots \,,\, 0 \,,\, \gamma_\ell^{(\ell)} \,,\, \ldots \,,\, \gamma_m^{(\ell)} \,\right]', \quad \text{(I2)}$$

In general, if A is a matrix with elements $A_{i,j}$ for $i, j = 0, \ldots, m$, we denote by $A_{\geq \ell}$ the matrix with elements $A_{i,j}$ for $i, j \geq \ell$; and similarly for the vectors $B_{\geq \ell}$ and $\rho_{\geq \ell}$. A sum of vectors $X_{\geq \ell}$ over different ℓ is carried by padding $X_{\geq \ell}$ with zeros for the first ℓ entries.

We normalize some variables by dividing them with the forecast variance, σ_w^2. We denote this by placing a tilde above the variable. For instance, we define the normalized instantaneous order flow variance $\tilde{\sigma}_y^2$, as well as the normalized instantaneous noise

trader variance $\tilde{\sigma}_u^2$ as follows:

$$\tilde{\sigma}_{y,t}^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t}, \qquad \tilde{\sigma}_u^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}u_t)}{\sigma_w^2 \mathrm{d}t} \;=\; \frac{\sigma_u^2}{\sigma_w^2}. \tag{I3}$$

Nevertheless, when necessary we denote by $\gamma^{(\ell)}$ also the vector above, but padded with zeros in positions $0, \ldots, \ell-1$. if we take for instantce

We follow the usual convention that sums from a larger index to a smaller index are equal to zero. For instance, if $m = 0$, for any variable X_i the sum from the index 1 to m is by convention equal to zero:

$$m \;=\; 0 \qquad \Longrightarrow \qquad \sum_{i=1}^{m} X_i \;=\; 0. \tag{I4}$$

Similarly, an enumeration from a larger index to a smaller index is by convention the empty set. For instance, if $m = 0$, saying that the condition $\mathcal{P}_i$ holds for $i = 1, \ldots, m$ is equivalent to imposing no condition at all.

I.2 The General Speed Case

In this section, we solve for the equilibrium of the model $\mathcal{M}_m$ when the number of lags $m \geq 0$ is fixed. Under an additional assumption stated below, we show that the equilibrium reduces to the solution of a system of equations (see Theorem I.1). This system can be solved in closed form in some particular cases of interest, and can in principle be solved numerically.

To proceed with our solution, we need to be more specific about how the dealer sets her expectation $z_{t-i,t} = \mathsf{E}(\mathrm{d}w_{t-i} \mid \{\mathrm{d}y_\tau\}_{\tau<t})$. Since $\mathrm{d}w_{t-i}$ is the speculator's signal from i trading rounds before (corresponding to calendar time $t - i\,\mathrm{d}t$), it is plausible to expect that (i) $z_{t-i,t}$ only involves the order flow from at most i periods before, and (ii) $z_{t-i,t}$ is linear in the order flow. Thus, we assume (and show it to be true in equilibrium) that

$$z_{t-i,t} \;=\; \rho_{0,t}\mathrm{d}y_{t-i} + \cdots + \rho_{i-1,t}\mathrm{d}y_{t-1}, \quad i = 0, 1, \ldots, m, \tag{I5}$$

where $\mathrm{d}y_{t-j}$ is the order flow from j trading rounds before. Define the *fresh signal* $\mathrm{d}_t w_{t-i}$ to be the unanticipated part of the signal at t:

$$\mathrm{d}_t w_{t-i} \;=\; \mathrm{d}w_{t-i} - z_{t-i,t}. \tag{I6}$$

For all lags $i, j = 0, \ldots, m$, denote by

$$A_{i,j,t} \;=\; \frac{\mathsf{Cov}\big(\mathrm{d}_t w_{t-i}, \mathrm{d}_t w_{t-j}\big)}{\sigma_w^2 \, \mathrm{d}t}, \qquad B_{j,t} \;=\; \frac{\mathsf{Cov}\big(w_t, \mathrm{d}_t w_{t-j}\big)}{\sigma_w^2 \, \mathrm{d}t}, \tag{I7}$$

Since A measures the instantaneous covariance of fresh signals at the relevant lags, we call A the *fresh covariance matrix*. The vector B measures the instantaneous contribution of each fresh signal to the profit, thus we call B the *benefit vector*. In Section 2, we have assumed that the speculator takes A and B as fixed, and considers them as set by the dealer (just as $\rho_{j,t}$ and λ_t).

The next result shows that a linear equilibrium exists if a certain system of equations is satisfied. We write the system of equations using the notations from Section I.1. For instance, the coefficients γ_j and ρ_j are collected in the following vectors:[1] with $\ell = 0, \ldots, m$:

$$\gamma^{(\ell)} \;=\; \big[\, \gamma_\ell^{(\ell)} \,,\; \ldots \;,\; \gamma_m^{(\ell)} \,\big]', \qquad \rho \;=\; \big[\, \rho_0 \,,\; \ldots \;,\; \rho_m \,\big]', \tag{I8}$$

Theorem I.1. *Let $m \geq 0$ be fixed, and consider the model $\mathcal{M}_m$ with m lags, and N_ℓ speculators of type $\ell = 0, \ldots, m$. Suppose there exists a linear equilibrium of the model with constant coefficients, of the form:*

$$\begin{aligned} \mathrm{d}x_t^{(\ell)} \;&=\; \gamma_\ell^{(\ell)} \mathrm{d}_t w_{t-\ell} \;+\; \cdots \;+\; \gamma_m^{(\ell)} \mathrm{d}_t w_{t-m}, \quad \ell = 0, \ldots, m, \\ \mathrm{d}_t w_{t-i} \;&=\; \mathrm{d}w_{t-i} - z_{t-i,t}, \quad i = 0, 1, \ldots, m, \\ z_{t-i,t} \;&=\; \rho_0 \mathrm{d}y_{t-i} + \cdots + \rho_{i-1} \mathrm{d}y_{t-1}, \quad i = 0, 1, \ldots, m, \\ \mathrm{d}p_t \;&=\; \lambda \mathrm{d}y_t. \end{aligned} \tag{I9}$$

Then, the constants λ, ρ_i, $\gamma_i^{(\ell)}$ satisfy the following system of equations ($i = 0, \ldots, m$):

$$\begin{aligned} \bar{\gamma} \;&=\; \sum_{\ell=0}^{m} N_\ell \, \gamma^{(\ell)}, \quad \textit{with} \quad \gamma^{(\ell)} \;=\; \big(A_{\geq \ell}\big)^{-1} \left(\frac{1}{\lambda} \, B_{\geq \ell} - \tilde{\sigma}_y^2 \rho_{\geq \ell} \right), \\ \tilde{\sigma}_y^2 \rho \;&=\; A \bar{\gamma}, \qquad \tilde{\sigma}_y^2 \lambda \;=\; B' \bar{\gamma}, \qquad \tilde{\sigma}_y^2 \;=\; \frac{\tilde{\sigma}_u^2}{1 - \bar{\gamma}' \rho}, \\ A_{i,j} \;&=\; \mathbf{1}_{i=j} - \tilde{\sigma}_y^2 \sum_{k=1}^{\min(i,j)} \rho_{i-k} \rho_{j-k}, \qquad B_i \;=\; 1 - \tilde{\sigma}_y^2 \lambda \sum_{k=1}^{i} \rho_{i-k}, \end{aligned} \tag{I10}$$

[1] Note that ρ_m, the last entry of the vector ρ, is not part the dealer's expectations $z_{t-i,t}$, but we introduce it in order to simplify notation. In particular, the last row of the equilibrium equation $\tilde{\sigma}_y^2 \rho = A \bar{\gamma}$ can be omitted.

where $\mathbf{1}_{\mathcal{P}}$ is the indicator function, which equals 1 if $\mathcal{P}$ is true, and 0 otherwise.

Conversely, suppose the constants λ, ρ_i, $\gamma_i^{(\ell)}$ satisfy (I10), and in addition the following conditions are satisfied: (i) $\lambda > 0$; (ii) for all $\ell = 0, \dots, m$, the matrix $A_{\geq \ell}$ is invertible; and (iii) the numbers $\beta_k = \sum_{i=0}^{m-k} \rho_i \bar{\gamma}_{k+i}$ satisfy $1 > \beta_1 > \cdots > \beta_m > 0$. Then, the equations in (I9) provide an equilibrium of the model.

In principle, the system of equations (I10) can be solved numerically as follows. To simplify notation, we make a change of variables and denote by $r_i = \tilde{\sigma}_y \rho_i$, $\Lambda = \tilde{\sigma}_y \lambda$, $g = \frac{\bar{\gamma}}{\tilde{\sigma}_y}$. Then, suppose we start with some values for r_i and Λ. Then, A can be expressed only in terms of r_i, and the equation $\tilde{\sigma}_y^2 \rho = A\bar{\gamma}$ implies that $g = A^{-1} r$ can also be expressed only in terms of r_i. Also, B can be expressed only in terms of r_i and Λ. Then, the first equation in (I10) and $\Lambda = B' g$ (which is the rescaled equation, $\tilde{\sigma}_y^2 \lambda = B' \bar{\gamma}$) become $m + 2$ equations in the variables r_i and Λ.

In practice, however, this procedure does not work well. Numerically, it turns out that the solution is badly behaved, especially when N or m are large.[2] Moreover, without more explicit formulas, it is difficult to study properties of the solution. In the next section, we provide a more explicit solution for the case when all speculators have the same speed, i.e., when there are only 0-speculators.

We now analyze the forecast error variance,

$$\Sigma_t \;=\; \mathsf{Var}\big((w_t - p_{t-1})^2\big). \tag{I11}$$

Note that Σ_t is inversely related to price informativeness. Indeed, when prices are informative, they stay close to the forecast w_t, which implies that the variance Σ_t is small. Define the instantaneous price variance by

$$\sigma_p^2 \;=\; \frac{\mathsf{Var}(\mathrm{d}p_t)}{\mathrm{d}t} \;=\; \frac{\lambda^2\ \mathsf{Var}(\mathrm{d}y_t)}{\mathrm{d}t} \;=\; \lambda^2 \sigma_y^2. \tag{I12}$$

The next result shows that growth rate of Σ is constant, and it is equal to the difference between the forecast variance σ_w^2 and the price variance σ_p^2.

Proposition I.1. *In the context of Theorem I.1, the growth in the forecast error variance is constant, and satisfies the following formula:*

$$\Sigma_t' \;=\; \sigma_w^2 - \sigma_p^2. \tag{I13}$$

[2]This is because in that case the matrix A is almost singular, and thus the equation $g = A^{-1} r$ produces unreliable solutions. Indeed, equation (I82) from Section I.4 shows that that the determinant of $(A^0)^{-1}$, a matrix close to A^{-1}, is equal to $\frac{(N+1)^{m+1}}{(m+1)N+1}$. This is a large number when m or N are large.

This result can be explained by the fact that competition among speculators increases price volatility, with an upper bound given by the forecast volatility σ_w^2. But competition also makes prices more informative, which implies that the forecast error variance grows more slowly. As we see in the next section, it is a feature of this equilibrium to have the forecast error variance grow at a positive rate. This result is in contrast to Kyle (1985), in which the forecast error variance *decreases* at a constant rate, so that it becomes zero at the end. The reason for this difference is that in our model traders in equilibrium only use their most recent signals, and thus do not trade on longer-lived information.[3]

I.3 The Equal Speed Case

In this section, we search for an equilibrium of the model $\mathcal{M}_m$ with $m \geq 0$ lags in the simpler case when there is no speed difference among speculators. This translates into all speculators being 0-speculators, i.e., $N_0 > 0$ and $N_1 = N_2 = \cdots = N_m = 0$. Because there are only 0-speculators, we write their number simply as $N = N_0$.

Theorem I.2 provides an efficient numerical procedure to solve for the equilibrium. When the number of speculators is large, we also obtain asymptotic formulas for the equilibrium trading strategies and pricing functions. Proposition I.3 then shows that the value of information decays exponentially. This result is proved rigorously only asymptotically, when the number of speculators is large. However, we have verified numerically that the result remains true for all values of the parameters we have checked.

The first result is a restatement of Theorem I.1 to the case when all speculators have the same speed.

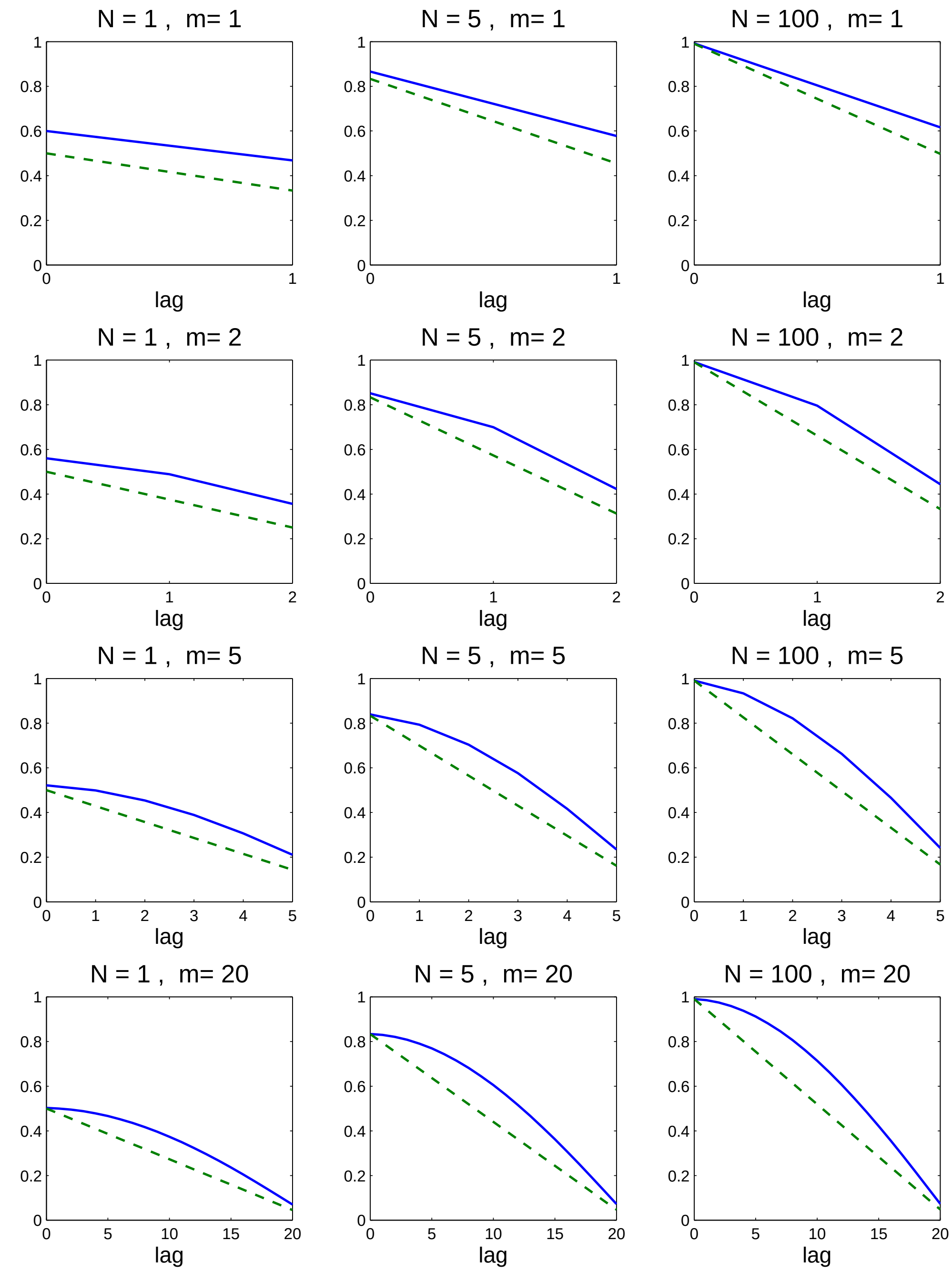
Proposition I.2. *Let $m \geq 0$ be fixed, and consider the model $\mathcal{M}_m$ with m lags, and N speculators with equal speed (of type $\ell = 0$). Suppose there exists a linear equilibrium of the model with constant coefficients, of the form:*

$$\begin{aligned}
\mathrm{d}x_t &= \gamma_0 \mathrm{d}_t w_t + \gamma_1 \mathrm{d}_t w_{t-1} + \cdots + \gamma_m \mathrm{d}_t w_{t-m}, \\
\mathrm{d}_t w_{t-i} &= \mathrm{d}w_{t-i} - z_{t-i,t}, \quad i = 0, 1, \ldots, m, \\
z_{t-i,t} &= \rho_0 \mathrm{d}y_{t-i} + \cdots + \rho_{i-1} \mathrm{d}y_{t-1}, \quad i = 0, 1, \ldots, m, \\
\mathrm{d}p_t &= \lambda \mathrm{d}y_t.
\end{aligned} \tag{I14}$$

Then, the constants λ, ρ_i, γ_i and $\bar{\gamma}_i = N\gamma_i$ satisfy the following system of equations

[3] For a discussion on why traders might not want to use longer-lived information, see Internet Appendix L.

Figure I.1: Optimal Trading Weights. Consider the model with $m \in \{1,2,5,20\}$ lags and $N \in \{1,5,100\}$ identical speculators. The figure plots the rescaled aggregate weight $g_i = N\gamma_i \frac{1}{\sqrt{N+1}} \frac{\sigma_w}{\sigma_u}$ (continuous line) against the lag $i = 0,\dots,m$, and compares it with the value $g_i^0 = \frac{N}{N+1} \frac{(m-i+1)N+1}{(m+1)N+1}$ (dashed line).



($i = 0, \ldots, m$):

$$B_i = \frac{1}{(N+1)^i}, \qquad \tilde{\sigma}_y = \tilde{\sigma}_u \sqrt{N+1}, \qquad \tilde{\sigma}_y^2 \rho_i = \frac{1}{\lambda} \frac{N}{(N+1)^{i+1}},$$
$$\tilde{\sigma}_y^2 \rho = A\bar{\gamma}, \qquad \tilde{\sigma}_y^2 \lambda = B'\bar{\gamma}, \qquad A_{i,j} = \mathbf{1}_{i=j} - \frac{1}{\tilde{\sigma}_y^2 \lambda^2} \frac{N}{N+2} \frac{(N+1)^{2\min(i,j)} - 1}{(N+1)^{i+j}}. \tag{I15}$$

Conversely, suppose the constants λ, ρ_i, γ_i satisfy (I15), and in addition the following conditions are satisfied: (i) $\lambda > 0$; (ii) the matrix A is invertible; and (iii) the numbers $\beta_k = \sum_{i=0}^{m-k} \rho_i \bar{\gamma}_{k+i}$ satisfy $1 > \beta_1 > \cdots > \beta_m > 0$. Then, the equations in (I14) provide an equilibrium of the model.

Note that the system of equations (I15) has a simpler form. Following the discussion after Theorem I.1, we make a change of variables and denote by $r_i = \tilde{\sigma}_y \rho_i$, $\Lambda = \tilde{\sigma}_y \lambda$, $g = \frac{\bar{\gamma}}{\tilde{\sigma}_y}$. In this case, we see that $r_i = \frac{1}{\Lambda} \frac{N}{(N+1)^{i+1}}$ can further be expressed in terms of Λ. This suggests the following procedure to search for a solution of (I15): Suppose we start with some value for Λ. From (I15) we see that all the constants of the model (g, A, B, r) can be expressed as a function of Λ. Then, the equation $\Lambda = g'B$ becomes the equation that determines Λ. In Section I.4, we show that the equation in Λ is an infinite polynomial equation, which in practice can be solved very accurately. Then, the conditions (i) and (ii) from Proposition I.2 follow from a certain condition (I77) on Λ from the proof of Theorem I.2.

The next result uses the procedure outlined above to find approximations for the equilibrium coefficients, which use the "big-O" notation.[4]

Theorem I.2. *Let $m \geq 0$ be fixed, and consider the model $\mathcal{M}_m$ with m lags, and N speculators with equal speed (of type $\ell = 0$). Define the following numbers:*

$$\gamma_i^0 = \frac{\sigma_u}{\sigma_w} \frac{1}{\sqrt{N+1}} \frac{m-i+1}{m+1}, \qquad i = 0, 1, \ldots, m,$$
$$\rho_i^0 = \frac{\sigma_w}{\sigma_u} \frac{N}{(N+1)^{i+1+1/2}}, \qquad i = 0, 1, \ldots, m-1, \tag{I16}$$
$$\lambda^0 = \frac{\sigma_w}{\sigma_u} \frac{1}{\sqrt{N+1}}.$$

[4] For $\alpha \in \mathbb{R}$, we say that the expression x_N is of the order of N^α, and write $x_N = O_N(N^\alpha)$, if there exists an integer N_* and a real number M such that $|x_N| \leq M |N^\alpha|$ for all $N \geq N_*$. In other words, x_M is of order N^α if $\frac{x_N}{N^\alpha}$ is bounded when N is sufficiently large.

Then if conditions (I76) and (I77) from Section I.4 are satisfied,[5] *there exists an equilibrium. In this equilibrium, the coefficients of the optimal strategy (γ_i) and of the pricing functions (λ, ρ_i) approximate the coefficients in (I16) as follows:*

$$\begin{aligned}
\gamma_i &= \gamma_i^0\Big(1 + O_N(1)\Big), \quad i = 0, 1, \ldots, m, \\
\rho_i &= \rho_i^0\Big(1 + O_N\big(\tfrac{1}{N}\big)\Big), \quad i = 0, 1, \ldots, m-1, \\
\lambda &= \lambda^0\Big(1 - O_N\big(\tfrac{1}{N}\big)\Big).
\end{aligned} \tag{I17}$$

Figure I.1 plots the optimal weights for various numbers of speculators N and various maximum signal lags m. For all the parameter values we have checked, the weights decrease with the lag. However, while the approximate weights, $\gamma_i^0 = \frac{\sigma_u}{\sigma_w} \frac{1}{\sqrt{N+1}} \frac{m-i+1}{m+1}$, decrease at the same rate, in the Figure we see that the actual weights decrease less quickly for smaller lags, and then decrease faster for larger lags. When m is large, one can also see that the initial decrease in the actual weights is very small.[6]

An important consequence of the theorem is that the expected profit from each additional signal decays exponentially.

Proposition I.3. *In the context of Theorem I.2, let π_0 be the expected profit at $t = 0$ of a speculator in equilibrium, and let γ_i be his optimal trading weight on the signal with lag $i = 0, \ldots, m$. Then the profit can be decomposed as follows:*

$$\pi_0 = \sigma_w^2 \sum_{j=0}^{m} \pi_{0,j} = \sigma_w^2 \left(\frac{\gamma_0}{N+1} + \frac{\gamma_1}{(N+1)^2} + \cdots \frac{\gamma_m}{(N+1)^{m+1}} \right). \tag{I18}$$

Moreover, the ratio of two consecutive components of the expected profit is

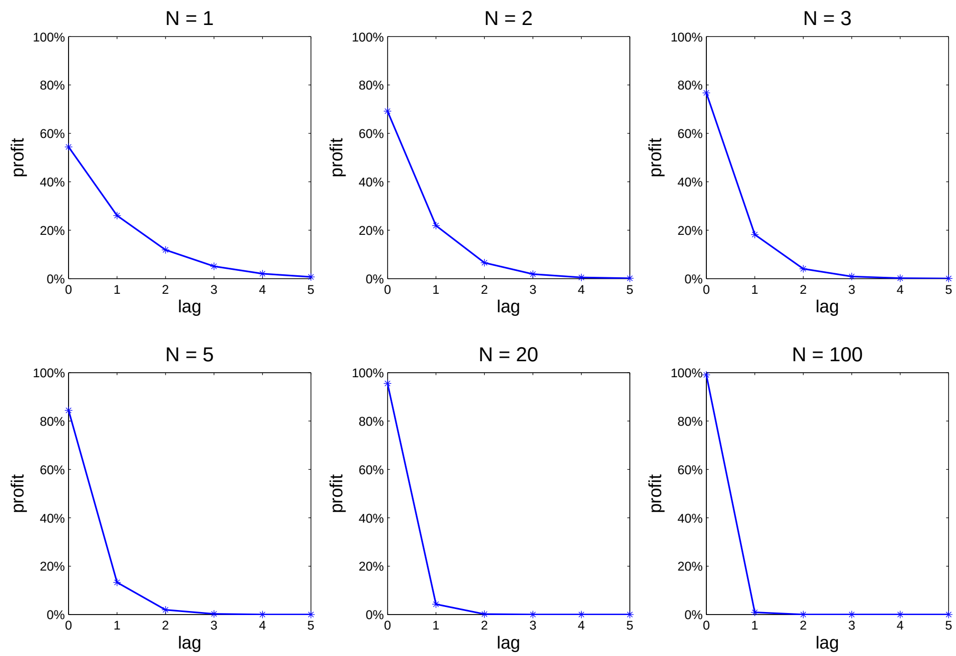
$$\frac{\pi_{0,j+1}}{\pi_{0,j}} = \frac{\gamma_{j+1}}{\gamma_j} \frac{1}{N+1} = O_N\big(\tfrac{1}{N}\big). \tag{I19}$$

A graphic illustration of this result can be found in Figure I.2, which plots the profits of a speculator who can trade on at most $m = 5$ lagged signals. The cases studied correspond to the number of speculators $N \in \{1, 2, 3, 5, 20, 100\}$. One sees that

[5]Numerically, these conditions are satisfied for all the values of N and m we have tried.

[6]Thus, in the limit when m approaches infinity, we conjecture that the weights become approximately equal. In that case, the informed traders behave as in Kyle (1985), by trading a multiple of the sum $\int_0^t \mathrm{d}w_\tau = w_t - w_0$. However, we can see from Theorem I.2 that in our model the weights do not become of the order of $\mathrm{d}t$, as in the Kyle model, but rather remain of the same order of magnitude as for the lower m. In our model therefore prices are very close to strong form efficient when m is large. This equilibrium resembles that of Caldentey and Stacchetti (2010).

Figure I.2: Profit from Lagged Signals. The figure plots the percentage of a speculator's profit from each his lagged signals when there is competition among $N \in \{1, 2, 3, 5, 20, 100\}$ identical speculators. In these examples, the speculators can trade up to $m = 5$ lagged signals.



indeed, when N is large, the profits coming even from a signal of lag 1 are small.

The next result analyzes in more detail price volatility, and shows that volatility has an upper bound, which makes rigorous the intuition for the general case, discussed after Proposition I.1. Moreover, Proposition I.4 provides a more thorough understanding about how information is revealed over time by trading. For this purpose, we define *signal revelation* as the covariance of a signal $\mathrm{d}w_t$ with $\mathrm{d}p_{t+k}$, the price change from k trading periods later:

$$SR_k = \frac{\mathsf{Cov}(\mathrm{d}w_t, \mathrm{d}p_{t+k})}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{Cov}(\mathrm{d}w_{t-k}, \mathrm{d}p_t)}{\sigma_w^2 \mathrm{d}t}, \quad k = 0, 1, \dots. \tag{I20}$$

Since $\sum_{k=0}^{\infty} \mathrm{d}w_{t-k} = w_t$ (speculator's initial forecast is $w_0 = 0$), the sum of all SR_k equals

$$\sum_{k=0}^{\infty} SR_k = \frac{\mathsf{Cov}(w_t, \mathrm{d}p_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\lambda\, \mathsf{Cov}(w_t, \mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\lambda^2 \sigma_y^2}{\sigma_w^2} = \frac{\sigma_p^2}{\sigma_w^2}, \tag{I21}$$

where we have used the formula $\mathsf{Cov}(w_t, \mathrm{d}y_t) = \lambda\, \mathsf{Var}(\mathrm{d}y_t) = \lambda \sigma_y^2$ from the dealer's pricing equation for λ, proved in (I46) in Section I.4.

Proposition I.4. *In the context of Theorem I.2, price volatility is always smaller than the forecast volatility. Their difference is small when the number of speculators is large:*

$$\sigma_w^2 - \sigma_p^2 \;=\; O_N\big(\tfrac{1}{N}\big). \tag{I22}$$

The signal revelation measure satisfies

$$SR_k \;=\; \frac{N}{(N+1)^{k+1}}, \quad k = 0, 1, \ldots, m, \quad \Longrightarrow \quad \sum_{k=0}^{m} SR_k \;=\; 1 - \frac{1}{(N+1)^{m+1}}. \tag{I23}$$

Therefore, the difference $\sigma_w^2 - \sigma_p^2$ is also small when m is large.

Thus, an interesting implication of the Proposition is that, when the number of lags m is large, each signal $\mathrm{d}w_t$ gets revealed by trading almost entirely. From Proposition I.1, this case coincides with the one in which the growth rate of Σ, the forecast error variance, is very small.

I.4 Proofs of Results

Before we proceed with the proofs of our equilibrium results, we recall the notations from Section I.1, and introduce a few more useful notations. As before, a tilde above a symbol denotes normalization by σ_w, while $\widetilde{\mathsf{Cov}}$ and $\widetilde{\mathsf{Var}}$ are the instantaneous covariance and variance (which already means normalized by $\mathrm{d}t$) normalized by σ_w^2. For instance,

$$\tilde{\sigma}_u \;=\; \frac{\sigma_u}{\sigma_w}, \qquad \sigma_y^2 \;=\; \widetilde{\mathsf{Var}}(\mathrm{d}y_t) \;=\; \frac{\mathsf{Var}(\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t}. \tag{I24}$$

We now denote by $M_{a,b}$ the set of matrices of real numbers with a rows and b columns, by $M_a = M_{a,a}$ the set of square matrices, and by $V_a = M_{a,1}$ the set of column vectors. Recall that X' denotes the transpose of X.

If $\ell \in \{0, 1, \ldots, m\}$, recall from Section I.1 that the vector $\gamma^{(\ell)}$ collects the ℓ-speculator's weight on $\mathrm{d}w_{t-i} - z_{t-i,t}$. By a slight abuse of notation, we also write $\gamma^{(\ell)}$ as a vector in V_{m+1} by padding with zeros for the entries $j = 0, \ldots, \ell - 1$. Define the aggregate speculator weights, $\bar{\gamma} \in V_{m+1}$:

$$\bar{\gamma} \;=\; \sum_{\ell=0}^{m} N_\ell \, \gamma^{(\ell)}. \tag{I25}$$

Let $\rho \in V_{m+1}$ be the vector which collects the coefficients involved in the dealer's expec-

tation $z_{t-i,t}$ of $\mathrm{d}w_{t-i}$:[7]

$$\rho \;=\; \left[\, \rho_0 \,,\; \ldots \;,\; \rho_m \,\right]', \tag{I26}$$

For $i = 0, \ldots, m$, let $\mathrm{d}_t w_{t-i}$ be the unpredictable part of the signal $\mathrm{d}w_{t-i}$ (computed before trading at t):

$$\mathrm{d}_t w_{t-i} \;=\; \mathrm{d}w_{t-i} - z_{t-i,t}. \tag{I27}$$

Define the matrices $A \in M_{m+1}$ and $B \in V_{m+1}$. For $i, j = 0, \ldots, m$, let

$$\begin{aligned} A_{i,j} \;&=\; \widetilde{\mathsf{Cov}}(\mathrm{d}_t w_{t-i}, \mathrm{d}_t w_{t-j}) \;=\; \frac{1}{\sigma_w^2 \mathrm{d}t}\, \mathsf{Cov}(\mathrm{d}_t w_{t-i}, \mathrm{d}_t w_{t-j}), \\ B_j \;&=\; \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}_t w_{t-j}) \;=\; \frac{1}{\sigma_w^2 \mathrm{d}t}\, \mathsf{Cov}(w_t, \mathrm{d}_t w_{t-j}). \end{aligned} \tag{I28}$$

We now rescale $\bar{\gamma}$, ρ, λ, by defining $r, g \in V_{m+1}$ and $\Lambda \in \mathbb{R}$ as follows:

$$g \;=\; \frac{\bar{\gamma}}{\tilde{\sigma}_y}, \qquad r \;=\; \tilde{\sigma}_y\, \rho, \qquad \Lambda \;=\; \tilde{\sigma}_y\, \lambda. \tag{I29}$$

Proof of Theorem I.1. We need to prove that a linear equilibrium exists if there exists a solution $(g, r, \Lambda, \tilde{\sigma}_y, A, B)$ to the following system of equations:

$$\begin{aligned} g \;&=\; \sum_{\ell=0}^{m} N_\ell \left(A_{\geq \ell}\right)^{-1} \left(\frac{1}{\Lambda}\, B_{\geq \ell} - r_{\geq \ell}\right), \\ r \;&=\; Ag, \qquad \Lambda \;=\; g'B, \qquad \tilde{\sigma}_y^2 \;=\; \frac{\tilde{\sigma}_u^2}{1 - g'r}, \\ A_{i,j} \;&=\; \mathbf{1}_{i=j} - \sum_{k=1}^{\min(i,j)} r_{i-k} r_{j-k}, \qquad B_i \;=\; 1 - \Lambda \sum_{k=1}^{i} r_{i-k}. \end{aligned} \tag{I30}$$

Recall that $\mathbf{1}_{\mathcal{P}}$ is the indicator function, which equals 1 if $\mathcal{P}$ is true, and 0 otherwise. Also, $A_{\geq \ell}$ is the matrix with elements $A_{i,j}$ for $i, j \geq \ell$; and similarly for the vectors $B_{\geq \ell}$ and $r_{\geq \ell}$. The sum of vectors $X_{\geq \ell}$ over different ℓ is carried by padding $X_{\geq \ell}$ with zeros for the first ℓ entries.

Speculators' Optimal Strategy (γ)

We begin by analyzing the optimal strategy of a ℓ-speculator, where $\ell \in \{0, \ldots, m\}$. This speculator takes as given (i) the dealer's pricing rules: $\mathrm{d}p_t = \lambda \mathrm{d}y_t$ and $z_{t-i,t} = \rho_0 \mathrm{d}y_{t-i} + \cdots + \rho_{i-1} \mathrm{d}y_{t-1}$ for $i = 0, 1, \ldots, m$; and (ii) the other speculators' trading strategies. For instance, if another speculator is of type k, he is assumed to trade

[7] The last entry, ρ_m, is not part of any of dealer's expectations $z_{t-i,t}$, but we introduce it in order to simplify notation.

according to $\mathrm{d}x_t^{(k)} = \sum_{j=k}^m \gamma_j^{(k)} \mathrm{d}_t w_{t-j}$. Also, as we have discussed, the ℓ-speculator chooses among trading strategies of the form: $\mathrm{d}x_t = \gamma_{\ell,t} \mathrm{d}_t w_{t-\ell} + \cdots + \gamma_{m,t} \mathrm{d}_t w_{t-m}$. Therefore, the ℓ-speculator assumes that the total order flow at t satisfies

$$\mathrm{d}y_t \;=\; \mathrm{d}u_t + \sum_{j=0}^{\ell-1} \bar{\gamma}_j \mathrm{d}_t w_{t-j} + \sum_{j=\ell}^{m} \big(\gamma_{j,t} + \gamma_j^-\big) \mathrm{d}_t w_{t-j}. \tag{I31}$$

where

$$\bar{\gamma}_j \;=\; \sum_{k=0}^{j} N_k \gamma_j^{(k)}, \quad j = 0, \ldots, m, \qquad \gamma_j^- \;=\; (N_\ell - 1)\gamma_j^{(\ell)} + \sum_{\substack{k=0 \\ k \neq \ell}}^{j} N_k \gamma_j^{(k)}, \quad j = \ell, \ldots, m. \tag{I32}$$

At $t = 0$, equation (10) implies that his normalized expected profit is

$$\tilde{\pi}_0 \;=\; \frac{\pi_0}{\sigma_w^2} \;=\; \frac{1}{\sigma_w^2}\, \mathsf{E}\left(\int_0^T \big(w_t - p_{t-1} - \lambda \mathrm{d}y_t\big) \mathrm{d}x_t \right). \tag{I33}$$

By construction, the terms $\mathrm{d}_t w_{t-j}$ are orthogonal to $\mathcal{I}_t$, hence also to p_{t-1}. Hence, $\mathrm{d}x_t$ is also orthogonal to p_{t-1}. We now use (I31) and the definitions: $A_{i,j} = \widetilde{\mathsf{Cov}}(\mathrm{d}_t w_{t-i}, \mathrm{d}_t w_{t-j})$, $B_j = \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}_t w_j)$ to compute:

$$\begin{aligned} \tilde{\pi}_0 \;&=\; \mathsf{E}\left(\int_0^T \Big(w_t - \lambda \sum_{i=0}^{\ell-1} \bar{\gamma}_i \mathrm{d}_t w_{t-i} - \lambda \sum_{i=\ell}^{m} \big(\gamma_{i,t} + \gamma_i^-\big) \mathrm{d}_t w_{t-i} \Big) \sum_{j=\ell}^{m} \gamma_{j,t} \mathrm{d}_t w_{t-j} \right) \\ &=\; \sum_{j=\ell}^{m} B_j \gamma_{j,t} - \lambda \sum_{i=0}^{\ell-1} \sum_{j=\ell}^{m} \bar{\gamma}_i A_{i,j} \gamma_{j,t} - \lambda \sum_{i,j=\ell}^{m} \big(\gamma_{i,t} + \gamma_i^-\big) A_{i,j} \gamma_{j,t}. \end{aligned} \tag{I34}$$

Thus, we have reduced the problem to a linear–quadratic optimization. The first order condition with respect to $\gamma_{k,t}$, for $k = \ell, \ldots, m$, is:

$$B_k - \lambda \sum_{i=0}^{\ell-1} \bar{\gamma}_i A_{i,k} - \lambda \sum_{i=\ell}^{m} \big(2\gamma_{i,t} + \gamma_i^-\big) A_{i,k} \;=\; 0. \tag{I35}$$

Denote by γ_t the $(m-\ell+1)$-column vector of trading weights at t. We divide the matrix A into four blocks, by restricting indices to be either $< \ell$ or $\geq \ell$. With matrix notation, the first order condition (I36) becomes

$$B_{\geq \ell} - \lambda\, A_{\geq \ell, < \ell}\, \bar{\gamma}_{<\ell} - \lambda\, A_{\geq \ell} \left(2\gamma_t + \gamma^-\right) \;=\; 0. \tag{I36}$$

We obtain that for any ℓ-speculator and any t,

$$2\gamma_t + \gamma^- \;=\; \big(A_{\geq\ell}\big)^{-1}\left(\frac{1}{\lambda}B_{\geq\ell} - A_{\geq\ell,<\ell}\,\bar{\gamma}_{<\ell}\right). \tag{I37}$$

This equation implies that the ℓ-speculators have identical weights in equilibrium (previously denoted by $\gamma^{(\ell)}$), and these weights do not depend on t. We then have $\gamma^{(\ell)} + \gamma^- = \bar{\gamma}_{\geq\ell}$, hence

$$\gamma^{(\ell)} \;=\; \big(A_{\geq\ell}\big)^{-1}\left(\frac{1}{\lambda}B_{\geq\ell} - A_{\geq\ell,<\ell}\,\bar{\gamma}_{<\ell}\right) - \bar{\gamma}_{\geq\ell}. \tag{I38}$$

Thus, equation (I38) reduces the computation of the optimal weights $\gamma^{(\ell)}$ to the computation of the aggregate weights $\bar{\gamma}$.

We now derive the equation that $\bar{\gamma}$ must satisfy in equilibrium. To simplify formulas, note from (I47) that $A\bar{\gamma} = \tilde{\sigma}_y^2\rho$, or in block matrix notation $A_{\geq\ell,<\ell}\,\bar{\gamma}_{<\ell} + A_{\geq\ell}\,\bar{\gamma}_{\geq\ell} = \tilde{\sigma}_y^2\,\rho_{\geq\ell}$. Using this, equation (I39) becomes

$$\gamma^{(\ell)} \;=\; \big(A_{\geq\ell}\big)^{-1}\left(\frac{1}{\lambda}B_{\geq\ell} - \tilde{\sigma}_y^2\,\rho_{\geq\ell}\right). \tag{I39}$$

To obtain the equation that determines $\bar{\gamma}$, we multiply (I39) by N_ℓ and sum over all $\ell = 0,\ldots,m$, padding with zeroes where necessary. We get

$$\bar{\gamma} \;=\; \sum_{\ell=0}^{m} N_\ell \gamma^{(\ell)} = \sum_{\ell=0}^{m} N_\ell \big(A_{\geq\ell}\big)^{-1}\left(\frac{1}{\lambda}B_{\geq\ell} - \tilde{\sigma}_y^2\,\rho_{\geq\ell}\right). \tag{I40}$$

After dividing this equation by $\tilde{\sigma}_y$, use $g = \frac{\bar{\gamma}}{\tilde{\sigma}_y}$, $r = \tilde{\sigma}_y\,\rho$, and $\Lambda = \tilde{\sigma}_y\,\lambda$ to obtain the corresponding equation in (I30).

So far, we have shown that equation (I39) is a necessary condition for equilibrium. We now prove that it is also sufficient condition for the speculator's problem, if one imposes two additional conditions: (i) $\lambda > 0$; and (ii) for all $\ell = 0,\ldots,m$, the matrix $A_{\geq\ell}$ is invertible. Indeed, the second order condition in the maximization problem above for the ℓ-speculator is:

$$\lambda\,\det(A_{\geq\ell}) \;>\; 0. \tag{I41}$$

Normally, we expect that $\det(A_{\geq\ell}) > 0$, since economically A is the covariance matrix of the fresh signals, and the signals $\mathrm{d}w_{t-i}$ are independent. But if A is just the solution of a system of equations, this condition needs to be checked. If $\det(A_{\geq\ell}) > 0$, then the second order condition from (I41) becomes $\lambda > 0$, which is just condition (i).

Dealer's Pricing Rules (λ, ρ, A, B)

The dealer takes as given that the aggregate order flow is of the form:

$$\mathrm{d}y_t \;=\; \mathrm{d}u_t + \bar{\gamma}_0 \mathrm{d}w_t + \bar{\gamma}_1 \mathrm{d}_t w_{t-1} + \cdots + \bar{\gamma}_m \mathrm{d}_t w_{t-m}, \tag{I42}$$

where the speculators set, for $k = 0, \ldots, m$,

$$\mathrm{d}_t w_{t-k} \;=\; \mathrm{d}w_{t-k} - \big(\rho_0^* \mathrm{d}y_{t-k} + \cdots + \rho_{k-1}^* \mathrm{d}y_{t-1}\big), \tag{I43}$$

with ρ_i^* constant. (Of course, in equilibrium the dealer will eventually set $\rho_i = \rho_i^*$.) We combine the two equations:

$$\begin{aligned} \mathrm{d}y_t \;&=\; \mathrm{d}u_t + \sum_{k=0}^{m} \bar{\gamma}_k \mathrm{d}w_{t-k} - \sum_{k=0}^{m} \bar{\gamma}_k \sum_{i=0}^{k-1} \rho_i^* \mathrm{d}y_{t-k+i} \\ &=\; \mathrm{d}u_t + \sum_{k=0}^{m} \bar{\gamma}_k \mathrm{d}w_{t-k} - \sum_{k=1}^{m} \sum_{i=0}^{m-k} \rho_i^* \bar{\gamma}_{k+i} \mathrm{d}y_{t-k}. \end{aligned} \tag{I44}$$

Because each speculator only trades on the unpredictable part of his signal, $\mathrm{d}y_t$ are orthogonal to each other. Thus, the dealer computes

$$\begin{aligned} z_{t-k,t} \;&=\; \mathsf{E}\big(\mathrm{d}w_{t-k} \mid \mathrm{d}y_{t-k}, \ldots, \mathrm{d}y_{t-1}\big) \;=\; \sum_{i=0}^{k-1} \rho_{i,t-k} \mathrm{d}y_{t-k+i}, \quad k = 0, \ldots, m, \\ \mathrm{d}p_t \;&=\; \lambda_t \mathrm{d}y_t, \end{aligned} \tag{I45}$$

where the coefficients $\rho_{i,t-k}$ and λ_t are given by:[8]

$$\begin{aligned} \rho_{i,t-k} \;&=\; \frac{\mathsf{Cov}(\mathrm{d}w_{t-k}, \mathrm{d}y_{t-k+i})}{\mathsf{Var}(\mathrm{d}y_{t-k+i})}, \quad k = 0, \ldots, m, \quad i = 0, \ldots, k-1, \\ \lambda_t \;&=\; \frac{\mathsf{Cov}(v_1, \mathrm{d}y_t)}{\mathsf{Var}(\mathrm{d}y_t)} \;=\; \frac{\mathsf{Cov}(w_t, \mathrm{d}y_t)}{\mathsf{Var}(\mathrm{d}y_t)}. \end{aligned} \tag{I46}$$

At the end of this proof, we show that the following numbers do not depend on t: λ_t, $\rho_{i,t}$, $A_{i,j,t} = \widetilde{\mathsf{Cov}}(\mathrm{d}_t w_{t-i}, \mathrm{d}_t w_{t-j})$, $B_{j,t} = \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}_t w_j)$.

Taking these numbers as constant, we now prove the rest of the equations in (I30). First, note that in equilibrium $\rho_i^* = \rho_i$. We rewrite the equation for ρ in (I46) by taking $k = i$. Thus, $\rho_i = \frac{\widetilde{\mathsf{Cov}}(\mathrm{d}w_{t-i}, \mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)}$, and note that $\mathrm{d}y_t$ is orthogonal on all other $\mathrm{d}y_{t-k}$ for

[8] Note that in principle $\rho_{i,t-k}$ might also depend on t, the time at which the expectation is computed. However, the formula shows that ρ only depends on i and $t-k$, and not on t separately.

$k > 0$. Hence, from (I42) and (I43), we get

$$\rho_i \;=\; \frac{\widetilde{\mathsf{Cov}}(\mathrm{d}_t w_{t-i}, \mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\widetilde{\mathsf{Cov}}\big(\mathrm{d}_t w_{t-i}, \sum_{j=0}^m \bar{\gamma}_j \mathrm{d}_t w_{t-j}\big)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\sum_{j=0}^m A_{i,j}\bar{\gamma}_j}{\tilde{\sigma}_y^2}. \tag{I47}$$

Since we have denoted by $g = \frac{\bar{\gamma}}{\tilde{\sigma}_y}$ and $r = \tilde{\sigma}_y\,\rho$, we get $r_i = (Ag)_i$, or in matrix notation $r = Ag$. This proves the corresponding equation in (I30). Also, from the equation from λ in (I46), we get

$$\lambda \;=\; \frac{\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\widetilde{\mathsf{Cov}}\big(w_t, \sum_{j=0}^m \bar{\gamma}_j \mathrm{d}_t w_{t-j}\big)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} \;=\; \frac{\sum_{j=0}^m \bar{\gamma}_j B_j}{\tilde{\sigma}_y^2}. \tag{I48}$$

Since $\Lambda = \lambda\,\tilde{\sigma}_y$, we obtain $\Lambda = \sum_{j=0}^m g_j B_j$, or in matrix notation $\Lambda = g'B$. This proves the corresponding equation in (I30).

By computing $\widetilde{\mathsf{Cov}}(\mathrm{d}y_t, \mathrm{d}y_t)$ and using (I42), it follows that $\tilde{\sigma}_y^2 = \widetilde{\mathsf{Var}}(\mathrm{d}y_t)$ satisfies $\tilde{\sigma}_y^2 = \tilde{\sigma}_u^2 + \sum_{i,j=0}^m A_{i,j}\bar{\gamma}_i\bar{\gamma}_j$, or in matrix notation $\tilde{\sigma}_y^2 = \tilde{\sigma}_u^2 + \bar{\gamma}' A \bar{\gamma}$. Since $\bar{\gamma} = g\,\tilde{\sigma}_y$, we compute

$$\tilde{\sigma}_y^2 \;=\; \bar{\gamma}' A \bar{\gamma} + \tilde{\sigma}_u^2 \;=\; g' A g\,\tilde{\sigma}_y^2 + \tilde{\sigma}_u^2. \tag{I49}$$

But we have already shown that $Ag = r$, hence $\tilde{\sigma}_y^2 = g' r\,\tilde{\sigma}_y^2 + \tilde{\sigma}_u^2$, which implies $\tilde{\sigma}_y^2 = \frac{\tilde{\sigma}_u^2}{1 - g'r}$. This proves the corresponding equation in (I30).

Now, consider the equation $A_{k,\ell} = \widetilde{\mathsf{Cov}}(\mathrm{d}_t w_{t-k}, \mathrm{d}_t w_{t-\ell})$. From (I43), $\mathrm{d}_t w_{t-k} = \mathrm{d}w_{t-k} - \big(\rho_0 \mathrm{d}y_{t-k} + \cdots + \rho_{k-1}\mathrm{d}y_{t-1}\big)$. But $\mathrm{d}_t w_{t-\ell}$ is orthogonal to the previous order flow, hence $A_{k,\ell} = \widetilde{\mathsf{Cov}}(\mathrm{d}w_{t-k}, \mathrm{d}_t w_{t-\ell})$. Because A is a symmetric matrix, without loss of generality assume $k \geq \ell$, which implies $\ell = \min(k,\ell)$. Since $\widetilde{\mathsf{Cov}}(\mathrm{d}w_{t-i}, \mathrm{d}y_t) = \rho_i \tilde{\sigma}_y^2\,\mathbf{1}_{i\geq 0}$, we get

$$\begin{aligned} A_{k,\ell} \;&=\; \widetilde{\mathsf{Cov}}\Big(\mathrm{d}w_{t-k}, \mathrm{d}w_{t-\ell} - \big(\rho_0 \mathrm{d}y_{t-\ell} + \cdots + \rho_{\ell-1}\mathrm{d}y_{t-1}\big)\Big) \\ &=\; \mathbf{1}_{k=l} \;-\; \sum_{j=0}^{\ell-1} \rho_j \rho_{k-\ell+j}\,\tilde{\sigma}_y^2 \;=\; \mathbf{1}_{k=l} \;-\; \sum_{i=1}^{\ell} \rho_{k-i}\rho_{\ell-i}\,\tilde{\sigma}_y^2. \end{aligned} \tag{I50}$$

Since $r = \rho\,\tilde{\sigma}_y$, we get $A_{k,\ell} = \mathbf{1}_{k=l} - \sum_{j=1}^{\ell} r_{k-i} r_{\ell-i}$, which proves the corresponding equation in (I30). We also compute $B_\ell = \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}_t w_\ell)$. From (I43),

$$\begin{aligned} B_\ell \;&=\; \widetilde{\mathsf{Cov}}\Big(w_t, \mathrm{d}w_{t-\ell} - \big(\rho_0 \mathrm{d}y_{t-\ell} + \cdots + \rho_{\ell-1}\mathrm{d}y_{t-1}\big)\Big) \\ &=\; 1 - \lambda\big(\rho_0 + \cdots + \rho_{\ell-1}\big) \end{aligned} \tag{I51}$$

Since $\Lambda = \lambda\,\tilde{\sigma}_y$ and $r = \rho\,\tilde{\sigma}_y$, we obtain $B_\ell = 1 - \Lambda \sum_{j=0}^{\ell-1} r_j = 1 - \Lambda \sum_{i=1}^{\ell} r_{i-k}$. This proves the corresponding equation in (I30).

We now prove that the various pricing coefficients do not depend on t. For this, we show that the following numbers are independent of t: $\mathsf{Cov}(\mathrm{d}w_t, \mathrm{d}y_{t+k})$ for all k; $\mathsf{Cov}(w_t, \mathrm{d}y_{t+k})$ for all k; $\mathsf{Var}(\mathrm{d}y_t)$; $\mathsf{Cov}(w_t, \mathrm{d}_t w_j)$ for $j = 0, \ldots, m$; and $\mathsf{Cov}(\mathrm{d}_t w_i, \mathrm{d}_t w_j)$ for $i, j = 0, \ldots, m$.

First, we prove by induction that $\widetilde{\mathsf{Cov}}(\mathrm{d}w_t, \mathrm{d}y_{t+k})$ does not depend on t for $k \geq 0$. (This is trivially true for $k < 0$.) The statement is true for $k = 0$, since equation (I42) implies $\widetilde{\mathsf{Cov}}(\mathrm{d}w_t, \mathrm{d}y_t) = \bar{\gamma}_0$. Assume that the statement is is true for all $i < k$. We now prove that $\widetilde{\mathsf{Cov}}(\mathrm{d}w_t, \mathrm{d}y_{t+k})$ does not depend on t. Equations (I44) implies that $\mathrm{d}y_{t+k}$ only involves three types of terms: (i) $\mathrm{d}u_{t+k}$, (ii) $\mathrm{d}w_{t+k-i}$ for $i = 0, \ldots, m$, and (iii) $\mathrm{d}y_{t+k-1-i}$ for $i = 0, \ldots, m-1$. Also, the coefficients ρ_i^* do not depend on time. Therefore, by the induction hypothesis all these terms have covariances with $\mathrm{d}w_t$ that do not depend on t.

Next, we prove that $a_t = \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_t)$ does not depend on t. Equation (I44) implies the following recursive formula for all t:

$$a_t \;=\; \sum_{k=0}^{m} \bar{\gamma}_k - \sum_{k=1}^{m} \sum_{i=0}^{m-k} \rho_i^* \bar{\gamma}_{k+i} a_{t-k}. \tag{I52}$$

But Lemma I.1 below implies that a_t does not depend on t, provided that

$$1 \;>\; \beta_1 \;>\; \cdots \;>\; \beta_m \;>\; 0, \quad \text{with} \quad \beta_k \;=\; \sum_{i=0}^{m-k} \rho_i^* \bar{\gamma}_{k+i}. \tag{I53}$$

Therefore, $\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_t)$ does not depend on t. This result also implies that $\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_{t+k})$ does not depend on t for any integer k. To see this, note first that the case $k > 0$ reduces to the case $k = 0$. Indeed, $\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_{t+k}) = \widetilde{\mathsf{Cov}}(w_{t+k}, \mathrm{d}y_{t+k}) - \sum_{i=1}^{k} \widetilde{\mathsf{Cov}}(\mathrm{d}w_{t+i}, \mathrm{d}y_{t+k})$, and we already proved that all these terms are independent of t. Also, the case $k < 0$ reduces to the case $k = 0$, since $\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_{t-i}) = \widetilde{\mathsf{Cov}}(w_{t-i}, \mathrm{d}y_{t-i})$ if $i \geq 0$.

We now prove that $\widetilde{\mathsf{Var}}(\mathrm{d}y_t) = \sum_{k=0}^{m} \bar{\gamma}_k \,\mathsf{Cov}\big(\mathrm{d}_t w_{t-k}, \mathrm{d}y_t\big)$ does not depend on t. Since $\mathrm{d}y_t$ is orthogonal to previous order flow, $\widetilde{\mathsf{Var}}(\mathrm{d}y_t) = \sum_{k=0}^{m} \bar{\gamma}_k \,\mathsf{Cov}\big(\mathrm{d}w_{t-k}, \mathrm{d}y_t\big)$. But these terms have already been proved to be independent of t.

Finally, one uses the results proved above to show that $B_{j,t} = \widetilde{\mathsf{Cov}}(w_t, \mathrm{d}_t w_j)$ and $A_{i,j,t} = \widetilde{\mathsf{Cov}}(\mathrm{d}_t w_i, \mathrm{d}_t w_j)$ do not depend on t. Indeed, we have already shown that $\widetilde{\mathsf{Cov}}(\mathrm{d}w_t, \mathrm{d}y_{t-k})$ and $\widetilde{\mathsf{Cov}}(w_t, \mathrm{d}y_{t-k})$ are independent of t, and all is left to do is to use the fact that $\mathrm{d}y_{t-k}$ are orthogonal to each other.

So far, we have provided necessary equations for the equilibrium. We now prove that the conditions in (I30) except for the first one are also sufficient to justify the dealer's pricing equations, if one imposes an additional condition: (iii) the numbers $\beta_k = \sum_{i=0}^{m-k} \rho_i \bar{\gamma}_{k+i}$ satisfy $1 > \beta_1 > \cdots > \beta_m > 0$. But we have already seen that condition (iii) ensures that the equilibrium pricing coefficients are well defined and constant. More generally, one can use Lemma I.1 below to replace condition (iii) with the condition that the (complex) roots of the polynomial $Q(z) = z^m + \beta_1 z^{m-1} + \cdots + \beta_{m-1} z + \beta_m$ lie in the open unit disk $D = \{z \in \mathbb{C} \mid |z| < 1\}$. The proof is now complete. □

Lemma I.1. *Let $X_1, \dots, X_m \in \mathbb{R}$, and consider a sequence $X_t \in \mathbb{R}$ which satisfies the following recursive equation:*

$$X_t + \beta_1 X_{t-1} + \cdots + \beta_m X_{t-m} \; = \; \alpha, \quad t \geq m+1. \tag{I54}$$

Then the sequence X_t converges to $\bar{X} = \frac{\alpha}{1+(\beta_1+\cdots+\beta_m)}$, regardless of the initial values $X_1, \dots, X_m$, if and only if all the (complex) roots of the polynomial $Q(z) = z^m + \beta_1 z^{m-1} + \cdots + \beta_{m-1} z + \beta_m$ lie in the open unit disk $D = \{z \in \mathbb{C} \mid |z| < 1\}$. For this, a sufficient condition is that the coefficients β_i satisfy

$$1 \; > \; \beta_1 \; > \; \cdots \; > \; \beta_m \; > \; 0. \tag{I55}$$

Furthermore, if $\alpha = \alpha_t$ is not constant, then under the same conditions on β_i, the difference $X_t - \frac{\alpha_t}{1+(\beta_1+\cdots+\beta_m)}$ converges to zero, regardless of the initial values for X.

Proof. First, note that $\bar{X}$ is well defined as long as $1 + \beta_1 + \cdots + \beta_m \neq 0$. Indeed, if $1 + \beta_1 + \cdots + \beta_m = 0$ then $Q(z)$ would have $z = 1$ as a root, which does not lie in D. Now, if we denote by $Y_t = X_t - \bar{X}$, the new sequence Y_t satisfies the recursive equation $Y_t + \beta_1 Y_{t-1} + \cdots + \beta_m Y_{t-m} = 0$. We now show that Y_t converges to 0 (and $\bar{X}$ is well defined) if and only if all the roots of Q lie in D. Denote these roots by $q_1, \dots, q_m$. Then, the difference equation in complex numbers: $Y_t + \beta_1 Y_{t-1} + \cdots + \beta_m Y_{t-m} = 0$ has the following general solution:

$$Y_t \; = \; C_1 q_1^t + \cdots C_m q_m^t, \quad t \geq 1, \tag{I56}$$

where $C_1, \dots, C_m$ are arbitrary complex constants.[9] But then, Y_t is convergent for any values of C_i if and only if all q_i lie in D, or if $q_i = 1$. But in the latter case,

[9] To obtain real values of Y_t, one needs to impose the following conditions: (i) if q_i is real, then so is C_i; and (ii) if q_i and q_j are complex conjugate, then so are C_i and C_j.

$1 + \beta_1 + \cdots + \beta_m = 0$, which makes $\bar{X}$ non-defined.

The statement that (I55) implies that all roots of Q lie in D is known as the Eneström–Kakeya theorem. For completeness, we include the proof here. First, we prove that all roots of Q must lie in $\bar{D} = \{z \in \mathbb{C} \mid |z| \leq 1\}$. By contradiction, suppose that there exists z_* a root of Q with $|z_*| > 1$. Then, we also have $(1 - z_*)Q(z_*) = 0$ which implies $z_*^{m+1} = \beta_m + \sum_{i=0}^{m-1}(\beta_i - \beta_{i+1})z_*^{m-i}$, where $\beta_0 = 1$. After taking absolute values, we obtain $|z_*^{m+1}| \leq \beta_m + \sum_{i=0}^{m-1}(\beta_i - \beta_{i+1})|z_*^{m-i}| < \beta_m|z_*^m| + \sum_{i=0}^{m-1}(\beta_i - \beta_{i+1})|z_*^m| = \big(\beta_m + \sum_{i=0}^{m-1}(\beta_i - \beta_{i+1})\big)|z_*^m| = |z_*^m|$. Thus, $|z_*^{m+1}| < |z_*^m|$, which is a contradiction. We have just proved that all the roots of Q must lie in $\bar{D}$. Finally, we show that the roots of any $Q(z) = z^m + \beta_1 z^{m-1} + \cdots + \beta_{m-1}z + \beta_m$ satisfying (I55) must actually lie in D. Let $r < 1$ be sufficiently close to 1 so that we have $r^m > \beta_1 r^{m-1} > \ldots > \beta_{m-1}r > \beta_m > 0$. Then, we have seen that the polynomial $Q_r(z) = Q(rz)$ must have all roots in $\bar{D}$. Let z_* be a root of Q. Then, $Q_r\big(\frac{z_*}{r}\big) = Q(z_*) = 0$, which implies that $\frac{z_*}{r} \in \bar{D}$, or equivalently $z_* \in r\bar{D}$. But $r\bar{D} \subset D$, and the proof is now complete. $\square$

Proof of Proposition I.1. Since the forecast error variance equals $\Sigma_t = \mathsf{Var}(w_t - p_{t-1}) = \mathsf{E}\big((w_t - p_{t-1})^2\big)$, we compute the derivative of Σ_t as follows:

$$\begin{aligned} \Sigma_t' &= \frac{1}{\mathrm{d}t}\,\mathsf{E}\Big(2(\mathrm{d}w_{t+1} - \mathrm{d}p_t)(w_t - p_{t-1}) + (\mathrm{d}w_{t+1} - \mathrm{d}p_t)^2\Big), \\ &= -2\frac{\mathsf{Cov}(w_t, \mathrm{d}p_t)}{\mathrm{d}t} + \sigma_w^2 + \frac{\mathsf{Var}(\mathrm{d}p_t)}{\mathrm{d}t}. \end{aligned} \tag{I57}$$

The pricing equation (I46) from the proof of Theorem I.1 shows that $\mathsf{Cov}(w_t, \mathrm{d}y_t) = \lambda\,\mathsf{Var}(\mathrm{d}y_t) = \lambda\sigma_y^2$, therefore using $\mathrm{d}p_t = \lambda \mathrm{d}y_t$ we get

$$\frac{\mathsf{Var}(\mathrm{d}p_t)}{\mathrm{d}t} = \frac{\mathsf{Cov}(w_t, \mathrm{d}p_t)}{\mathrm{d}t} = \lambda^2\sigma_y^2 = \sigma_p^2. \tag{I58}$$

The equation (I57) now implies that $\Sigma_t' = \sigma_w^2 - \sigma_p^2$, which finishes the proof. $\square$

Proof of Proposition I.2. Compared to the setup of Theorem I.1, here there are only speculators with zero lag ($\ell = 0$). Therefore, to finish the proof of this Proposition, all we need to do is to show that the system in (I10) reduces to the system in (I15). With the usual notation, the system in (I10) translates into (I30). In the particular case when all speculators are of type $\ell = 0$, and there are $N_0 = N$ of them, equation (I30)

becomes:[10]

$$
\begin{aligned}
A\left(\frac{1}{N}\,g + g\right) \;&=\; \frac{1}{\Lambda}B, \qquad r \;=\; Ag, \qquad \Lambda \;=\; g'B, \qquad \tilde{\sigma}_y^2 \;=\; \frac{\tilde{\sigma}_u^2}{1-g'r}, \\
A_{i,j} \;&=\; \mathbf{1}_{i=j} - \sum_{k=1}^{\min(i,j)} r_{i-k}r_{j-k}, \qquad B_i \;=\; 1-\Lambda\sum_{k=1}^{i} r_{i-k}.
\end{aligned}
\tag{I59}
$$

We now proceed by observing that the first two equations imply $Ag = r = \frac{N}{N+1}\,\frac{1}{\Lambda}\,B$. When $i = 0$, this equation implies $r_0 = \frac{1}{\Lambda}\,\frac{N}{N+1}$. When $i = 1$, we get $r_1 = \frac{1}{\Lambda}\,\frac{N}{N+1}(1 - \Lambda\, r_0) = \frac{1}{\Lambda}\,\frac{N}{(N+1)^2}$. By induction, we obtain ($i = 0, \ldots, m$):

$$
r_i \;=\; \frac{1}{\Lambda}\,\frac{N}{(N+1)^{i+1}}. \tag{I60}
$$

This proves the equation for ρ_i in (I15). We also get that

$$
B_i \;=\; \frac{1}{(N+1)^i}, \tag{I61}
$$

which proves the equation for B_i in (I15). Moreover, we compute $g'r = g'B\,\frac{N}{N+1}\,\frac{1}{\Lambda}$. But $g'B = \Lambda$, hence

$$
g'r \;=\; \frac{N}{N+1}. \tag{I62}
$$

This implies

$$
\tilde{\sigma}_y^2 \;=\; \frac{\tilde{\sigma}_u^2}{1-g'r} \;=\; (N+1)\tilde{\sigma}_u^2, \qquad \text{or} \qquad \tilde{\sigma}_y \;=\; \sqrt{N+1}\,\tilde{\sigma}_u, \tag{I63}
$$

which proves the equation for $\tilde{\sigma}_y$ in (I15). Using the formula (I60) for r, we use (I59) to compute ($i, j = 0, \ldots, m$):

$$
A_{i,j} \;=\; \mathbf{1}_{i=j} - \frac{1}{\Lambda^2}\,\frac{N}{N+2}\,\frac{(N+1)^{2\min(i,j)}-1}{(N+1)^{i+j}}, \tag{I64}
$$

which proves the equation for $A_{i,j}$ in (I15). Finally, the two equations, $r = Ag$ and $\Lambda = g'B$, after rescaling are equivalent to $\tilde{\sigma}_y^2\rho \;=\; A\bar{\gamma}$ and $\tilde{\sigma}_y^2\lambda \;=\; B'\bar{\gamma}$, which finishes the proof of this Proposition. $\square$

Proof of Theorem I.2. We use the notations from the proofs of Theorem I.1 and Proposition I.2. The idea is to study the behavior of (A, Λ, r, g) around $\Lambda = 1$, but

[10]Note that except for the first equation, the other equations in (I59) are the same as in (I30). We also provide a direct derivation of the first equation in the proof of Proposition I.3.

without yet imposing the condition $\Lambda = g'B$. To do that, define the following numbers:

$$\begin{aligned} A^0_{i,j} &= \mathbf{1}_{i=j} - \frac{N}{N+2}\,\frac{(N+1)^{2\min(i,j)}-1}{(N+1)^{i+j}}, & \Lambda_0 &= 1, \\ r^0_i &= \frac{N}{(N+1)^{i+1}}, & g^0_i &= \frac{N}{N+1}\,\frac{(m-i+1)N+1}{(m+1)N+1}. \end{aligned} \tag{I65}$$

One verifies the formula:[11]

$$A^0 g^0 = r^0, \quad \text{or, equivalently,} \quad g^0 = \left(A^0\right)^{-1} r^0. \tag{I66}$$

For $\varepsilon < 1$, define the following variables:

$$\begin{aligned} A^\varepsilon_{i,j} &= \mathbf{1}_{i=j} - \frac{1}{1-\varepsilon}\,\frac{N}{N+2}\,\frac{(N+1)^{2\min(i,j)}-1}{(N+1)^{i+j}}, & \Lambda^\varepsilon &= \sqrt{1-\varepsilon} \\ r^\varepsilon_i &= \frac{1}{\sqrt{1-\varepsilon}}\,\frac{N}{(N+1)^{i+1}}, & g^\varepsilon &= \left(A^\varepsilon\right)^{-1} r^\varepsilon, \end{aligned} \tag{I67}$$

whenever A^ε is invertible. Using (I66), one sees that the variables defined in (I65) are the same as the variables in (I67) in the particular case when $\varepsilon = 0$. In other words, given a solution $(A, B, \Lambda, r, g, \tilde{\sigma}_y)$ to the system (I30), if we let

$$\varepsilon = 1 - \Lambda^2, \tag{I68}$$

it must be that (Λ, A, r, g) satisfy

$$\Lambda = \Lambda^\varepsilon, \quad A = A^\varepsilon, \quad r = r^\varepsilon, \quad g = g^\varepsilon. \tag{I69}$$

We now multiply the equation for $A = A^\varepsilon$ from (I67) by $\Lambda^2 = 1 - \varepsilon$, to obtain

$$\Lambda^2 A = A^0 - \varepsilon I, \tag{I70}$$

where I is the identity matrix ($I_{i,j} = \mathbf{1}_{i=j}$). This implies $\frac{1}{\Lambda^2} A^{-1} = \left(A^0 - \varepsilon I\right)^{-1}$. Multiplying this equation to the right with $r = \frac{1}{\Lambda} r^0$, we obtain

$$\frac{g}{\Lambda} = \left(A^0 - \varepsilon I\right)^{-1} r^0. \tag{I71}$$

[11] This can be done either directly, or by using the method described below which involves recursive computation of the inverse matrix, $(A^0)^{-1}$. Then, one verifies by induction that $(A^0)^{-1} r^0 = g^0$.

If we multiply this equation to the left with B', and use $B'g = \Lambda$, we obtain

$$1 \;=\; B'\big(A^0 - \varepsilon I\big)^{-1} r^0. \tag{I72}$$

This equation determines ε, or equivalently $\Lambda = \sqrt{1-\varepsilon}$. We make this equation more explicit by observing that the inverse matrix $\big(A^0 - \varepsilon I\big)^{-1}$ has the following series expansion:

$$\big(A^0 - \varepsilon I\big)^{-1} \;=\; (A^0)^{-1} + \varepsilon\,(A^0)^{-2} + \varepsilon^2 (A^0)^{-3} + \cdots . \tag{I73}$$

By multiplying this equation to the left by B' and to the right by r^0, we get

$$1 \;=\; B'\big(A^0 - \varepsilon I\big)^{-1} r^0 \;=\; B'(A^0)^{-1} r^0 + \varepsilon\, B'(A^0)^{-2} r^0 + \varepsilon^2 B'(A^0)^{-3} r^0 + \cdots . \tag{I74}$$

We compute $1 - B'g^0 = \frac{1}{(m+1)N+1}$. Since $\big(A^0\big)^{-1} r^0 = g^0$, we obtain

$$\frac{1}{(m+1)N+1} \;=\; \varepsilon\, B'\big(A^0\big)^{-1} g^0 + \varepsilon^2 B'\big(A^0\big)^{-2} g^0 + \varepsilon^3 B'\big(A^0\big)^{-3} g^0 + \cdots . \tag{I75}$$

We now determine sufficient conditions for the existence of an equilibrium. From the previous discussion, we need the following conditions: (i) $\varepsilon < 1$ (Λ is well defined and $\Lambda > 0$); (ii) $(A^0 - \varepsilon I)$ is invertible or equivalently $A = A^{\varepsilon}$ is invertible (g is well defined); and (iii) the numbers β_k from the proof of Theorem I.1 satisfy (I53) for $k = 1, \ldots, m$ (which implies $\mathsf{Cov}(w_t, \mathrm{d}y_t)$ is independent of t). With the current notation, condition (iii) requires that[12]

$$1 \;>\; \beta_1 \;>\; \cdots \;>\; \beta_m \;>\; 0, \quad \text{with} \quad \beta_k \;=\; \sum_{i=0}^{m-k} r_i g_{k+i}. \tag{I76}$$

We also introduce the following condition that implies (i) and (ii):

$$\text{Equation (I75) has a solution } \varepsilon \in \left(0\,,\; \frac{1}{\frac{N(m+2)^2}{8} + \frac{(m+2)^2}{m+1}}\right). \tag{I77}$$

Since $\frac{N(m+2)^2}{8} + \frac{(m+2)^2}{m+1} > 1$, clearly (I77) implies $\varepsilon < 1$, which proves (i). The difficult part is to show that (I77) also implies that $(A^0 - \varepsilon I)$ is invertible, which proves (ii). For this, we need a better understanding of the inverse matrix $\big(A^0\big)^{-1}$. Denote by $A^0_{(m)} \in M_{m+1}$ the matrix A^0 from (I65) by making explicit the dependence on m. Since A^0 satisfies $A^0_{i,j} = \mathbf{1}_{i=j} - \sum_{k=1}^{\min(i,j)} r_{i-k} r_{j-k}$, it follows that the block $\big(A^0_{(m)}\big)^{11}$ which is

[12] One can check that condition (iii) is implied by the condition $g_1 > g_2 > \cdots > g_m > 0$.

obtained $A^0_{(m)}$ by removing the last row and the last column is the same as $A^0_{(m-1)}$. We then have

$$A^0_{(m)} \;=\; \begin{bmatrix} A^0_{(m-1)} & a_{(m)} \\ a'_{(m)} & \alpha_{(m)} \end{bmatrix}, \tag{I78}$$

for some m-column vector $a_{(m)}$, and scalar $\alpha_{(m)}$. Write the inverse matrix $H_{(m)} = (A^0_{(m)})^{-1}$ also in block format:

$$H_{(m)} \;=\; \begin{bmatrix} H^{11}_{(m)} & h_{(m)} \\ h'_{(m)} & \eta_{(m)} \end{bmatrix}. \tag{I79}$$

From the theory of block matrices, we have the following formulas:

$$\begin{aligned} \eta_{(m)} \;&=\; \frac{1}{\alpha_{(m)} - a'_{(m)}\big(A^0_{(m-1)}\big)^{-1} a_{(m)}}, \\ h_{(m)} \;&=\; -\eta_{(m)}\big(A^0_{(m-1)}\big)^{-1} a_{(m)}, \\ H^{11}_{(m)} \;&=\; \big(A^0_{(m-1)}\big)^{-1} + \frac{h_{(m)}h'_{(m)}}{\eta_{(m)}}. \end{aligned} \tag{I80}$$

By induction, one verifies that

$$\begin{aligned} \eta_{(m)} \;&=\; \frac{(mN+1)(N+1)}{(m+1)N+1}, \\ h_{(m)} \;&=\; \frac{N^2}{(m+1)N+1}\Big[\,0\;,\;1\;,\;\cdots\;,\;m-1\,\Big]'. \end{aligned} \tag{I81}$$

Using the equations above, one can now prove various useful formulas. As a first result, we prove by induction that

$$\det\big(A^0_{(m)}\big) \;=\; \frac{(m+1)N+1}{(N+1)^{m+1}}. \tag{I82}$$

For $m = 0$, the equality is true, since in this case $A^0_{(0)} = 1$. Suppose it is true for $m-1$. From the theory of block matrices, $\det\big(A^0_{(m)}\big) = \det\big(A^0_{(m-1)}\big)\ \det\Big(\alpha_{(m)} - a'_{(m)}\big(A^0_{(m-1)}\big)^{-1}a_{(m)}\Big) = \frac{\det\left(A^0_{(m-1)}\right)}{\eta_{(m)}}$, which together with the formula for $\eta_{(m)}$ from (I81) proves the induction step. Another useful result is:

$$H_{i,j} \;\geq\; 0, \quad i,j = 0,\ldots,m. \tag{I83}$$

Indeed, one uses the recursive formula $H^{11}_{(m)} = H_{(m-1)} + \frac{h_{(m)}h'_{(m)}}{\eta_{(m)}}$ and the explicit formulas

in (I81) to verify by induction that all entries of $H = H_{(m)}$ are positive.[13]

In order to prove that the matrix $(A^0 - \varepsilon I)$ is invertible, we rewrite equation (I84):

$$\big(A^0 - \varepsilon I\big)^{-1} \; = \; H\,\Big(1 + \varepsilon H + \varepsilon^2 H^2 + \varepsilon^3 H^3 + \cdots\Big). \tag{I84}$$

Thus, if we can show that the right hand side is a convergent series (in the space of matrices), then its limit is a matrix which coincides with the matrix inverse $\big(A^0 - \varepsilon I\big)^{-1}$. To prove convergence, we use the infinity norm, $\|H\|_\infty$, which is the maximum absolute row sum of the matrix, i.e., $H = \max\limits_{i=0,\dots,m} \sum_{j=0}^{m} |H_{i,j}|$, . Thus, if we can show that $\|\varepsilon H\|_\infty < 1$, this proves condition (ii).

We now search for an upper bound for $\|H\|_\infty$. For instance, we show that $\big\|\frac{H}{N(m+2)^2}\big\|_\infty \leq \frac{1}{4}\big(1 + O_N\frac{1}{N})\big)$. For this, define $\bar{h}_{(m)}$ the $(m+1)$-column vector given by $\big(\bar{h}_{(m)}\big)_i = (N+1)\big((m+1)N+1\big)\,\sum_{j=0}^{m}\big(H_{(m)}\big)_{i,j}$. This is proved by induction to be a polynomial in N of degree 3. Denote by $C_{(m)}$ the vector of coefficients of N^3 in $\bar{h}_{(m)}$. Note that $\max\limits_{i=0,\dots,m} \bar{h}_{(m)} = N^2(m+1)\|H\|_\infty\big(1 + O_N\frac{1}{N})\big)$. At the same time, we have $\max\limits_{i=0,\dots,m} \bar{h}_{(m)} = N^3 \max\limits_{i=0,\dots,m} C_{(m)}$, which implies $\big\|\frac{H}{N(m+2)^2}\big\|_\infty = \frac{1}{(m+1)(m+2)^2} \max\limits_{i=0,\dots,m} C_{(m)}\,\big(1 + O_N\frac{1}{N})\big)$. Now one computes $C_{(0)} = 0$, and for $m > 1$ one uses the recursive formulas above for H to get a recursive formula for C. More precisely, $\Big(\frac{C_{(m)}}{m+1}\Big)_i = \Big(\frac{C_{(m-1)}}{m}\Big)_i + \frac{i}{2}$ for $i = 0, \dots, m-1$, and $\Big(\frac{C_{(m)}}{m+1}\Big)_m = \frac{m}{2}$. By induction then, one shows that $\max\limits_i\big(C_{(m)}\big)_i \leq \frac{(m+1)^2(m+2)}{4}$, which implies the upper bound stated above for $\|H\|_\infty$. By similar methods, one verifies a sharper estimate:

$$\left\|\frac{H}{N(m+2)^2}\right\|_\infty \; \leq \; \frac{1}{8} + \frac{1}{(m+1)N}. \tag{I85}$$

Note that condition (I77) implies $\varepsilon\, N(m+2)^2 < \frac{1}{\frac{1}{8} + \frac{1}{(m+1)N}}$, which together with (I85) implies

$$\varepsilon\,\|H\|_\infty \; < \; 1. \tag{I86}$$

This proves that the series $1 + \varepsilon H + \varepsilon^2 H^2 + \varepsilon^3 H^3 + \cdots$ is convergent, and that the limit coincides with $\big(A^0 - \varepsilon I\big)^{-1}$.

Next, we analyze how well (Λ, r, g) approximate (Λ^0, r^0, g^0). Recall that

$$\gamma \; = \; \frac{\bar{\gamma}}{N} \; = \; \frac{g\tilde{\sigma}_y}{N}, \qquad \rho \; = \; \frac{r}{\tilde{\sigma}_y}, \qquad \lambda \; = \; \frac{\Lambda}{\tilde{\sigma}_y}, \tag{I87}$$

where from (I63),

$$\tilde{\sigma}_y \; = \; \sqrt{N+1}\,\tilde{\sigma}_u \; = \; \sqrt{N+1}\,\frac{\sigma_u}{\sigma_w}. \tag{I88}$$

[13]The inequality is strict except that $H_{i,0} = H_{0,i} = 0$ for $i > 0$.

Note that in the statement of the Theorem we have defined ($i = 0, \dots, m$):

$$\gamma_i^0 \;=\; \frac{\tilde{\sigma}_y}{N+1}\,\frac{m-i+1}{m+1}, \qquad \rho_i^0 \;=\; \frac{1}{\tilde{\sigma}_y}\,\frac{N}{(N+1)^{i+1}}, \qquad \lambda^0 \;=\; \frac{1}{\tilde{\sigma}_y}. \tag{I89}$$

Condition (I77) implies $\varepsilon < \frac{1}{\frac{N(m+2)^2}{8} + \frac{(m+2)^2}{m+1}}$, which shows that $\varepsilon = O_N\big(\frac{1}{N}\big)$. Also, since $\Lambda = \sqrt{1-\varepsilon}$, it follows that $\Lambda = 1 - O_N\big(\frac{1}{N}\big)$. Thus, we obtain

$$\varepsilon \;=\; O_N\big(\tfrac{1}{N}\big), \qquad \Lambda \;=\; \sqrt{1-\varepsilon} \;=\; 1 - O_N\big(\tfrac{1}{N}\big). \tag{I90}$$

Now, from (I87) and (I89), we obtain $\frac{\lambda}{\lambda^0} = \Lambda = 1 - O_N\big(\frac{1}{N}\big)$. This proves the approximate equation for λ in (I17). From (I87) and (I89), we also compute $\frac{\rho_i}{\rho_i^0} = \frac{r_i}{r_i^0}$, since $r_i^0 = \frac{N}{(N+1)^{i+1}}$. But $r = \frac{1}{\Lambda} r^0$, which implies $\frac{\rho_i}{\rho_i^0} = \frac{1}{\Lambda} = 1 + O_N\big(\frac{1}{N}\big)$. This proves the approximate equation for ρ_i in (I17). Finally, from (I87) and (I89), we get $\frac{\gamma_i}{\gamma_i^0} = \frac{g_i}{\frac{N}{N+1}\frac{m-i+1}{m+1}} = \frac{g_i}{g_i^0}\big(1 + O_N\big(\frac{1}{N}\big)\big)$. We now show that $\frac{g_i}{g_i^0} = O_N(1)$, which proves the approximate equation for γ_i in (I17). Since $\gamma_i^0 = O_N(1)$, it is enough to show that $g_i - g_i^0 = O_N(1)$, or from (I90) it is enough to show $\frac{g_i}{\Lambda} - g_i^0 = O_N(1)$. If we combine equations (I71) and (I84), and use $\big(A^0\big)^{-1} r^0 = g^0$, we get

$$\frac{g}{\Lambda} \;=\; \Big(1 + \varepsilon H + \varepsilon^2 H^2 + \varepsilon^3 H^3 + \cdots\Big)\, g^0. \tag{I91}$$

Therefore, we get $\frac{g}{\Lambda} - g^0 = \big(\varepsilon H + \varepsilon^2 H^2 + \varepsilon^3 H^3 + \cdots\big)\, g^0$. But (I86) implies that the convergent series of matrices is of the order $O_N(1)$, hence it remains of order $O_N(1)$ when multiplied with $g^0 = O_N(1)$. □

Proof of Proposition I.3. Following the proof of Theorem I.1, consider a speculator who must choose the weights $\gamma_{i,t}$ on $\mathrm{d}_t w_{t-i}$. He assumes that all the other speculators use γ_i^*, hence with an aggregate weight of $(N-1)\gamma_i^*$ on $\mathrm{d}_t w_{t-i}$. Then, equation (I33) for the speculator's normalized expected profit at $t = 0$ becomes

$$\begin{aligned}
\tilde{\pi}_0 \;&=\; \frac{\pi_0}{\sigma_w^2} \;=\; \frac{1}{\sigma_w^2}\, \mathsf{E}\bigg(\int_0^T (w_t - p_{t-1} - \lambda \mathrm{d}y_t)\mathrm{d}x_t\bigg) \\
&=\; \mathsf{E}\Bigg(\int_0^T \Big(w_t - \lambda \sum_{i=0}^m \big(\gamma_{i,t} + (N-1)\gamma_i^*\big)\mathrm{d}_t w_{t-i}\Big) \sum_{j=0}^m \gamma_{j,t} \mathrm{d}_t w_{t-j}\Bigg) \\
&=\; \sum_{j=0}^m B_j \gamma_{j,t} - \lambda \sum_{i,j=0}^m \big(\gamma_{i,t} + (N-1)\gamma_i^*\big) A_{i,j} \gamma_{j,t}.
\end{aligned} \tag{I92}$$

The first order condition with respect to $\gamma_{k,t}$, for $k = 0, \ldots, m$, is:

$$B_k - \lambda \sum_{i=0}^{m} \big(2\gamma_{i,t} + (N-1)\gamma_i^*\big) A_{i,k} \ = \ 0. \tag{I93}$$

Since this equation is true for all speculators, we obtain that all $\gamma_{i,t}$ are equal and independent on t, i.e., $\gamma_i = \gamma_i^* = \frac{\bar{\gamma}}{N}$. Using matrix notation, we have $B = \lambda(N+1)A\gamma$, hence $B = \lambda \frac{N+1}{N} A\bar{\gamma}$.[14] Thus, $\frac{N}{N+1}B = \lambda A \bar{\gamma}$, which implies $B - \lambda A\bar{\gamma} = \frac{B}{N+1}$. Thus, in equilibrium the normalized expected profit is equal to

$$\tilde{\pi}_0 \ = \ \sum_{j=0}^{m} \left(B_j - \lambda \sum_{i=0}^{m} A_{j,i}\bar{\gamma}_i \right) \gamma_j \ = \ \sum_{j=0} \frac{B_j}{N+1}\, \gamma_j. \tag{I94}$$

From equation (I61), $B_j = \frac{1}{(N+1)^j}$. We compute

$$\tilde{\pi}_0 \ = \ \sum_{j=0}^{m} \tilde{\pi}_{0,j} \ = \ \sum_{j=0}^{m} \frac{\gamma_j}{(N+1)^{j+1}}, \tag{I95}$$

which proves (I18). The ratio of two consecutive components is

$$\frac{\tilde{\pi}_{0,j+1}}{\tilde{\pi}_{0,j}} \ = \ \frac{\gamma_{j+1}}{\gamma_j} \frac{1}{N+1} \ = \ \frac{g_{j+1}}{g_j} \frac{1}{N+1}. \tag{I96}$$

But in the proof of Theorem I.2 we have shown that $g_j = O_N(1)$. Thus, $\frac{\tilde{\pi}_{0,j+1}}{\tilde{\pi}_{0,j}} = O_N\big(\frac{1}{N}\big)$, which finishes the proof of the Proposition. $\square$

Proof of Propositions I.4. Let $\sigma_w^2 - \sigma_p^2 = \sigma_w^2(1 - \lambda^2 \tilde{\sigma}_y^2) = \sigma_w^2(1 - \Lambda^2)$. In the proof of Theorem I.2, we have seen that $1 - \Lambda^2 = \varepsilon$, and this is strictly positive according to the condition (I77). Moreover, from (I90), we have $\varepsilon = O_N\big(\frac{1}{N}\big)$, which finishes the first part of the Proposition.

For the second part, we need only to prove the equation $SR_k = \frac{N}{(N+1)^{k+1}}$ when $k = 0, \ldots, m$. From the definition of SR_k, we obtain

$$SR_k \ = \ \frac{\lambda\, \mathsf{Cov}(\mathrm{d}w_{t-k}, \mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} \ = \ \frac{\lambda \rho_k\, \mathsf{Var}(\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} \ = \ \lambda \rho_k \tilde{\sigma}_y^2 \ = \ \frac{N}{(N+1)^{k+1}}, \tag{I97}$$

where the second equality comes from (I46), and the last equation comes from (I15). $\square$

[14] Since $\Lambda = \lambda\,\tilde{\sigma}_y$ and $g = \frac{\bar{\gamma}}{\tilde{\sigma}_y}$, we obtain $\frac{1}{\Lambda}B = \frac{N+1}{N}Ag$, which provides a direct proof of the first equation in (I59).

J Equilibrium with Inventory Management

In this section, we discuss the general equilibrium of the model with inventory management from Section 5, and provide the proofs that have been left out of the paper. We recall that in the model with inventory management the agents are:

- One IFT, who chooses a trading strategy of the form $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$, with $\Theta \in [0, 2)$ and $G \in \mathbb{R}$. The IFT maximizes the expected utility U given by (42):

$$U \;=\; \mathsf{E}\left(\int_0^T (v_1 - p_t)\mathrm{d}x_t\right) \;-\; C_I\,\mathsf{E}\left(\int_0^T x_t^2 \mathrm{d}t\right), \tag{J1}$$

  where $T = 1$, and $C_I > 0$ is the IFT's inventory aversion coefficient;

- N_F risk-neutral fast traders, who choose a trading strategy $\mathrm{d}x_{i,t} = \gamma_i \mathrm{d}w_t$, with $\gamma_i \in \mathbb{R}$;

- N_L risk-neutral slow traders, who choose a trading strategy $\mathrm{d}x_{j,t} = \mu_j \widetilde{\mathrm{d}w}_{t-1}$, with $\mu_j \in \mathbb{R}$; the term $\widetilde{\mathrm{d}w}_{t-1}$ is of the form $\widetilde{\mathrm{d}w}_{t-1} = \mathrm{d}w_{t-1} - z_{t-1,t}$, where $z_{t-1,t}$ is the dealer's expectation of $\mathrm{d}w_{t-1}$ given past order flow;

- A dealer who sets a linear pricing rule $\mathrm{d}p_t = \lambda \mathrm{d}y_t$ such that the dealer's expected profit is zero; she also computes $z_{t-1,t} = \mathsf{E}_t(\mathrm{d}w_{t-1}) = \rho \mathrm{d}y_{t-1}$;

- Exogenous noise traders, who on aggregate submit at each t a market order $\mathrm{d}u_t$.

We introduce the following model coefficients:

$$\begin{aligned}
&R \;=\; \frac{\lambda}{\rho}, \qquad \gamma^- \;=\; \sum_{i=1}^{N_F} \gamma_i, \qquad \bar{\gamma} \;=\; \gamma^- + G, \qquad \bar{\mu} \;=\; \sum_{j=1}^{N_L} \mu_j, \\
&a^- \;=\; \rho\gamma^-, \qquad a \;=\; \rho\bar{\gamma}, \qquad b \;=\; \rho\bar{\mu}.
\end{aligned} \tag{J2}$$

The next three sections describe the equilibrium of the model, by considering one agent at a time and taking the behavior of the other agents as given. Section J.4 puts together all the equilibrium conditions, and derives a single system of equations that the model coefficients should satisfy. In doing so, we also prove Theorem 3.

J.1 Optimal Inventory Management

In this section, we describe the optimal choice of the IFT, while taking the behavior of the FTs, the STs, and the dealer as given. Since we wish to prove a more general result

than Theorem 2, we also analyze the case when the slow trading coefficient b is below the threshold $\leq \frac{\sqrt{17}-1}{8}$. Proposition J.1 below shows when b is below the threshold, a sufficiently averse IFT optimally chooses Θ positive but as small as possible. We denote this case by

$$\Theta \;=\; 0_{+}. \tag{J3}$$

This case is different from $\Theta = 0$. Indeed, at $\Theta = 0$ the IFT's inventory follows a random walk, while at $\Theta > 0$ the IFT's inventory is negligible, as one can see for instance in equation (A32). This shows that inventory management has a discontinuity at $\Theta = 0$.

Corollary K.1 in Internet Appendix K shows that the cases $\Theta = 0$ and $\Theta = 0_{+}$ are joined continuously by a *smooth regime*, in which the IFT has a strategy of the form $\mathrm{d}x_t = -\theta x_{t-1}\mathrm{d}t + G\mathrm{d}w_t$, with $\theta \in [0, \infty]$. (Continuity here means that the IFT's expected utility varies continuously across the regimes.) When $\theta = \infty$, the IFT has the same expected utility as in the case $\Theta = 0_{+}$. Hence, proposition J.1 suggests that when $\Theta = 0_{+}$ is optimal, this is because we have not included the smooth regime in the analysis. If we did include it the smooth regime, we conjecture that $\Theta = 0_{+}$ means that the IFT optimally chooses an interior point $\theta \in (0, \infty)$ in the smooth regime.

Proposition J.1. *Consider the behavior of the other speculators and the dealer as given, and fix the coefficients $\gamma \geq 0$, $\mu > 0$, $\lambda > 0$, $\rho > 0$. Define $\bar{\gamma}$, $\bar{\mu}$, a^{-}, b and R as in (J2). Moreover, suppose that $b = \rho\bar{\mu} < 1$. Then, if $\bar{C}_I$ and $\bar{C}_I^0$ as in (J34), the optimal strategy of the IFT is as follows:*

(i) If $b \leq \frac{\sqrt{17}-1}{8} = 0.3904$, and $C_I \leq \bar{C}_I^0$, the IFT sets

$$\Theta \;=\; 0, \qquad G \;=\; \frac{1 - Ra^{-}}{2\lambda + C_I}. \tag{J4}$$

If $b \leq \frac{\sqrt{17}-1}{8}$, and $C_I > \bar{C}_I^0$, the IFT sets

$$\Theta \;=\; 0_{+}, \qquad G \;=\; \frac{b(1 - a^{-})}{2\rho\big(b + \frac{1}{2}\big)}. \tag{J5}$$

In the latter case, the maximum expected utility of the IFT is

$$U^{\max}_{\Theta=0_{+}} \;=\; \frac{\big(Rb(1 - a^{-})\big)^2}{4\lambda(1 + b)\big(b + \frac{1}{2}\big)}\sigma_w^2. \tag{J6}$$

(ii) If $b > \frac{\sqrt{17}-1}{8}$, and $C_I \leq \bar{C}_I$, the IFT sets

$$\Theta \;=\; 0, \qquad G \;=\; \frac{1-Ra^{-}}{2\lambda + C_I}. \tag{J7}$$

If $b > \frac{\sqrt{17}-1}{8}$, and $C_I > \bar{C}_I$, the IFT sets

$$\Theta \;=\; 2 - \frac{\sqrt{1-b}}{b}, \qquad G \;=\; \frac{1-a^{-}}{2\rho\left(1+\frac{1}{\sqrt{1-b}}\right)}. \tag{J8}$$

In the latter case, the maximum expected utility of the IFT is

$$U^{\max}_{\Theta>0} \;=\; \frac{\big(Rb(1-a^{-})\big)^2}{4\lambda b(1+\sqrt{1-b})^2}\sigma_w^2. \tag{J9}$$

Proof. See Section J.5. □

J.2 Optimal Strategies of Fast and Slow Traders

In this section, we describe the optimal choice of the FTs and STs, while taking the behavior of the IFT and the dealer as given.

Proposition J.2. *Consider the behavior of the IFT and the dealer as given, and fix the coefficients $G \in \mathbb{R}$, $\Theta \in (0,2)$, $\rho \geq 0$, $\lambda > 0$. Suppose there exists a solution to the following system of equations ($\phi = 1 - \Theta$):*

$$\begin{aligned}
\bar{\gamma} \;&=\; \frac{N_F}{\lambda(N_F+1)} + \frac{G}{N_F+1}, \qquad \bar{\mu} \;=\; \frac{E+\lambda\Theta X}{\lambda W}\,\frac{N_L}{N_L+1}, \\
X \;&=\; \frac{\frac{\rho\phi G^2}{1+\phi} + G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}}, \qquad Z \;=\; \frac{\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}} - \frac{G^2}{(1+\phi)(1+\phi\rho\bar{\mu})}, \\
Y \;&=\; \frac{-\frac{\Theta G^2}{1+\phi} - 2\Theta Z + \bar{\gamma}^2 + \bar{\mu}^2(1-2\rho\bar{\gamma}) + \tilde{\sigma}_u^2}{1-\rho^2\bar{\mu}^2}, \\
W \;&=\; 1 - 2\rho\bar{\gamma} + \rho^2 Y \;=\; \frac{-\frac{\Theta\rho^2 G^2}{1+\phi} - 2\rho^2\Theta Z + (1-\rho\bar{\gamma})^2 + \rho^2\tilde{\sigma}_u^2}{1-\rho^2\bar{\mu}^2}, \\
E \;&=\; \frac{1-\lambda\bar{\gamma} - \rho\bar{\gamma} + \rho G(1-\lambda\bar{\gamma}) - \lambda\rho\phi Z + \rho\lambda Y}{1+\rho\bar{\mu}}.
\end{aligned} \tag{J10}$$

and that this solution satisfies

$$0 \;\leq\; \rho\bar{\mu} \;<\; 1, \qquad W \;>\; 0. \tag{J11}$$

Define the coefficients γ and μ by:

$$\gamma \;=\; \frac{1-\lambda G}{\lambda(N_F+1)}, \qquad \mu \;=\; \frac{E+\lambda\Theta X}{\lambda W(N_L+1)}. \tag{J12}$$

Then, the optimal trading strategies of the FTs and the STs satisfy $\gamma_i = \gamma$ for all $i = 1,\dots,N_F$, and all $\mu_j = \mu$ for $j = 1,\dots,N_L$.

Proof. See Section J.5. □

For future reference, in Corollary J.1 we describe the profit function of the FTs and STs if their trading strategy is symmetric (the same for the FTs and the same for the STs), but not necessarily the optimal one.

Corollary J.1. *Consider the behavior of the IFT and the dealer as given, and suppose the trading strategies of the FTs and the STs satisfy, respectively,*

$$\mathrm{d}x_t^F \;=\; \gamma \mathrm{d}w_t + \mu\widetilde{\mathrm{d}w}_{t-1}, \qquad \mathrm{d}x_t^S \;=\; \mu\widetilde{\mathrm{d}w}_{t-1}, \tag{J13}$$

with $\widetilde{\mathrm{d}w}_{t-1} = \mathrm{d}w_{t-1} - \rho \mathrm{d}y_{t-1}$. Denote by $\gamma^- = N_F\gamma$, $\bar{\gamma} = \gamma^- + G$, $\bar{\mu} = N_L\mu$. Then, the expected profits of the FTs and STs satisfy, respectively,

$$\pi^F \;=\; \gamma(1-\lambda\bar{\gamma})\sigma_w^2, \qquad \pi^S \;=\; \mu\big(E+\lambda\Theta X-\lambda W\bar{\mu}\big)\sigma_w^2, \tag{J14}$$

where E, X, W are defined as in (J10).

Proof. See equation (J53) from the proof of Proposition J.2. □

J.3 Dealer's Pricing Rules with Inventory Management

In this section, we describe the dealer's pricing functions, while taking the behavior of the IFT and of the FTs and STs as given.

Proposition J.3. *Consider the behavior of the speculators as given, and fix the coefficients $G \in \mathbb{R}$, $\Theta \in (0,2)$, $\gamma \geq 0$, $\mu > 0$. Denote by $\phi = 1-\Theta$, and by $\gamma^- = N_F\gamma$, $\bar{\gamma} = G + \gamma^-$, $\bar{\mu} = N_L\mu$. Suppose the following third degree equation in ρ has a solution $\rho > 0$:*

$$\begin{aligned}\bar{\gamma}(1-\rho^2\bar{\mu}^2)(1+\phi\rho\bar{\mu}) \;=\; \Theta\, & \frac{G^2(1-\phi\rho\bar{\mu}) - 2\bar{\mu}G(1+\phi)(1-\rho\bar{\gamma})}{(1+\phi)} \\ & + \big(\bar{\gamma}^2 + \bar{\mu}^2(1-2\rho\bar{\gamma}) + \tilde{\sigma}_u^2\big)(1+\phi\rho\bar{\mu}).\end{aligned} \tag{J15}$$

Define Z and Y by the formulas:

$$Z = \frac{\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}} - \frac{G^2}{(1+\phi)(1+\phi\rho\bar{\mu})}, \qquad Y = \frac{\bar{\gamma}}{\rho}. \tag{J16}$$

Then, the dealer sets ρ equal to the solution of (J15), and sets λ as follows:

$$\lambda = \frac{\bar{\mu} + (\bar{\gamma} - G)}{Y + \bar{\gamma}\bar{\mu} - \bar{\gamma}G - \phi Z}. \tag{J17}$$

Proof. See Section J.5. □

J.4 Equilibrium Conditions

In this section, we solve for the equilibrium of the inventory management model with one IFT, N_F fast traders, and N_L slow traders. Let

$$n_F = \frac{N_F}{N_F+1}, \qquad n_L = \frac{N_L}{N_L+1}. \tag{J18}$$

We now collect all the equilibrium conditions from the partial equilibrium results from Sections J.1–J.3, to generate the full equilibrium conditions. We restate Theorem 3 from the paper as Theorem J.1 below. It provides necessary and sufficient conditions for an equilibrium of the model.

Theorem J.1 (Theorem 3)**.** *Suppose there is an equilibrium in which the speculators's strategies are:* $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$ *(the IFT),* $\mathrm{d}x_t^F = \gamma \mathrm{d}w_t$ *(the FTs),* $\mathrm{d}x_t^S = \mu\widetilde{\mathrm{d}w}_{t-1}$ *(the STs); and the dealer's pricing rules are:* $\mathrm{d}p_t = \lambda \mathrm{d}y_t$, $\widetilde{\mathrm{d}w}_t = \mathrm{d}w_t - \rho \mathrm{d}y_t$. *Denote the model coefficients R, a^-, b as follows:*

$$R = \frac{\lambda}{\rho}, \qquad \gamma^- = N_F\gamma, \qquad \bar{\mu} = N_L\mu, \qquad a^- = \rho\gamma^-, \qquad b = \rho\bar{\mu}. \tag{J19}$$

Suppose $b > \frac{\sqrt{17}-1}{8}$. Then, the equilibrium coefficients satisfy the following equations:[15]

$$
\begin{aligned}
\frac{2b(1+b)(2B+1)}{n_L} &= \frac{Q}{B^2(a^-+b)} + \frac{3bB+2b^2B-1-b}{b}(1-a^-) - 2,\\
B &= \frac{1}{\sqrt{1-b}}, \qquad q = (B+1)\Big(2(B^2-1) - n_F(3B^2-2)\Big),\\
a^- &= \frac{-q \pm \sqrt{q^2 + n_F B^5\big((4-n_F)B + 2(2-n_F)\big)}}{B^2\big((4-n_F)B+2(2-n_F)\big)},\\
Q &= B^3(a^-)^2 + 2(3B^3+3B^2-2B-1)a^- + (B^3+2B^2-2),\\
R &= \frac{4(B+1)B^2\big(a^-+b\big)}{Q}, \qquad a = \frac{(2B+1)a^-+1}{2(B+1)}\\
\rho^2 &= \left((a-b^2) + \frac{2bB-1}{2B+1}(1-a)\right)(1-a)\frac{\sigma_w^2}{\sigma_u^2}, \qquad \lambda = R\rho\\
\Theta &= 2 - \frac{\sqrt{1-b}}{b}, \qquad G = \frac{1-a}{\rho(2B+1)}, \qquad \gamma = \frac{a^-}{\rho N_F}, \qquad \mu = \frac{b}{\rho N_L}.
\end{aligned}
\tag{J20}
$$

Conversely, suppose the equations (J20) have a real solution such that $\frac{\sqrt{17}-1}{8} < b < 1$, $a < 1$, $\lambda > 0$. Then, the speculators' strategies and the dealer's pricing rules with these coefficients provide an equilibrium of the model.

In equilibrium, the expected profits of the IFT, FT, ST are respectively,

$$
\pi = \frac{Rb}{\rho}\frac{(1-a^-)^2}{(1+\sqrt{1-b})^2}\sigma_w^2, \qquad \pi^F = \lambda\gamma^2\sigma_w^2, \qquad \pi^S = \lambda\mu^2(1-a)\sigma_w^2. \tag{J21}
$$

Note that Theorem 3 suggest a procedure to search numerically for an equilibrium, once the parameters N_F, N_L, σ_w, and σ_u are given. Indeed, if we substitute the formulas for a^-, q and B in the first equation of (J20), this becomes a non-linear equation in one variable, b. This equation can be solved numerically very efficiently. Then, one needs to verify that the conditions $\frac{\sqrt{17}-1}{8} < b < 1$, $a < 1$, $\lambda > 0$ are satisfied. Then, the equations in (J20) provide formulas for all the equilibrium coefficients of the model.

J.5 Proofs of Results

Proof of Proposition J.1. We follow the proof of Theorem 2, with a few modifications.

When $\Theta = 0$, the trading strategy of the IFT is $\mathrm{d}x_t = G\,\mathrm{d}w_t$. As in the proof of

[15] To be rigorous, we have included the case when a^- is negative. However, numerically this case never occurs in equilibrium, because it leads to $\lambda < 0$, which contradicts the FT's second order condition (J56) in Internet Appendix J.

Theorem 2, the normalized expected utility of the IFT is $\tilde{U}_{\Theta=0} = G(1-\lambda\bar{\gamma}) - \frac{C_I}{2}G^2$. Since $\bar{\gamma} = \gamma^- + G$, we have $\tilde{U}_{\Theta=0} = G(1-\lambda\gamma^-) - \lambda G^2 - \frac{C_I}{2}G^2$. Since $\lambda\gamma^- = \frac{\lambda}{\rho}\rho\gamma^- = Ra^-$, we get:

$$\tilde{U}_{\Theta=0} \;=\; G\big(1 - Ra^-\big) - G^2\left(\lambda + \frac{C_I}{2}\right). \tag{J22}$$

The function $\tilde{U}_{\Theta=0}$ attains its maximum at

$$G \;=\; \frac{1 - Ra^-}{2\lambda + C_I}, \tag{J23}$$

as stated in the Proposition. The maximum value is:

$$\tilde{U}^{\max}_{\Theta=0} \;=\; \frac{\big(1 - Ra^-\big)^2}{2(2\lambda + C_I)}. \tag{J24}$$

When $\Theta > 0$, the trading strategy of the IFT is $\mathrm{d}x_t = -\Theta\, x_{t-1} + G\,\mathrm{d}w_t$. As in the proof of Theorem 2, the IFT's inventory costs are zero, hence the IFT's expected utility is the same as his expected profit. Then, equation (A31) shows that the IFT's normalized expected utility/profit is:

$$\tilde{U}_{\Theta>0} \;=\; G\,\frac{Rb(1-a^-)}{1+\phi b} - G^2\,\frac{\lambda\big(b + \frac{1}{1+\phi}\big)}{1+\phi b}. \tag{J25}$$

Fix ϕ. Then, the first order condition with respect to G implies that the optimum G satisfies

$$G \;=\; \frac{Rb(1-a^-)}{2\lambda\big(b+\frac{1}{1+\phi}\big)} \;=\; \frac{b(1-a^-)}{2\rho\big(b+\frac{1}{1+\phi}\big)}, \tag{J26}$$

as stated in the Proposition. For this G, the normalized expected utility (profit) of the IFT is:

$$\tilde{U}_{\Theta>0} \;=\; \frac{\big(Rb(1-a^-)\big)^2}{4\lambda(1+\phi b)\big(b+\frac{1}{1+\phi}\big)}. \tag{J27}$$

We now analyze the function

$$f(\phi) \;=\; (1+\phi b)\left(b + \frac{1}{1+\phi}\right) \quad\Longrightarrow\quad f'(\phi) \;=\; \frac{b^2(1+\phi)^2 + b - 1}{(1+\phi)^2}. \tag{J28}$$

The polynomial in the numerator has two roots

$$\phi_1 \;=\; -1 + \frac{\sqrt{1-b}}{b} \qquad \phi_2 \;=\; -1 - \frac{\sqrt{1-b}}{b}. \tag{J29}$$

Note that $b = \rho\bar{\mu} > 0$, and since we have assumed that $b < 1$, the two roots are real and distinct.

Clearly, $\phi_2 < -1$ and $\phi_1 > -1$. Since the numerator of $f'(\phi)$ is a quadratic function of ϕ, it follows that $f'(\phi) < 0$ for $\phi \in (\phi_2, \phi_1)$, and positive everywhere else. We are only interested in $\phi \in (-1, 1]$, hence we have the following two cases:

(i) If $\phi_1 \geq 1$, f is strictly decreasing on $(-1, 1]$, hence it attains its minimum at $\phi = 1$. Thus, the maximum normalized expected utility $\tilde{U}_{\Theta>0}$ from (J27) attains its maximum at $\phi = 1$, or equivalently at $\Theta = 0_+$ (recall that there is a discontinuity at $\Theta = 0$). This maximum value is:

$$\tilde{U}^{\max}_{\Theta=0_+} = \frac{\big(Rb(1-a^-)\big)^2}{4\lambda(1+b)\big(b+\frac{1}{2}\big)}. \tag{J30}$$

To determine the cutoff value for C_I, we set $\tilde{U}^{\max}_{\Theta=0_+} = \tilde{U}^{\max}_{\Theta=0}$. We get:

$$\bar{C}^0_I = 2\lambda\left(\frac{(1-Ra^-)^2(1+b)(b+\frac{1}{2})}{R^2b^2(1-a^-)^2} - 1\right). \tag{J31}$$

(ii) If $\phi_1 \in (-1, 1)$, f is strictly decreasing on $(-1, \phi_1)$ and strictly increasing on $(\phi_1, 1)$, hence it attains its minimum at $\phi = \phi_1$. Thus, the maximum normalized expected utility $\tilde{U}_{\Theta>0}$ from (J27) attains its maximum at $\phi = \phi_1$, or equivalently at $\Theta = 2 - \frac{\sqrt{1-b}}{b} \in (0, 2)$. This maximum value is:

$$\tilde{U}^{\max}_{\Theta>0} = \frac{\big(Rb(1-a^-)\big)^2}{4\lambda b(1+\sqrt{1-b})^2}. \tag{J32}$$

To determine the cutoff value for C_I, we set $\tilde{U}^{\max}_{\Theta>0} = \tilde{U}^{\max}_{\Theta=0}$. We get:

$$\bar{C}_I = 2\lambda\left(\frac{(1-Ra^-)^2(1+\sqrt{1-b})^2}{R^2b(1-a^-)^2} - 1\right). \tag{J33}$$

But $\phi_1 \geq 1$ is equivalent to $b \leq \frac{\sqrt{17}-1}{8}$, hence the two cases (i) and (ii) described here are the same as the cases described in the Theorem.

Finally, we collect the equations for $\bar{C}_I$ and $\bar{C}_I^0$:

$$\begin{aligned} \bar{C}_I &= 2\lambda\left(\frac{(1-Ra^-)^2(1+\sqrt{1-b})^2}{R^2b(1-a^-)^2}-1\right), \\ \bar{C}_I^0 &= 2\lambda\left(\frac{(1-Ra^-)^2(1+b)(b+\frac{1}{2})}{R^2b^2(1-a^-)^2}-1\right), \end{aligned} \tag{J34}$$

For future reference, we also compute the normalized expected utility at $\Theta = 0_+$. By using $\bar{\gamma} = \gamma^- + G$, some algebraic manipulation of (J25) shows that

$$\tilde{U}_{\Theta>0} = \frac{\lambda\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}} - \frac{\lambda G^2}{(1+\phi)(1+\phi\rho\bar{\mu})}. \tag{J35}$$

Taking the limit when $\Theta \to 0$ (or equivalently when $\phi \to 1$), we obtain:

$$\tilde{U}_{\Theta=0_+} = \frac{\lambda\bar{\mu}G(1-\rho\bar{\gamma})}{1+\rho\bar{\mu}} - \frac{\lambda G^2}{2(1+\rho\bar{\mu})}. \tag{J36}$$

□

Proof of Proposition J.2. As in Theorem 2, we define some normalized covariances that are used throughout this proof. If x_t is the IFT's inventory in the risky asset, denote by

$$\begin{aligned} \Omega_t^{xx} &= \frac{\mathsf{E}\big(x_t^2\big)}{\sigma_w^2\mathrm{d}t}, \quad \Omega_t^{xw} = \frac{\mathsf{E}\big(x_t w_t\big)}{\sigma_w^2\mathrm{d}t}, \quad \Omega_t^{xp} = \frac{\mathsf{E}\big(x_t p_t\big)}{\sigma_w^2\mathrm{d}t}, \quad \Omega_t^{xe} = \frac{\mathsf{E}\big(x_t(w_t-p_t)\big)}{\sigma_w^2\mathrm{d}t}, \\ E_t &= \frac{\mathsf{E}\big((w_t-p_t)\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\mathrm{d}t}, \quad X_t = \frac{\mathsf{E}\big(x_t\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\mathrm{d}t}, \quad Y_t = \frac{\mathsf{E}\big((\mathrm{d}y_t)^2\big)}{\sigma_w^2\mathrm{d}t}, \quad Z_t = \frac{\mathsf{E}\big(x_{t-1}\mathrm{d}y_t\big)}{\sigma_w^2\mathrm{d}t}, \\ W_t &= \frac{\mathsf{E}\big((\widetilde{\mathrm{d}w_t})^2\big)}{\sigma_w^2\mathrm{d}t}, \quad H_t = \frac{\mathsf{E}\big((w_t-p_t)\mathrm{d}y_t\big)}{\sigma_w^2\mathrm{d}t}, \quad H_t^w = \frac{\mathsf{E}\big(w_t\mathrm{d}y_t\big)}{\sigma_w^2\mathrm{d}t}, \quad H_t^p = \frac{\mathsf{E}\big(p_t\mathrm{d}y_t\big)}{\sigma_w^2\mathrm{d}t}. \end{aligned} \tag{J37}$$

Recall that in Theorem 2 we have proved the following formulas

$$\begin{aligned} \Theta\Omega^{xx} &= \frac{G^2}{1+\phi}, \qquad \Theta\Omega^{xw} = G, \qquad \Theta\Omega^{xp} = \lambda G\bar{\gamma} + \lambda\phi Z, \\ \Theta\Omega^{xe} &= G(1-\lambda\bar{\gamma}) - \lambda\phi Z, \qquad X = \frac{\frac{\rho\phi G^2}{1+\phi} + G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}}, \\ Z &= -\Theta\Omega^{xx} + \bar{\mu}X = \frac{\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}} - \frac{G^2}{(1+\phi)(1+\phi\rho\bar{\mu})}. \end{aligned} \tag{J38}$$

The formula for W_t is (recall that $\widetilde{\mathrm{d}w_t} = \mathrm{d}w_t - \rho \mathrm{d}y_t$):

$$W_t \;=\; \frac{\mathsf{E}\big((\widetilde{\mathrm{d}w_t})^2\big)}{\sigma_w^2 \mathrm{d}t} \;=\; 1 - 2\rho\bar{\gamma} + \rho^2 Y_t. \tag{J39}$$

The formula for Y_t is (recall that $\tilde{\sigma}_u = \frac{\sigma_u}{\sigma_w}$):

$$\begin{aligned} Y_t \;&=\; \frac{\mathsf{E}\big((\mathrm{d}y_t)^2\big)}{\sigma_w^2 \mathrm{d}t} \;=\; \Theta^2 \Omega^{xx}_{t-1} + \bar{\gamma}^2 + \bar{\mu}^2 W_{t-1} - 2\Theta\bar{\mu} X_{t-1} + \tilde{\sigma}_u^2 \\ &=\; \Theta^2 \Omega^{xx}_{t-1} + \bar{\gamma}^2 + \bar{\mu}^2 \big(1 - 2\rho\bar{\gamma} + \rho^2 Y_{t-1}\big) - 2\Theta\bar{\mu} X_{t-1} + \tilde{\sigma}_u^2. \end{aligned} \tag{J40}$$

Because $\rho\bar{\mu} \in [0,1)$, we apply Lemma A.1 to deduce that Y_t is constant, and equal to:

$$Y \;=\; \frac{\Theta^2 \Omega^{xx} + \bar{\gamma}^2 + \bar{\mu}^2 (1 - 2\rho\bar{\gamma}) - 2\Theta\bar{\mu} X + \tilde{\sigma}_u^2}{1 - \rho^2 \bar{\mu}^2}. \tag{J41}$$

From (J38), we compute $\Theta^2\Omega^{xx} - 2\Theta\bar{\mu}X = -\Theta^2\Omega^{xx} - 2\Theta(-\Theta\Omega^{xx} + \bar{\mu}X) = -\frac{\Theta G^2}{1+\phi} - 2\Theta Z$, hence

$$Y \;=\; \frac{-\frac{\Theta G^2}{1+\phi} - 2\Theta Z + \bar{\gamma}^2 + \bar{\mu}^2 (1 - 2\rho\bar{\gamma}) + \tilde{\sigma}_u^2}{1 - \rho^2 \bar{\mu}^2}, \tag{J42}$$

as desired. Therefore, W_t is constant and equal to:

$$W \;=\; 1 - 2\rho\bar{\gamma} + \rho^2 Y \;=\; \frac{-\frac{\Theta \rho^2 G^2}{1+\phi} - 2\Theta\rho^2 Z + (1 - \rho\bar{\gamma})^2 + \rho^2 \tilde{\sigma}_u^2}{1 - \rho^2 \bar{\mu}^2}, \tag{J43}$$

as desired. Since $H^w_t = \frac{\mathsf{E}(w_t \mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t}$ and $\Theta\Omega^{xw} = G$, we compute

$$\begin{aligned} H^w_t \;&=\; \frac{\mathsf{E}\big((w_{t-1} + \mathrm{d}w_t)(-\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1})\big)}{\sigma_w^2 \mathrm{d}t} \\ &=\; \bar{\gamma} + \frac{\mathsf{E}\big(w_{t-1}(-\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1})\big)}{\sigma_w^2 \mathrm{d}t} \\ &=\; \bar{\gamma} - \Theta\Omega^{xw} + \bar{\mu} E^w_{t-1} \\ &=\; \bar{\gamma} - G + \bar{\mu} E^w_{t-1}. \end{aligned} \tag{J44}$$

We get the following recursive equation for E^w:

$$\begin{aligned} E^w_t \;&=\; \frac{\mathsf{E}\big(w_t \widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2 \mathrm{d}t} \;=\; \frac{\mathsf{E}\big(w_t(\mathrm{d}w_t - \rho\mathrm{d}y_t)\big)}{\sigma_w^2 \mathrm{d}t} \;=\; 1 - \rho H^w_t \\ &=\; 1 - \rho(\bar{\gamma} - G) - \rho\bar{\mu} E^w_{t-1}. \end{aligned} \tag{J45}$$

As long as $b = \rho\bar{\mu} < 1$ (and $b \geq 0$), we apply Lemma A.1 to deduce that E_t^w is constant, and equal to:

$$E^w \;=\; \frac{1-\rho(\bar{\gamma}-G)}{1+\rho\bar{\mu}}. \tag{J46}$$

Equation (J45) implies $H^w = \frac{1-E^w}{\rho}$, from which we compute:

$$H^w \;=\; \frac{\bar{\mu}+(\bar{\gamma}-G)}{1+\rho\bar{\mu}}. \tag{J47}$$

Since $\Theta\Omega^{xp} = \lambda G\bar{\gamma} + \lambda\phi Z$ and $H_t^p = \frac{\mathsf{E}(p_t \mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}((p_{t-1}+\lambda \mathrm{d}y_t)\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} = \frac{\mathsf{E}(p_{t-1}\mathrm{d}y_t)}{\sigma_w^2 \mathrm{d}t} + \lambda Y$, we compute

$$\begin{aligned} H_t^p \;&=\; \lambda Y + \frac{\mathsf{E}\big(p_{t-1}(-\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1})\big)}{\sigma_w^2 \mathrm{d}t} \\ &=\; \lambda Y - \Theta\Omega^{xp} + \bar{\mu}E_{t-1}^p \\ &=\; \lambda\big(Y - G\bar{\gamma} - \phi Z\big) + \bar{\mu}E_{t-1}^p. \end{aligned} \tag{J48}$$

We get the following recursive equation for E^p:

$$\begin{aligned} E_t^p \;&=\; \frac{\mathsf{E}\big(p_t\widetilde{\mathrm{d}w}_t\big)}{\sigma_w^2 \mathrm{d}t} \;=\; \frac{\mathsf{E}\big(p_t(\mathrm{d}w_t - \rho\mathrm{d}y_t)\big)}{\sigma_w^2 \mathrm{d}t} \;=\; \lambda\bar{\gamma} - \rho H_t^p \\ &=\; \lambda\big(\bar{\gamma} - \rho Y + \rho G\bar{\gamma} + \rho\phi Z\big) - \rho\bar{\mu}E_{t-1}^p. \end{aligned} \tag{J49}$$

As long as $b = \rho\bar{\mu} < 1$ (and $b \geq 0$), we apply Lemma A.1 to deduce that E_t^w is constant, and equal to:

$$E^p \;=\; \lambda\,\frac{\bar{\gamma} - \rho Y + \rho G\bar{\gamma} + \rho\phi Z}{1+\rho\bar{\mu}}. \tag{J50}$$

Equation (J49) implies $H^p = \frac{\lambda\bar{\gamma}-E^p}{\rho}$, from which we compute:

$$H^p \;=\; \lambda\,\frac{\bar{\gamma}\bar{\mu} + Y - \bar{\gamma}G - \phi Z}{1+\rho\bar{\mu}}. \tag{J51}$$

Putting together (J46) and (J50), we obtain:

$$E \;=\; E^w - E^p \;=\; \frac{1-\rho(\bar{\gamma}-G) - \lambda(\bar{\gamma} - \rho Y + \rho G\bar{\gamma} + \rho\phi Z)}{1+\rho\bar{\mu}}, \tag{J52}$$

which is equivalent to the desired equation for E.

To simplify presentation, we combine the fast and the slow traders by considering a

speculator with trading strategy of the form $\mathrm{d}x_{i,t} = \gamma_i \mathrm{d}w_t + \mu_i \widetilde{\mathrm{d}w}_{t-1}$. If the speculator is a FT, we set $\mu_i = 0$; and if the speculator is a ST, we set$\gamma_i = 0$. The normalized expected profit of this speculator is:

$$\begin{aligned}
\tilde{\pi}_i &= \frac{1}{\sigma_w^2} \mathsf{E} \int_0^T \big(w_t - p_t\big)\big(\gamma_i \mathrm{d}w_t + \mu_i \widetilde{\mathrm{d}w}_{t-1}\big) \\
&= \frac{1}{\sigma_w^2} \mathsf{E} \int_0^T \Big(w_{t-1} - p_{t-1} + \mathrm{d}w_t - \lambda\big(-\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1}\big)\Big)\big(\gamma_i \mathrm{d}w_t + \mu_i \widetilde{\mathrm{d}w}_{t-1}\big) \\
&= \frac{1}{\sigma_w^2} \mathsf{E} \int_0^T \Big(w_{t-1} - p_{t-1} + \mathrm{d}w_t(1 - \lambda\bar{\gamma}) + \lambda\Theta x_{t-1} - \lambda\bar{\mu}\widetilde{\mathrm{d}w}_{t-1}\big)\Big)\big(\gamma_i \mathrm{d}w_t + \mu_i \widetilde{\mathrm{d}w}_{t-1}\big) \\
&= \gamma_i(1 - \lambda\bar{\gamma}) + \mu_i\big(E + \lambda\Theta X - \lambda\bar{\mu}W\big).
\end{aligned} \tag{J53}$$

Recall that by assumption (see equation (13)) the covariances E, X, W do not depend on speculators' strategies. That is, the speculator regards them as constant and not as functions of γ_i, μ_i.

We compute the optimal weight of a FT indexed by $i = 1, \ldots, N_F$. From (J53) with $\mu_i = 0$, his normalized expected profit is

$$\tilde{\pi}_i^F = \gamma_i(1 - \lambda G) - \gamma_i(\gamma_i + \gamma_{-i})\lambda, \tag{J54}$$

where γ_{-i} is the aggregate weight on $\mathrm{d}w_t$ of the other FTs. The first order condition with respect to γ_i implies

$$1 - \lambda G = \lambda(2\gamma_i + \gamma_{-i}), \tag{J55}$$

and the second order condition for a maximum is

$$\lambda > 0. \tag{J56}$$

Note that this second order condition is satisfied by assumption. The first order condition is true for all FTs, hence all γ_i are equal to γ given by

$$\gamma = \frac{1 - \lambda G}{\lambda(N_F + 1)}. \tag{J57}$$

From this, we compute $\gamma^- = N_F\gamma = \left(\frac{1}{\lambda} - G\right)\frac{N_F}{N_F+1}$, which implies $\bar{\gamma} = \gamma^- + G = \frac{N_F}{\lambda(N_F+1)} + \frac{G}{N_F+1}$. This proves the desired formula for $\bar{\gamma}$.

We compute the optimal weight of a ST indexed by $i = 1, \ldots, N_L$. From (J53) with

$\gamma_i = 0$, his normalized expected profit is

$$\tilde{\pi}_i^S \;=\; \mu_i(E + \lambda\Theta X) - \mu_i(\mu_i + \mu_{-i})\lambda W, \tag{J58}$$

where μ_{-i} is the aggregate weight on $\widetilde{\mathrm{d}w}_{t-1}$ of the other STs. The first order condition with respect to μ_i implies

$$E + \lambda\Theta X - (2\mu_i + \mu_{-i})\lambda W \;=\; 0, \tag{J59}$$

and the second order condition for a maximum is

$$\lambda W \;>\; 0. \tag{J60}$$

From (J56), this condition is equivalent to $W > 0$, which is assumed true. The first order condition is true for all STs, hence all μ_i are equal to μ given by

$$\mu \;=\; \frac{E + \lambda\Theta X}{\lambda W (N_L + 1)}. \tag{J61}$$

From this, $\bar{\mu} = N_L \mu = \frac{E+\lambda\Theta X}{\lambda W}\,\frac{N_L}{N_L+1}$. This proves the desired formula for $\bar{\mu}$. □

Proof of Proposition J.3. We compute the pricing functions set by the dealer. As in the proof of Theorem 1, the definition $\widetilde{\mathrm{d}w}_t = \mathrm{d}w_t - \mathsf{E}_{t+1}(\mathrm{d}w_t)$ implies $\rho = \frac{\widetilde{\mathsf{Cov}}(\mathrm{d}w_t, \mathrm{d}y_t)}{\widetilde{\mathsf{Var}}(\mathrm{d}y_t)} = \frac{\bar{\gamma}}{Y}$. Hence, we obtain

$$\rho \;=\; \frac{\bar{\gamma}}{Y} \quad\Longrightarrow\quad \rho Y \;=\; \bar{\gamma},. \tag{J62}$$

Note that this is equivalent to the second equation in (J16).

To compute λ, we impose the zero expected profit condition for the dealer. Recall the notations from equation (J37):

$$\begin{aligned} \Omega_t^{xe} \;&=\; \frac{\mathsf{E}\big(x_t(w_t - p_t)\big)}{\sigma_w^2 \mathrm{d}t}, \qquad Y_t \;=\; \frac{\mathsf{E}\big((\mathrm{d}y_t)^2\big)}{\sigma_w^2 \mathrm{d}t}, \qquad Z_t \;=\; \frac{\mathsf{E}\big(x_{t-1}\mathrm{d}y_t\big)}{\sigma_w^2 \mathrm{d}t} \\ H_t^w \;&=\; \frac{\mathsf{E}\big(w_t \mathrm{d}y_t\big)}{\sigma_w^2 \mathrm{d}t}, \qquad H_t^p \;=\; \frac{\mathsf{E}\big(p_t \mathrm{d}y_t\big)}{\sigma_w^2 \mathrm{d}t}. \end{aligned} \tag{J63}$$

In the proof of Proposition J.2, we see that all these numbers are constant. Equations (J47) and (J51) imply that

$$H^w \;=\; \frac{\bar{\mu} + (\bar{\gamma} - G)}{1 + \rho\bar{\mu}}, \qquad H^p \;=\; \lambda\,\frac{Y + \bar{\gamma}\bar{\mu} - \bar{\gamma}G - \phi Z}{1 + \rho\bar{\mu}}. \tag{J64}$$

The dealer's normalized expected profit at $t=0$ is given by:

$$\begin{aligned}\tilde{\pi}_d &= \frac{1}{\sigma_w^2}\,\mathsf{E}\int_0^T (p_t - w_t)\mathrm{d}y_t \;=\; H^p - H^w \\ &= \lambda\,\frac{Y+\bar{\gamma}\bar{\mu}-\bar{\gamma}G-\phi Z}{1+\rho\bar{\mu}} \;-\; \frac{\bar{\mu}+(\bar{\gamma}-G)}{1+\rho\bar{\mu}}.\end{aligned} \tag{J65}$$

Setting the dealer's expected profit to zero is then equivalent to

$$\lambda \;=\; \frac{\bar{\mu}+(\bar{\gamma}-G)}{Y+\bar{\gamma}\bar{\mu}-\bar{\gamma}G-\phi Z}, \tag{J66}$$

which proves equation (J17).

We now compute Y. From Proposition J.2, we have:

$$Z \;=\; \frac{\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}} - \frac{G^2}{(1+\phi)(1+\phi\rho\bar{\mu})}, \qquad Y \;=\; \frac{-\frac{\Theta G^2}{1+\phi} - 2\Theta Z + \bar{\gamma}^2 + \bar{\mu}^2(1-2\rho\bar{\gamma}) + \tilde{\sigma}_u^2}{1-\rho^2\bar{\mu}^2}. \tag{J67}$$

Note that the equation for Z is identical to the first equation in (J16). By substituting Z in the equation for Y, we get

$$Y \;=\; \frac{\Theta\,\frac{G^2(1-\phi\rho\bar{\mu})-2\bar{\mu}G(1+\phi)(1-\rho\bar{\gamma})}{(1+\phi)(1+\phi\rho\bar{\mu})} + \bar{\gamma}^2 + \bar{\mu}^2(1-2\rho\bar{\gamma}) + \tilde{\sigma}_u^2}{1-\rho^2\bar{\mu}^2}. \tag{J68}$$

We multiply this equation by $\rho(1-\rho^2\bar{\mu}^2)(1+\phi\rho\bar{\mu})$. Because (J62) implies $\rho Y = \bar{\gamma}$, we obtain:

$$\begin{aligned}\bar{\gamma}(1-\rho^2\bar{\mu}^2)(1+\phi\rho\bar{\mu}) \;=\; &\Theta\,\frac{G^2(1-\phi\rho\bar{\mu})-2\bar{\mu}G(1+\phi)(1-\rho\bar{\gamma})}{(1+\phi)} \\ &+ \big(\bar{\gamma}^2+\bar{\mu}^2(1-2\rho\bar{\gamma})+\tilde{\sigma}_u^2\big)(1+\phi\rho\bar{\mu}).\end{aligned} \tag{J69}$$

This is the third degree equation in ρ stated in (J15). □

Proof of Theorem J.1 (Theorem 3). To find necessary conditions, suppose we are in equilibrium. Since the IFT is sufficiently averse and $b > \frac{\sqrt{17}-1}{8}$, according to Proposition J.1, he chooses optimally $\Theta > 0$. Then, we can put together all the equations from Propositions J.1–J.3. To simplify the equations, note that (see equation (J62)):

$$\rho Y \;=\; \bar{\gamma}. \tag{J70}$$

Then, equation (J39) becomes:

$$W \;=\; 1 - \rho\bar{\gamma} \;=\; 1 - a. \tag{J71}$$

We now assume that the IFT is sufficiently inventory averse, so that his inventory mean reversion is strictly positive ($\Theta > 0$). The relevant equations from Propositions J.1–J.3 are:

$$\begin{aligned}
&\gamma^{-} \;=\; N_F\gamma \;=\; \frac{1-\lambda G}{\lambda}\,\frac{N_F}{N_F+1}, \qquad \bar{\mu} \;=\; N_L\mu \;=\; \frac{E+\lambda\Theta X}{\lambda W}\,\frac{N_L}{N_L+1},\\
&G \;=\; \frac{1-a^{-}}{2\rho\left(1+\frac{1}{\sqrt{1-b}}\right)}, \qquad \Theta \;=\; 2-\frac{\sqrt{1-b}}{b} \;\in\; (0,1), \qquad \phi \;=\; 1-\Theta,\\
&W \;=\; 1-a, \qquad \rho Y \;=\; \bar{\gamma}, \qquad \lambda \;=\; \frac{\bar{\mu}+\bar{\gamma}-G}{Y+\bar{\gamma}\bar{\mu}-\bar{\gamma}G-\phi Z},\\
&X \;=\; \frac{\frac{\rho\phi G^2}{1+\phi}+G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}}, \qquad Y \;=\; \frac{-\frac{\Theta G^2}{1+\phi}-2\Theta Z+\bar{\gamma}^2+\bar{\mu}^2(1-2\rho\bar{\gamma})+\tilde{\sigma}_u^2}{1-\rho^2\bar{\mu}^2},\\
&Z \;=\; \frac{\bar{\mu}G(1-\rho\bar{\gamma})}{1+\phi\rho\bar{\mu}}-\frac{G^2}{(1+\phi)(1+\phi\rho\bar{\mu})},\\
&E \;=\; \frac{1-\rho\bar{\gamma}+\rho G}{1+\rho\bar{\mu}}-\lambda\,\frac{\bar{\gamma}-\rho Y+\rho G\bar{\gamma}+\rho\phi Z}{1+\rho\bar{\mu}} \;=\; \frac{1-\rho\gamma^{-}}{1+\rho\bar{\mu}}-\lambda\,\frac{\rho G\bar{\gamma}+\rho\phi Z}{1+\rho\bar{\mu}}.
\end{aligned} \tag{J72}$$

The goal is to express all the equations in (J72) as functions of a^{-} and b, and then use the first two equations to solve numerically for a^{-} and b. Denote by

$$R \;=\; \frac{\lambda}{\rho}, \qquad B \;=\; \frac{1}{\sqrt{1-b}}, \qquad a^{-} \;=\; \rho\gamma^{-}, \qquad a \;=\; \rho\bar{\gamma}, \qquad b \;=\; \rho\bar{\mu}. \tag{J73}$$

Rewrite the equations in (J72), except for the first two, as functions of a^{-}, b (or B),

and R:

$$\begin{aligned}
\rho G \;&=\; \frac{1-a^-}{2(B+1)}, \qquad \Theta \;=\; 2-\frac{1}{bB}, \qquad \phi \;=\; \frac{1}{bB}-1,\\
W \;&=\; 1-a \;=\; (1-a^-)\frac{2B+1}{2(B+1)},\\
\rho^2 Y \;&=\; a \;=\; a^- + \rho G \;=\; \frac{1+a^-(2B+1)}{2(B+1)},\\
R \;&=\; \frac{a^-+b}{\rho^2 Y + ab - a\rho G - \phi\rho^2 Z} \;=\; \frac{a^-+b}{a+ab-a\rho G-\phi\rho^2 Z},\\
\rho X \;&=\; \frac{(1-a^-)^2 B}{4(B+1)}, \qquad \rho^2 Z \;=\; \frac{(1-a^-)^2}{4(B+1)^2}\, bB^2,\\
(1-b^2)\rho^2 Y \;&=\; (1-b^2)a \;=\; -\frac{\Theta\rho^2 G^2}{1+\phi} - 2\Theta\rho^2 Z + a^2 + b^2(1-2a) + \rho^2\tilde{\sigma}_u^2,\\
E \;&=\; \frac{1-a^-}{1+b} - R\,\frac{2(1-a^-)-(1-a^-)^2 B}{4(B+1)(1+b)},
\end{aligned} \tag{J74}$$

From the corresponding equation for R in (J74), we compute (after some algebraic manipulation):

$$R \;=\; \frac{4(B+1)B^2\big(a^-+b\big)}{B^3(a^-)^2 + 2(3B^3+3B^2-2B-1)a^- + (B^3+2B^2-2)}. \tag{J75}$$

By setting

$$Q \;=\; B^3(a^-)^2 + 2(3B^3+3B^2-2B-1)a^- + (B^3+2B^2-2), \tag{J76}$$

we have just proved the equation for R in (J20).

Now (J72) implies $a = \frac{1+a^-(2B+1)}{2(B+1)}$, which proves the corresponding equation in (J20). Recall that

$$n_F \;=\; \frac{N_F}{N_F+1}, \qquad n_L \;=\; \frac{N_L}{N_L+1}. \tag{J77}$$

Equation $\gamma^- = \frac{1-\lambda G}{\lambda}\,\frac{N_F}{N_F+1}$ from (J72) can be written as $Ra^- = (1-R\rho G)n_F$, or equivalently $R\big(a^- + \rho G n_F\big) = n_F$. Using the formula for R from (J75), (after some algebraic manipulation) we obtain the following second degree equation in a^-:

$$B^2\Big(4(B+1)-n_F(B+2)\Big)(a^-)^2 + 2(B+1)\Big(2(B^2-1)-n_F(3B^2-2)\Big)a^- - B^3 n_F \;=\; 0. \tag{J78}$$

This second degree polynomial has two real roots, given by:

$$
\begin{aligned}
a^- &= \frac{-q \pm \sqrt{q^2 + n_F B^5\big((4-n_F)B + 2(2-n_F)\big)}}{B^2\big((4-n_F)B + 2(2-n_F)\big)}, \quad \text{with} \\
q &= (B+1)\Big(2(B^2-1) - n_F(3B^2-2)\Big).
\end{aligned} \tag{J79}
$$

This proves the equation for a^- in (J20).

From (J74), we compute

$$
\frac{E}{1-a^-} = \frac{1}{1+b} - R\,\frac{2-(1-a^-)B}{4(B+1)(1+b)}, \qquad \frac{\rho X}{1-a^-} = \frac{(1-a^-)B}{4(B+1)}, \tag{J80}
$$

which implies

$$
\begin{aligned}
\frac{E + R\Theta\rho X}{R(1-a^-)} &= \frac{1}{R(1+b)} + \frac{\big(\Theta(1+b)+1\big)(1-a^-)B - 2}{4(B+1)(1+b)} \\
&= \frac{Q}{4(B+1)(1+b)B^2(a^-+b)} + \frac{\big(\Theta(1+b)+1\big)(1-a^-)B - 2}{4(B+1)(1+b)},
\end{aligned} \tag{J81}
$$

where the last equation comes from (J75).

Now, multiply equation $\bar{\mu} = \frac{E+\lambda\Theta X}{\lambda W}\,\frac{N_L}{N_L+1}$ from (J72) by ρ to obtain:

$$
b = \frac{E + R\Theta\rho X}{R(1-a^-)}\,\frac{2(B+1)}{2B+1}\,n_L \tag{J82}
$$

Multiplying this equation by $\frac{2(1+b)(2B+1)}{n_L}$ and using (J81), we get

$$
\frac{2b(1+b)(2B+1)}{n_L} = \frac{Q}{B^2(a^-+b)} + \big(\Theta(1+b)+1\big)(1-a^-)B - 2. \tag{J83}
$$

Since $\Theta(1+b)+1 = \frac{3bB+2b^2B-1-b}{b}$, we have just proved the first equation in (J20).

It remains just to prove the equation for ρ. The penultimate equation in (J74) implies $\rho^2\tilde{\sigma}_u^2 = (1-a)(a-b^2) + \frac{\Theta\rho^2 G^2}{1+\phi} + 2\Theta b\rho^2 Z$, hence

$$
\begin{aligned}
\rho^2 &= \frac{(1-a)(a-b^2) + \frac{\Theta\rho^2 G^2}{1+\phi} + 2\Theta\rho^2 Z}{\tilde{\sigma}_u^2} \\
&= \frac{(1-a)(a-b^2) + \frac{\Theta}{1+\phi}\,\frac{(1-a^-)^2}{4(B+1)^2}\,\big(1 + 2bB^2(1+\phi)\big)}{\tilde{\sigma}_u^2}.
\end{aligned} \tag{J84}
$$

Using $1-a=(1-a^-)\frac{2B+1}{2(B+1)}$, we compute (after some algebraic manipulation):

$$\frac{\Theta}{1+\phi}\frac{(1-a^-)^2}{4(B+1)^2}\big(1+2bB^2(1+\phi)\big) \;=\; \frac{2bB-1}{2B+1}(1-a)^2, \tag{J85}$$

which proves the corresponding formula for ρ in (J20).

We have just finished the proof that the equations in (J20) are necessary for the existence of an equilibrium. We now show that they are sufficient if we assume that the solution to (J20) also satisfies $\frac{\sqrt{17}-1}{8}<b<1$, $a<1$, $\lambda>0$. We now follow the proofs of Propositions J.1–J.3 to show that the strategies defined by using these coefficients provide an equilibrium. The condition $b<1$ is used to perform the computations in Proposition J.1. The condition $b>\frac{\sqrt{17}-1}{8}$ is used in showing that the IFT chooses $\Theta>0$. The condition $\lambda>0$ is used as the second order condition for maximization for all three types of speculators (see in particular the second order condition (J56) for the FT). The condition $a<1$ or equivalently $W=1-a>0$ is used as a second order condition for the ST (see equation (J60)).

Finally, we compute the equilibrium expected profits of the IFT, FTs and STs, denoted respectively by π,π^F,π^S. From equation (J32), the normalized IFT's profit is:

$$\tilde{\pi} \;=\; \frac{Rb(1-a^-)^2}{4\rho(1+\sqrt{1-b})^2}, \tag{J86}$$

as stated in the Thorem. From (J14), the normalized expected profits of the FTs and STs are respectively:

$$\tilde{\pi}^F \;=\; \gamma(1-\lambda G-\lambda\gamma^-), \qquad \tilde{\pi}^S \;=\; \mu\big(E+\lambda\Theta X-\lambda W\bar{\mu}\big). \tag{J87}$$

From (J12),

$$\gamma \;=\; \frac{1-\lambda G}{\lambda(N_F+1)}, \qquad \mu \;=\; \frac{E+\lambda\Theta X}{\lambda W(N_L+1)}. \tag{J88}$$

We compute $\lambda\gamma^-=N_F\lambda\gamma=\frac{N_F}{N_F+1}(1-\lambda G)$. Therefore, $\tilde{\pi}^F=\gamma\,\frac{1-\lambda G}{N_F+1}=\lambda\gamma^2$. Similarly, $\tilde{\pi}^S=\lambda W\mu^2$. But in equilibrium $W=1-a$, hence $\tilde{\pi}^S=\lambda\mu^2(1-a)$. We obtain

$$\tilde{\pi}^F \;=\; \lambda\gamma^2, \qquad \tilde{\pi}^S \;=\; \lambda\mu^2(1-a), \tag{J89}$$

as stated in the Theorem. The proof is now complete. □

Proof of Proposition 6. The asymptotic notation in this proof is:

$$X \;\approx\; X_\infty \qquad \Longleftrightarrow \qquad \lim_{N_F, N_L \to \infty} \frac{X}{X_\infty} \;=\; 1. \tag{J90}$$

(Note that $N_L \to \infty$ is also included as part of the definition.) Denote by

$$b_\infty \;=\; \frac{\sqrt{5}-1}{2}, \qquad B_\infty \;=\; 1 + b_\infty \;=\; \frac{\sqrt{5}+1}{2}, \qquad a_\infty \;=\; 1. \tag{J91}$$

We begin by proving that

$$b \;\approx\; b_\infty, \qquad a \;\approx\; a_\infty, \qquad 1 - a \;\approx\; \frac{1+b_\infty}{N_F+1}, \qquad 1 - a^- \;\approx\; \frac{2}{N_F+1}. \tag{J92}$$

Define the function of two variables f by:

$$\begin{aligned} f(B,\varepsilon) \;&=\; \frac{-q + \sqrt{q^2 + (1-\varepsilon)B^5\big((3+\varepsilon)B + 2(1+\varepsilon)\big)}}{B^2\big((3+\varepsilon)B + 2(1+\varepsilon)\big)}, \\ \text{with} \quad q \;&=\; (B+1)\big(-B^2 + \varepsilon(3B^2-2)\big). \end{aligned} \tag{J93}$$

Also, define the function of two variables g by:

$$\begin{aligned} g(B,\varepsilon) \;=\; &-\frac{2b(1+b)(2B+1)}{n_L} + \frac{Q}{B^2(a^- + b)} + \frac{3bB + 2b^2B - 1 - b}{b}(1-a^-) - 2, \\ &\text{with} \quad n_L \;=\; 1, \qquad b \;=\; 1 - \frac{1}{B^2}, \qquad a^- \;=\; f(B,\varepsilon), \\ &\text{and} \quad Q \;=\; B^3(a^-)^2 + 2(3B^3 + 3B^2 - 2B - 1)a^- + (B^3 + 2B^2 - 2). \end{aligned} \tag{J94}$$

We now use the formulas $B_\infty = 1 + b_\infty$ and $b_\infty(1+b_\infty) = 1$ to compute the values of f and g and of their partial derivatives at $B = B_\infty$ and $\varepsilon = 0$. After some algebraic manipulation, we compute:

$$g(B_\infty, 0) \;=\; 0, \quad f(B_\infty, 0) \;=\; 1, \quad \frac{\partial f}{\partial B}(B_\infty, 0) \;=\; 0, \quad \frac{\partial f}{\partial \varepsilon}(B_\infty, 0) \;=\; -2. \tag{J95}$$

Denote by $B(\varepsilon)$ the solution of $g(B,\varepsilon) = 0$:

$$B(\varepsilon) \qquad \Longleftrightarrow \qquad g(B,\varepsilon) \;=\; 0. \tag{J96}$$

From (J95), $g(B_\infty, 0) = 0$, therefore

$$B(0) \;=\; B_\infty. \tag{J97}$$

Denote by $a^-(\varepsilon)$ the function:

$$a^-(\varepsilon) \;=\; f(B(\varepsilon), \varepsilon). \tag{J98}$$

Using (J95), we compute the derivative of a^- at $\varepsilon = 0$:

$$\begin{aligned}\frac{\mathrm{d}a^-}{\mathrm{d}\varepsilon}(0) \;&=\; \frac{\partial f}{\partial B}(B_\infty, 0)B'(\varepsilon) \;+\; \frac{\partial f}{\partial \varepsilon}(B_\infty, 0)\\ &=\; 0 \times B'(\varepsilon) \;+\; (-2)\\ &=\; -2.\end{aligned} \tag{J99}$$

Fix $N_F \geq 0$. Let a^-_* and B_* be, respectively, the equilibrium values of a^- and $B = \frac{1}{\sqrt{1-b}}$ when N_L approaches infinity:

$$a^-_* \;=\; \lim_{N_L \to \infty} a^-, \qquad B_* \;=\; \lim_{N_L \to \infty} B. \tag{J100}$$

Theorem 3 shows that the equations (J20) are necessary conditions for an equilibrium, hence a^- and B satisfy (J20). Taking the limit when $N_L \to \infty$ ($n_L \to 1$), it follows that a^-_* and B_* satisfy equations (J20) with $n_L = 1$. But, by definition, the numbers $a^-(\varepsilon)$ and $B(\varepsilon)$ satisfy the same equations when

$$\varepsilon \;=\; \frac{1}{N_F + 1}. \tag{J101}$$

Therefore, we have

$$a^-_* \;=\; a^-(\varepsilon), \qquad B_* \;=\; B(\varepsilon). \tag{J102}$$

From (J95) and (J99), we have $B(0) = B_\infty$, $a^-(0) = 1$ and $\frac{\mathrm{d}a^-}{\mathrm{d}\varepsilon}(0) = -2$. Therefore, $B(\varepsilon) \approx B_\infty$, $a^-(\varepsilon) \approx 1$, and $1 - a^-(\varepsilon) \approx 2\varepsilon$. From (J102), this translates into $B_* \approx B_\infty$, $a^-_* \approx 1$, and $1 - a^-_* \approx \frac{2}{N_F+1}$. But B_* and a^-_* are limits when $N_L \to \infty$, therefore we get the following asymptotic formulas:

$$B \;\approx\; B_\infty, \qquad a^- \;\approx\; 1, \qquad 1 - a^- \;\approx\; 2\varepsilon. \tag{J103}$$

From (J20), we have $a = a^- + \frac{1-a^-}{2(B+1)}$, which implies $1 - a = (1 - a^-)\frac{2B+1}{2(B+1)}$. But

$\frac{2B_\infty+1}{2(B_\infty+1)} = \frac{1+b_\infty}{2}$. Therefore, we get:

$$a \;\approx\; 1, \qquad 1-a \;\approx\; (1+b_\infty)\varepsilon. \tag{J104}$$

We have $a = a^- + \rho G$, hence $\rho G = a - a^- = (1-a^-) - (1-a) \approx 2\varepsilon - (1+b_\infty)\varepsilon = (1-b_\infty)\varepsilon$. We thus get

$$\rho G \;\approx\; (1-b_\infty)\varepsilon. \tag{J105}$$

We now analyze $R = \frac{\lambda}{\rho}$. From (J72), we have $\gamma^- = \frac{1-\lambda G}{\lambda}\frac{N_F}{N_F+1}$, which multiplying by ρ becomes $a^- = \big(\frac{1}{R} - \rho G\big)(1-\varepsilon)$. From this, $\frac{1}{R} = \frac{a^-}{1-\varepsilon} + \rho G$, which implies $1 - \frac{1}{R} = \frac{1-a^- - \varepsilon}{1-\varepsilon} - \rho G$. Using $1 - a^- \approx 2\varepsilon$ and $\rho G \approx (1-b_\infty)\varepsilon$, we get $1 - \frac{1}{R} \approx b_\infty \varepsilon$. From this, we get $R \approx 1$ and $\frac{R-1}{R} \approx b_\infty \varepsilon$, hence

$$R \;\approx\; 1, \qquad R-1 \;\approx\; b_\infty \varepsilon. \tag{J106}$$

We compute $1 - Ra^- \;\approx\; 1 - (1+b_\infty\varepsilon)(1-2\varepsilon) \;\approx\; (2-b_\infty)\varepsilon$. Similarly, $1 - Ra \approx 1 - (1+b_\infty\varepsilon)\big(1-(1+b_\infty)\varepsilon\big) \approx \varepsilon$. We obtain:

$$1 - Ra^- \;\approx\; (2-b_\infty)\varepsilon, \qquad 1 - Ra \;\approx\; \varepsilon. \tag{J107}$$

Since $b = 1 - \frac{1}{B^2}$, $b_\infty = 1 - \frac{1}{B_\infty^2}$, $\Theta = 2 - \frac{\sqrt{1-b}}{b}$, we obtain

$$b \;\approx\; b_\infty, \qquad \Theta \;\approx\; 2 - \frac{\sqrt{1-b_\infty}}{b_\infty} \;=\; 1, \qquad \phi \;\approx\; 0. \tag{J108}$$

From (J20), we have $\rho^2\tilde{\sigma}_u^2 = (1-a)(a-b^2) + \frac{\Theta}{1+\phi}\frac{(1-a^-)^2}{4(B+1)^2}\big(1+2bB^2(1+\phi)\big)$, which implies $\frac{\rho^2\tilde{\sigma}_u^2}{1-a} \approx a_\infty - b_\infty^2$. Using $1-a \approx \frac{1+b_\infty}{N_F+1}$ we get $\rho^2\tilde{\sigma}_u^2 \approx \frac{(1-b_\infty^2)(1+b_\infty)}{N_F+1}$, hence

$$\rho^2\tilde{\sigma}_u^2 \;\approx\; \frac{1}{N_F+1}, \qquad \text{or} \quad \rho \;\approx\; \frac{\sigma_w}{\sigma_u}\frac{1}{\sqrt{N_F+1}}. \tag{J109}$$

Since $R \approx 1$, we have $\lambda \approx \rho$. Therefore,

$$\lambda \;\approx\; \rho \;\approx\; \frac{\sigma_w}{\sigma_u}\frac{1}{\sqrt{N_F+1}}. \tag{J110}$$

We now compare the asymptotic results with the corresponding results in the benchmark model. Denote by γ_0, μ_0, λ_0, ρ_0, a_0, b_0 the equilibrium coefficients from Theorem 1, and by γ_∞, μ_∞, λ_∞, ρ_∞, a_∞, b_∞, respectively, their asymptotic limits. We have already shown that $\lambda \approx \rho \approx \lambda_\infty = \rho_\infty = \frac{\sigma_w}{\sigma_u}\frac{1}{\sqrt{N_F+1}}$; also $a \approx a_\infty = 1$, and $b \approx b_\infty = \frac{\sqrt{5}-1}{2}$. More-

over, in the inventory management equilibrium we have $\gamma = \frac{\gamma^-}{N_F} = \frac{a^-}{\rho N_F} \approx \frac{1}{\rho_0(N_F+1)} \approx \frac{a_0}{\rho_0(N_F+1)} = \frac{\bar{\gamma}_0}{N_F+1} = \gamma_0$; and $\mu = \frac{\bar{\mu}}{N_L} = \frac{b}{\rho N_L} \approx \frac{b_\infty}{\rho_0 N_L} \approx \frac{b_0}{\rho_0 N_L} = \frac{\bar{\mu}_0}{N_L} = \mu_0$. Thus, $\gamma \approx \gamma_0$ and $\mu \approx \mu_0$. We have just proved that:

$$\gamma \;\approx\; \gamma_0, \qquad \mu \;\approx\; \mu_0, \qquad \lambda \;\approx\; \lambda_0, \qquad \rho \;\approx\; \rho_0, \qquad a \;\approx\; a_0, \qquad b \;\approx\; b_0. \tag{J111}$$

We also report the asymptotic results for $1 - a_0$, $1 - a_0^-$, $1 - R_0 a_0$, $1 - R_0 a_0^-$. From Theorem 1 (with $N_F + 1$ fast traders), $a_0 = \frac{(N_F+1)-b_0}{(N_F+1)+1}$, hence $1 - a_0 = \frac{1+b_0}{(N_F+1)+1} \approx (1 + b_\infty)\varepsilon$. Also, $a_0^- = \frac{N_F}{N_F+1} a_0 = (1-\varepsilon)a_0$, hence $1 - a_0^- \approx 1 - (1-\varepsilon)\big(1 - (1 + b_\infty)\varepsilon\big) \approx (2 + b_\infty)\varepsilon$. From Corollary 1, $\lambda_0\bar{\gamma}_0 = \frac{N_F+1}{N_F+2}$, hence $1 - R_0 a_0 = 1 - \lambda_0\bar{\gamma}_0 = \frac{1}{N_F+2} \approx \varepsilon$. Also, $R_0 a_0^- = R_0 \frac{N_F}{N_F+1} a_0 = \frac{N_F}{N_F+1}\lambda_0\bar{\gamma}_0 = \frac{N_F}{N_F+2}$, hence $1 - R_0 a_0^- = \frac{2}{N_F+2} \approx 2\varepsilon$. Putting together these formulas, it follows that in the benchmark model:

$$1 - a_0 \;\approx\; (1 + b_\infty)\varepsilon, \quad 1 - a_0^- \;\approx\; (2 + b_\infty)\varepsilon, \quad 1 - R_0 a_0 \;\approx\; \varepsilon, \quad 1 - R_0 a_0^- \;\approx\; 2\varepsilon. \tag{J112}$$

By contrast, in the inventory management model:

$$1 - a \;\approx\; (1 + b_\infty)\varepsilon, \quad 1 - a^- \;\approx\; 2\varepsilon, \quad 1 - Ra \;\approx\; \varepsilon, \quad 1 - Ra^- \;\approx\; (2 - b_\infty)\varepsilon. \tag{J113}$$

The difference comes from the fact that the IFT's equilibrium weight G is not equal asymptotically to the FT's weight $\gamma \approx \gamma_0$ in either the inventory management model or the benchmark model. To see this, note that in the inventory management model we have $1 - a = 1 - a^- - \rho G$, while in the benchmark model $1 - a_0 = 1 - a_0^- - \rho_0\gamma_0$. But $\rho G \approx (1 - b_\infty)\varepsilon$ (from equation (J105)), while $\rho_0\gamma_0 \approx \lambda_0\gamma_0 \approx \varepsilon$, where the last approximation follows from $\lambda_0\bar{\gamma}_0 = \frac{N_F+1}{N_F+2}$, which implies $\lambda_0\gamma_0 = \frac{1}{N_F+2} \approx \varepsilon$. We record this result for future reference:

$$\lambda_0\gamma_0 \;\approx\; \varepsilon. \tag{J114}$$

If we now use $\rho G \approx (1 - b_\infty)\varepsilon$ and $\rho_0\gamma_0 \approx \varepsilon$, by taking their ratio we obtain:

$$\frac{G}{\gamma} \;\approx\; \frac{G}{\gamma_0} \;\approx\; 1 - b_\infty \;=\; 0.3820. \tag{J115}$$

From Corollary J.1, the normalized expected profit of a FT in the inventory management model is $\tilde{\pi}^F = \gamma(1 - \lambda\bar{\gamma}) = \gamma(1 - Ra)$. From (J113), $1 - Ra \approx \varepsilon$. Since $\gamma \approx \gamma_0$, we get

$$\tilde{\pi}^F \;=\; \gamma_0\,\varepsilon. \tag{J116}$$

From Proposition 1, the normalized expected profit of a FT in the benchmark model is

$\tilde{\pi}_0^F = \frac{\gamma_0}{N_F+2}$, which implies

$$\tilde{\pi}_0^F \;\approx\; \gamma_0\,\varepsilon. \tag{J117}$$

Therefore, the FTs in the inventory management model make asymptotically the same profits as the FTs in the benchmark model:

$$\frac{\tilde{\pi}^F}{\tilde{\pi}_0^F} \;\approx\; 1. \tag{J118}$$

For the IFT, equation (J32) implies that the normalized expected utility (or profit) is $\tilde{\pi} = \frac{b\left(R(1-a^-)\right)^2}{4\lambda(1+\sqrt{1-b})^2}$. Since $R \approx 1$ and $1 - a^- \approx 2\varepsilon$, we have $\tilde{\pi} \approx \frac{b_\infty \varepsilon^2}{\lambda_0(1+b_\infty)^2}$. From (J114), $\frac{\varepsilon}{\lambda_0} = \gamma_0$, which implies $\tilde{\pi} \approx \frac{b_\infty}{(1+b_\infty)^2}\,\gamma_0\,\varepsilon$, or

$$\tilde{\pi} \;\approx\; (2b_\infty - 1)\,\gamma_0\,\varepsilon. \tag{J119}$$

Asymptotically, the ratio of IFT's profit to the FT's profit is given by:

$$\frac{\tilde{\pi}}{\tilde{\pi}^F} \;\approx\; 2b_\infty - 1 \;=\; 0.2361. \tag{J120}$$

Denote by $\tilde{\pi}_{C_I=0}$ the IFT's maximum normalized expected profit when $C_I = 0$. Equation (A39) implies that

$$\tilde{\pi}_{C_I=0} \;=\; \frac{\left(1 - Ra^-\right)^2}{4\lambda}. \tag{J121}$$

Since $1 - Ra^- \approx (2 - b_\infty)\varepsilon$ and $\lambda \approx \lambda_0 \approx \frac{\varepsilon}{\gamma_0}$ (equation (J114)), we get $\tilde{\pi}_{C_I=0} \approx \frac{(2-b_\infty)^2}{4}\gamma_0\varepsilon$, or

$$\tilde{\pi}_{C_I=0} \;\approx\; \frac{5}{4}(1 - b_\infty)\,\gamma_0\,\varepsilon \;=\; 0.4775\,\gamma_0\,\varepsilon. \tag{J122}$$

Asymptotically, the ratio of $\tilde{\pi}$ to $\tilde{\pi}_{C_I=0}$ is $\frac{2b_\infty-1}{\frac{5}{4}(1-b_\infty)} = \frac{4}{5}b_\infty$.

$$\frac{\tilde{\pi}}{\tilde{\pi}_{C_I=0}} \;\approx\; \frac{4}{5}\,b_\infty \;=\; 0.4944. \tag{J123}$$

Thus, inventory management generates a profit loss of about 50% for the IFT.

Equation (J34) implies that the threshold inventory aversion for the IFT is given by $1 + \frac{\bar{C}_I}{2\lambda} = \frac{(1-Ra^-)^2(1+\sqrt{1-b})^2}{R^2 b(1-a^-)^2}$. We have $R \approx 1$, $1 - Ra^- \approx (2 - b_\infty)\varepsilon$, $1 - a^- \approx 2\varepsilon$, hence $1 + \frac{\bar{C}_I}{2\lambda} \approx \frac{(2-b_\infty)^2(1+b_\infty)^2}{4b_\infty} = \frac{5}{4}(1 + b_\infty)$, which implies $\frac{\bar{C}_I}{2\lambda} \approx \frac{1+5b_\infty}{4}$. Since $\lambda \approx \frac{\sigma_w}{\sigma_u}\,\frac{1}{\sqrt{N_F+1}}$,

$$\bar{C}_I \;\approx\; \frac{1 + 5b_\infty}{2}\,\lambda \;\approx\; 2.0451\,\frac{\sigma_w}{\sigma_u}\,\frac{1}{\sqrt{N_F + 1}}. \tag{J124}$$

$\square$

Proof of Proposition 7. As in Section 4, we identify trading volume with (instantaneous) order flow variance. From (A25), $\Theta\Omega^{xx} = \frac{G^2}{1+\phi}$, where $\phi = 1 - \Theta$. Since $\mathrm{d}x_t = -\Theta x_{t-1} + G\mathrm{d}w_t$, the IFT's normalized order flow variance (or normalized trading volume) satisfies:

$$\frac{\mathsf{Var}(\mathrm{d}x_t)}{\sigma_w^2 \mathrm{d}t} = \frac{TV_x}{\sigma_w^2} = \Theta^2\Omega^{xx} + G^2 = \frac{1-\phi}{1+\phi}\,G^2 + G^2 = \frac{2G^2}{1+\phi}. \tag{J125}$$

Since $x_t = \phi x_{t-1} + G\mathrm{d}w_t$, the IFT's order flow autocovariance satisfies $\mathsf{Cov}\big(\mathrm{d}x_t, \mathrm{d}x_{t+1}\big) = \mathsf{Cov}\big(-\Theta x_{t-1} + G\mathrm{d}w_t, -\Theta\phi x_{t-1} - \Theta G\mathrm{d}w_t + G\mathrm{d}w_{t+1}\big)$, hence

$$\frac{\mathsf{Cov}\big(\mathrm{d}x_t, \mathrm{d}x_{t+1}\big)}{\sigma_w^2 \mathrm{d}t} = \Theta^2\phi\Omega^{xx} - \Theta G^2 = -\Theta\left(-\frac{\phi}{1+\phi}\,G^2 + G^2\right) = -\frac{\Theta G^2}{1+\phi}. \tag{J126}$$

Therefore, the IFT's order flow autocorrelation is

$$\rho_x = \mathsf{Corr}\big(\mathrm{d}x_t, \mathrm{d}x_{t+1}\big) = \frac{\mathsf{Cov}\big(\mathrm{d}x_t, \mathrm{d}x_{t+1}\big)}{\mathsf{Var}(\mathrm{d}x_t)} = -\frac{\Theta}{2}, \tag{J127}$$

which proves the corresponding formula in (57). Asymptotically, since $\Theta \approx 1$, $\rho_x \approx -\frac{1}{2}$.

The individual and the aggregate trading volume of FTs satisfies, respectively:

$$\frac{TV_{x^F}}{\sigma_w^2} = \gamma^2, \qquad \frac{TV_{\bar{x}^F}}{\sigma_w^2} = (\gamma^-)^2. \tag{J128}$$

From (J125) and (J128), we get $\frac{TV_x}{TV_{x^F}} = \frac{2G^2}{(1+\phi)\gamma^2}$, which proves the corresponding formula in (57). Asymptotically, Proposition 6 shows that $\frac{G}{\gamma} \approx 1 - b_\infty$ and $\phi \approx 0$, hence $\frac{TV_x}{TV_{x^F}} \approx 2(b_\infty - 1)^2 = 2(2 - 3b_\infty) = 0.2918$. The aggregate trading volume of STs satisfies

$$\frac{TV_{\bar{x}^S}}{\sigma_w^2} = \bar{\mu}^2\,\frac{\mathsf{Var}\big(\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2 \mathrm{d}t} = \bar{\mu}^2\,W = \bar{\mu}^2(1-a), \tag{J129}$$

where we use the equilibrium formula $W = 1 - a$ in (J74). From (J128) and (J129), we compute

$$\frac{TV_{\bar{x}^S}}{TV_{\bar{x}^F}} = \frac{\bar{\mu}^2(1-a)}{(\gamma^-)^2} = \frac{b^2(1-a)}{(a^-)^2}, \tag{J130}$$

as stated in the Proposition. Asymptotically, from (J113) we have $1 - a \approx \frac{1+b_\infty}{N_F+1}$, and we use $b \approx b_\infty$, $a \approx 1$, $b_\infty^2(1+b_\infty) = b_\infty$ to get $\frac{TV_{\bar{x}^S}}{TV_{\bar{x}^F}} \approx \frac{b_\infty}{N_F+1}$, as stated.

Recall that $X = \frac{\mathsf{Cov}(\widetilde{\mathrm{d}w_t} x_t)}{\sigma_w^2 \mathrm{d}t}$ and $W = \frac{\mathsf{Var}(\widetilde{\mathrm{d}w_t})}{\sigma_w^2 \mathrm{d}t}$. From (J74) and $1 - a = (1-a^-)\frac{2B+1}{2(B+1)}$,

we compute ($B = \frac{1}{\sqrt{1-b}}$):

$$\rho X \;=\; (1-a)^2\,\frac{B(B+1)}{(2B+1)^2}. \tag{J131}$$

The regression coefficient of the IFT's strategy ($\mathrm{d}x_t$) on the slow trading component ($\mathrm{d}\bar{x}_t^S$) satisfies

$$\beta_{x,\bar{x}^S} \;=\; \frac{\mathsf{Cov}(\mathrm{d}x_t, \mathrm{d}\bar{x}_t^S)}{\mathsf{Var}(\mathrm{d}\bar{x}_t^S)} \;=\; \frac{-\Theta\bar{\mu}\,\mathsf{Cov}\big(x_{t-1}, \widetilde{\mathrm{d}w}_{t-1}\big)}{\bar{\mu}^2\,\mathsf{Var}\big(\widetilde{\mathrm{d}w}_{t-1}\big)} \;=\; \frac{-\Theta X}{\bar{\mu}W} \;=\; \frac{-\Theta\,(\rho X)}{b(1-a)}. \tag{J132}$$

Using (J131) and $W = 1-a$, we compute:

$$\beta_{x,\bar{x}^S} \;=\; -\frac{\Theta B(B+1)}{b(2B+1)^2}\,(1-a) \;=\; -\frac{\Theta B}{2b(2B+1)}\,(1-a^-). \tag{J133}$$

Asymptotically, $b \approx b_\infty$, $B \approx \frac{1}{b_\infty}$, $\Theta \approx 1$, and from (J113) $1-a^- \approx \frac{2}{N_F+1}$. Hence

$$\beta_{x,\bar{x}^S} \;\approx\; -\frac{1}{b_\infty(1+2b_\infty)}\,\frac{1}{N_F+1} \;=\; -\frac{3+b_\infty}{5(N_F+1)} \;=\; -\frac{0.7236}{N_F+1}, \tag{J134}$$

which proves the stated formula.

Since the trading strategy of a FT is $\gamma \mathrm{d}w_t$, the order flow autocorrelation of the FTs is $\rho_{\bar{x}^F} = 0$. For the STs, since $\widetilde{\mathrm{d}w}_t = \mathrm{d}w_t - \rho\mathrm{d}y_t$ and $\mathrm{d}y_t = -\Theta x_{t-1} + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w}_{t-1} + \mathrm{d}u_t$, we compute the normalized order flow autocoviance:

$$\frac{\mathsf{Cov}\big(\mathrm{d}\bar{x}_{t-1}^S, \mathrm{d}\bar{x}_t^S\big)}{\sigma_w^2\mathrm{d}t} \;=\; -\rho\bar{\mu}^2\,\frac{\mathsf{Cov}\big(\widetilde{\mathrm{d}w}_{t-1}, \mathrm{d}y_t\big)}{\sigma_w^2\mathrm{d}t} \;=\; -\rho\bar{\mu}^2\Big(-\Theta X + \bar{\mu}W\Big). \tag{J135}$$

From (J129) and (J135), we compute the autocorrelation of slow trading $\mathrm{d}\bar{x}_t^S$,

$$\begin{aligned}\rho_{\bar{x}^S} \;&=\; \frac{\mathsf{Cov}\big(\mathrm{d}\bar{x}_{t-1}^S, \mathrm{d}\bar{x}_t^S\big)}{\mathsf{Var}(\mathrm{d}\bar{x}_t^S)} \;=\; \frac{-\rho\bar{\mu}^2\big(-\Theta X + \bar{\mu}(1-a)\big)}{\bar{\mu}^2(1-a)} \;=\; -b + \frac{\Theta\rho X}{(1-a)} \\ &=\; -b - b\beta_{x,\bar{x}^S},\end{aligned} \tag{J136}$$

where for the last equality we use (J133). Asymptotically, the term $b + b\beta_{x,\bar{x}^S} \approx b$, since $\beta_{x,\bar{x}^S}$ is of the order of $\frac{1}{N_F+1}$. Hence, $\rho_{\bar{x}^S} \approx -b_\infty$. □

K Smooth Inventory Management

In this section, we solve for a partial equilibrium of the model with inventory management from Section 5, in which the IFT chooses to trade in the *smooth regime* (see the discussion at the end of Section 5.1). In the smooth regime, the IFT has a trading strategy of the form:

$$\mathrm{d}x_t \;=\; -\theta\, x_t \mathrm{d}t + G\,\mathrm{d}w_t, \quad \text{with} \quad \theta \;\in\; [0, \infty). \tag{K1}$$

Recall that in the usual *fast regime* the IFT has a trading strategy of the form:

$$\mathrm{d}x_t \;=\; -\Theta x_{t-1} + G\,\mathrm{d}w_t, \quad \text{with} \quad \Theta \;\in\; [0, 2). \tag{K2}$$

We call strategies of type (K1) *smooth strategies*, and strategies of type (K2) *fast strategies*. A smooth strategy can be considered a particular case of a fast strategy if the coefficient Θ is infinitesimal: $\Theta = \theta \mathrm{d}t$.[16]

A result that we prove below is that the IFT's expected utility changes smoothly from the smooth regime to the fast regime. The connection is made by the right limit of $\theta \in [0, \infty)$, which coincides with the left limit of $\Theta \in (0, 2)$, which we write as $\Theta = 0_+$. Therefore, we make the equivalence:

$$\theta \;=\; +\infty \qquad \Longleftrightarrow \qquad \Theta \;=\; 0_+. \tag{K3}$$

The agents in the model are:

- One IFT, who chooses a smooth strategy of the form (K1) with $\theta \in [0, \infty)$ and $G \in \mathbb{R}$. The IFT maximizes the expected utility U given by (42):

$$U \;=\; \mathsf{E}\left(\int_0^T (v_T - p_t) \mathrm{d}x_t \right) \;-\; C_I\, \mathsf{E}\left(\int_0^T x_t^2 \mathrm{d}t \right), \tag{K4}$$

  where $T = 1$, and $C_I > 0$ is the IFT's inventory aversion coefficient;

- N_F fast traders, with trading strategy $\mathrm{d}x_t^F = \gamma \mathrm{d}w_t$, with $\gamma \geq 0$;

- N_L slow traders, with trading strategy $\mathrm{d}x_t^S = \mu(\mathrm{d}w_{t-1} - \rho \mathrm{d}y_{t-1})$, with $\mu \geq 0$;

- A dealer who sets a linear pricing rule $\mathrm{d}p_t = \lambda \mathrm{d}y_t$;

[16] In calculus, $\mathrm{d}t$ is considered positive but smaller than any positive real number (and with $\mathrm{d}t^2 = 0$).

- Exogenous noise traders, whose order flow is $\mathrm{d}u_t$.

We introduce the following model coefficients:

$$
\begin{aligned}
R &= \frac{\lambda}{\rho}, \qquad \gamma^- = N_F \gamma, \qquad \bar{\gamma} = \gamma^- + G, \qquad \bar{\mu} = N_L \mu, \\
a^- &= \rho \gamma^-, \qquad a = \rho \bar{\gamma}, \qquad b = \rho \bar{\mu}.
\end{aligned}
\tag{K5}
$$

The model coefficients satisfy $\gamma^- \geq 0$, $\bar{\mu} \geq 0$, $\rho > 0$, $\lambda > 0$.

As usual, tilde notation denotes normalization by σ_w or σ_w^2. For instance, the normalized expected utility of the IFT is denoted by

$$
\tilde{U} = \frac{U}{\sigma_w^2}. \tag{K6}
$$

K.1 The Expected Utility of the IFT

For any smooth strategy of the IFT (not necessarily optimal), we compute the IFT's expected utility, while taking the behavior of the others as given.

Proposition K.1. *In the model described above, suppose $b = \rho\bar{\mu} < 1$. Then, the normalized expected utility of the IFT with a trading strategy as in (K1) is:*

$$
\begin{aligned}
\tilde{U}_\theta &= G(1 - \lambda\bar{\gamma}) \int_0^T \mathrm{e}^{-\theta t}\, \mathrm{d}t + \bar{\mu} \frac{\lambda G (1 - \rho\bar{\gamma})}{1 + \rho\bar{\mu}} \int_0^T \big(1 - \mathrm{e}^{-\theta t}\big) \mathrm{d}t - \frac{\lambda G^2}{2(1 + \rho\bar{\mu})} \int_0^T \big(1 - \mathrm{e}^{-2\theta t}\big) \mathrm{d}t \\
&\quad - \frac{C_I G^2}{2\theta} \int_0^T \big(1 - \mathrm{e}^{-2\theta t}\big) \mathrm{d}t.
\end{aligned}
\tag{K7}
$$

Proposition K.1 shows that the normalized maximum utility of the IFT in the smooth regime ($\tilde{U}_\theta$) varies continuously from $\theta = 0$ to $\theta = \infty$. The next result shows that:

- The limit when $\theta \to 0$ of $\tilde{U}_\theta$ coincides with $\tilde{U}_{\Theta=0}$, the normalized maximum utility of the IFT in the neutral regime ($\Theta = 0$).
- The limit when $\theta \to \infty$ of $\tilde{U}_\theta$ coincides with $\tilde{U}_{\Theta=0_+}$, the left limit when $\Theta \to 0$ of the normalized maximum utility of the IFT in the fast regime ($\Theta > 0$).

Corollary K.1. *In the context of Proposition K.1, the normalized expected utility of the IFT in the smooth regime various continuously from $\theta = 0$ to $\theta = \infty$. It has the*

following limits at the endpoints:

$$\begin{aligned}
\lim_{\theta\to 0} \tilde{U}_{\theta} \; &= \; \tilde{U}_{\theta=0} \; = \; \tilde{U}_{\Theta=0} \; = \; G(1-\lambda\bar{\gamma}), \\
\lim_{\theta\to \infty} \tilde{U}_{\theta} \; &= \; \tilde{U}_{\Theta=0_+} \; = \; \bar{\mu}\frac{\lambda G(1-\rho\bar{\gamma})}{1+\rho\bar{\mu}} - \frac{\lambda G^2}{2(1+\rho\bar{\mu})}.
\end{aligned} \tag{K8}$$

Also, when $\theta \to \infty$, the IFT's (normalized) inventory costs converge to zero:

$$\lim_{\theta\to\infty} \frac{1}{\sigma_w^2} C_I \, \mathsf{E}\left(\int_0^T x_t^2 \mathrm{d}t \right) \; = \; 0. \tag{K9}$$

K.2 Optimal Smooth Inventory Management

In this section, we take a partial equilibrium approach, and solve for the optimal behavior of the IFT in the smooth regime, while taking the behavior of other agents as given. We show that this problem translates into an optimization problem in one variable, which can be solved numerically. The main conclusion of this section is that the optimal trading strategy of the IFT in the smooth regime occurs either at $\theta = 0$ or at $\theta = \infty$. This result is obtained in two steps.

In the first step, we fix $\theta \in [0, \infty]$ and compute the maximum expected utility of the IFT when the coefficient G varies. Denote this utility by $U_{\theta}^{\max}$. In the second step, we numerically search for the $\theta \in [0, \infty]$ that maximizes $U_{\theta}^{\max}$, and we find that the optimum θ is either 0 or ∞.

The next result computes $U_{\theta}^{\max}$. Define the following function of θ (recall $T = 1$):

$$F_{\theta} \; = \; \int_0^T \left(1 - \mathrm{e}^{-\theta t}\right) \mathrm{d}t \; = \; 1 - \frac{1 - \mathrm{e}^{-\theta}}{\theta} \; \in \; [0,1]. \tag{K10}$$

Proposition K.2. *In the context of Proposition K.1, for a fixed $\theta \in [0, \infty)$ denote by $U_{\theta}^{\max} = \tilde{U}_{\theta}^{\max}(C_I)$ the maximum normalized expected utility of the IFT in the smooth regime when G varies. Then,*

$$\tilde{U}_{\theta}^{\max} \; = \; \frac{1}{2} \, \frac{\left(\left(1 - Ra^-\right) - F_{\theta}\left(1 - R\frac{a^-+b}{1+b}\right)\right)^2}{2\lambda\left(1 - \frac{F_{\theta}}{1+b}\right) + F_{2\theta}\left(\frac{\lambda}{1+b} + \frac{C_I}{\theta}\right)}. \tag{K11}$$

We also provide formulas for $\tilde{U}_{\theta}^{\max}$ when $\theta = 0$ and $\theta = \infty$.

Corollary K.2. *In the context of Proposition K.2, we have*

$$\tilde{U}^{\max}_{0} = \frac{1}{2}\frac{\left(1-Ra^{-}\right)^2}{2\lambda+C_I}, \qquad \tilde{U}^{\max}_{\infty} = \frac{\left(Rb(1-a^{-})\right)^2}{4\lambda(1+b)\left(b+\frac{1}{2}\right)}. \tag{K12}$$

The value of C_I that makes $\tilde{U}^{\max}_{0} = \tilde{U}^{\max}_{\infty}$ is:

$$C_I^s = 2\lambda\left(\frac{(1-Ra^{-})^2(1+b)\left(b+\frac{1}{2}\right)}{R^2b^2(1-a^{-})^2} - 1\right). \tag{K13}$$

Moreover, when $C_I = 0$ and $\theta = 0$, the maximum expected of the IFT is:

$$\tilde{U}_0 = \tilde{U}^{\max}_{0,C_I=0} = \frac{\left(1-Ra^{-}\right)^2}{4\lambda}. \tag{K14}$$

In the second step, we show numerically that the maximum $U^{\max}_{\theta}$ occurs either at $\theta = 0$ or at $\theta = \infty$, but not at an interior point in $(0,\infty)$. This results holds for all the parameter values we have checked.

Result K.1. *Suppose the model coefficients arise from the inventory management equilibrium of Theorem 3 for which $N_F \geq 1$ and $N_L \geq 2$. Then, the expression $U^{\max}_{\theta}$ in (K11) never attains its maximum value at an interior point $\theta \in (0,\infty)$.*

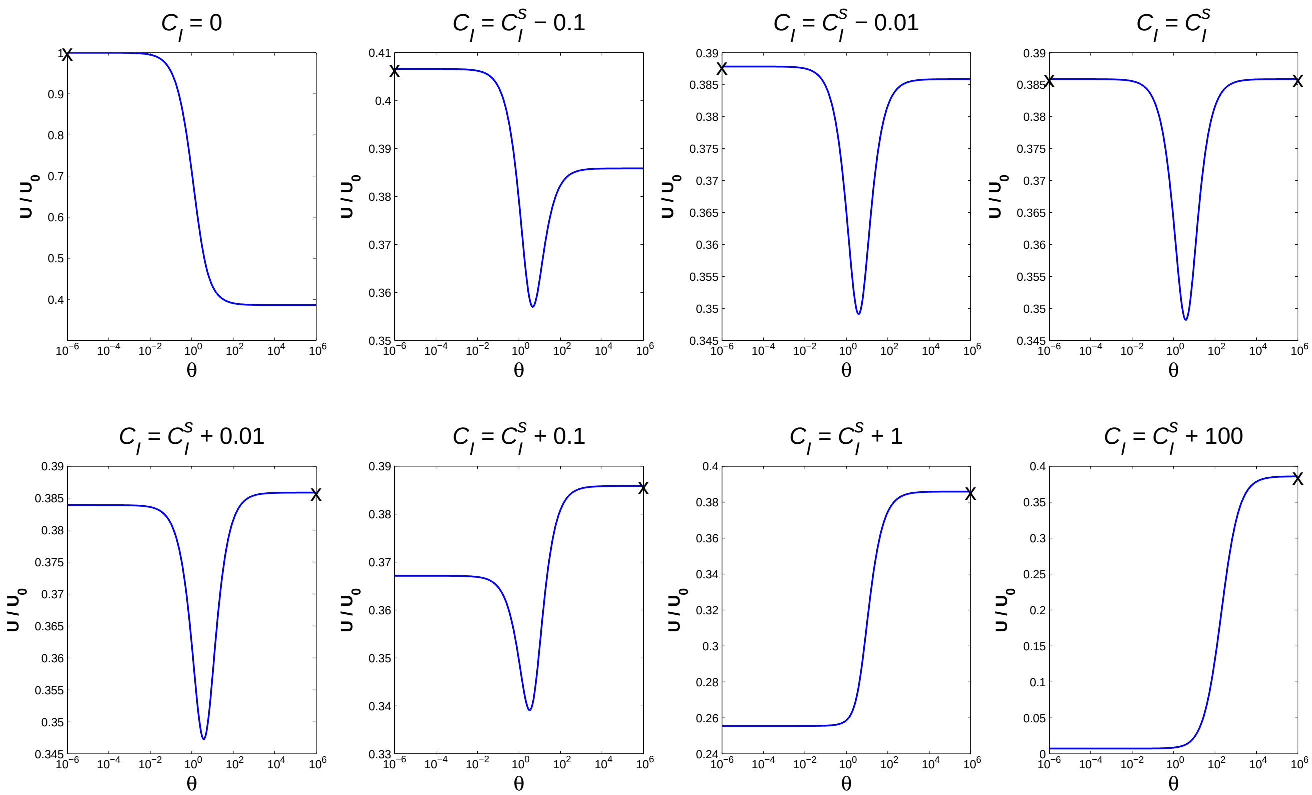
To understand this numerical result, we consider a particular example, with $N_F = 5$ fast traders, and $N_S = 5$ slow traders. In this case, equation (K13) implies that the value of the cutoff is $C_I^s = 1.2038$. This means that when $C_I = C_I^s$, the expected utility difference at the two endpoints $(U_0 - U_\infty)$ switches sign. Figure K.1 plots the IFT's maximum expected utility as a function of θ, given several values of C_I around the cutoff. The maximum expected utility U is computed according to equation (K11), and reported in the graph as a ratio $\frac{U}{U_0}$, where U_0 is the expected utility in (K14) that corresponds to $\theta = 0$ and $C_I = 0$ (no inventory management, and zero inventory costs).

As we can see from the figure, there are two sharply distinctly regimes, depending on whether the inventory aversion coefficient C_I is above the threshold value C_I^s:

- If $C_I < C_I^s$, the IFT optimally chooses $\theta = 0$;
- If $C_I > C_I^s$, the IFT optimally chooses $\theta = \infty$.

Thus, when there is sufficient slow trading so that Theorem 3 holds (and $b > \frac{\sqrt{17}-1}{8} = 0.3904$), the smooth regime is not optimal, and we can just study what happens at the extremities, $\theta = 0$ and $\theta = \infty$.

Figure K.1: Optimal IFT Smooth Trading Strategies. This figure plots $\tilde{U}_\theta^{\max}(C_I)$, the maximum expected utility of the IFT in the smooth regime as a function of θ, for various values of the inventory aversion C_I. Each graph corresponds to an inventory aversion coefficient C_I, which in certain cases is reported relative to the cutoff value $C_I^s = 1.2038$. In each graph, the expected utility U is normalized by the value U_0 that coresponds to $\theta = 0$ and $C_I = 0$. In each graph, the maximum utility is marked with an "x". The parameter values are $N_F = 5$, $N_L = 5$, $\sigma_w = 1$, $\sigma_u = 1$.



K.3 Proofs of Results

Proof of Proposition K.1. Denote by

$$
\begin{aligned}
\Omega_t^{xx} &= \frac{\mathsf{E}\big(x_t^2\big)}{\sigma_w^2}, \qquad \Omega_t^{xe} = \frac{\mathsf{E}\big(x_t(w_t - p_t)\big)}{\sigma_w^2}, \qquad E_t = \frac{\mathsf{E}\big((w_t - p_t)\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\,\mathrm{d}t}, \\
\Omega_t^{xw} &= \frac{\mathsf{E}\big(x_t w_t\big)}{\sigma_w^2}, \qquad \Omega_t^{xp} = \frac{\mathsf{E}\big(x_t p_t\big)}{\sigma_w^2}.
\end{aligned}
\tag{K15}
$$

Note first that

$$
\Omega_t^{xp} = \Omega_t^{xw} - \Omega_t^{xe}. \tag{K16}
$$

From the definition of Ω^{xw}, we get

$$\begin{aligned}\frac{\mathrm{d}\Omega_t^{xw}}{\mathrm{d}t} &= \frac{1}{\sigma_w^2 \mathrm{d}t}\,\mathsf{E}\big(\mathrm{d}x_t w_{t-1} + x_{t-1}\mathrm{d}w_t + \mathrm{d}x_t \mathrm{d}w_t\big) \\ &= -\theta\Omega_{t-1}^{xw} + G.\end{aligned} \tag{K17}$$

As there is no initial inventory, $\Omega_0^{xw} = 0$. Thus, the solution for the differential equation (K17) is:

$$\Omega_t^{xw} = G\,\frac{1-\mathrm{e}^{-\theta t}}{\theta}. \tag{K18}$$

In order to compute Ω_t^{xx} and Ω_t^{xe}, we need to define additional covariances.[17] Denote by

$$X_t = \frac{\mathsf{E}\big(x_t\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\,\mathrm{d}t}, \quad W_t = \frac{\mathsf{E}\big(w_t\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\,\mathrm{d}t}, \quad P_t = \frac{\mathsf{E}\big(p_t\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\,\mathrm{d}t}, \quad E_t = \frac{\mathsf{E}\big((w_t-p_t)\widetilde{\mathrm{d}w_t}\big)}{\sigma_w^2\,\mathrm{d}t}. \tag{K19}$$

To compute these covariances, we derive recursive formulas for them. Note that the aggregate order flow at t is of the form:

$$\mathrm{d}y_t = -\theta x_{t-1}\mathrm{d}t + \bar{\gamma}\mathrm{d}w_t + \bar{\mu}\widetilde{\mathrm{d}w_{t-1}} + \mathrm{d}u_t. \tag{K20}$$

To simplify notation, denote by

$$a = \rho\bar{\gamma}, \qquad b = \rho\bar{\mu}, \qquad A = 1-a, \qquad B = \frac{1-a}{1+b}. \tag{K21}$$

Then, write $\widetilde{\mathrm{d}w_t} = \mathrm{d}w_t - \rho\mathrm{d}y_t$ as follows:

$$\widetilde{\mathrm{d}w_t} = \rho\theta x_{t-1}\mathrm{d}t + \mathrm{d}w_t(1-a) - b\widetilde{\mathrm{d}w_{t-1}} - \rho\mathrm{d}u_t. \tag{K22}$$

The recursive formula for X_t is:

$$\begin{aligned}X_t &= \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\Big(\big(x_{t-1}+\mathrm{d}x_t\big)\big(\rho\theta x_{t-1}\mathrm{d}t + \mathrm{d}w_t(1-a) - b\widetilde{\mathrm{d}w_{t-1}} - \rho\mathrm{d}u_t\big)\Big) \\ &= \rho\theta\frac{\mathsf{E}\big((x_{t-1})^2\big)}{\sigma_w^2\mathrm{d}t} - b\frac{\mathsf{E}\big(x_{t-1}\widetilde{\mathrm{d}w_{t-1}}\big)}{\sigma_w^2\mathrm{d}t} + (1-a)\frac{\mathsf{E}\big(\mathrm{d}x_t\mathrm{d}w_t\big)}{\sigma_w^2\mathrm{d}t} - b\frac{\mathsf{E}\big(\mathrm{d}x_t\widetilde{\mathrm{d}w_{t-1}}\big)}{\sigma_w^2\mathrm{d}t} \\ &= \rho\theta\Omega_{t-1}^{xx} - bX_{t-1} + (1-a)G.\end{aligned} \tag{K23}$$

[17] The inventory management term is of the order of $\mathrm{d}t$, and thus it does not affect any instantaneous covariances with infinitesimal terms, such as $\frac{\mathsf{E}\left((\widetilde{\mathrm{d}w_t})^2\right)}{\sigma_w^2\mathrm{d}t} = A$. However, covariances with aggregate measures such as x_t, w_t, p_t are affected by the slow accumulation of the $\mathrm{d}t$ term. For instance, as we see from the formula (K26) for W_t, the equation $\frac{\mathsf{E}\left(w_t\widetilde{\mathrm{d}w_t}\right)}{\sigma_w^2\mathrm{d}t} = B$ is no longer true here.

Thus, $X_t + bX_{t-1} = \rho\theta\Omega^{xx}_{t-1} + (1-a)G$. Since by assumption $b < 1$ (and $b \geq 0$), we use Lemma A.1 to obtain the following formula:[18]

$$X_t \;=\; \frac{\rho\theta\Omega^{xx}_t + (1-a)G}{1+b} \;=\; \frac{\rho\theta\Omega^{xx}_t}{1+b} + BG. \tag{K24}$$

Similarly, the recursive formula for W_t is:

$$\begin{aligned} W_t \;&=\; \frac{1}{\sigma_w^2 \mathrm{d}t}\,\mathsf{E}\Big(\big(w_{t-1} + \mathrm{d}w_t\big)\big(\rho\theta x_{t-1}\mathrm{d}t + \mathrm{d}w_t(1-a) - b\widetilde{\mathrm{d}w}_{t-1} - \rho\mathrm{d}u_t\big)\Big) \\ &=\; \rho\theta\frac{\mathsf{E}\big(w_{t-1}x_{t-1}\big)}{\sigma_w^2\mathrm{d}t} - b\frac{\mathsf{E}\big(w_{t-1}\widetilde{\mathrm{d}w}_{t-1}\big)}{\sigma_w^2\mathrm{d}t} + (1-a)\frac{\mathsf{E}\big(\mathrm{d}w_t\mathrm{d}w_t\big)}{\sigma_w^2\mathrm{d}t} \\ &=\; \rho\theta\Omega^{xw}_{t-1} - bW_{t-1} + (1-a). \end{aligned} \tag{K25}$$

Thus, $W_t + bW_{t-1} = \rho\theta\Omega^{xw}_{t-1} + (1-a)$. As above, we use Lemma A.1 to get:

$$W_t \;=\; \frac{\rho\theta\Omega^{xw}_t + (1-a)}{1+b} \;=\; \frac{\rho\theta\Omega^{xw}_t}{1+b} + B. \tag{K26}$$

The recursive formula for P_t is:

$$\begin{aligned} P_t \;&=\; \frac{1}{\sigma_w^2 \mathrm{d}t}\,\mathsf{E}\Big(\big(p_{t-1} + \lambda\mathrm{d}y_t\big)\big(\rho\theta x_{t-1}\mathrm{d}t + \mathrm{d}w_t(1-a) - b\widetilde{\mathrm{d}w}_{t-1} - \rho\mathrm{d}u_t\big)\Big) \\ &=\; \rho\theta\frac{\mathsf{E}\big(p_{t-1}x_{t-1}\big)}{\sigma_w^2\mathrm{d}t} - b\frac{\mathsf{E}\big(p_{t-1}\widetilde{\mathrm{d}w}_{t-1}\big)}{\sigma_w^2\mathrm{d}t} + (1-a)\lambda\frac{\mathsf{E}\big(\mathrm{d}y_t\mathrm{d}w_t\big)}{\sigma_w^2\mathrm{d}t} - b\lambda\frac{\mathsf{E}\big(\mathrm{d}y_t\widetilde{\mathrm{d}w}_{t-1}\big)}{\sigma_w^2\mathrm{d}t} - \lambda\rho\tilde{\sigma}_u^2 \\ &=\; \rho\theta\Omega^{xp}_{t-1} - bP_{t-1} + (1-a)\lambda\bar{\gamma} - b\lambda\bar{\mu}A - \lambda\rho\tilde{\sigma}_u^2. \end{aligned} \tag{K27}$$

Define

$$M \;=\; (1-a)\lambda\bar{\gamma} - b\lambda\bar{\mu}A - \lambda\rho\tilde{\sigma}_u^2. \tag{K28}$$

One can check that $M = 0$ when a and b have the equilibrium values from Theorem 1. This simply reflects the fact that $\widetilde{\mathrm{d}w}_t$ is orthogonal to p_t in the absence of inventory management, i.e., $P_t = 0$ when $\theta = 0$.[19] As in the case of X_t, we use Lemma A.1 to obtain

$$P_t \;=\; \frac{\rho\theta\Omega^{xp}_t + M}{1+b} \;=\; \frac{\rho\theta(\Omega^{xw}_t - \Omega^{xe}_t) + M}{1+b}. \tag{K29}$$

[18] The difference between Ω^{xx}_t and Ω^{xx}_{t-1} is infinitesimal, hence it can be ignored. In other words, one can use Lemma A.1 either for α_t or for α_{t-1}, and obtain the same result.

[19] Indeed, using $\rho^2\tilde{\sigma}_u^2 = (1-a)(a-b^2)$, we compute $\frac{\lambda}{\rho}\big((1-a)a - (1-a)b^2 - (1-a)(a-b^2)\big) = 0$.

From (K26) and (K29) we also get

$$E_t \;=\; W_t - P_t \;=\; \frac{\rho\theta\Omega_t^{xe} - M}{1+b} + B. \tag{K30}$$

We now compute Ω_t^{xx}. From its definition, we have

$$\begin{aligned}
\frac{\mathrm{d}\Omega_t^{xx}}{\mathrm{d}t} \;&=\; \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\big(2x_{t-1}\mathrm{d}x_t + (\mathrm{d}x_t)^2\big)\\
&=\; \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\Big(2x_{t-1}\big(-\theta x_{t-1}\mathrm{d}t + \gamma\mathrm{d}w_t + \mu\widetilde{\mathrm{d}w}_{t-1}\big) + (\mathrm{d}x_t)^2\Big).\\
&=\; -2\theta\Omega_{t-1}^{xx} + G^2.
\end{aligned} \tag{K31}$$

This is a first order ODE with solution:

$$\Omega_t^{xx} \;=\; G^2\,\frac{1-\mathrm{e}^{-2\theta t}}{2\theta}. \tag{K32}$$

Finally, we compute Ω_t^{xe}. Since $\mathrm{d}w_t - \mathrm{d}p_t = \lambda\theta x_{t-1}\mathrm{d}t + (1-\lambda\bar{\gamma})\mathrm{d}w_t - \lambda\bar{\mu}\widetilde{\mathrm{d}w}_{t-1} - \lambda\mathrm{d}u_t$, from the definition of Ω_t^{xe}, we have

$$\begin{aligned}
\frac{\mathrm{d}\Omega_t^{xe}}{\mathrm{d}t} \;&=\; \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\big((w_{t-1}-p_{t-1})\mathrm{d}x_t + x_{t-1}(\mathrm{d}w_t - \mathrm{d}p_t) + (\mathrm{d}w_t - \mathrm{d}p_t)\mathrm{d}x_t\big)\\
&=\; -\theta\Omega_{t-1}^{xe} + \lambda\theta\Omega_{t-1}^{xx} - \lambda\bar{\mu}X_{t-1} + (1-\lambda\bar{\gamma})G.
\end{aligned} \tag{K33}$$

From (K24), we have $X_{t-1} = \frac{\rho\theta\Omega_{t-1}^{xx}}{1+b} + BG$. We obtain

$$\begin{aligned}
\frac{\mathrm{d}\Omega_t^{xe}}{\mathrm{d}t} \;&=\; -\theta\Big(1 - \frac{\rho\mu}{1+b}\Big)\Omega_{t-1}^{xe} + \lambda\theta\Big(1 - \frac{\rho\bar{\mu}}{1+b}\Big)\Omega_{t-1}^{xx} - \lambda\bar{\mu}BG + (1-\lambda\bar{\gamma})G\\
&=\; -\theta\Omega_{t-1}^{xe} + \lambda\theta\frac{1}{1+b}\,\Omega_{t-1}^{xx} - \lambda\bar{\mu}BG + \tilde{\pi}_0,
\end{aligned} \tag{K34}$$

where $\tilde{\pi}_0$ is the IFT's normalized expected profit when $\theta = 0$:

$$\tilde{\pi}_0 \;=\; (1-\lambda\bar{\gamma})G. \tag{K35}$$

From (K32), $\lambda\theta\frac{1}{1+b}\,\Omega_{t-1}^{xx} = \frac{\lambda\theta}{1+b}\,G^2\,\frac{1-\mathrm{e}^{-2\theta t}}{2\theta} = \frac{\lambda G^2}{2(1+b)}\,(1-\mathrm{e}^{-2\theta t})$. The differential equation for Ω_t^{xe} becomes:

$$\begin{aligned}
&\frac{\mathrm{d}\Omega_t^{xe}}{\mathrm{d}t} \;=\; -\theta\Omega_{t-1}^{xe} + D_1\big(1-\mathrm{e}^{-2\theta t}\big) + D_2, \quad \text{with}\\
&D_1 \;=\; \frac{\lambda G^2}{2(1+b)}, \qquad D_2 \;=\; \tilde{\pi}^0 - \lambda\bar{\mu}BG.
\end{aligned} \tag{K36}$$

This is a first order ODE with solution:

$$\Omega_t^{xe} \;=\; \Big(D_1 + D_2\Big)\frac{1 - \mathrm{e}^{-\theta t}}{\theta} + D_1\,\frac{\mathrm{e}^{-\theta t} - \mathrm{e}^{-2\theta t}}{\theta}. \tag{K37}$$

Next, we compute the IFT's expected profit in the smooth regime:

$$\begin{aligned}
\tilde{\pi}_{\theta} \;&=\; \frac{1}{\sigma_w^2}\,\mathsf{E}\int_0^T (w_t - p_t)\mathrm{d}x_t \\
&=\; \frac{1}{\sigma_w^2}\,\mathsf{E}\int_0^T \Big(w_{t-1} - p_{t-1} + \mathrm{d}w_t - \lambda\mathrm{d}y_t\Big)\Big(G\mathrm{d}w_t - \theta x_{t-1}\mathrm{d}t\Big) \\
&=\; \int_0^T \Big(-\theta\Omega_{t-1}^{xe} + G - \lambda G\bar{\gamma}\Big)\mathrm{d}t.
\end{aligned} \tag{K38}$$

Therefore,

$$\tilde{\pi}_{\theta} \;=\; \tilde{\pi}_{0} - \int_0^T \theta\Omega_t^{xe}\mathrm{d}t. \tag{K39}$$

From (K37),

$$\theta\Omega_t^{xe} \;=\; \frac{\lambda G^2}{2(1+b)}\Big(1 - \mathrm{e}^{-2\theta t}\Big) + \big(\tilde{\pi}_{0} - \lambda\bar{\mu}BG\big)\Big(1 - \mathrm{e}^{-\theta t}\Big), \tag{K40}$$

with D_1 and D_2 as in (K36). We compute

$$\tilde{\pi}_{\theta} \;=\; \tilde{\pi}_{0}\int_0^T \mathrm{e}^{-\theta t}\,\mathrm{d}t + \lambda\bar{\mu}BG\int_0^T \big(1 - \mathrm{e}^{-\theta t}\big)\mathrm{d}t - \frac{\lambda G^2}{2(1+b)}\int_0^T \big(1 - \mathrm{e}^{-2\theta t}\big)\mathrm{d}t. \tag{K41}$$

This is the first line in equation (K7). Since the normalized expected utility of the IFT satisfies:

$$\tilde{U}_{\theta} \;=\; \tilde{\pi}_{\theta} - \frac{1}{\sigma_w^2}\,C_I\,\mathsf{E}\left(\int_0^T x_t^2\mathrm{d}t\right), \tag{K42}$$

to prove the second part of equation (K7) we only have to show that

$$\frac{1}{\sigma_w^2}\,C_I\,\mathsf{E}\left(\int_0^T x_t^2\mathrm{d}t\right) \;=\; C_I\int_0^T \Omega_t^{xx}\mathrm{d}t \;=\; \frac{C_I G^2}{2\theta}\int_0^T \big(1 - \mathrm{e}^{-2\theta t}\big)\mathrm{d}t. \tag{K43}$$

But equation (K32) implies that $\Omega_t^{xx} = G^2\,\frac{1 - \mathrm{e}^{-2\theta t}}{2\theta}$. This completes the proof of Proposition K.1. □

Proof of Corollary K.1. Note that for $t > 0$,

$$\lim_{\theta\to 0}\frac{1 - \mathrm{e}^{-2\theta t}}{2\theta} \;=\; t \quad\Longrightarrow\quad \lim_{\theta\to 0}\frac{1}{2\theta}\int_0^T \big(1 - \mathrm{e}^{-2\theta t}\big)\mathrm{d}t \;=\; \int_0^T t\mathrm{d}t \;=\; \frac{1}{2}. \tag{K44}$$

Equation (K7) implies that when $\theta \to 0$, $\tilde{U}_{\theta}$ converges to $G(1-\lambda\bar{\gamma}) - \frac{C_I}{2}G^2$. But by summing (A36) and (A37) from the proof of Theorem 2, we obtain that $\tilde{U}_{\Theta=0} = G(1-\lambda\bar{\gamma}) - \frac{C_I}{2}G^2$. Therefore, $\tilde{U}_{\Theta=0} = \tilde{U}_{\theta=0}$.

Equation (K7) also implies that when $\theta \to \infty$, $\tilde{U}_{\theta}$ converges to $\bar{\mu}\frac{\lambda G(1-\rho\bar{\gamma})}{1+\rho\bar{\mu}} - \frac{\lambda G^2}{2(1+\rho\bar{\mu})}$. But equation (J36) from the proof of Proposition J.1 computes $\tilde{U}_{\Theta=0_+} = \frac{\lambda\bar{\mu}G(1-\rho\bar{\gamma})}{1+\rho\bar{\mu}} - \frac{\lambda G^2}{2(1+\rho\bar{\mu})}$. Therefore, $\tilde{U}_{\Theta=0_+} = \tilde{U}_{\theta=\infty}$.

To show that the inventory costs approach zero when $\theta \to \infty$, note that in equation (K7), $\mathrm{e}^{-\theta t}$ converges uninformly to zero (for $t \in [0,T]$). $\square$

Proof of Proposition K.2. In Proposition K.1, we have already computed the normalized expected utility of the IFT:

$$\tilde{U}_{\theta} = G(1-\lambda\bar{\gamma})\int_0^T \mathrm{e}^{-\theta t}\,\mathrm{d}t + \bar{\mu}\frac{\lambda G(1-\rho\bar{\gamma})}{1+\rho\bar{\mu}}\int_0^T \big(1-\mathrm{e}^{-\theta t}\big)\mathrm{d}t - \frac{\lambda G^2}{2(1+\rho\bar{\mu})}\int_0^T \big(1-\mathrm{e}^{-2\theta t}\big)\mathrm{d}t - \frac{C_I G^2}{2\theta}\int_0^T \big(1-\mathrm{e}^{-2\theta t}\big)\mathrm{d}t. \tag{K45}$$

Recall that

$$F_\theta = \int_0^T \big(1-\mathrm{e}^{-\theta t}\big)\mathrm{d}t = 1 - \frac{1-\mathrm{e}^{-\theta}}{\theta}. \tag{K46}$$

One can check that $\frac{F_\theta}{\theta}$ is a well define analytical function, and

$$\lim_{\theta\to 0}\frac{F_\theta}{\theta} = \frac{1}{2}. \tag{K47}$$

Equation (K45) becomes

$$\tilde{U}_{\theta} = G(1-\lambda\bar{\gamma})(1-F_\theta) + \frac{RbG(1-\rho\bar{\gamma})}{1+b}F_\theta - \frac{\lambda G^2}{2(1+b)}F_{2\theta} - \frac{C_I G^2}{2\theta}F_{2\theta}. \tag{K48}$$

Since

$$\bar{\gamma} = \gamma^- + G, \tag{K49}$$

we compute

$$\tilde{U}_{\theta} = G\left(\big(1-Ra^-\big) - F_\theta\left(1 - R\frac{a^- + b}{1+b}\right)\right) - \frac{G^2}{2}\left(2\lambda\left(1-\frac{F_\theta}{1+b}\right) + F_{2\theta}\left(\frac{\lambda}{1+b} + \frac{C_I}{\theta}\right)\right). \tag{K50}$$

Fix $\theta \in [0, \infty]$. Then, the first order condition in G implies

$$G = \frac{\left(1 - Ra^-\right) - F_\theta \left(1 - R\frac{a^- + b}{1+b}\right)}{2\lambda \left(1 - \frac{F_\theta}{1+b}\right) + F_{2\theta} \left(\frac{\lambda}{1+b} + \frac{C_I}{\theta}\right)}. \tag{K51}$$

The second order condition for a maximum is also clearly satisfied. Hence, for a given θ, the maximum normalized expected utility of the IFT when G varies is:

$$\tilde{U}_\theta^{\max} = \frac{1}{2} \frac{\left(\left(1 - Ra^-\right) - F_\theta \left(1 - R\frac{a^- + b}{1+b}\right)\right)^2}{2\lambda \left(1 - \frac{F_\theta}{1+b}\right) + F_{2\theta} \left(\frac{\lambda}{1+b} + \frac{C_I}{\theta}\right)}. \tag{K52}$$

This proves equation (K11). $\square$

Proof of Corollary K.2. We use the formula for $U_\theta^{\max}$ from Proposition K.2. When $\theta = 0$, we have $\lim\limits_{\theta \to 0} \frac{F_{2\theta}}{2\theta} = \frac{1}{2}$, hence we obtain the first equation in (K12). When $\theta \to \infty$, we have $\lim\limits_{\theta \to \infty} F_\theta = 1$, hence

$$\tilde{U}_\infty = \frac{1}{2} \frac{R^2 b^2 \frac{(1-a^-)^2}{(1+b)^2}}{2\lambda \frac{b}{1+b} + \frac{\lambda}{1+b}} = \frac{\left(Rb(1 - a^-)\right)^2}{4\lambda(1+b)\left(b + \frac{1}{2}\right)}, \tag{K53}$$

which proves the second equation in (K12). One can now solve directly for the C_I that makes $U_0^{\max} = U_\infty^{\max}$.

Finally, equation (K14) follows from the formula for $\tilde{U}_0^{\max}$ in (K12) by setting $C_I = 0$. $\square$

L Robust Trading Strategies

The purpose of this section is twofold. First, we verify that the intuition of our baseline model from Section 2 extends to a setup in which the fundamental value has more than one component. For simplicity, we focus on extending the model $\mathcal{M}_0$ in which all speculators trade only on their current signal (with no lags). Thus, if $\mathrm{d}w_t$ is the current signal about only one of two orthogonal components of the fundamental value, we verify that trading strategies of the form

$$\mathrm{d}x_t \ = \ \gamma_t \mathrm{d}w_t \tag{L1}$$

remain profitable.

Second, when the fundamental value has two components, we study the decision of speculators to use smooth strategies of the Kyle (1985) type::[20]

$$\mathrm{d}x_t \ = \ \beta_t (w_t - p_t)\mathrm{d}t, \tag{L2}$$

We call these strategies *smooth*. In our model of Section 2, these strategies are not allowed because they involve using an infinite number of lags:

$$w_t \ = \ \mathrm{d}w_t + \mathrm{d}w_{t-1} + \mathrm{d}w_{t-2} + \cdots \tag{L3}$$

(Recall that we denote $X_{t-1} = X_{t-\mathrm{d}t}$.) In this section, we show that using any smooth strategy as in equation (L2) would produce an expected loss for certain parameter values. In this sense, smooth strategies are not robust to the alternate model in which the asset value is multidimensional.

L.1 Motivation

In most trading models with asymmetric information, speculators learn only about one component of the asset's fundamental value. For instance, in Kyle (1985), the unique informed trader (the "insider") uses private information to generate profits smoothly over time, using a strategy as in equation (L2). Thus, the insider compares his forecast with the price, and then buys slowly if his forecast is above the price, and sells otherwise. The implicit assumption in Kyle (1985) is that the price only contains information about

[20]In Kyle (1985) w_t is in fact constant. Back and Pedersen (1998), however, show that the same type of strategies are optimal even if the fundamental value changes over time.

his signal, and thus the insider has no inference problem: he knows his information to be superior to that of the public's.

We now introduce a second component of the fundamental value, as in Subrahmanyam and Titman (1999), and allow a different group of speculators to learn about this second component.[21] We show that in this case the smooth strategy in (L2) starts losing money if the parameters related to the other component of the asset value are large enough. In other words, smooth strategies are not robust. By contrast, fast strategies as in equation (L1) are robust. Indeed, Proposition L.1 below shows that the expected profit from this strategy is positive, and stays constant under all these specifications (taking the price impact coefficient λ as given).

Intuitively, when the fundamental value has multiple components, a speculator who specializes in only one component is potentially adversely selected when using the price to decide his strategy. For instance, suppose the value of IBM has both a domestic and an international component. Then, suppose that a hedge fund that specializes only in the IBM's domestic component uses a smooth strategy as in (L2). Then, by buying and selling at the public price, the hedge fund essentially behaves as a noise trader with respect to the international component, and can therefore make losses on average. If instead, the hedge fund uses a fast strategy and buys if its signal about the domestic component is positive, its average profit is not affected by what happens in the international component.

L.2 Multidimensional Asset Value

We now describe formally the model with two components of the fundamental value. Suppose the liquidation value of the risk asset v_T (at $T = 1$) can be decomposed as a sum:

$$v_T \;=\; w_T + e_T, \qquad T \;=\; 1. \tag{L4}$$

We consider a model similar to $\mathcal{M}_0$ from Section 2 in which speculators only use their current signal (see also Proposition 3). There are $N_w \geq 1$ speculators, called the *w-speculators*, who learn about w_T by observing at each t the increment $\mathrm{d}w_t$ of a diffusion process with terminal value w_T. Also, there are $N_e \geq 1$ speculators, called the *e-speculators*, who learn about e_T by observing at each t the increment $\mathrm{d}e_t$ of a diffusion

[21] Subrahmanyam and Titman (1999) have a one-period model with information acquisition. They find that multidimensional asset values generate liquidity complementarity, in the sense that informed traders in one component of the asset value behave as noise traders in the second component, and thus encourage information production in that component.

process with terminal value e_T.[22] Recall that in our baseline model, the speculators receive signals of the form $\mathrm{d}s_t = \mathrm{d}v_t + \mathrm{d}\eta_t$, such that the increment of their forecast w_t is equal to $\mathrm{d}w_t = a\mathrm{d}s_t$, with $a = \frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}$. In this section, however, we do not need these explicit formulas, and assume instead that the w-speculators directly observe $\mathrm{d}w_t$. Thus, the fundamental value increment $\mathrm{d}v_t$ has the following orthogonal decomposition:

$$\mathrm{d}v_t \;=\; \mathrm{d}w_t + \mathrm{d}e_t, \quad \text{with} \quad \sigma_e^2 \;=\; \sigma_v^2 - \sigma_w^2 \;=\; \frac{\sigma_v^2 \sigma_\eta^2}{\sigma_v^2 + \sigma_\eta^2}. \tag{L5}$$

Proposition L.1. *Consider $N_w + N_e$ speculators, of which N_w speculators learn about w_t, and N_e speculators learn about e_t. Denote the position in the risky asset of an w- or e-speculator, respectively, by $x_{w,t}$ and $x_{e,t}$. Assume that the speculators can only trade on their most recent signal, $\mathrm{d}w_t$ or $\mathrm{d}e_t$, respectively. Then, there exists a unique linear equilibrium, in which the speculators' trading strategies, and the dealer's pricing functions are of the form:*

$$\mathrm{d}x_{w,t} \;=\; \gamma_w \mathrm{d}w_t, \qquad \mathrm{d}x_{e,t} \;=\; \gamma_e \mathrm{d}e_t, \qquad \mathrm{d}p_t \;=\; \lambda \mathrm{d}y_t, \tag{L6}$$

with equilibrium coefficients given by

$$\gamma_w \;=\; \frac{1}{\lambda}\frac{1}{N_w+1}, \quad \gamma_e \;=\; \frac{1}{\lambda}\frac{1}{N_e+1}, \quad \lambda \;=\; \left(\frac{N_w}{(N_w+1)^2}\frac{\sigma_w^2}{\sigma_u^2} + \frac{N_e}{(N_e+1)^2}\frac{\sigma_e^2}{\sigma_u^2} \right)^{1/2}. \tag{L7}$$

The expected profits at $t=0$ of the w- and e-speculators, are given, respectively, by:

$$\pi_w \;=\; \frac{\sigma_w^2}{\lambda(N_w+1)^2}, \quad \pi_e \;=\; \frac{\sigma_w^2}{\lambda(N_e+1)^2}. \tag{L8}$$

Proof. See Section L.4 below. □

Proposition L.1 shows that, taking λ as given, the w-speculators have indeed the same strategy, and the same expected profits regardless of the structure of the e-component. That is to say, the strategy and profits of the w speculators do not depend on N_e or σ_e. The magnitude of the price impact coefficient λ, however, does change with the specification, because of the increase in adverse selection in the other component.

[22]We do not model the information acquisition explicitly. Subrahmanyam and Titman (1999) solve a one-period model with endogenous information acquisition, and analyze the liquidity externalities that result from this choice.

L.3 Smooth Trading with Multidimensional Value

The next result analyzes the expected profits that arise from combining smooth strategies as in (L2) with fast strategies as in (L1).

Proposition L.2. *In the context of Proposition L.1, suppose now that one of the w-speculators, now called the β-*speculator*, adds a smooth component to his equilibrium trading strategy:*

$$\mathrm{d}x_t^{\beta} \;=\; \beta_t(w_{t-1} - p_{t-1})\mathrm{d}t + \gamma_w \mathrm{d}w_t, \tag{L9}$$

while the other traders and the dealer maintain their equilibrium strategies. Then, the expected profit of the β-speculator at $t = 0$ equals:

$$\pi^{\beta} \;=\; \pi^{\beta=0} + \frac{1}{2\lambda}\int_0^T \big(1-\varepsilon_t\big)\left((1+\varepsilon_t)\frac{1}{N_w+1}\sigma_w^2 - (1-\varepsilon_t)\frac{N_e}{N_e+1}\sigma_e^2\right)\mathrm{d}t, \tag{L10}$$

where

$$\pi^{\beta=0} \;=\; \frac{\sigma_w^2}{\lambda(N_w+1)^2}, \qquad \varepsilon_t \;=\; \mathrm{e}^{-\lambda\left(\int_t^1 \beta_\tau \mathrm{d}\tau\right)}. \tag{L11}$$

***Proof**.* See Section L.4 below. □

An important implication of Proposition L.2 is that the profit π^{β} depends on the specification of the model. Consider the following cases:

- If $\sigma_e = 0$, the fundamental value has only one component. Then, the β-speculator increases his profit by β-trading: $\frac{\sigma_w^2}{2\lambda(N_w+1)}\int_0^T(1-\varepsilon_t^2)\mathrm{d}t > 0$. Moreover, if as in Kyle (1985) the β-speculator sets $\beta_t = \frac{\beta_0}{1-t} > 0$, then $\varepsilon_t = 0$, and the profit is maximized.

- If $\sigma_e > 0$, there is more than one component of the fundamental value. Then, we show that $\beta_t > 0$ produces a loss for certain values of σ_e (and N_e). Indeed, the condition $\beta_t > 0$ translates into ε_t not being identically equal to 1, or equivalently $\int_0^T(1-\varepsilon_t)^2\mathrm{d}t > 0$. Choose a value σ_e such that

$$\sigma_e^2 \;>\; \frac{N_e+1}{N_e}\,\frac{\frac{2}{(N_w+1)^2} + \frac{\int_0^T(1-\varepsilon_t^2)\mathrm{d}t}{N_w+1}}{\int_0^T(1-\varepsilon_t)^2\mathrm{d}t}\,\sigma_w^2. \tag{L12}$$

  Then, one can easily verify that $\pi^{\beta} < 0$.

In other words, the β-strategy in equation (L9) is not robust to the fundamental value having more than one component.

L.4 Proofs of Results

Proof of Proposition L.1. We first determine the optimal strategies of the speculators, taking the dealer's pricing rule as given. The expected profit at $t = 0$ of the i'th w-speculator is

$$\begin{aligned}\pi^i_{w,0} &= \mathsf{E}\int_0^T \left(w_t + e_t - p_{t-1} - \lambda_t\Big((\gamma^i_{w,t} + \gamma^{-i}_{w,t})\mathrm{d}w_t + \sum_{j=1}^{N_e} \gamma^j_{e,t}\mathrm{d}e_t + \mathrm{d}u_t \Big) \right) \gamma^i_{w,t}\mathrm{d}w_t \\ &= \int_0^T \gamma^i_{w,t}\sigma^2_w \mathrm{d}t - \lambda_t\gamma^i_{w,t}\big(\gamma^i_{w,t} + \gamma^{-i}_{w,t}\big)\sigma^2_w\mathrm{d}t,\end{aligned} \tag{L13}$$

where $\gamma^{-i}_{w,t}$ denotes the aggregate coefficient of the other $N_w - 1$ w-speculators. This is a pointwise quadratic optimization problem, with solution $\lambda_t\gamma^i_{w,t} = \frac{1-\lambda_t\gamma^{-i}_{w,t}}{2}$. Since this is true for each w-speculator, we obtain that all $\gamma^i_{w,t}$ are equal to $\gamma_{w,t} = \frac{1}{\lambda_t(N_w+1)}$. Similarly, all $\gamma^j_{e,t}$ are equal to $\gamma_{e,t} = \frac{1}{\lambda_t(N_e+1)}$. Combining these two equations, we get

$$\gamma^i_{w,t} = \gamma_{w,t} = \frac{1}{\lambda_t(N_w+1)}, \qquad \gamma^j_{e,t} = \gamma_{e,t} = \frac{1}{\lambda_t(N_e+1)}. \tag{L14}$$

We now determine the dealer's pricing rule, taking the behavior of the speculators as given. The dealer assumes that the aggregate order flow is $\mathrm{d}y_t = N_w\gamma_t\mathrm{d}w_t + N_e\phi_t\mathrm{d}e_t + \mathrm{d}u_t$. To set λ_t, the dealer sets p_t such that the market is efficient, which implies

$$\mathrm{d}p_t = \lambda_t\mathrm{d}y_t, \quad \text{with} \quad \lambda_t = \frac{\mathsf{Cov}(w_t + e_t, \mathrm{d}y_t)}{\mathsf{Var}(\mathrm{d}y_t)} = \frac{N_w\gamma_{w,t}\sigma^2_w + N_e\gamma_{e,t}\sigma^2_e}{N^2_w\gamma^2_{w,t}\sigma^2_w + N^2_e\gamma^2_{t,e}\sigma^2_e + \sigma^2_u}. \tag{L15}$$

This implies $(N_w\lambda_t\gamma_{w,t})^2\sigma^2_w + (N_e\lambda_t\gamma_{e,t})^2\sigma^2_e + \lambda^2_t\sigma^2_u = N_w\lambda_t\gamma_{w,t}\sigma^2_w + N_e\lambda_t\gamma_{e,t}\sigma^2_e$. But $N_w\lambda_t\gamma_{w,t} = \frac{N_w}{N_w+1}$ and $N_e\lambda_t\gamma_{e,t} = \frac{N_e}{N_e+1}$. Hence, $\lambda^2_t\sigma^2_u = \frac{N_w}{(N_w+1)^2}\sigma^2_w + \frac{N_e}{(N_e+1)^2}\sigma^2_e$, which implies

$$\lambda_t = \lambda = \left(\frac{N_w}{(N_w+1)^2}\frac{\sigma^2_w}{\sigma^2_u} + \frac{N_e}{(N_e+1)^2}\frac{\sigma^2_e}{\sigma^2_u} \right)^{1/2}. \tag{L16}$$

This proves the stated formulas. □

Proof of Proposition L.2. The trading strategy of the β-speculator is of the form:

$$\mathrm{d}x_t = \beta_t(w_{t-1} - p_{t-1})\mathrm{d}t + \gamma_w\mathrm{d}w_t \tag{L17}$$

where γ_w is the equilibrium (constant) value. Denote the aggregate coefficients by

$$\bar{\gamma}_w \;=\; N_w\gamma_w, \qquad \bar{\gamma}_e \;=\; N_e\gamma_e. \tag{L18}$$

Define the following (normalized) covariances

$$\Sigma_t \;=\; \frac{\mathsf{E}\big((w_t-p_t)^2\big)}{\sigma_w^2}, \qquad \Omega_t \;=\; \frac{\mathsf{E}\big(e_t p_t\big)}{\sigma_w^2}, \qquad \tilde{\sigma}_e^2 \;=\; \frac{\sigma_e^2}{\sigma_w^2}. \tag{L19}$$

Since $w_0 = e_0 = p_0$, we have

$$\Sigma_0 \;=\; \Omega_0 \;=\; 0. \tag{L20}$$

The normalized expected profit of the β-speculator at $t=0$ is:

$$\begin{aligned}
\tilde{\pi}^\beta \;&=\; \frac{1}{\sigma_w^2}\,\mathsf{E}\int_0^T \bigg(w_t+e_t-p_{t-1}-\lambda\Big(\beta_t(w_{t-1}-p_{t-1})\mathrm{d}t+\bar{\gamma}_w\mathrm{d}w_t+\bar{\gamma}_e\mathrm{d}e_t+\mathrm{d}u_t\Big)\bigg)\times \\
&\qquad\qquad \times\Big(\beta_t(w_{t-1}-p_{t-1})\mathrm{d}t+\gamma_w\mathrm{d}w_t\Big) \\
&=\; \int_0^T \Big(\gamma_w-\lambda\gamma_w\bar{\gamma}_w+\beta_t\Sigma_{t-1}-\beta_t\Omega_{t-1}\Big)\mathrm{d}t \\
&=\; \tilde{\pi}^0 \;+\; \int_0^T \Big(\beta_t\Sigma_{t-1}-\beta_t\Omega_{t-1}\Big)\mathrm{d}t,
\end{aligned} \tag{L21}$$

where $\tilde{\pi}^0$ is the normalized profit of the β-speculator when $\beta=0$, i.e.,

$$\tilde{\pi}^0 \;=\; \gamma_w-\lambda\gamma_w\bar{\gamma}_w \;=\; \frac{\gamma_w}{N_w+1} \;=\; \frac{1}{\lambda(N_w+1)^2}. \tag{L22}$$

Since $\mathrm{d}w_t-\mathrm{d}p_t = -\lambda\beta_t(w_{t-1}-p_{t-1})\mathrm{d}t+(1-\lambda\bar{\gamma}_w)\mathrm{d}w_t-\lambda\bar{\gamma}_e\mathrm{d}e_t-\lambda\mathrm{d}u_t$, Σ_t satisfies:

$$\begin{aligned}
\frac{\mathrm{d}\Sigma_t}{\mathrm{d}t} \;&=\; \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\big(2(w_{t-1}-p_{t-1})(\mathrm{d}w_t-\mathrm{d}p_t)+(\mathrm{d}w_t-\mathrm{d}p_t)^2\big) \\
&=\; -2\lambda\beta_t\Sigma_{t-1}+(1-\lambda\bar{\gamma}_w)^2+(\lambda\bar{\gamma}_e)^2\tilde{\sigma}_e^2+\lambda^2\tilde{\sigma}_u^2.
\end{aligned} \tag{L23}$$

This is a first order ODE with solution:

$$\begin{aligned}
\Sigma_t \;&=\; D_\Sigma\,\mathrm{e}^{-2\lambda B_t}\int_0^t \mathrm{e}^{2\lambda B_\tau}\,\mathrm{d}\tau, \quad \text{with} \\
B_t \;&=\; \int_0^t \beta_\tau\mathrm{d}\tau, \qquad D_\Sigma \;=\; (1-\lambda\bar{\gamma}_w)^2+(\lambda\bar{\gamma}_e)^2\tilde{\sigma}_e^2+\lambda^2\tilde{\sigma}_u^2.
\end{aligned} \tag{L24}$$

We recall now a formula we derived in the computation of λ:

$$(\lambda\bar{\gamma}_w)^2 + (\lambda\bar{\gamma}_e)^2\tilde{\sigma}_e^2 + \lambda^2\tilde{\sigma}_u^2 \;=\; \lambda\bar{\gamma}_w + \lambda\bar{\gamma}_e\tilde{\sigma}_e^2. \tag{L25}$$

Then, we compute

$$D_\Sigma \;=\; 1 - \lambda\bar{\gamma}_w + \lambda\bar{\gamma}_e\tilde{\sigma}_e^2 \;=\; \frac{1}{N_w+1} + \frac{N_e}{N_e+1}\tilde{\sigma}_e^2. \tag{L26}$$

By integrating (L23) over $[0,T]$ (and using $\Sigma_0 = 0$), we also compute

$$\int_0^T \beta_t\Sigma_{t-1}\mathrm{d}t \;=\; \frac{D_\Sigma - \Sigma_1}{2\lambda}. \tag{L27}$$

Since $\mathrm{d}p_t = \lambda\beta_t(w_{t-1} - p_{t-1})\mathrm{d}t + \lambda\bar{\gamma}_w\mathrm{d}w_t + \lambda\bar{\gamma}_e\mathrm{d}e_t + \lambda\mathrm{d}u_t$, Ω_t satisfies:

$$\begin{aligned}\frac{\mathrm{d}\Omega_t}{\mathrm{d}t} \;&=\; \frac{1}{\sigma_w^2\mathrm{d}t}\,\mathsf{E}\big(p_{t-1}\mathrm{d}e_t + e_{t-1}\mathrm{d}p_t + \mathrm{d}e_t\mathrm{d}p_t\big)\\ &=\; -\lambda\beta_t\Omega_{t-1} + \lambda\bar{\gamma}_e\tilde{\sigma}_e^2.\end{aligned} \tag{L28}$$

This is a first order ODE with solution:

$$\begin{aligned}\Omega_t \;&=\; D_\Omega\,\mathrm{e}^{-\lambda B_t}\int_0^t \mathrm{e}^{\lambda B_\tau}\,\mathrm{d}\tau, \quad \text{with}\\ D_\Omega \;&=\; \lambda\bar{\gamma}_e\tilde{\sigma}_e^2 \;=\; \frac{N_e}{N_e+1}\tilde{\sigma}_e^2.\end{aligned} \tag{L29}$$

By integrating (L28) over $[0,T]$ (and using $\Omega_0 = 0$), we also compute

$$\int_0^T \beta_t\Omega_{t-1}\mathrm{d}t \;=\; \frac{D_\Omega - \Omega_1}{\lambda}. \tag{L30}$$

We compute

$$\Sigma_1 \;=\; D_\Sigma\int_0^T \mathrm{e}^{-2\lambda(B_1 - B_t)}\,\mathrm{d}t, \qquad \Omega_1 \;=\; D_\Omega\int_0^T \mathrm{e}^{-\lambda(B_1 - B_t)}\,\mathrm{d}t. \tag{L31}$$

Combining the formulas above, we compute

$$\begin{aligned}\tilde{\pi}^\beta \;&=\; \tilde{\pi}^0 + \frac{D_\Sigma}{2\lambda}\left(1 - \int_0^T \mathrm{e}^{-2\lambda(B_1-B_t)}\,\mathrm{d}t\right) - \frac{D_\Omega}{\lambda}\left(1 - \int_0^T \mathrm{e}^{-\lambda(B_1-B_t)}\,\mathrm{d}t\right),\\ B_t \;&=\; \int_0^t \beta_\tau\mathrm{d}\tau, \qquad D_\Sigma \;=\; \frac{1}{N_w+1} + \frac{N_e}{N_e+1}\tilde{\sigma}_e^2, \qquad D_\Omega \;=\; \frac{N_e}{N_e+1}\tilde{\sigma}_e^2.\end{aligned} \tag{L32}$$

If we define

$$\varepsilon_t \;=\; \mathrm{e}^{-\lambda(B_1-B_t)} \;=\; \mathrm{e}^{-\lambda(\int_t^1 \beta_\tau \mathrm{d}\tau)} \;\in\; [0,T], \tag{L33}$$

we compute

$$\begin{aligned}\tilde{\pi}^{\beta} \;&=\; \tilde{\pi}^0 + \frac{1}{\lambda}\int_0^T \big(1-\varepsilon_t\big)\left(D_\Sigma \frac{1+\varepsilon_t}{2} - D_\Omega\right)\mathrm{d}t \\ &=\; \tilde{\pi}^0 + \frac{1}{2\lambda}\int_0^T \big(1-\varepsilon_t\big)\left(\frac{1+\varepsilon_t}{N_w+1} - \frac{(1-\varepsilon_t)N_e\tilde{\sigma}_e^2}{N_e+1}\right)\mathrm{d}t.\end{aligned} \tag{L34}$$

This completes the proof. $\square$

M Discrete Time Models

In this section, we analyze a discrete time version of our model with m lags from Section 2. We denote this discrete time version by $\mathcal{D}_m$, just as in continuous time we denote its counterpart by $\mathcal{M}_m$. The goal of this section is twofold. First, in a general framework we show that optimal linear trading strategies in $\mathcal{D}_m$ must be orthogonal to the public information, as we have assumed for the continuous time model $\mathcal{M}_m$ in Section 2. This is a plausible assumption in continuous time, but we prove it as a result in discrete time.

Second, we study how the discrete time model $\mathcal{D}_m$ compares in the limit to the continuous time model $\mathcal{M}_m$. For that purpose, we analyze the particular case when $m = 1$. In Section M.3, we see that the model $\mathcal{D}_1$ does not converge to its continuous time counterpart $\mathcal{M}_1$, although the difference is very small (see Figure M.1 for numerical results).

We recall that in $\mathcal{M}_m$ the speculators's choice of weights is assumed to have no effect on the covariance structure of the dealer's expected signals (see equation (13)). By contrast, we conjecture that in the continuous time limit of $\mathcal{D}_m$ the speculators take this effect into account.[23] If this conjecture is correct, the results of Section M.3 allow us to analyze the equilibrium effect of changing this assumption. As mentioned before, this effect is very small, and therefore making the assumption (13) is justified.

M.1 General Model

We consider a model similar to the model $\mathcal{M}_m$ with m lags from Section 2, but set in discrete time. For simplicity of presentation, we only consider the case in which speculators can use all their signals, without any restriction on the number of lags. To keep consistent with previous notation, we denote this model by $\mathcal{D}_\infty$. Similar results are true for any $\mathcal{D}_m$, but the equations are more complicated for the general m. The particular case $m = 1$ is presented in detail in Section M.3.

Trading occurs at intervals of length Δt apart, at times $t_1 = \Delta t$, $t_2 = 2\Delta t$, ..., $t_T = T\Delta t$. To simplify notation, we refer to these times as $1, 2, \ldots, T$. The liquidation value of the asset is

$$v_T = \sum_{t=1}^{T} \Delta v_t, \quad \text{with} \quad \Delta v_t = v_t - v_{t-1} = \sigma_v \Delta B_t^v, \tag{M1}$$

[23]Some evidence that our conjecture is correct is the speculator's behavior in the continuous version of Kyle (1985). Indeed, in that model the speculator chooses his optimal weight by taking into account his effect on the covariance matrix $\Sigma_t = \mathsf{Var}(v - p_t)$. In our model, signals are used only for a finite number of lags, and therefore we conjecture that the this effect is not as strong as in Kyle (1985).

where B_t^v is a Brownian motion. The risk-free rate is assumed to be zero.

There are three types of market participants: (a) $N \geq 1$ risk neutral speculators, who observe the flow of information at different speeds, as described below; (b) noise traders; and (c) one competitive risk neutral dealer, who sets the price at which trading takes place.

Information. At $t = 0$, there is no information asymmetry between the speculators and the dealer. Subsequently, each speculator receives the following flow of signals:

$$\Delta s_t \;=\; \Delta v_t + \Delta \eta_t, \quad \text{with} \quad \Delta \eta_t \;=\; \sigma_\eta \mathrm{d}B_t^\eta, \tag{M2}$$

where $t = 1, \ldots, T$ and B_t^η is a Brownian motion independent from all other variables. Define the speculators' forecast by

$$w_t \;=\; \mathsf{E}(v_T \;\big|\; \{s_\tau\}_{\tau \leq t}). \tag{M3}$$

Its increment is $\Delta w_t = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2} \Delta s_t$. Also, $w_0 = 0$.

Speculators obtain their signal with a lag $\ell = 0, 1, \ldots$. A *ℓ-speculator* is a trader who at $t = \ell + 1, \ldots, T$ observes the lagged signal $\Delta s_{t-\ell}$, from ℓ periods before. For each lag $\ell = 0, 1, 2, \ldots$, denote by N_ℓ the number of ℓ-speculators.

Trading. At each $t \in (0, T]$, denote by Δx_t^i the market order submitted by speculator $i = 1, \ldots, N$ at t, and by Δu_t the market order submitted by the noise traders, which is of the form $\Delta u_t = \sigma_u \Delta B_t^u$, where B_t^u is a Brownian motion independent from all other variables. Then, the aggregate order flow executed by the dealer at t is

$$\Delta y_t \;=\; \sum_{i=1}^{N} \Delta x_t^i + \Delta u_t. \tag{M4}$$

Because the dealer is risk neutral and competitive, she executes the order flow at a price equal to her expectation of the liquidation value conditional on her information. Let $\mathcal{I}_t = \{y_\tau\}_{\tau < t}$ be the dealer's information set just before trading at t. Thus, the order flow at date t, Δy_t, executes at:

$$p_t \;=\; \mathsf{E}\big(v_T \mid \mathcal{I}_t \cup \Delta y_t\big). \tag{M5}$$

Equilibrium Definition. A trading strategy for a ℓ-speculator is a process for his position in the risky asset, x_t, measurable with respect to his information set $\mathcal{J}_t^{(\ell)} =$

$\{y_\tau\}_{\tau<t} \cup \{s_\tau\}_{\tau \le t-\ell}$. Denote by E_t^ℓ the expectation of a ℓ-speculator, conditional on $\mathcal{J}_t^{(\ell)}$; and denote by E_t the expectation of the dealer, conditional on the public information:

$$\mathsf{E}_t^\ell(\,\cdot\,) \;=\; \mathsf{E}(\,\cdot \mid \mathcal{J}_t^{(\ell)}), \qquad \mathsf{E}_t(\,\cdot\,) \;=\; \mathsf{E}(\,\cdot \mid \mathcal{I}_t). \tag{M6}$$

For a given trading strategy, the speculator's expected profit π_τ, from date τ onwards, is:

$$\pi_\tau \;=\; \mathsf{E}_t^\ell\left(\sum_{t=\tau}^{T}(v_T - p_t)\Delta x_t\right). \tag{M7}$$

Define the *fresh k-signal* at t to be ($k = 1, \dots, t$)

$$\Delta_t w_k \;=\; \Delta w_k - z_{k,t}, \tag{M8}$$

where $z_{k,t}$ is the dealer's expectation of Δw_k conditional on her information set at t:

$$z_{k,t} \;=\; \mathsf{E}_t\big(\Delta w_k\big). \tag{M9}$$

We focus on trading strategies of the ℓ-speculator which are linear in the fresh signals:

$$\Delta x_t \;=\; \gamma_{1,t}\Delta_t w_1 \;+\; \cdots \;+\; \gamma_{t-\ell,t}\Delta_t w_{t-\ell}. \tag{M10}$$

A linear equilibrium is such that: (i) at every date t, each speculator's trading strategy (12) maximizes his expected trading profit (10) given the dealer's pricing policy, and (ii) the dealer's pricing policy given by (M5) and (M9) is consistent with the equilibrium speculators' trading strategies.

M.2 Equilibrium of the General Model

In this section, we show that the equilibrium of the discrete time model $\mathcal{D}_\infty$ (with no restriction of the number of lagged signals used by the speculators) reduces to a set of equations that the optimal weights and the pricing coefficients must satisfy.

Theorem M.1. *Consider the discrete time model $\mathcal{D}_\infty$, with T trading rounds. Suppose speculators can trade on all their signals. Then, there exists a linear equilibrium of the*

form

$$\begin{aligned}
\Delta x_t^\ell &= \gamma_{1,t}^\ell \Delta_t w_1 + \cdots + \gamma_{t-\ell,t}^\ell \Delta_t w_{t-\ell}, \\
\Delta_t w_j &= \Delta w_j - z_{j,t}, \\
z_{j,t+1} &= z_{j,t} + \rho_{j,t} \Delta y_t, \\
\Delta p_t &= \lambda_t \Delta y_t.
\end{aligned} \tag{M11}$$

if there exist numbers λ_t, $\rho_{i,t}$, $\gamma_{k,t}^\ell$, $A_{i,j,t}$, with $t = 1, \dots, T$, $i, j = 1, \dots, t-1$, $k = 1, \dots, t-\ell$, such that the variables satisfy a system of equations described in Section M.4.

Just like the discrete version of Kyle (1985), the system of equations can only be solved numerically. In the next section, we consider the particular case when speculators trade on signals with lag at most one.

M.3 Fast and Slow Traders in Discrete Time

In this section, we consider the discrete time model $\mathcal{D}_1$, in which speculators can trade only on lagged signal with lag at most $m = 1$. This resembles the model with fast and slow traders from Section 3, but it is set in discrete time. Therefore, we call it the *discrete model with fast and slow traders.*

For this model, we see that the system of equations that arises from Theorem M.1 can be solved numerically, and we verify numerically that the solution converges to a constant solution. Ignoring the initial conditions, a constant solution of the system of equations can be computed by an efficient numerical procedure. Then, we verify numerically that the difference between the model $\mathcal{M}_1$ and the continuous time limit of $\mathcal{D}_1$ is very small (see Figure M.1).

We introduce more notation for the model $\mathcal{D}_1$. Trading takes place at $t = 1, \dots, T$. There are N_F fast traders (FTs) who can trade at t using their current signal Δw_t and their lagged signal Δw_{t-1} (minus its predictable part); and N_S slow traders (STs) who can only use their lagged signals Δw_{t-1}. Denote the unpredictable part of the lagged signal by

$$\widetilde{\Delta w}_{t-1} = \Delta w_{t-1} - \mathsf{E}_t\big(\Delta w_{t-1}\big). \tag{M12}$$

Thus, the FTs and STs have trading strategies of the form:

$$\Delta x_t = \gamma_t \Delta w_t + \mu_t \widetilde{\Delta w}_{t-1}, \tag{M13}$$

where γ_t must be zero for the STs. The dealer sets linear pricing rules of the form:

$$\Delta p_t \;=\; \lambda_t \Delta y_t, \qquad \mathsf{E}_t\big(\Delta w_{t-1}\big) \;=\; \rho_t \Delta y_{t-1}, \tag{M14}$$

where Δy_t is the total order flow at t.

As in continuous time, we simplify notation, and normalize covariances and variances using the tilde notation. For instance,

$$\widetilde{\mathsf{Cov}}(\Delta w_t, \Delta w_t) \;=\; \frac{\mathsf{Cov}(\Delta w_t, \Delta w_t)}{\sigma_w^2 \Delta t} \;=\; 1. \tag{M15}$$

The main result reduces the equilibrium in this model to a system of equations. As in the continuous time case, we denote the total number of traders by N_L:

$$N_L \;=\; N_F + N_S. \tag{M16}$$

This is also the number of "lag traders," i.e., the number of traders that use their lagged signals.

Theorem M.2. *Consider the discrete model with N_F fast traders and N_S slow traders, and let $N_L = N_F + N_S$. Then, the equilibrium reduces to the following system of equations:*

$$\begin{aligned}
a_t \;&=\; N_F \rho_t \gamma_t, \qquad b_t \;=\; N_L \rho_t \mu_t, \qquad R_t \;=\; \frac{\lambda_t}{\rho_t}, \\
a_t \;&=\; \frac{1 - 2\alpha_t \rho_t}{\frac{N_F+1}{N_F} R_t - 2\alpha_t \rho_t}, \qquad b_t \;=\; \frac{C_t}{\frac{N_L+1}{N_L} R_t - 2\alpha_t \rho_t}, \\
C_t \;&=\; \frac{B_t}{1 - a_{t-1}}, \qquad a_t \;=\; a_t^2 + b_t^2 (1 - a_{t-1}) + \rho_t^2 \tilde{\sigma}_u^2, \\
R_t \;&=\; 1 + \frac{b_t B_t}{a_t}, \qquad B_{t+1} \;=\; 1 - a_t - b_t B_t \;=\; 1 - R_t a_t, \\
\alpha_{t-1} \;&=\; b_t^2 \left(\frac{R_t}{N_L^2 \rho_t} + \left(1 - \frac{2}{N_L}\, \alpha_t \right) \right).
\end{aligned} \tag{M17}$$

Proof. See Section M.4. □

The system of equations (M17) can be solved numerically. For all the parameter values we have checked, the solutions are numerically very close to a constant, except when t is either close to 0, or close to T. This suggests that it is a good idea to analyze the behavior of these coefficients when the number of trading rounds becomes large. In this continuous time limit, using Lemma I.1, one expects that all these coefficients

become constant.

Therefore, we consider a constant solution of (M17) with all coefficients constant. For instance, from the recursive equation for B, we have $B = \frac{1-a}{1+b}$, which coincides with the value of B from the continuous time version. We obtain the following equations:

$$\begin{aligned} B \;&=\; \frac{1-a}{1+b}, \quad C \;=\; \frac{1}{1+b}, \quad R \;=\; \frac{a+b}{a(1+b)}, \quad \rho^2\tilde{\sigma}_u^2 \;=\; (a-b^2)(1-a), \\ \alpha \;&=\; \frac{b^2 R}{N_L^2 \rho} + b^2\left(1 - \frac{2}{N_L}\right)\alpha, \quad a \;=\; \frac{1-2\alpha\rho}{\frac{N_F+1}{N_F}R - 2\alpha\rho}, \quad b \;=\; \frac{C}{\frac{N_L+1}{N_L}R - 2\alpha\rho}. \end{aligned} \tag{M18}$$

Solving the equation for α, and multiplying by 2ρ, we obtain

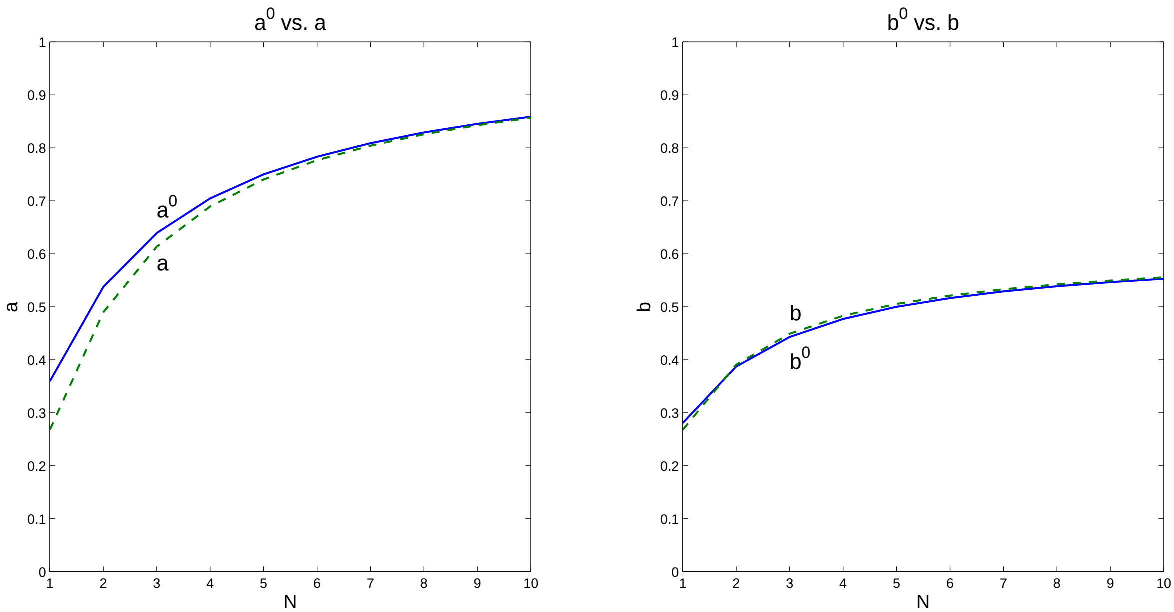
$$2\alpha\rho \;=\; \frac{2b^2 R}{1 - b^2\big(1 - \frac{2}{N_L}\big)}\,\frac{1}{N_L^2}. \tag{M19}$$

Note that the first 4 equations in (M19) coincide with the corresponding ones in the continuous time case. However, the last 2 equations, for γ and μ differ from the continuous time value by the term $2\alpha\rho$. From (M19), and using the fact that all the other terms are of order one, we see that the term $2\alpha\rho$ is of the order of $\frac{1}{N_L^2}$:

$$2\alpha\rho \;=\; O_{N_L}\big(\tfrac{1}{N_L^2}\big). \tag{M20}$$

We now describe a numerical procedure which computes with high accuracy a solution (a, b) of the system above. Denote by a^0 and b^0 the corresponding equilibrium values from Theorem 1, in which the choice of weights does not affect the covariance structure. Then, starting with (a^0, b^0), we compute $R^0 \;=\; \frac{a^0+b^0}{a^0(1+b^0)}$, and then $2\alpha^0\rho^0 \;=\; \frac{2(b^0)^2 R^0}{1-(b^0)^2\left(1-\frac{2}{N_L}\right)}\,\frac{1}{N_L^2}$ using (M19). Using (M18), we recompute the values of (a, b), and denote them (a^1, b^1). Iterate the procedure until it converges. Then, define $(a, b) \;=\; \lim_{n\to\infty}(a^n, b^n)$. Then, (a, b) satisfy the system of equations in (M18). Figure M.1 plots the solution for the case when there are only FTs, and their number is $N \in \{1, \ldots, 10\}$. (The introduction of STs makes the approximation even better, since $N_L = N_F + N_S$ increases.) From the figure, we see that the approximation is good even for low N.

Figure M.1: Comparison of Equilibrium Weights. The figure compares the equilibrium modified weights a and b, which are a solution of the system of equations (M18), with the modified weights a^0 and b^0 from the baseline model, in which the choice of weights does not affect the covariance structure (see Theorem 1). In each model, there are $N = 1, 2, \ldots, 10$ identical speculators.



M.4 Proofs of Results

We first prove a lemma that helps compute the expected profit of a ℓ-speculator. Define the following numbers ($t = 1, \ldots, T$, $i, j = 1, \ldots, t$):

$$A_{i,j,t} \;=\; \mathsf{Cov}_t\big(\Delta w_i, \Delta w_j\big) \;=\; \mathsf{Cov}\big(\Delta_t w_i, \Delta_t w_j\big). \tag{M21}$$

This is called the *fresh covariance* matrix. For a given t, it is a $t \times t$-matrix. We write it in block format, with blocks above and below $t - \ell$:

$$A \;=\; \begin{bmatrix} A^{\ell}_{11,t} & A^{\ell}_{21,t} \\ A^{\ell}_{12,t} & A^{\ell}_{22,t} \end{bmatrix}. \tag{M22}$$

Thus, note that $A^{\ell}_{11,t}$ is $(t-\ell) \times (t-\ell)$ and $A^{\ell}_{12,t}$ is $\ell \times (t-\ell)$.

Lemma M.1. *Consider a ℓ-speculator, and let $k = t - \ell + 1, \ldots, t$. Then,*

$$\mathsf{E}^{\ell}_t\big(\Delta_t w_k\big) \;=\; \sum_{j=1}^{t-\ell} c^{\ell}_{k,j,t} \Delta_t w_j, \tag{M23}$$

where the constants $c^\ell_{k,j,t} \in \mathcal{I}_t$ form a $\ell \times (t-\ell)$-matrix c^ℓ_t which satisfies

$$c^\ell_t \;=\; A^\ell_{12,t}\big(A^\ell_{11,t}\big)^{-1}, \tag{M24}$$

and A is the fresh covariance matrix.

Proof of Lemma M.1. Since $\mathcal{J}^\ell_t = \mathcal{I}_t \cup \Delta w_1 \cup \cdots \cup \Delta w_{t-\ell}$, we write

$$\mathsf{E}^\ell_t\big(\Delta_t w_k\big) \;=\; \sum_{j=1}^{t-\ell} c^\ell_{k,j,t}\Delta_t w_j + h, \quad \text{with} \quad h, c^\ell_{k,j,t} \in \mathcal{I}_t. \tag{M25}$$

Since $\mathsf{E}_t\big(\Delta_t w_j\big) = 0$, if we take $\mathsf{E}_t(\cdot)$ on both sides, by the law of iterated expectations we obtain $0 = h$.

We now write:

$$\Delta_t w_k \;=\; \sum_{i=1}^{t-\ell} c^\ell_{k,i,t}\Delta_t w_i + \varepsilon_k, \tag{M26}$$

with ε_k orthogonal to $\mathcal{J}^\ell_t$. Taking covariance with $\Delta_t w_j$ ($j = 1, \ldots, t-\ell$) on both sides of the equation above, we get

$$A_{k,j,t} \;=\; \sum_{i=1}^{t-\ell} c^\ell_{k,i,t} A_{i,j,t}. \tag{M27}$$

In block format this equation becomes $c^\ell_t = A^\ell_{12,t}\big(A^\ell_{11,t}\big)^{-1}$, which completes the proof.

□

Proof of Theorem M.1. We work backwards, starting from $t = T$.

Speculators' Trading Strategies

Consider the ℓ-speculator's decision in the last trading round, at $t = T$. He takes as given the dealer's linear pricing rule:

$$\Delta p_T \;=\; \lambda_T \Delta y_T, \tag{M28}$$

and the other speculators' trading strategies:

$$\Delta x^k_t \;=\; \gamma^k_{1,t}\Delta_t w_1 + \cdots + \gamma^k_{t-k,t}\Delta_t w_{t-k}, \quad k = 1, \ldots, T-1. \tag{M29}$$

where $\Delta_T w_j = \Delta w_j - z_{j,T}$, and $z_{j,T} = \mathsf{E}_T(\Delta w_j)$. Thus, if he submits Δx at T, he assumes the aggregate order flow at T to be of the form:

$$\begin{aligned} \Delta y_T &= \Delta x + \Delta u_T + \sum_{j=1}^{T} g^{\ell}_{j,T} \Delta_T w_j, \quad \text{with} \\ g^{\ell}_{j,T} &= \sum_{k=0}^{T-j} N_k^{-\ell} \gamma^k_{j,T}, \quad \text{and} \quad N_k^{-\ell} \;=\; N_k - \mathbf{1}_{k=\ell}. \end{aligned} \tag{M30}$$

Then, he computes the expected profit as follows:

$$\pi_T \;=\; \mathsf{E}^{\ell}_T \left(\Big(w_T - p_{T-1} - \lambda_T \Big(\Delta x + \Delta u_T + \sum_{k=1}^{T} g^{\ell}_{k,T} \Delta_T w_k \Big) \Big) \Delta x \right). \tag{M31}$$

Since $p_{T-1} = \mathsf{E}_T(w_T)$, $z_{k,T} = \mathsf{E}_T(\Delta w_k)$, and $w_T = \sum_{k=1}^{T} \Delta w_k$ (recall $w_0 = 0$), we have

$$w_T - p_{T-1} \;=\; \sum_{k=1}^{T} \Delta_T w_k. \tag{M32}$$

From Lemma M.1 above,

$$\begin{aligned} \mathsf{E}^{\ell}_T \big(w_T - p_{T-1} \big) &= \sum_{k=1}^{T-\ell} \Delta_T w_k + \sum_{k=T-\ell+1}^{T} \sum_{j=1}^{T-\ell} c^{\ell}_{k,j,T} \Delta_T w_j \\ &= \sum_{k=1}^{T-\ell} \Delta_T w_k \left(1 + \sum_{j=T-\ell+1}^{T} c^{\ell}_{j,k,T} \right). \end{aligned} \tag{M33}$$

We compute also

$$\begin{aligned} \mathsf{E}^{\ell}_T \left(\sum_{k=1}^{T} g^{\ell}_{k,T} \Delta_T w_k \right) &= \sum_{k=1}^{T-\ell} g^{\ell}_{k,T} \Delta_T w_k + \sum_{k=T-\ell+1}^{T} g^{\ell}_{k,T} \sum_{j=1}^{T-\ell} c^{\ell}_{k,j,T} \Delta_T w_j \\ &= \sum_{k=1}^{T-\ell} \Delta_T w_k \left(g^{\ell}_{k,T} + \sum_{j=T-\ell+1}^{T} g^{\ell}_{j,T} c^{\ell}_{j,k,T} \right). \end{aligned} \tag{M34}$$

Putting together the formulas above, we get

$$\begin{aligned} \mathsf{E}^{\ell}_T \Big(w_T - p_{T-1} - \lambda_T \sum_{k=1}^{T} g^{\ell}_{k,T} \Delta_T w_k \Big) &= \sum_{k=1}^{T-\ell} C^{\ell}_{k,T} \Delta_T w_k, \\ C^{\ell}_{k,T} &= 1 - \lambda_T g^{\ell}_{k,T} + \sum_{j=T-\ell+1}^{T} c^{\ell}_{j,k,T} \big(1 - \lambda_T g^{\ell}_{j,T} \big). \end{aligned} \tag{M35}$$

From (M31) and (M35), we compute

$$\pi_T \;=\; \left(\sum_{k=1}^{T-\ell} C_{k,T}^{\ell} \Delta_T w_k - \lambda_T \Delta x\right) \Delta x. \tag{M36}$$

The first order condition implies

$$\Delta x_T^{\ell} \;=\; \frac{1}{2\lambda_T} \sum_{k=1}^{T-\ell} C_{k,T}^{\ell} \Delta_T w_k, \tag{M37}$$

from which we obtain the equation ($k = 1, \ldots, T-\ell$):

$$\begin{aligned} \gamma_{k,T}^{\ell} \;&=\; \frac{C_{k,T}^{\ell}}{2\lambda_T} \;=\; \frac{1 - \lambda_T g_{k,T}^{\ell} + \sum_{j=T-\ell+1}^{T} c_{j,k,T}^{\ell} \big(1 - \lambda_T g_{j,T}^{\ell}\big)}{2\lambda_T}, \\ \text{with} \quad g_{j,T}^{\ell} \;&=\; \sum_{k=0}^{T-j} N_k^{-\ell} \gamma_{j,T}^{k}. \end{aligned} \tag{M38}$$

The second order condition for a maximum is simply

$$\lambda_T \;>\; 0. \tag{M39}$$

We now compute the value function at T of the ℓ-speculator. This is the maximum expected profit, which corresponds to $\Delta x = \Delta x_T^{\ell}$:

$$\begin{aligned} V_T^{\ell} \;&=\; \frac{1}{4\lambda_T} \left(\sum_{k=1}^{T-\ell} C_{k,T}^{\ell} \Delta_T w_k\right)^2 \;=\; \sum_{i,j=1}^{T-\ell} \alpha_{i,j,T-1}^{\ell} \, \Delta_T w_i \, \Delta_T w_j, \\ \text{with} \quad \alpha_{i,j,T-1}^{\ell} \;&=\; \frac{1}{4\lambda_T} C_{i,T}^{\ell} C_{j,T}^{\ell}. \end{aligned} \tag{M40}$$

Now, we indicate how to find the equations that arise from the ℓ-speculator's optimization at $t < T$. We guess a quadratic value function at $t+1$ of the form

$$V_{t+1}^{\ell} \;=\; \alpha_{0,t} + \sum_{i,j=1}^{t+1-\ell} \alpha_{i,j,t}^{\ell} \, \Delta_{t+1} w_i \, \Delta_{t+1} w_j. \tag{M41}$$

The ℓ-speculator assumes that Δy_t is of the form

$$\Delta y_t \;=\; \Delta x + \Delta u_t + \sum_{j=1}^{t} g_{j,t}^{\ell} \Delta_t w_j, \quad \text{with} \quad g_{j,t}^{\ell} \;=\; \sum_{k=0}^{t-j} N_k^{-\ell} \gamma_{j,t}^{k}. \tag{M42}$$

Using the same method as above, we get the following formula:

$$\mathsf{E}_t^\ell\Big(\big(\Delta_t w_i - \rho_{i,t}\Delta y_t\big)\,\big(\Delta_t w_j - \rho_{j,t}\Delta y_t\big)\Big) \;=\; \rho_{i,t}\rho_{j,t}\Delta x^2 + L_{i,j,t}\Delta x + Q_{i,j,t}, \tag{M43}$$

where as functions of the fresh signals $\Delta_t w_k$ ($k = 1, \ldots, t-\ell$), $L_{i,j,t}$ is linear and $Q_{i,j,t}$ is quadratic. Then, the value function at t satisfies the Bellman equation:

$$\begin{aligned} V_t^\ell \;&=\; \max_{\Delta x} \mathsf{E}_t^\ell\Big(\big(w_t - p_{t-1} - \lambda_t \Delta y_t\big)\Delta x \;+\; V_{t+1}^\ell\Big) \\ &=\; \max_{\Delta x}\left(\sum_{k=1}^{t-\ell} C_{k,t}^\ell \Delta_t w_k - \lambda_t \Delta x\right)\Delta x + \alpha_{0,t} + \sum_{i,j=1}^{t+1-\ell} \alpha_{i,j,t}^\ell\Big(\rho_{i,t}\rho_{j,t}\Delta x^2 + L_{i,j,t}\Delta x + Q_{i,j,t}\Big). \end{aligned} \tag{M44}$$

The first order condition for Δx implies

$$\Delta x_t^\ell \;=\; \frac{\sum_{k=1}^{T-\ell} C_{k,T}^\ell \Delta_T w_k + \sum_{i,j=1}^{t+1-\ell} \alpha_{i,j,t}^\ell L_{i,j,t}}{2\big(\lambda_t - \sum_{i,j=1}^{t+1-\ell}\alpha_{i,j,t}^\ell \rho_{i,t}\rho_{j,t}\big)}, \tag{M45}$$

and the second order condition for a maximum is

$$\lambda_t - \sum_{i,j=1}^{t+1-\ell} \alpha_{i,j,t}^\ell \rho_{i,t}\rho_{j,t} \;>\; 0. \tag{M46}$$

Now we identify the coefficients of Δx_t^ℓ to obtain the equations for the optimal weights $\gamma_{k,t}^\ell$. By substituting $\Delta x = \Delta x_t^\ell$ in the Bellman equation, we get a quadratic formula in the fresh signals, from which we obtain equations for $\alpha_{i,j,t}^\ell$. Furthermore, using similar formulas, one computes the recursive formula for the fresh covariance matrix A. We give more explicit formulas in Section M.2.

Dealer's Pricing Rules

The dealer assumes that the aggregate order flow is of the form:

$$\Delta y_t \;=\; \Delta u_t + \sum_{j=1}^{t} \bar{\gamma}_{j,t}\Delta_t w_j, \quad \text{with} \quad \bar{\gamma}_{j,t} \;=\; \sum_{k=0}^{t-j} N_k \gamma_{j,t}^k. \tag{M47}$$

Because each speculator only trades on the unpredictable part of his signal, Δy_t are orthogonal to each other. Thus, the dealer computes

$$\begin{aligned} z_{k,t} &= \mathsf{E}\big(\Delta w_k \mid \Delta y_1, \ldots, \Delta y_{t-1}\big) = \sum_{j=k}^{t-1} \rho_{k,j} \Delta y_j, \quad k = 0, \ldots, m, \\ \Delta p_t &= \lambda_t \Delta y_t, \end{aligned} \tag{M48}$$

where the coefficients $\rho_{k,t}$ and λ_t are given by:

$$\begin{aligned} \rho_{k,t} &= \frac{\mathsf{Cov}(\Delta w_k, \Delta y_t)}{\mathsf{Var}(\Delta y_t)}, \\ \lambda_t &= \frac{\mathsf{Cov}_t(v_1, \Delta y_t)}{\mathsf{Var}_t(\Delta y_t)} = \frac{\mathsf{Cov}\big(w_t - p_{t-1}, \Delta y_t\big)}{\mathsf{Var}(\Delta y_t)}. \end{aligned} \tag{M49}$$

Since the order flow is orthogonal, we can write

$$\begin{aligned} \rho_{k,t} &= \frac{\mathsf{Cov}(\Delta_t w_k, \Delta y_t)}{\mathsf{Var}(\Delta y_t)} = \frac{\sum_{j=1}^t A_{k,j,t} \bar{\gamma}_{j,t}}{\sum_{i,j=1}^t A_{i,j,t} \bar{\gamma}_{i,t} \bar{\gamma}_{j,t}}, \\ \lambda_t &= \sum_{k=1}^t \frac{\mathsf{Cov}\big(\Delta_t w_k, \Delta y_t\big)}{\mathsf{Var}(\Delta y_t)} = \frac{\sum_{k,j=1}^t A_{k,j,t} \bar{\gamma}_{j,t}}{\sum_{k,j=1}^t A_{k,j,t} \bar{\gamma}_{k,t} \bar{\gamma}_{j,t}}. \end{aligned} \tag{M50}$$

These equations complete the system of equations that must be satisfied in equilibrium. Note that here $\lambda_t = \sum_{k=1}^t \rho_{k,t}$. This is not true if instead of taking $m = T$ we take $m < T$. □

Before proving Theorem M.2, we prove a Lemma that computes the speculators' expected profit.

Lemma M.2. *In the context of Theorem M.2, the FT computes*

$$\mathsf{E}\Big(w_t - p_{t-1} \;\Big|\; \mathcal{I}_t, \Delta w_t, \widetilde{\Delta w}_{t-1}\Big) = \Delta w_t + C_t \widetilde{\Delta w}_{t-1}; \tag{M51}$$

the coefficient C_t is given by

$$C_t = \frac{B_t}{A_t}, \tag{M52}$$

where B_t, D_t, A_t satisfy the following recursive formulas:

$$\begin{aligned} B_{t+1} &= 1 - N_F \lambda_t \gamma_t^* - N_F \rho_t \gamma_t^* - \rho_t (N_F \mu_t^* + N_S \nu_t^*) B_t + \lambda_t \rho_t D_t, \\ D_{t+1} &= (N_F \gamma_{t+1}^*)^2 + (N_F \mu_{t+1}^* + N_S \nu_{t+1}^*)^2 A_{t+1} + \tilde{\sigma}_u^2, \\ A_{t+1} &= 1 - 2 N_F \rho_t \gamma_t^* + \rho_t^2 D_t, \end{aligned} \tag{M53}$$

and γ^*, μ^*, ν^* are the equilibrium values of the corresponding coefficients.

The ST computes

$$\mathsf{E}\Big(w_t - p_{t-1} \,\Big|\, \mathcal{I}_t, \widetilde{\Delta w}_{t-1}\Big) \;=\; C_t \widetilde{\Delta w}_{t-1}, \tag{M54}$$

with C_t as above.

***Proof**.* Since all the variables involved are jointly multivariate normal, the conditional expectation in (M51) is of the form

$$\mathsf{E}\Big(w_t - p_{t-1} \,\big|\, \mathcal{I}_t, \Delta w_t, \widetilde{\Delta w}_{t-1}\Big) \;=\; c_{1,t}\Delta w_t + c_{2,t}\widetilde{\Delta w}_{t-1} + c_{0,t}, \quad \text{with} \quad c_{i,t} \in \mathcal{I}_t. \tag{M55}$$

Because $w_t - p_{t-1}$, Δw_t and $\widetilde{\Delta w}_{t-1}$ are orthogonal to $\mathcal{I}_t$, we have

$$c_{0,t} \;=\; 0, \quad c_{1,t} \;=\; \frac{\mathsf{Cov}\big(w_t - p_{t-1}, \Delta w_t\big)}{\mathsf{Var}\big(\Delta w_t\big)}, \quad c_{2,t} \;=\; \frac{\mathsf{Cov}\big(w_t - p_{t-1}, \widetilde{\Delta w}_{t-1}\big)}{\mathsf{Var}\big(\widetilde{\Delta w}_{t-1}\big)}. \tag{M56}$$

Since $p_{t-1} \in \mathcal{I}_t$, we have $c_{1,t} = 1$. Denote by

$$B_t \;=\; \widetilde{\mathsf{Cov}}\big(w_t - p_{t-1}, \widetilde{\Delta w}_{t-1}\big), \quad D_t \;=\; \widetilde{\mathsf{Var}}\big(\Delta y_t\big), \quad A_t \;=\; \widetilde{\mathsf{Var}}\big(\widetilde{\Delta w}_{t-1}\big). \tag{M57}$$

We now give a recursive formula for

$$C_t \;=\; c_{2,t} \;=\; \frac{B_t}{A_t}. \tag{M58}$$

The aggregate order flow has the form:

$$\Delta y_t \;=\; N_F \gamma_t^* \Delta w_t + (N_F \mu_t^* + N_S \nu_t^*)\widetilde{\Delta w}_{t-1} + \Delta u_t. \tag{M59}$$

Therefore, we compute ($\tilde{\sigma}_u^2 = \frac{\sigma_u^2}{\sigma_w^2}$):

$$\begin{aligned}
A_{t+1} \;&=\; \widetilde{\mathsf{Var}}\big(\Delta w_t - \rho_t \Delta y_t\big) \;=\; 1 - 2N_F \rho_t \gamma_t + \rho_t^2 D_t \\
D_{t+1} \;&=\; (N_F \gamma_{t+1}^*)^2 + (N_F \mu_{t+1}^* + N_S \nu_{t+1}^*)^2 A_{t+1} + \tilde{\sigma}_u^2, \\
B_{t+1} \;&=\; \widetilde{\mathsf{Cov}}\big(w_t - p_{t-1} - \lambda_t \Delta y_t, \Delta w_t - \rho_t \Delta y_t\big) \\
&=\; 1 - \rho_t N_F \gamma_t^* - \rho_t (N_F \mu_t^* + N_S \nu_t^*) B_t - \lambda_t N_F \gamma_t^* + \lambda_t \rho_t D_t.
\end{aligned} \tag{M60}$$

These proves the desired formulas.

For the ST, we have the same computation as for the FT. □

Proof of Theorem M.2. We start by computing the speculators' optimal strategies, taking the dealer's pricing rules as given. Then, we derive the dealer's pricing rules taking the speculator's optimal strategies as given. Finally, we put together the equilibrium conditions to determine the system of equations satisfied by the equilibrium coefficients.

Speculators' Optimal Strategies

We now proceed with computing the FT's value function. Denote by

$$\mathsf{E}_t^F(X) \;=\; \mathsf{E}(X|\mathcal{I}_t, \Delta w_t, \widetilde{\Delta w}_{t-1}), \tag{M61}$$

the expectation from the point of view of the FT at t. Then, the FT's value function at t satisfies the Bellman equation:

$$V_t^F \;=\; \max_{\Delta x}\Big(\mathsf{E}_t^F\big((v_T - p_t)\Delta x\big) \;+\; V_{t+1}^F\Big). \tag{M62}$$

As in the general case, we conjecture a value function for the FT that is quadratic in the current signals:

$$V_t^F \;=\; \alpha_{t-1}^0 + \alpha_{t-1}\big(\widetilde{\Delta w}_{t-1}\big)^2 + 2\alpha_{t-1}'\big(\widetilde{\Delta w}_{t-1}\big)\big(\Delta w_t\big) + \alpha_{t-1}''\big(\Delta w_t\big)^2. \tag{M63}$$

Then, the Bellman equation becomes

$$\begin{aligned} V_t^F \;=\; \max_{\Delta x} \mathsf{E}_t^F \Bigg(&\big(w_t - p_{t-1} - \lambda_t \Delta y_t\big)\Delta x \\ &+ \alpha_t^0 + \alpha_t\big(\Delta w_t - \rho_t \Delta y_t\big)^2 + 2\alpha_t'\big(\Delta w_t - \rho_t \Delta y_t\big)\big(\Delta w_{t+1}\big) + \alpha_t''\big(\Delta w_{t+1}\big)^2\Bigg), \end{aligned} \tag{M64}$$

where Δy_t is assumed by the FT to be of the form

$$\Delta y_t \;=\; \Delta x + (N_F - 1)\gamma_t^* \Delta w_t + \big((N_F - 1)\mu_t^* + N_S \nu_t^*\big)\widetilde{\Delta w}_{t-1} + \Delta u_t. \tag{M65}$$

From equation (M51), we compute $\mathsf{E}_t^F\big(w_t - p_{t-1}\big) = \Delta w_t + C_t \widetilde{\Delta w}_{t-1}$, with C_t satisfying certain equations described in Lemma M.2 above. Therefore,

$$\begin{aligned} V_t^F \;=\; \max_{\Delta x} \mathsf{E}_t^F \Bigg(&\big(\Delta w_t + C_t \widetilde{\Delta w}_{t-1} - \lambda_t \Delta y_t\big)\Delta x + \\ &\alpha_t^0 + \alpha_t\big(\Delta w_t - \rho_t \Delta y_t\big)^2 + \alpha_t'' \sigma_w^2 \Delta t\Bigg), \end{aligned} \tag{M66}$$

The terms can be rearranged to write

$$V_t^F \;=\; \max_{\Delta x}\big(W_1 - \lambda_t \Delta x\big) + \alpha_t\big(W_2 - \rho_t \Delta x)^2 + Z, \tag{M67}$$

where

$$\begin{aligned}
W_1 \;&=\; W_{11}\Delta w_t + W_{12}\widetilde{\Delta w}_{t-1}, \quad \text{with} \\
&\qquad W_{11} \;=\; 1-(N_F-1)\lambda_t\gamma_t^*, \qquad W_{12} \;=\; C_t - \lambda_t\big((N_F-1)\mu_t^* + N_S\nu_t^*\big), \\
W_2 \;&=\; W_{21}\Delta w_t + W_{22}\widetilde{\Delta w}_{t-1}, \quad \text{with} \\
&\qquad W_{21} \;=\; 1-(N_F-1)\rho_t\gamma_t^*, \qquad W_{22} \;=\; -\rho_t\big((N_F-1)\mu_t^* + N_S\nu_t^*\big), \\
Z \;&=\; \alpha_t^0 + \alpha_t''\sigma_w^2\Delta t + \alpha_t\sigma_u^2\Delta t.
\end{aligned} \tag{M68}$$

The first order condition with respect to Δx is

$$W_1 - 2\lambda_t\Delta x - 2\alpha_t\rho_t(W_2 - \rho_t\Delta x) \;=\; 0. \tag{M69}$$

If we denote by

$$\tilde{\lambda}_t \;=\; \lambda_t - \alpha_t\rho_t^2, \tag{M70}$$

the first order condition implies

$$\Delta x \;=\; \frac{W_1 - 2\alpha_t\rho_t W_2}{2\tilde{\lambda}_t} \;=\; \frac{W_{11} - 2\alpha_t\rho_t W_{21}}{2\tilde{\lambda}_t}\Delta w_t + \frac{W_{21} - 2\alpha_t\rho_t W_{22}}{2\tilde{\lambda}_t}\widetilde{\Delta w}_{t-1}. \tag{M71}$$

The second order condition for a maximum is

$$\tilde{\lambda}_t \;>\; 0. \tag{M72}$$

By identifying the coefficients of V_t^F, we obtain

$$\begin{aligned}
\alpha_{t-1}^0 \;&=\; Z, \\
\alpha_{t-1} \;&=\; \frac{\big(W_{12} - 2\alpha_t\rho_t W_{22}\big)^2}{4\tilde{\lambda}_t} + \alpha_t W_{22}^2, \\
\alpha_{t-1}' \;&=\; \frac{\big(W_{11} - 2\alpha_t\rho_t W_{21}\big)\big(W_{12} - 2\alpha_t\rho_t W_{22}\big)}{4\tilde{\lambda}_t} + \alpha_t W_{11}W_{21}, \\
\alpha_{t-1}'' \;&=\; \frac{\big(W_{11} - 2\alpha_t\rho_t W_{21}\big)^2}{4\tilde{\lambda}_t} + \alpha_t W_{11}^2.
\end{aligned} \tag{M73}$$

Note that only α_t is involved in a recursive equation, while all the other coefficients can be computed using α_t (and the equilibrium coefficients). We write the equation for α_t more explicitly:

$$\alpha_{t-1} \;=\; \frac{\Big(C_t - \big(\lambda_t - 2\alpha_t\rho_t^2\big)\big((N_F-1)\mu_t^* + N_S\nu_t^*\big)\Big)^2}{4\tilde{\lambda}_t} + \alpha_t\rho_t^2\big((N_F-1)\mu_t^* + N_S\nu_t^*\big)^2. \tag{M74}$$

From (M71), we obtain the equations for the coefficients γ_t and μ_t for the FT:

$$\begin{aligned}\gamma_t \;&=\; \frac{1 - 2\alpha_t\rho_t - \big(\lambda_t - 2\alpha_t\rho_t^2\big)(N_F-1)\gamma_t^*}{2\tilde{\lambda}_t},\\ \mu_t \;&=\; \frac{C_t - \big(\lambda_t - 2\alpha_t\rho_t^2\big)\big((N_F-1)\mu_t^* + N_S\nu_t^*\big)}{2\tilde{\lambda}_t}.\end{aligned} \tag{M75}$$

We now proceed in a similar way to compute the ST's value function. Denote by $\mathsf{E}_t^S(X) = \mathsf{E}(X|\mathcal{I}_t, \widetilde{\Delta w}_{t-1})$, the expectation from the point of view of the ST at t. Then, the ST's value function at t satisfies the Bellman equation:

$$V_t^S \;=\; \max_{\Delta x}\Big(\mathsf{E}_t^S\big((v_T - p_t)\Delta x\big) \;+\; V_{t+1}^S\Big). \tag{M76}$$

We conjecture a value function for the ST that is quadratic in the current signal:

$$V_t^F \;=\; \beta_{t-1}^0 + \beta_{t-1}\big(\widetilde{\Delta w}_{t-1}\big)^2. \tag{M77}$$

With a similar computation as for the FT, β_t satisfies the recursive equation

$$\beta_{t-1} \;=\; \frac{\Big(C_t - \big(\lambda_t - 2\beta_t\rho_t^2\big)\big(N_F\mu_t^* + (N_S-1)\nu_t^*\big)\Big)^2}{4\lambda_t'} + \beta_t\rho_t^2\big(N_F\mu_t^* + (N_S-1)\nu_t^*\big)^2, \tag{M78}$$

where $\lambda_t' = \lambda_t - \beta_t\rho_t^2$. We also obtain

$$\nu_t \;=\; \frac{C_t - \big(\lambda_t - 2\beta_t\rho_t^2\big)\big(N_F\mu_t^* + (N_S-1)\nu_t^*\big)}{2\lambda_t'}. \tag{M79}$$

Note that if $\mu_t = \nu_t$, α_t and β_t satisfy the same equation. Thus, we can take

$$\mu_t \;=\; \nu_t, \qquad \alpha_t \;=\; \beta_t. \tag{M80}$$

Thus, we look for a symmetric equilibrium in which $\gamma_t = \gamma_t^*$, $\mu_t = \nu_t = \mu_t^* = \nu_t^*$, and

$\alpha_t = \beta_t$. We get the following equations

$$
\begin{aligned}
\gamma_t &= \frac{1 - 2\alpha_t \rho_t}{(N_F + 1)\lambda_t - 2N_F \alpha_t \rho_t^2}, \\
\mu_t &= \frac{C_t}{(N_L + 1)\lambda_t - 2N_L \alpha_t \rho_t^2}, \\
\alpha_{t-1} &= \mu_t^2 \Big(\lambda_t + (N_L^2 - 2N_L)\alpha_t \rho_t^2 \Big).
\end{aligned}
\tag{M81}
$$

Dealer's Pricing Rules

The dealer takes the speculator's strategies as given, which means that she assumes the aggreate order flow to be of the form

$$
\Delta y_t = N_F \gamma_t \Delta w_t + N_L \mu_t \widetilde{\Delta w}_{t-1} + \mathrm{d}u_t. \tag{M82}
$$

Therefore, she sets λ_t and ρ_t according to the usual formulas:

$$
\begin{aligned}
\lambda_t &= \frac{\mathsf{Cov}_t(v_T, \Delta y_t)}{\mathsf{Var}_t(\Delta y_t)} = \frac{\widetilde{\mathsf{Cov}}\big(w_t - p_{t-1}, \Delta y_t\big)}{\widetilde{\mathsf{Var}}(\Delta y_t)} = \frac{N_F \gamma_t + N_L \mu_t C_t}{D_t}, \\
\rho_t &= \frac{\mathsf{Cov}_t(\Delta v_t, \Delta y_t)}{\mathsf{Var}_t(\Delta y_t)} = \frac{\widetilde{\mathsf{Cov}}\big(\Delta w_t, \Delta y_t\big)}{\widetilde{\mathsf{Var}}(\Delta y_t)} = \frac{N_F \gamma_t}{D_t}.
\end{aligned}
\tag{M83}
$$

We now rewrite the equations from Lemma M.2 above, using the equation we derived above: $\rho_t D_t = N_F \gamma_t$:

$$
\begin{aligned}
B_{t+1} &= 1 - N_F \rho_t \gamma_t - N_L \rho_t \mu_t B_t, \\
A_{t+1} &= 1 - N_F \rho_t \gamma_t, \\
D_{t+1} &= (N_F \gamma_{t+1})^2 + (N_L \mu_{t+1})^2 (1 - N_F \rho_t \gamma_t) + \tilde{\sigma}_u^2, \\
C_t &= \frac{B_t}{A_t} = \frac{B_t}{1 - N_F \rho_{t-1} \gamma_{t-1}}.
\end{aligned}
\tag{M84}
$$

Equilibrium Conditions

Following the continuous time version of the model, we define the following variables:

$$
a_t = N_F \rho_t \gamma_t, \qquad b_t = N_L \rho_t \mu_t, \qquad R_t = \frac{\lambda_t}{\rho_t}. \tag{M85}
$$

From $\rho D_t = N_F \gamma_t$, we obtain $N_F \rho_t \gamma_t = \rho_t^2 D_t = (N_F \rho_t \gamma_t)^2 + (N_L \rho_t \mu_t)^2 (1 - N_F \rho_{t-1} \gamma_{t-1}) + \rho^2 \tilde{\sigma}_u^2$. With the new notation, this equation becomes

$$a_t \;=\; a_t^2 + b_t^2(1 - a_{t-1}) + \rho_t^2 \tilde{\sigma}_u^2. \tag{M86}$$

Also, we compute

$$R_t \;=\; \frac{\lambda_t}{\rho_t} \;=\; \frac{a_t + b_t B_t}{\rho^2 D_t} \;=\; \frac{a_t + b_t B_t}{a_t} \;=\; 1 + \frac{b_t B_t}{a_t}. \tag{M87}$$

We put together the equations that determine the equilibrium:

$$\begin{aligned}
a_t &= N_F \rho_t \gamma_t, \qquad b_t = N_L \rho_t \mu_t, \qquad R_t = \frac{\lambda_t}{\rho_t}, \\
a_t &= \frac{1 - 2\alpha_t \rho_t}{\frac{N_F+1}{N_F} R_t - 2\alpha_t \rho_t}, \qquad b_t = \frac{C_t}{\frac{N_L+1}{N_L} R_t - 2\alpha_t \rho_t}, \\
C_t &= \frac{B_t}{1 - a_{t-1}}, \qquad a_t = a_t^2 + b_t^2(1 - a_{t-1}) + \rho_t^2 \tilde{\sigma}_u^2, \\
R_t &= 1 + \frac{b_t B_t}{a_t}, \qquad B_{t+1} = 1 - a_t - b_t B_t = 1 - R_t a_t, \\
\alpha_{t-1} &= b_t^2 \left(\frac{R_t}{N_L^2 \rho_t} + \left(1 - \frac{2}{N_L}\, \alpha_t \right) \right).
\end{aligned} \tag{M88}$$

This proves (M17). $\square$