Segmented Trading Markets*

Kerry Back, Rice University **Oğuzhan Çelebi,** Stanford University
Ali Kakhbod, UC Berkeley **A. Max Reppen,** Boston University

November 2023

Abstract

We study competition and endogenous fragmentation among heterogenous trading venues that differ in technology (fast vs. slow), where traders can dynamically choose which venue to trade in. We show that technological improvements increase trading speed, but may also heighten differentiation, which reduces competition, leads to higher trading fees, and potentially reduces trading volume and welfare. Improvements in the slower venue lead to increased trading speed, decreased differentiation, and thus increased trading volume and welfare. Conversely, the effect of improvements in the faster venue is generally ambiguous and depends on the extent of traders' patience, the frequency of their preference shocks, and the competition between venue owners. We further study the effect of technological improvement in one of the venues when both initially have the same trading speed. We find that if the trading speeds are initially slow enough, the technological improvement will increase trading volume and trader welfare. Conversely, if the trading speeds are initially fast, the increase in trading fees outweighs the speed advantage that comes with technological improvement, leading to decreased trading volume and trader welfare.

Keywords: OTC vs Exchange. Heterogeneous venues. Fragmentation. Competition.
JEL Classification: G12, G2, D4, L13.

*First Draft: April 2022. This version November 2023. We thank Will Cong, Amir Kermani, Giacomo Lanzani, Kevin Li, Dmitry Livdan, Andrey Malenko, Christine Parlour, Marzena Rostek, Farzad Saidi, Tarik Umar, Pierre-Olivier Weill, Hao Xing and numerous seminar and conference participants for helpful comments. Kerry Back is with the Jesse H. Jones School of Business at Rice University. Oğuzhan Çelebi is with the Department of Economics at Stanford University. Ali Kakhbod is with the Haas School of Business at UC Berkeley. A. Max Reppen is with the Questrom School of Business at Boston University. Contact: Kerry Back (kerry.e.back@rice.edu), Oğuzhan Çelebi (ocelebi@stanford.edu), Ali Kakhbod (akakhbod@berkeley.edu), A. Max Reppen (amreppen@bu.edu).

1 Introduction

In the last decade, there has been a notable shift in trading landscapes. With significant fragmentation and heterogeneity spanning trading venues, we have observed varying market behaviors. Differences in execution speed, trading fees, and even asset prices emerge across these venues (Biais, Hombert and Weill (2021); Chen and Duffie (2021); Pagnotta and Philippon (2018)). As traders decide between fast and slow trading venues over time, the decision is not just about the speed but also the associated costs. With evolving technology, how do trading fees adjust? What implications do these technological shifts have on trading volume? Does faster invariably equate to higher welfare? This paper studies these questions. We probe the competition between trading venues that strategically set their trading fees while differentiating based on technology. Our aim is to decipher how such technological variances and strategic decisions shape disparities across marketplaces. Notably, the exploration of a segmented market, especially one where investors dynamically choose between technologically diverse vendors, is, to our knowledge, new to the literature.

In this context, we study dynamic interactions between traders and trading venues using a model in which venues with different technologies compete on fees. The technology of the venue is referred to and measured as *speed*. Speed is understood here in a general sense to comprise not only trade execution latency but also convenience and reliability properties such as user interfaces, data feeds, etc. In short, speed encompasses any factors that impact the time between the decision to trade and when the asset changes hands.

Investors over time observe random shocks to their marginal utility of holding the asset. Trading takes place in two venues: one slow and one fast. An investor with high marginal holding utility is a natural buyer, whereas those with low marginal utility are natural sellers. These diverse incentives give rise to different speed demands, and traders choose whether to trade in the fast or slow market accordingly. Venues maximize their profit from trading, which is generated by the demand of traders to maximize their utility from holding the asset. The choice of venue is dynamic, and traders may update their choice at any time.

We first characterize when: (i) no trading occurs, (ii) the market is segmented (fragmented), meaning that both venues are active, and (iii) traders lack a strong enough preference for speed to engage in the fast venue. This result serves as our primary building block to examine how speed differentiation influences fee competition, trading volumes, and welfare, the central focus of our study.

How does speed differentiation affect fee competition? We allow venues to strategically set their fees and characterize the equilibrium trading fees. In equilibrium, the slow venue undercuts the faster venue, both venues are always active, and the market is always segmented. Moreover, when the differentiation between the two venues decreases (i.e., the slower venue becomes faster or the faster venue becomes slower), the trading fees in both venues decrease. Conversely, when the differentiation between the two venues increases (i.e., the slower venue becomes slower or the faster venue becomes faster), the trading fees in both venues increase. The intuition is simple: with more speed differentiation, venues generate their own market and move away from the zero-fee homogeneous case.

When venues compete by transaction fees, how does speed differentiation affect trading volume? First, a change in the transaction speed of a venue directly affects the instantaneous trading volume in that venue. Second, it affects the fees charged by the venues in equilibrium and thus the market structure itself. Both effects are positive for the increase in speed in the slower venue, and total trading volume increases with the speed of the slower venue. However, the effect of an increase in the speed of the fast venue is ambiguous. This is because faster trading speeds in the fast venue have two effects: a direct effect that increases trading volume and an indirect effect that increases differentiation, which in turn increases fees charged in equilibrium and decreases trading volume. In fact, in this case, the outcome depends on the extent of traders' patience, the frequency of their preference shocks, and the competition between the venues.

To understand the effect of the fast venue speed on trading volume, we consider the special case of *full competition*, where the speed difference between the venues is arbitrarily small. We find that the effect of this improvement depends on the rate of the preference shock, the discount factor, and the initial trading speed. If the traders are patient relative to the rate of the preference shock, the value created by trade is higher for the traders, so the higher fees that come with differentiation cause less distortion, and trading volume increases due to the improvement in transaction speed. Conversely, if the traders are impatient, then the effect of the improvement depends on the initial speed: trading volume increases after an improvement if and only if the initial trading speed is low enough. Thus, if both venues are already fast enough, the negative effect of competition dominates the positive effect of the speed improvement, while the converse is true when both venues are initially slow.

Finally, we study how speed differentiation affects welfare. We focus on the full competition benchmark and consider two different measures of welfare. First, we study total welfare, which includes the profits of the venues as well as the utility of traders. When profits are included in the calculation, higher fees caused by differentiation are not an issue, and a speed improvement always increases welfare. Second, we consider only the traders' welfare, without including the venue profits. As speed improvement of a venue creates differentiation and increases fees paid by the traders, the effect of an improvement is ambiguous and echoes our result on trading volume: trader welfare increases after an improvement if and only if the initial trading speed is low enough. These two results emphasize how the effect of a speed improvement on different statistics may depend on the relative initial speeds of the venues.

A major twist in our model and a point of departure from the existing literature is that we study competition and endogenous fragmentation in segmented and technologically diverse venues where traders can dynamically choose between the venues at any time. Our paper contributes to three major themes in the literature. First, it is related to the literature on market fragmentation. This literature is vast (see, e.g., Mendelson (1987); Pagano (1989); Malamud and Rostek (2017); Chen and

Duffie (2021); Biais, Hombert and Weill (2021)).[1] For example, Mendelson (1987) studied the tradeoff between market consolidation and fragmentation. Malamud and Rostek (2017) and Chen and Duffie (2021) assume each exchange operates a double auction and focus on welfare and allocative efficiency, respectively. Biais, Hombert and Weill (2021) obtain endogenous fragmentation of asset markets with heterogeneous agents in a dynamic exchange economy. Based on our model, we demonstrate how fragmentation can arise endogenously due to transaction fees, competition, costs, and speeds in heterogeneous venues.

Second, our paper contributes to a growing body of literature on competition between multiple venues, which has various focuses. For instance, Santos and Scheinkman (2001) consider competition in margin requirements, while Parlour and Seppi (2003) model competition for order flow between exchanges based on liquidity provision, and Foucault and Parlour (2004) examine competition in IPO listings. Pagnotta and Philippon (2018) consider a dynamic subscription model to study competition in heterogeneous venues in which each trader initially chooses a venue and stays there forever. Based on our model, we demonstrate that technological improvements increase trading speed but may also increase differentiation, which reduces competition, leads to higher trading fees, potentially reducing trading volume and welfare; in fact, the outcome depends on the extent of traders' patience, the frequency of their preference shocks, and the competition between venue owners.

Third, the dynamic trading framework of our paper is related to the literature on integrated (equity) markets (e.g., Kyle (1985*a*,*b*); Back (1992); Wang (1994); Back, Cao and Willard (2000)). Similar to these works, we also present a fully dynamic trading model. However, in contrast to these papers, we instead focus on the impacts of having multiple venues with different technologies on competition, trading volume, transaction fees, and welfare.

The rest of the paper is organized as follows. Section 2 presents the model. In Section 3, we analyze the decision of the traders for a given market structure

[1] See Rostek and Yoon (2020) for an excellent review on market fragmentation.

(taking the trading speed and transaction fees as given) and characterize the trading equilibrium with two venues. Section 4 characterizes the trading volume under this trading equilibrium. Section 5 analyzes how venues with different trading speeds compete in transaction fees to maximize their profit. In Section 6, we analyze the effect of differentiation on welfare. The paper concludes with Section 7. All proofs are provided in the Appendix.

2 Model

We consider a continuous time model of a unit measure of traders with time-discount factor $\rho > 0$ and a long-lived indivisible asset with supply $Z \in (0, 1)$. Traders have unit demand, and a trader who owns the asset is called an *owner*, whereas a trader who does not is called a *non-owner*. Traders have an individual desire $\eta \in [\eta_l, \eta_h]$ to hold the asset, which changes over time.

The instantaneous utility of an owner with value η is $u(\eta)$ (which is increasing in η), while a non-owner gains zero instantaneous utility (regardless of their desire to hold). The value η changes independently across traders at exponential inter-arrival times with rate γ. Conditional on the arrival of a shock, the new desire to hold takes a value in $[\eta_l, \eta_h]$ according to the cumulative distribution function $F(\cdot)$, with the probability distribution function $f(\cdot)$.

At any given time, each owner decides whether to sell or keep holding the asset, and each non-owner decides whether to buy the asset. If a trader decides to transact, she must also decide in which venue to trade.

The two venues are indexed by $v \in s, f$ and are differentiated by their transaction speeds σ_v and transaction fees c_v. The speed is modeled by letting σ_v represent the rate (of exponential random times) at which trades are executed. As the index suggests, we assume that venue f is faster, i.e., $\sigma_f \geqslant \sigma_s$.[2] In addition to transaction fees, traders may incur a fixed cost $\theta \geqslant 0$ per trade. Unlike the venue-specific fees

[2]When transactions fees c_v are endogenized, this has natural consequences for the equilibrium fees, because the slower venue must undercut the faster in order to stay competitive.

c_s and c_f, which venues may strategically set (c.f. Section 5), θ is an uncontrolled cost common to both venues. We interpret $\theta \geqslant 0$ as a fixed cost independent of venues, covering expenses such as transaction costs (excluding the venue fees), regulatory taxes, gas fees on a crypto asset's blockchain network, and so on.

3 Analysis

3.1 Traders' Decisions

We first characterize the behavior of traders, given the speed and transaction fees of the venues. Let $m \in \{no, o\}$ denote a trader's position, where $m = no$ represents a non-owner and $m = o$ an owner. At each time t, an owner either Sells (S) or Holds (H), and a non-owner Buys (B) or does Nothing (N). The action set of a trader is thus $\{H, S_s, S_f\}$ if $m = o$, and $\{N, B_s, B_f\}$ if $m = no$, where the subscripts denote the venue choice for trading.

We focus on the stationary equilibrium and let p_ν denote the equilibrium asset price in venue ν. Let $V_m(\eta)$ denote the expected payoff of a trader with current trading position $m \in \{no, o\}$ and desire to hold $\eta \in [\eta_l, \eta_h]$.

To define the problem, consider three independent and exponential times: τ_γ, τ_s, and τ_f, which respectively denote the arrival of a preference shock (rate γ) or the execution of a trade on the slow or fast market (rates σ_s and σ_f). Define

$$\tau = \begin{cases} \tau_\gamma & \text{if } A = H \text{ or } A = N, \\ \tau_\gamma \wedge \tau_\nu & \text{if } A = S_\nu \text{ or } A = B_\nu, \text{ for } \nu \in \{s, f\}. \end{cases}$$

The optimization problem for an asset owner is then

$$V_o(\eta) = \max_{A \in \{H, S_s, S_f\}} \mathbb{E}\Bigg[\int_0^\tau e^{-\rho t} u(\eta)\, \mathrm{d}t + 1_{\{\tau=\tau_\gamma\}} e^{-\rho\tau} \int_{\eta_l}^{\eta_h} V_o(\eta')\, \mathrm{d}F(\eta') \\ + \sum_{\nu \in \{s,f\}} 1_{\{\tau=\tau_\nu\}} e^{-\rho\tau} \Big(V_{no}(\eta) + p_\nu - c_\nu - \theta \Big) \Bigg], \tag{1}$$

and, similarly, for a non-owner the problem is

$$V_{no}(\eta) = \max_{A \in \{N, B_s, B_f\}} \mathbb{E}\left[1_{\{\tau=\tau_\gamma\}} e^{-\rho\tau} \int_{\eta_l}^{\eta_h} V_{no}(\eta')\, \mathrm{d}F(\eta') \right.$$
$$\left. + \sum_{\nu \in \{s,f\}} 1_{\{\tau=\tau_\nu\}} e^{-\rho\tau} \Big(V_o(\eta) - p_\nu - c_\nu - \theta \Big) \right].$$

As there are only three available actions, we define the following auxiliary functions for each of them, in order to write down the Hamilton–Jacobi–Bellman (HJB) equations:

$$\begin{aligned}
\rho V_H(\eta) &= \underbrace{u(\eta)}_{\text{flow gain of holding the asset}} + \underbrace{\gamma \Big(\mathbb{E}_{\eta'}[V_o(\eta')] - V_H(\eta) \Big)}_{\text{net gain of the pref. shock}}, \\
\rho V_S^s(\eta) &= u(\eta) + \gamma \Big(\mathbb{E}_{\eta'}[V_o(\eta')] - V_S^s(\eta) \Big) + \sigma_s \underbrace{\Big(V_{no}(\eta) + p_s - c_s - \theta - V_S^s(\eta) \Big)}_{\text{net gain of selling in the slow venue}}, \\
\rho V_S^f(\eta) &= u(\eta) + \gamma \Big(\mathbb{E}_{\eta'}[V_o(\eta')] - V_S^f(\eta) \Big) + \sigma_f \underbrace{\Big(V_{no}(\eta) + p_f - c_f - \theta - V_S^f(\eta) \Big)}_{\text{net gain of selling in the fast venue}}.
\end{aligned} \tag{2}$$

For a trader who currently owns the asset, $V_H(\eta)$ denotes the value of continuing to hold for the next instant, and $V_S^\nu(\eta)$, for $\nu \in \{s, f\}$, denotes the value of intending to sell the asset in venue ν. Since each of equation in (2) represents (1) for a fixed action, it is natural that

$$V_o(\eta) = \max\{V_H(\eta), V_S^s(\eta), V_S^f(\eta)\}.$$

We similarly define

$$
\begin{aligned}
\rho V_N(\eta) &= \gamma \underbrace{\left(\mathbb{E}_{\eta'}[V_{no}(\eta')] - V_N(\eta)\right)}_{\text{net gain of the pref. shock}}, \\
\rho V_B^s(\eta) &= \gamma \left(\mathbb{E}_{\eta'}[V_{no}(\eta')] - V_B^s(\eta)\right) + \sigma_s \underbrace{\left(V_o(\eta) - p_s - c_s - \theta - V_B^s(\eta)\right)}_{\text{net gain of buying in the slow venue}}, \\
\rho V_B^f(\eta) &= \gamma \left(\mathbb{E}_{\eta'}[V_{no}(\eta')] - V_B^f(\eta)\right) + \sigma_f \underbrace{\left(V_o(\eta) - p_f - c_f - \theta - V_B^f(\eta)\right)}_{\text{net gain of buying in the fast venue}}.
\end{aligned} \tag{3}
$$

For a trader who currently does not own the asset, $V_N(\eta)$ is the value of doing nothing and $V_S^\nu(\eta)$, for $\nu \in \{s, f\}$ is the value of buying the asset in venue ν. Like for V_o, we have that

$$V_{no}(\eta) = \max\{V_N(\eta), V_B^s(\eta), V_B^f(\eta)\}.$$

Finally, with some abuse of notation, let N, B_s, B_f, H, S_s, and S_f denote the sets of agents for whom these respective actions are optimal, i.e., $\eta \in N$ if and only if $V_{no}(\eta) = V_N(\eta)$; $\eta \in B_s$ if and only if $V_{no}(\eta) = V_B^s(\eta)$, etc.

We next define the stationary equilibrium, for which we need to characterize the stationary distribution of η of both owners and non-owners, whose densities we denote as $f_o(\eta)$ and $f_{no}(\eta)$.

Definition 1. *A stationary equilibrium consists of the sets N, B_s, B_f, H, S_s, and S_f; prices p_s and p_f; and f_o and f_{no} such that:*

- $f_o(\eta) + f_{no}(\eta) = f(\eta)$.
- *Traders behave optimally.*

- *Asset market clears:*

$$\int_{\eta_l}^{\eta_h} f_o(\eta)d\eta = Z. \tag{4}$$

- *Fast venue clears:*

$$\int_{B_f} f_{no}(\eta)d\eta = \int_{S_f} f_o(\eta)d\eta. \tag{5}$$

- *Slow venue clears:*

$$\int_{B_s} f_{no}(\eta)d\eta = \int_{S_s} f_o(\eta)d\eta. \tag{6}$$

For the rest of the paper, we make the following normalization:

Assumption 1. *$\eta_l = 0$ and $u(0) = 0$.*

We can now characterize the equilibrium. There are three main cases: *no trade*, where no trader buys or sells the asset; *market segmentation* (fragmentation), where both venues are active; and *no segmentation*, where only the slow (i.e., cheaper) venue is active.

Intuitively, the first case is obtained when the transaction fees, c_s, c_f, and/or the fixed cost, θ, are prohibitively high, so that trading is never profitable, and the third case occurs when the speed advantage of the fast venue is small relative to the difference in transaction fees. In Theorem 1, we derive conditions under which each of these three cases occurs and characterize the resulting equilibria.

Let us define the following function, which parameterizes the relative advantage the fast venue enjoys over the slow venue, as it will be useful in the following theorem statement:

$$g(\sigma_f, \sigma_s, c_s, c_f, \theta) = \frac{\sigma_f(\sigma_s + \gamma + \rho)(c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(c_s + \theta)}{\sigma_f - \sigma_s}. \tag{7}$$

The function g increases whenever the fast venue becomes better for traders (when σ_f increases or c_f decreases) and decreases whenever the slow venue becomes better for traders (when σ_s increases or c_s decreases).

Theorem 1. *For any $c_s < c_f$, $\sigma_s < \sigma_f$, θ, Z, γ, ρ, and u, the following hold.*

(i) **No trade.** *If $u(\eta_h) \leqslant 2(c_s + \theta)(\gamma + \rho)$, then there is no equilibrium at which a positive measure of traders trade.*

(ii) **Market segmentation (fragmentation).** *If $u(\eta_h) > 2g(\sigma_f, \sigma_s, c_s, c_f, \theta)$, then a positive measure of traders trade in both venues. In particular, the traders' actions (depending on their desire to hold η) are uniquely characterized by the following intervals:*

$$N = [\eta_l, \eta_1],\ B_s = [\eta_1, \eta_2],\ B_f = [\eta_2, \eta_h],$$
$$S_f = [\eta_l, \eta_3],\ S_s = [\eta_3, \eta_4],\ H = [\eta_4, \eta_h],$$

where the equilibrium cutoffs η_1, η_2, η_3, and η_4 satisfy $\eta_l < \eta_3 < \eta_4 < \eta_1 < \eta_2 < \eta_h$, and are uniquely pinned down by the following equations:

$$(1 - Z)F(\eta_1) + ZF(\eta_4) = 1 - Z, \tag{8}$$

$$(1 - Z)F(\eta_2) + ZF(\eta_3) = 1 - Z, \tag{9}$$

$$u(\eta_1) - u(\eta_4) = 2(\gamma + \rho)(c_s + \theta), \tag{10}$$

$$u(\eta_2) - u(\eta_3) = 2g(\sigma_f, \sigma_s, c_s, c_f, \theta). \tag{11}$$

(iii) **No segmentation.** *If $u(\eta_h) > 2(c_s+\theta)(\gamma+\rho)$ and $u(\eta_h) \leqslant 2g(\sigma_f, \sigma_s, c_s, c_f, \theta)$, then there is no segmentation and traders only trade in the slow venue. The equilibrium is characterized by cutoffs $\eta_l = \eta_3 < \eta_4 < \eta_1 < \eta_2 = \eta_h$, where*

$$N = [\eta_l, \eta_1], \ B_s = [\eta_1, \eta_h], \ S_s = [\eta_l, \eta_4], \ H = [\eta_4, \eta_h],$$

and the cutoffs η_1 and η_4 are uniquely pinned down by

$$(1-Z)F(\eta_1) + ZF(\eta_4) = 1 - Z, \tag{12}$$

$$\frac{u(\eta_1) - u(\eta_4)}{\gamma+\rho} = 2(c_s+\theta). \tag{13}$$

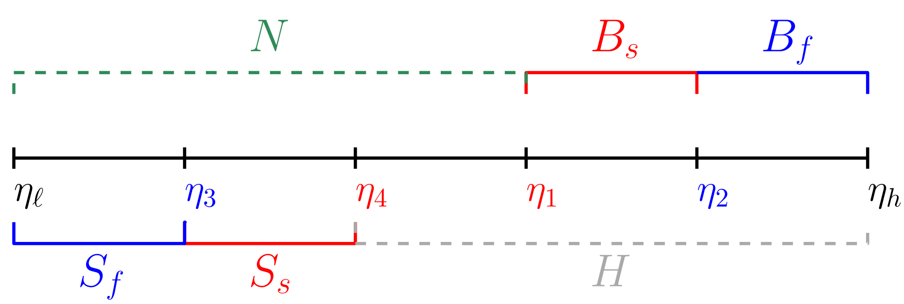

Figure 1: Threshold values for η in Theorem 1. The symbols S_f and S_s denote types that will sell in the fast and slow venue, respectively; B_f and B_s are those who will buy in the fast and slow venue, respectively. Types in N will not buy, and types in H will hold.

Theorem 1 characterizes the cutoffs in terms of $u(\eta_h)$ and serves as the building block for our analysis. Before we proceed with the analysis, we discuss the conditions that give rise to the three main cases, starting with the first case. We have the following equality:

$$u(\eta_h) = u(\eta_h) - u(0) = \Big(V_H(\eta_h) - V_H(0)\Big)(\gamma+\rho). \tag{14}$$

This quantity corresponds to the value of a transaction between traders with holding desires η_h and 0. Thus, $u(\eta_h)$ is a measure of the maximum value of a transaction, obtained when a trader with the highest possible valuation η_h buys from a trader with the lowest possible valuation $\eta_l = 0$. Intuitively, if the slow venue is too costly for traders with types η_h and 0 to trade, then it is also the case for all other traders, and there is no trade in any equilibria. In particular, whenever $2(\gamma + \rho)(c_s + \theta) \geqslant u(\eta_h)$, the trading fee is very high compared to the flow payoff of the asset, and no trader is willing to trade. It is instructive to express the condition for no trade as:

$$2(c_s + \theta) > V_H(\eta_h) - V_H(0). \tag{15}$$

In this form, the equation simply says that the cost $2(c_s + \theta)$ is higher than the value of the most profitable transaction, so there exists no price that makes a positive measure of traders on both sides of the market willing to trade.

In addition, whenever there is trade, the slow venue is always active. The reason behind this observation is simple: whenever a trader is indifferent between trading fast and holding (or doing nothing), she breaks even when the trade happens. However, as the slow venue is cheaper than the fast venue, if that trader trades in the slow venue, she pays a lower transaction fee and, thus, strictly prefers that outcome to trading in the fast venue or holding/doing nothing. Recall that there is positive trading in the fast venue whenever the following condition holds:

$$(\gamma + \rho)(V_H(\eta_h) - V_H(0)) > 2g(\sigma_f, \sigma_s, c_s, c_f, \theta). \tag{16}$$

As $c_f > c_s$, the numerator of $g(\cdot)$ is always positive and bounded away from zero (see Equation 7), while the denominator goes to zero as the speed difference between the venues vanishes. Thus, the existence of trading in the fast venue depends on the speed advantage of the fast venue and the difference in transaction fees. A segmented market is illustrated in Figure 1.

Theorem 1 characterizes $\{\eta_i\}_{i=1,\ldots,4}$ in terms of σ_f, σ_s, c_f, and c_s. We suppress the dependence of $\eta_i(\sigma_f, \sigma_s, c_f, c_s)$ on these parameters to simplify the notation in the rest of the paper.

3.2 How do fees and speeds affect market structure?

In this section, we discuss how the equilibrium market structure changes as a function of the model parameters; here, the term "market structure" refers to the state of market segmentation (i.e., one of the three cases in Theorem 1) and the corresponding cutoff thresholds as defined in Theorem 1.

3.2.1 Illustration of the market structure

For an example set of parameters, the typical thresholds behave as in Figure 2, where we plot the four thresholds as a function of the fixed cost, θ. With $\eta_l = 0$ and $\eta_h = 1$, the types $\eta \in [0, \eta_3] \cup [\eta_2, 1]$ trade in the fast venue. In other words, when $\eta_3 > 0$ and $\eta_2 < 1$, the market is segmented. This is the case in Figure 2 when the fixed cost θ is less than 1 (approximately). Once θ is sufficiently high, the market is no longer segmented: all traders use only the slow venue. Therefore, we obtain a "corner solution" for the thresholds: $\eta_3 = 0$ and $\eta_2 = 1$, and the fast venue disappears. When the θ is even higher, the no-trade condition in Theorem 1 is satisfied, and thus no trader engages in any trading. In Figure 2, we see that even the slow venue thresholds hit their corner solutions (i.e., when $\theta \geqslant 2.3$ in the figure). Thus, the figure showcases all three cases of Theorem 1.

3.2.2 Comparative statics of the market structure

In Figure 2, we illustrated how the market structure changes as the fixed cost θ increases. In general, the relevant exogenous parameters in Theorem 1 are (in addition to θ) the transaction fees c_s and c_f and the speeds σ_s and σ_f. Figure 3 illustrates the type thresholds as a function of these parameters. The behavior of the market structure with respect to each of the parameters is intuitive. Take, for

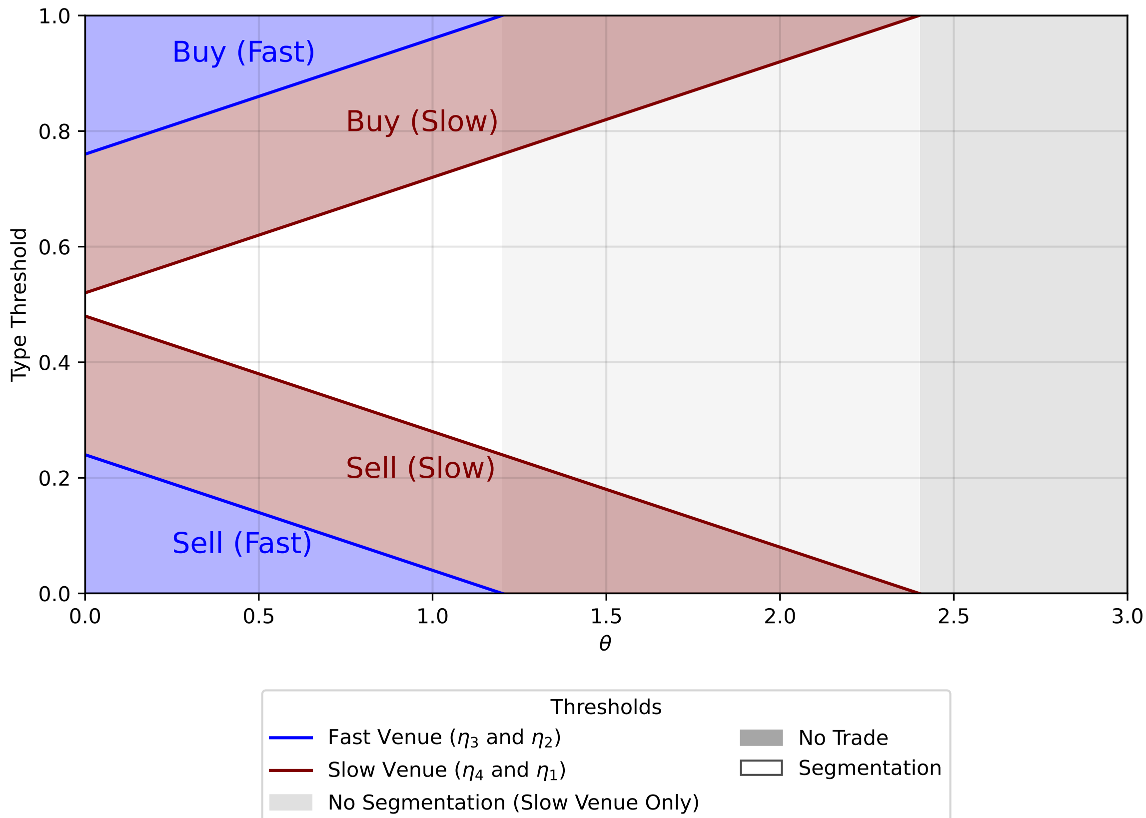

Figure 2: Type thresholds $\eta_1, \ldots, \eta_4$ as in Theorem 1 as a function of the fixed cost θ. Here η is uniformly distributed on $[0, 1]$ and the utility function $u(\eta) = \eta$ is linear. The other parameters are $Z = \frac{1}{2}$; $\rho = \gamma = 0.1$; $c_s = 0.1$ and $\sigma_s = 1$; and $c_f = 0.3$ and $\sigma_f = 10$.

example, the slow-venue transaction fee c_s; as it increases (and holding all other parameters constant), the slow venue becomes less attractive and the fast venue more attractive. When c_s is low enough, there is no segmentation because the fast venue is simply too expensive compared to the slow venue.

The behavior of the thresholds as the fast-venue transaction fee c_f changes is also intuitive; as c_f increases, fewer types trade in the fast venue (i.e., η_3 decreases and η_2 increases). When the fee is sufficiently high, all traders use the slow venue, and the market no longer exhibits segmentation. In contrast to changing c_s, changing c_f does not affect the cutoffs for types of traders who engage in the slow venue. This is because the traders who are on the cutoff are actually indifferent

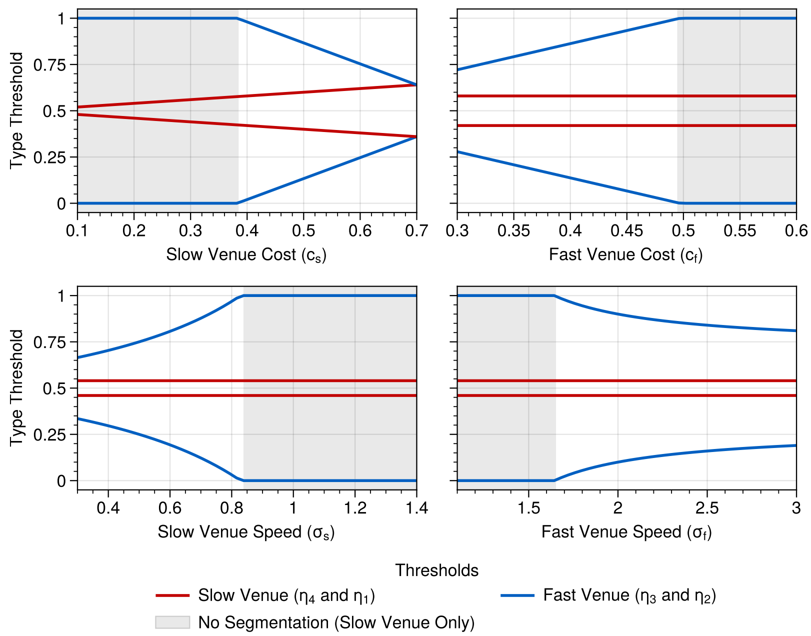

Figure 3: Comparative statics for Theorem 1. The type thresholds (y-axis) are plotted as a function of changing a single parameter at a time (x-axis); the parameters of interest are the transaction fees and speed offered by the slow and fast venues. As usual, the thresholds are ordered: $\eta_3 < \eta_4 < \eta_1 < \eta_2$. Thresholds for the fast venue are in blue, and those for the slow venue are in red.

between trading in the slow venue versus not trading at all: the alternative option of trading in the fast venue is not being considered since transaction fee concerns dominate speed concerns. Therefore, these traders remain marginal even as the conditions in the fast venue change. On the other hand, there are no marginal traders who are not affected by a change in the slow venue transaction fee: if c_s increases, the scale is tipped in favor of the fast venue for marginal traders between the fast and slow venues. Similarly, the scale is tipped in favor of not trading for marginal traders between trading slowly and not trading.

Finally, increasing the slow-venue speed has a similar effect as increasing the fast-venue transaction fee in that the fast venue becomes less attractive. Interestingly, the thresholds for marginal traders using the slow venue do not change: They care only about the cost of trading, as explained above. Increasing the speed of the fast venue has the opposite effect, though we see that it exhibits *diminishing* *returns* in that even an infinite speed advantage will not allow the fast venue to capture certain types of traders.

4 Trading Volume

In this section, we characterize the trading volume for a given market structure. The measure of traders in each venue is given by the following equations:

$$\begin{aligned} m_f(\sigma_s,\sigma_f,c_s,c_f,\theta) &= \int_{\eta_l}^{\eta_3} f_o(\eta)d\eta + \int_{\eta_2}^{\eta_h} f_{no}(\eta)d\eta \\ &= F(\eta_3)\frac{\gamma Z}{\gamma+\sigma_f} + (1-F(\eta_2))\frac{\gamma(1-Z)}{\gamma+\sigma_f}, \end{aligned} \tag{17}$$

and

$$\begin{aligned} m_s(\sigma_s,\sigma_f,c_s,c_f,\theta) &= \int_{\eta_3}^{\eta_4} f_o(\eta)d\eta + \int_{\eta_1}^{\eta_2} f_{no}(\eta)d\eta \\ &= (F(\eta_4)-F(\eta_3))\frac{\gamma Z}{\gamma+\sigma_s} + (F(\eta_2)-F(\eta_1))\frac{\gamma(1-Z)}{\gamma+\sigma_s}. \end{aligned} \tag{18}$$

The **trading volume** in the slow and fast venues is given by $TV_s = \sigma_s m_s$ and $TV_f = \sigma_f m_f$, respectively. We define the total trading volume as $TV = TV_s + TV_f$. The following proposition shows how trading volume in venues depends on fees and speeds.[3]

Proposition 1. *The trading volume in the fast venue is increasing in c_s,σ_f and is decreasing in c_f,σ_s,θ. The trading volume in the slow venue is increasing in c_f,σ_s and is decreasing in c_s,σ_f. The total trading volume is decreasing in θ.*

As expected, the trading volume in a venue is increasing in the trading speed of that venue and the transaction fee of the other venue, while it is decreasing in the trading speed of the other venue and the transaction fee of that venue.

[3] A full set of comparative statics of cutoffs and measures of traders is provided in the appendix.

Increasing the transaction fee reduces the trading volume while increasing the fee charged per transaction, which is the main trade-off for the firms when they compete on fees. We later allow firms to compete by setting c_s and c_f to maximize their revenues and analyze the effect that competition has on trading volume.

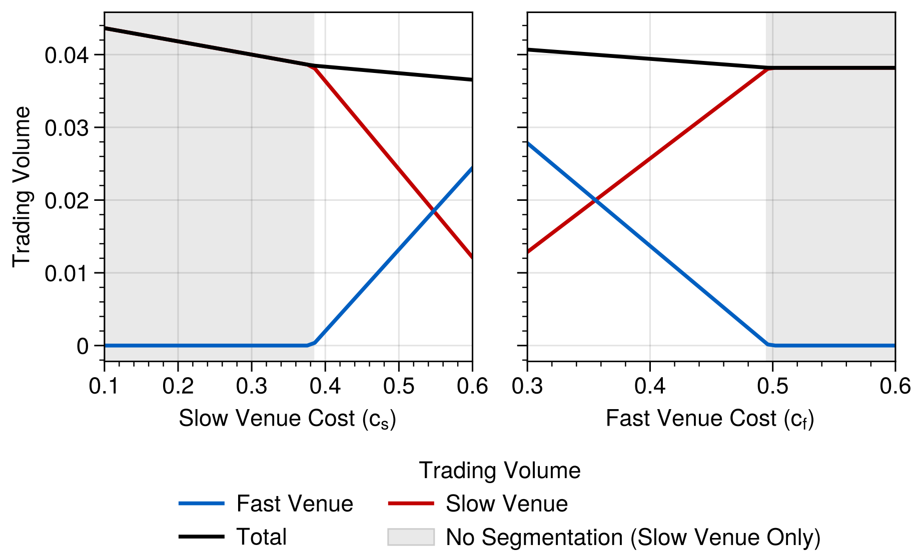

Figure 4: Trading volume in the two venues as a function of transaction fees c_s and c_f with $\eta \sim \text{Unif}[0, 1]$ and $u(\eta) = \eta$. The gray regions indicate cost regimes in which only the slow venue is active. The primary driver of the trading volume in either venue is whether the fees exceed the threshold values needed for segmentation.

Before endogenizing the transaction fees c_s and c_f with competition between venue owners, we illustrate the results of Proposition 1 in Figures 4 and 5. Figure 4 plots the trading volumes in the two venues as the transaction fees in the venues change. In the left pane, we observe that the fast venue is initially inactive due to the low fees in the slow venue, but eventually becomes active when the slow venue fees are high enough. Similarly, in the right pane, we observe that the fast venue is active whenever c_f is low enough compared to c_s. Figure 5 plots the trading volumes in the two venues as the venue speed changes and illustrates that for given transaction fees, the trading volume in each venue is increasing in its speed. In the next section, we introduce competition among the venues and characterize this behavior in greater detail.

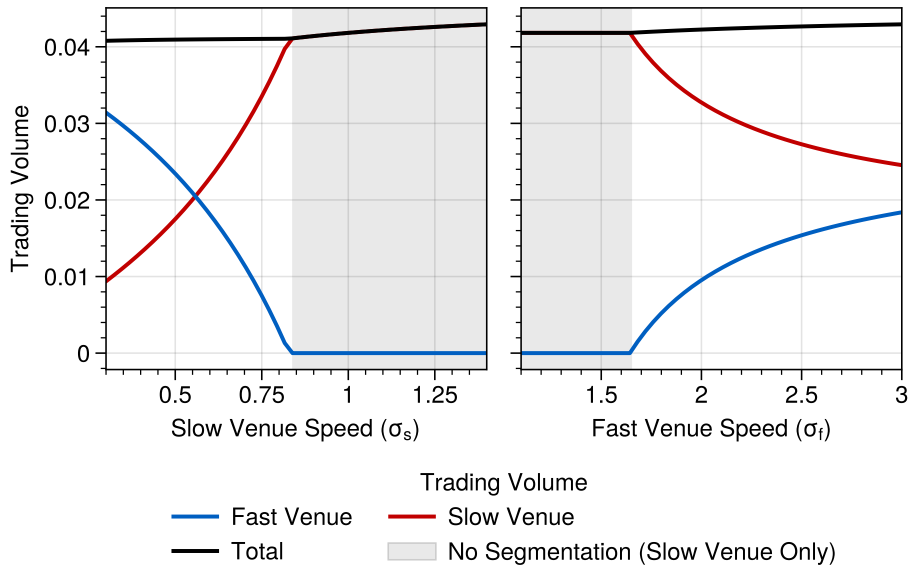

Figure 5: Trading volume as a function of venue speed with $\eta \sim \text{Unif}[0, 1]$ and $u(\eta) = \eta$. The gray region indicates the thresholds beyond which only the slow venue is active.

5 How does fee competition affect market structure?

Having characterized the trading volume, we next analyze the competition between the two venues by setting transaction fees. The revenues of the fast and slow venues for a specific transaction fee are given by the following expressions:

$$R_f(\sigma_f, c_f, \sigma_s, c_s) = \sigma_f m_f(\sigma_s, \sigma_f, c_s, c_f) c_f, \tag{19}$$

$$R_s(\sigma_f, c_f, \sigma_s, c_s) = \sigma_s m_s(\sigma_s, \sigma_f, c_s, c_f) c_s. \tag{20}$$

For the rest of the paper, we make the following assumptions to keep the analysis tractable.

Assumption 2. $u(\eta) = a\eta$.

Assumption 3. *F is uniform over* $[0, 1]$.

We also make the following assumption to guarantee that the fixed cost θ is not so high as to prohibit trading *a priori*.

Assumption 4. *$a > 2\theta(\gamma + \rho)$.*

An equilibrium is a set of fees c_s^*, c_f^* such that $c_s^* \in \arg\max_{c_s} R_s(\sigma_f, c_f^*, \sigma_s, c_s)$ and $c_f^* \in \arg\max_{c_f} R_f(\sigma_f, c_f, \sigma_s, c_s^*)$. The following proposition characterizes the equilibrium fees of two competing venues.

Proposition 2. *There exists a unique equilibrium. The fees charged by the fast and slow markets are given by*

$$c_f^*(\sigma_f, \sigma_s) = (a - 2\theta(\gamma + \rho)) \frac{\sigma_f - \sigma_s}{(4\sigma_f - \sigma_s)(\gamma + \rho) + 3\sigma_f \sigma_s}, \tag{21}$$

$$c_s^*(\sigma_f, \sigma_s) = \frac{c_f^*}{2}. \tag{22}$$

Proposition 2 characterizes how the venues compete to attract traders. The slower venue always undercuts the fast venue as otherwise, no trader would trade in it and the venue would make zero revenue. The slow venue attracts speed-insensitive traders, while the fast venue sets a higher transaction fee and attracts speed-sensitive traders.

More precisely, for any fixed set of exogenous parameters Z, γ, ρ, and θ, the fees c_s^* and c_f^* induced by any pair of speeds $0 < \sigma_s < \sigma_f$ will lead to a segmented market. In other words, the segmentation condition in Theorem 1 is satisfied *a posteriori* when the costs are optimally chosen given the other parameters. This captures the obvious fact that when costs are endogenous, the venues would never set uncompetitive fees. Hence, both venues are active in equilibrium.

Corollary 1. *Fix parameters $Z \in (0, 1)$, $\gamma > 0$, $\rho > 0$, and $\theta \geqslant 0$. For any pairs of venue speeds satisfying $0 < \sigma_s < \sigma_f$, the segmentation condition*

$$a = u(\eta_h) = u(1) > 2g(\sigma_f, \sigma_s, c_s^*, c_f^*, \theta) \tag{23}$$

is satisfied, where c_s^* and c_f^* are given by Proposition 2.

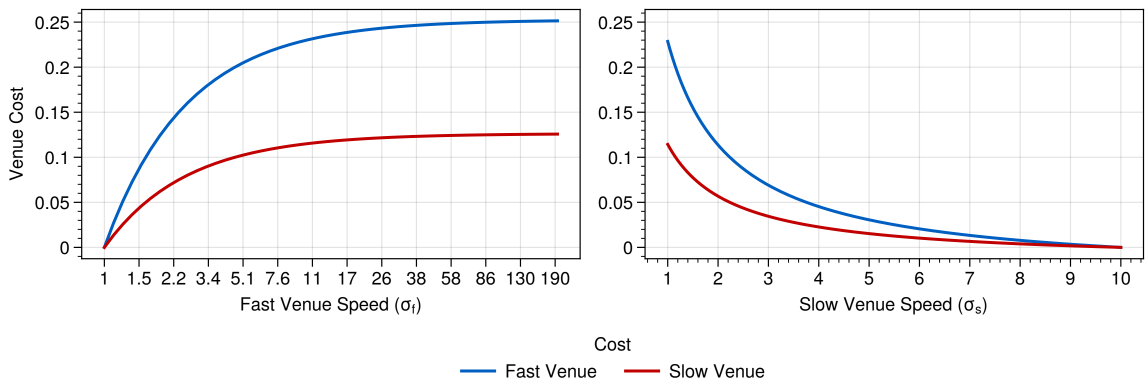

Figure 6: Optimal transaction fees (cost) as a function of venue speeds.

Moreover, as the differentiation between the venues becomes smaller, the competition between them disappears.

Corollary 2. $\lim_{|\sigma_s - \sigma_f| \to 0} c_s^*(\sigma_f, \sigma_s) = \lim_{|\sigma_s - \sigma_f| \to 0} c_f^*(\sigma_f, \sigma_s) = 0.$

The following proposition shows how competition is affected by the speed of each venue.

Proposition 3. *Transaction fees c_s^* and c_f^* are increasing in σ_f and decreasing in σ_s.*

These results show the importance of differentiation. When the fast venue becomes faster, the level of differentiation between the venues increases and the effect of competition decreases. This results in higher fees across venues. Conversely, when the slow venue becomes faster, the differentiation between the venues decreases and the effect of competition increases, which results in lower fees. As $\sigma_s \to \sigma_f$, i.e., the venues become similar, the effect of Bertrand competition drives down the transaction fees; hence, revenues go to zero. A similar insight is also found in Pagnotta and Philippon (2018), which considers competition between venues in a subscription model.

The results above are illustrated in Figure 6, where c_s^* and c_f^* are plotted as functions of σ_s and σ_f. The left panel uses $\sigma_s = 1$ and moves σ_f to a very large value to visualize the behavior when $\sigma_f \to \infty$. As in Proposition 2, the transaction fees are driven to zero as σ_s approaches σ_f. Moreover, the function c_f^* (and hence

c_s^*) is concave in σ_f: increasing σ_f has diminishing effects on the optimal fee. The intuitive explanation is that σ_f itself has *diminishing effects* on the market structure as $\sigma_f \to \infty$, giving us

$$\lim_{\sigma_f \to \infty} c_f^*(\sigma_f, \sigma_s = 1) = \frac{a - 2\theta(\gamma + \rho)}{4(\gamma + \rho) + 3}. \tag{24}$$

For the same reason, fees are convex and decreasing with respect to the slow venue speed σ_s.

Another interesting implication is that traders with different valuations are affected differently by the changes in speeds. For example, an increase in σ_s reduces transaction fees, making all traders better off. On the other hand, an increase in σ_f increases transaction fees and makes all traders, except those with the most extreme valuations, worse off.

Lastly, we analyze the effect of speed on trading volume. The next proposition shows the effect of the speed of the slow venue.

Proposition 4. *The trading volume in both venues is increasing in σ_s.*

Increasing σ_s has two effects. First, it increases the trading speed at the slow venue, which, in turn, increases the trading volume (the *improvement channel*). Second, it makes the slow venue more competitive. This competition reduces trading fees, which also increases the trading volume (the *differentiation channel*). A speed increase for the slow venue positively affects both channels by decreasing differentiation and improving the overall trading speed in the market, leading to higher trading volume.

Figure 7 plots the trading volumes in each venue as a function of σ_s. Note that trading volume is not only increasing but also a concave function of σ_s, indicating that increasing σ_s has a diminishing effect.

In fact, the two components of the competition channel that drive up the trading volume exhibit diminishing returns. We have already seen in the right panel of Figure 6 that increasing σ_s^* has diminishing effects on lowering costs.

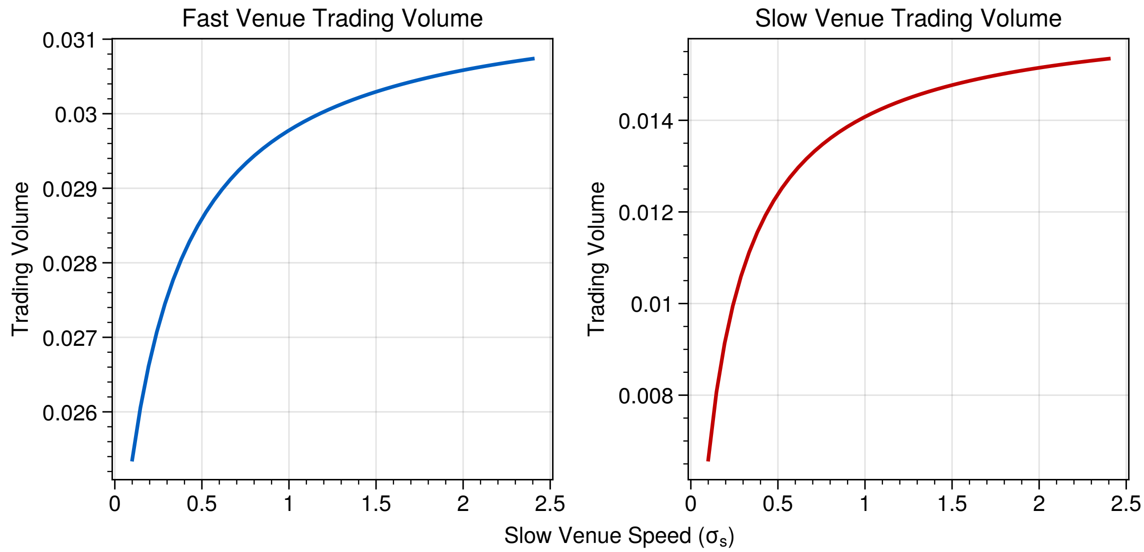

Figure 7: Trading volumes in both venues increase as σ_s increases.

Moreover, Figure 8 shows that the impact of competition—namely, inducing former nontraders to trade in the slow venue and formerly slow venue traders to trade in the fast venue—is also diminishing. The diminishing effect is captured by the thresholds $\eta_1, \ldots, \eta_4$ being concave; η_3 and η_4 are decreasing, while η_1 and η_2 are increasing. *Recall that exogenously, σ_s has no effect on slow-venue thresholds η_1 and η_4 (see Figure 3); it is the effect of σ_s on c_s and c_f that induces the change in market structure.*

The effect of speed in the fast venue is more complicated. On the one hand, a speed improvement in the fast venue increases the transaction rate there, which boosts trading volume. However, such a speed improvement also increases differentiation, thus raising the transaction fees charged in equilibrium. Increased transaction fees reduce traders' incentives to pay the fee and execute a trade, which lowers trading volume. Thus, the improvement and differentiation channels are working in opposite directions. In general, the effect of σ_f on trading volume is ambiguous. To better understand this, we consider an instructive special case, which we call the *full competition benchmark*. We compare the trading volume when there is no differentiation ($\sigma_s = \sigma_f = \sigma$) and when the fast venue exhibits a local

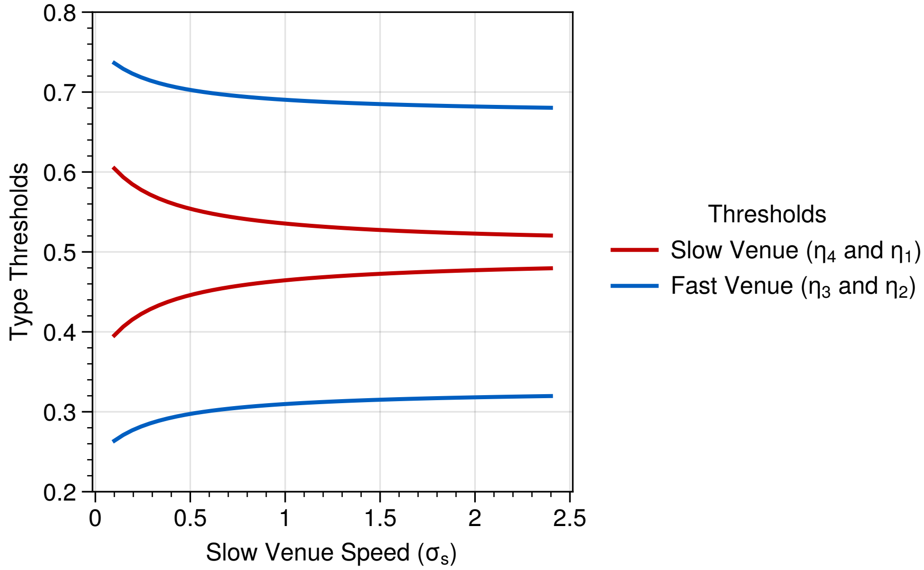

Figure 8: Increasing σ_s has a diminishing impact on type thresholds.

speed improvement ($\sigma_s = \sigma$ and $\sigma_f = \sigma + \epsilon$, for small $\epsilon > 0$).

Proposition 5. *Let $\sigma_s = \sigma$ and $\sigma_f = \sigma + \epsilon$, with $\epsilon > 0$. Then $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon}$ has the sign of $\gamma^2 + \gamma\sigma + \rho(\gamma - \sigma)$. In particular,*

- *If $\gamma \geqslant \rho$, then $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0$,*
- *If $\gamma < \rho$, then $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0$ if and only if $\sigma < \frac{\gamma(\rho+\gamma)}{\rho-\gamma}$.*

Proposition 5 shows that the rate of the preference shock (γ), the patience of the traders (ρ) and the initial trading speed (σ) are important determinants of the effect of differention on the trading volume. First, when $\gamma \geqslant \rho$, the trading volume increases after a local speed improvement. This is because when ρ is low, traders are patient and the value created by a trade is greater for them, so the higher fees associated with differentiation cause less distortion. Similarly, when γ is high, we observe the same effect, but this time because traders want to trade quickly to realize the value of having the asset until the preference shock occurs and, therefore, are less affected by the higher fees. In these cases, traders continue to trade even with increased fees, allowing the venues to extract more revenue from them.

When $\gamma < \rho$, the effect on trading volume depends on σ, the initial speed of the venues. In this case, a speed increase boosts trading volume if and only if the initial trading speed is low enough. This is intuitive as when trade is already fast enough initially, the benefit of faster trading speed is less significant and is dominated by the higher fees caused by differentiation. Moreover, the condition $\sigma < \frac{\gamma(\rho+\gamma)}{\rho-\gamma}$ is more easily satisfied when the preference shock is more frequent and traders are more patient, echoing the intuition for the first condition.[4] Thus, for a given σ, differentiation increases trading volume when γ is high and ρ is low.

6 Welfare Analysis

In this section, we study the relationship between venue characteristics and welfare. We assume the fixed cost $\theta = 0$, which causes the financial transactions in this model to preserve the financial value within the system.[5] As a result, they are purely distributional and do not affect total welfare. More precisely, from Theorem 1, we can characterize the venue welfare, which equals the total trading fees:

$$\begin{aligned} W_{\text{venue}} &= \int_{\eta_l}^{\eta_3} \sigma_f c_f f_o(\eta) d\eta + \int_{\eta_2}^{\eta_h} \sigma_f c_f f_{no}(\eta) d\eta + \int_{\eta_3}^{\eta_4} \sigma_s c_s f_o(\eta) d\eta + \int_{\eta_1}^{\eta_2} \sigma_s c_s f_{no}(\eta) d\eta \\ &= c_f TV_f + c_s TV_s = R_f + R_s, \end{aligned}$$

[4]To see this, observe that $\frac{\partial \frac{\gamma(\rho+\gamma)}{\rho-\gamma}}{\partial \gamma} = \frac{\rho^2-\gamma^2+2\rho\gamma}{(\gamma-\rho)^2}$ is positive whenever $\rho > \gamma$, and $\frac{\partial \frac{\gamma(\rho+\gamma)}{\rho-\gamma}}{\partial \rho} = \frac{-2\gamma^2}{(\gamma-\rho)^2}$ is negative.

[5]In other words, when $\theta = 0$, there is no cost in the model that would cause financial value to be lost in transactions.

and the trader welfare, which is the value of holding the asset minus fees:

$$\begin{aligned}
W_{\text{trader}} &= \int_0^1 a\eta f_o(\eta)d\eta - \int_{\eta_l}^{\eta_3} \sigma_f c_f f_o(\eta)d\eta - \int_{\eta_2}^{\eta_h} \sigma_f c_f f_{no}(\eta)d\eta \\
&\quad - \int_{\eta_3}^{\eta_4} \sigma_s c_s f_o(\eta)d\eta - \int_{\eta_1}^{\eta_2} \sigma_s c_s f_{no}(\eta)d\eta \\
&= \int_0^1 a\eta f_o(\eta)d\eta - W_{\text{venue}}.
\end{aligned}$$

Therefore, the total (equally weighted) welfare becomes:

$$\begin{aligned}
W &= W_{\text{trader}} + W_{\text{venue}} \\
&= \int_0^1 a\eta f_o(\eta)d\eta.
\end{aligned}$$

We now focus on the effect that differentiation has on welfare and extend Proposition 5 to welfare analysis. First, we concentrate on total welfare.

Proposition 6. *Let $\sigma_s = \sigma$ and $\sigma_f = \sigma + \epsilon$, with $\epsilon > 0$. Then $\lim_{\epsilon \to 0} \frac{\partial W}{\partial \epsilon} > 0$.*

Thus, when venue profits are included in the welfare calculation, the improvement channel dominates the differentiation channel, and increased trading speed enhances welfare. This is intuitive, as any fees that venues extract from traders due to differentiation are counted in the welfare calculation. However, the effect on trader welfare remains ambiguous.

Proposition 7. *Let $\sigma_s = \sigma$ and $\sigma_f = \sigma + \epsilon$, with $\epsilon > 0$. Then $\lim_{\epsilon \to 0} \frac{\partial W_{\text{trader}}}{\partial \epsilon} > 0$ if and only if $\sigma < 4\rho - \gamma$.*

The intuition behind Proposition 7 is similar to the one behind Proposition 5. The importance of a local speed improvement is decreasing in the initial speed of the venues and such an improvement increases trader welfare if and only if the venues are initially slow enough.

However, the effect that the preference shock and the discount factor have on the threshold is the opposite. In Proposition 5, we show that the trading volume increases with speed differentiation when traders are patient and the preference shock is frequent. In such cases, venues can extract more fees from traders due to both the increasing trading volume and the higher fees associated with differentiation. This lowers traders welfare. Conversely, in cases with impatient traders and infrequent preference shocks, venues cannot extract those rents, and a speed increase results in an increase in trader welfare.

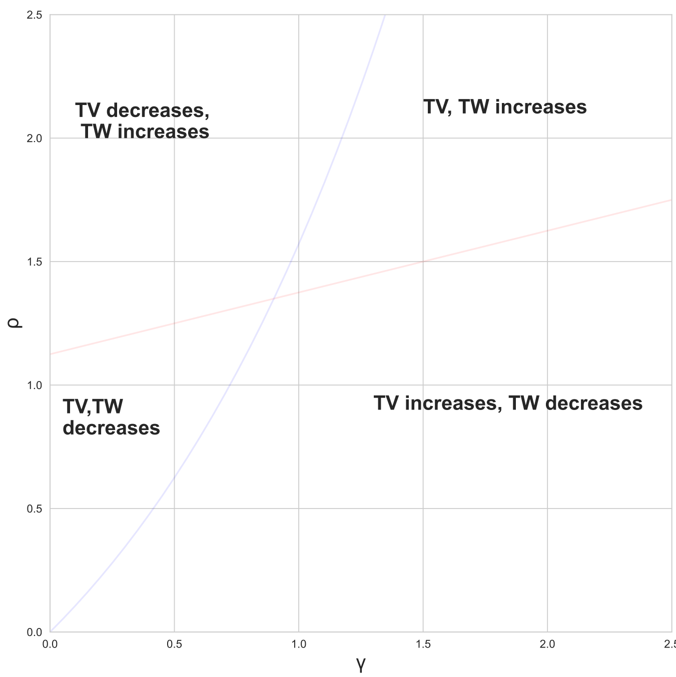

Figure 9: Behavior of trading volume (TV) and trader welfare (TW) as a function of γ (x-axis) and ρ (y-axis) under a local increase in trading speeds. We set $\sigma = 1$.

Propositions 5 and 7 illustrate how the effect of a local speed increase on trading volume and trader welfare depends on the initial speed, discount factor, and rate of preference shock. In Figure 9, we plot the regions where this effect (i) is positive for both the trading volume and trader welfare, (ii) is negative for both, (iii) is positive for the trading volume and negative for trader welfare, and (iv) is positive for trader welfare and negative for the trading volume. The red region with high values of γ and low values of ρ corresponds to patient traders and frequent preference shocks; in this region, trading volume increases while trader welfare decreases. Conversely, the blue region with low values of γ and high values of ρ corresponds to impatient traders and infrequent preference shocks, yielding the opposite effect. In the purple region, characterized by impatient traders and frequent preference shocks, both trading volume and trader welfare increase following a speed improvement. In contrast, in the white region, both values decline after such an improvement. Moreover, the qualitative features of the figure do not depend on the value that σ takes. For each σ, the point where all four regions intersect corresponds to the $\gamma^*(\sigma)$ and $\rho^*(\sigma)$ values that jointly solve the conditions given in Propositions 5 and 7, and $(\gamma^*(\sigma), \rho^*(\sigma))$ are increasing in σ.

7 Conclusion

We developed a dynamic model to study the strategic interactions between trading venues and the dynamic choices of traders when venues differ in technology (fast vs. slow) and clear trades separately (e.g., exchange vs. OTC, core vs. periphery dealers in OTC markets, crypto exchanges). We characterize the equilibrium in terms of the trading speeds and fees of the venues and demonstrate when it exhibits fragmentation. We then examine the equilibrium of fee competition among trading venues. In this equilibrium, both venues are active, leading to market segmentation (fragmentation).

We demonstrate that the equilibrium trading fees increase when the venues are technologically differentiated. We then explore trading volume and trader welfare. Improvements in the slow venue's technology correlate with increased trading volumes and enhanced trader welfare. In contrast, the impact of improvements in the fast venue's speed is ambiguous and hinges on the trading speed, the rate of preference shock, and the traders' patience. Specifically, an improvement in the fast venue's speed positively affects trading volume and trader welfare if and only if the initial trading speeds are relatively low compared to the rate of preference shock and the traders' discount factor.

In conclusion, our model highlights how technological differences impact trading strategies and market welfare. From a policy perspective, this underscores the need for regulatory considerations around technology disparities in trading venues. Policymakers should be cognizant of the potential market segmentation and varied trader welfare outcomes arising from technological advancement in this sector.

References

Back, Kerry. 1992. "Insider trading in continuous time." *Review of Financial Studies*, 5(3): 387–409.

Back, Kerry, C Henry Cao, and Gregory A Willard. 2000. "Imperfect competition among informed traders." *Journal of Finance*, 55(5): 2117–2155.

Biais, Bruno, Johan Hombert, and Pierre-Olivier Weill. 2021. "Incentive constrained risk sharing, segmentation, and asset pricing." *American Economic Review*, 111(11): 3575–3610.

Chen, Daniel, and Darrell Duffie. 2021. "Market fragmentation." *American Economic Review*, 111(7): 2247–74.

Foucault, Thierry, and Christine A Parlour. 2004. "Competition for listings." *Rand Journal of Economics*, 329–355.

Kyle, Albert S. 1985*a*. "Continuous auctions and insider trading." *Econometrica*, 1315–1335.

Kyle, Albert S. 1985*b*. "Long-lived information and intraday patterns." *Econometrica*, 53(6): 1315–1335.

Malamud, Semyon, and Marzena Rostek. 2017. "Decentralized exchange." *American Economic Review*, 107(11): 3320–62.

Mendelson, Haim. 1987. "Consolidation, fragmentation, and market performance." *Journal of Financial and Quantitative Analysis*, 22(2): 189–207.

Pagano, Marco. 1989. "Trading volume and asset liquidity." *Quarterly Journal of Economics*, 104(2): 255–274.

Pagnotta, Emiliano S., and Thomas Philippon. 2018. "Competing on speed." *Econometrica*, 122: 1067–1115.

Parlour, Christine A, and Duane J Seppi. 2003. "Liquidity-based competition for order flow." *Review of Financial Studies*, 16(2): 301–343.

Rostek, Marzena J, and Ji Hee Yoon. 2020. "Equilibrium theory of financial markets: Recent developments." *Available at SSRN 3710206*.

Santos, Tano, and Jose A Scheinkman. 2001. "Competition among exchanges." *Quarterly Journal of Economics*, 116(3): 1027–1061.

Wang, Jiang. 1994. "A Model of Competitive Stock Trading Volume." *Journal of Political Economy*, 102: 127–168.

Appendix

Throughout the appendix, we use the more compact notation $V_{m,\eta}$ for the quantity $V_m(\eta)$.

Proof of Theorem 1

The proof will proceed as follows: First, we show in Lemma 1 that in any equilibrium with positive trade, selling in the slow venue leaves the seller with higher revenue than selling in the fast venue and that buying in the slow venue is cheaper for the buyer compared to buying in the fast venue. Using these facts, Lemmas 2 and 3 characterize actions of traders after a trade and show that buyers hold the asset after buying and sellers do nothing after selling. Given these actions, Lemma 4 characterizes the value functions after any decision (selling, holding, buying or doing nothing) by a trader. Lemmas 5, 6, and 7 characterize the structure of traders' venue choices using a simple cutoff structure. Lemma 8 gives the sufficient (which we later show to be necessary) condition for no trade in both venues, proving part (i) of Theorem 1. This condition requires the present value of the gains from the most profitable trade (between a trader with valuation η_h and a trader with valuation 0) to be larger than total fees and fixed costs paid by the traders for the trade. Lemma 9 gives an analogous condition for existence of trade in the fast venue, which simply requires the value of most profitable (thus most speed sensitive) trade to be larger than some measure of differentiation among the venues. Lemma 10 characterizes the equilibrium type distributions of asset owners and non-owners using inflow and outflow equations. Using the distributions characterized in Lemma 10 and the venue clearing conditions, in Lemma 11 we arrive at two of the four main conditions in part (ii) of Theorem 1. Assuming both venues are active (i.e. the condition given in Lemma 9 holds), Lemma 12 finishes the the characterization of cutoffs structure when the market is segmented. Lemma 13 then finishes the proof of part (ii) of Theorem 1 by showing the existence and uniqueness of prices. Lastly, we prove part (iii) by using Lemmas 10 and 11 and showing the existence of prices where there is only demand for the slow venue whenever the condition in Lemma 9 does not hold.

Lemma 1. *In any equilibrium with positive trading, the following are satisfied:*

1. $p_s - c_s \geqslant p_f - c_f$,

2. $p_s + c_s \leqslant p_f + c_f$.

Proof. Assume for a contradiction that $p_s - c_s \geqslant p_f - c_f$ is not satisfied. Then we have $p_s - c_s < p_f - c_f$. For a contradiction, assume that there exists η such that $V_S^s(\eta) \geqslant V_H(\eta)$ and $V_S^s(\eta) \geqslant V_S^f(\eta)$. The former implies that $V_{no,\eta} + p_s - c_s - \theta - V_S^s(\eta) > 0$. Note that the values for $V_S^s(\eta)$ and $V_S^f(\eta)$ are given by following equations:

$$\rho V_S^s(\eta) = u(\eta) + \gamma\left(\mathbb{E}_{\eta'}[V_{o,\eta'}] - V_S^s(\eta)\right) + \sigma_s\left(V_{no,\eta} + p_s - c_s - \theta - V_S^s(\eta)\right), \tag{25}$$

$$\rho V_S^f(\eta) = u(\eta) + \gamma\left(\mathbb{E}_{\eta'}[V_{o,\eta'}] - V_S^f(\eta)\right) + \sigma_f\left(V_{no,\eta} + p_f - c_f - \theta - V_S^f(\eta)\right). \tag{26}$$

We compare these expressions term by term: Using that $V_S^s(\eta) \geqslant V_S^f(\eta)$, we have

$$\gamma\left(\mathbb{E}_{\eta'}[V_{o,\eta'}] - V_S^f(\eta)\right) \geqslant \gamma\left(\mathbb{E}_{\eta'}[V_{o,\eta'}] - V_S^s(\eta)\right).$$

Moreover, $V_S^s(\eta) \geqslant V_S^f(\eta)$, $\sigma_f > \sigma_s$, $V_{no,\eta} + p_s - c_s - \theta - V_S^s(\eta) > 0$ together with $p_s - c_s < p_f - c_f$ implies that

$$\sigma_f\left(V_{no,\eta} + p_f - c_f - \theta - V_S^f(\eta)\right) > \sigma_s\left(V_{no,\eta} + p_s - c_s - \theta - V_S^s(\eta)\right).$$

Thus, we have $V_S^s(\eta) < V_S^f(\eta)$, which is a contradiction. As a result, there is no η such that $V_S^s(\eta) \geqslant V_H(\eta)$ and $V_S^s(\eta) \geqslant V_S^f(\eta)$. This means that there is measure 0 of traders who prefer to sell slow. As we assumed there is positive trade, then there must be non-zero measure of traders who prefer to sell fast.

Next, we will show that under $p_s - c_s < p_f - c_f$, there positive demand in selling slow, which is a contradiction as slow venue clearing condition cannot hold in that case. Note that $p_s - c_s < p_f - c_f$ and $c_s < c_f$ implies $p_s + c_s < p_f + c_f$.

In any equilibrium with positive trade, there is a type η that prefers buying fast or slow to doing nothing. If type η prefers buying slow, i.e., $V_B^S(\eta) > \max\{V_S^s(\eta), V_N(\eta)\}$, the continuity of V_N, V_S^s and V_S^f in η implies that there is a positive measure of types around η that prefer to buy slow.[6] However, this will be

[6]The continuity of V_N and V_H in η follows directly from continuity of $u(\eta)$. The continuity of

a contradiction to the slow venue clearing condition and cannot happen under any equilibrium.

Next, assume for a contradiction there are no types that prefer to buy slow. Then each type either prefers doing nothing or buying fast. Let η^* denote the type that is indifferent between buying fast and doing nothing, i.e. $V_N(\eta^*) = V_b^f(\eta^*)$. This means that $V_{o,\eta^*} - p_f - c_f - \theta - V_B^f(\eta^*) = 0$. But as $p_s + c_s < p_f + c_f$ and $V_B^s \leqslant V_s^f$ we have $V_{o,\eta^*} - p_s - c_s - \theta - V_b^f(\eta^*) > 0$. This implies that $V_B^s(\eta^*) > V_B^f(\eta^*)$. But then, there is a positive measure of traders that prefer to buy slowly. As we have shown there are no traders that prefer to sell slowly, slow venue clearing condition cannot hold and this is a contradiction to the assertion that that we have an equilibrium.

The proof of the second part is analogous. Assume for a contradiction that $p_s + c_s \leqslant p_f + c_f$ is not satisfied. Then we have $p_s + c_s > p_f + c_f$. For a contradiction, assume that there exists η such that $V_B^s(\eta) \geqslant V_N(\eta)$ and $V_B^s(\eta) \geqslant V_B^f(\eta)$. The former implies that $V_{o,\eta} - p_s - c_s - \theta - V_B^s(\eta) > 0$. Note that the values for $V_B^s(\eta)$ and $V_B^f(\eta)$ are given by following equations:

$$\rho V_B^s(\eta) = \gamma \left(\mathbb{E}_{\eta'}[V_{no,\eta'}] - V_B^s(\eta) \right) + \sigma_s \left(V_{o,\eta} - p_s - c_s - \theta - V_B^s(\eta) \right), \tag{27}$$

and

$$\rho V_B^f(\eta) = \gamma \left(\mathbb{E}_{\eta'}[V_{no,\eta'}] - V_B^f(\eta) \right) + \sigma_f \left(V_{o,\eta} - p_f - c_f - \theta - V_B^f(\eta) \right). \tag{28}$$

Looking at term by term, as $V_B^s(\eta) \geqslant V_B^f(\eta)$, we have

$$\gamma \left(\mathbb{E}_{\eta'}[V_{no,\eta'}] - V_B^f(\eta) \right) \geqslant \gamma \left(\mathbb{E}_{\eta'}[V_{no,\eta'}] - V_B^s(\eta) \right).$$

Moreover, $V_B^s(\eta) \geqslant V_B^f(\eta)$, $\sigma_f > \sigma_s$, $V_{o,\eta} - p_s - c_s - \theta - V_B^s(\eta) > 0$ together with

V_S^s, V_S^f, V_B^s and V_B^f follows from continuity of $u(\eta)$, $V_{o,\eta}$ and $V_{no,\eta}$. To show the continuity of $V_{o,\eta}$ and $V_{no,\eta}$ in any equilibrium, let η and $\eta' = \eta + \epsilon$ with $\epsilon > 0$ denote two different types. First, note that the equilibrium strategy of η is available to η' and $u(\eta') > u(\eta)$, thus η' can guarantee herself a payoff of at least $V_{o,\eta}$ when she owns the asset and $V_{no,\eta}$ when she does not by playing the same strategy. Thus we have $V_{o,\eta'} \geqslant V_{o,\eta}$ and $V_{no,\eta'} > V_{no,\eta}$. Next, if η plays the equilibrium strategy of η', then $V_{o,\eta'} - V_{o,\eta}$ and $V_{no,\eta'} - V_{no,\eta}$ is bounded above by the expected utility derived from owning the asset until the preference shock strikes. As u is continuous, this bound converges to 0 as ϵ goes to 0, yielding the desired continuity.

$p_s + c_s > p_f + c_f$ implies that

$$\sigma_f\left(V_{o,\eta} - p_f - c_f - \theta - V_B^f(\eta)\right) > \sigma_s\left(V_{o,\eta} - p_s - c_s - \theta - V_B^s(\eta)\right).$$

Thus, we have $V_B^s(\eta) < V_B^f(\eta)$, which is a contradiction. As a result, There is no η such that $V_B^s(\eta) \geqslant V_N(\eta)$ and $V_B^s(\eta) \geqslant V_B^f(\eta)$. This means that there is measure 0 of traders who prefer to buy slow. As we assumed there is positive trade, then there must be non-zero measure of traders who prefer to buy fast.

Next, we will show that under $p_s + c_s > p_f + c_f$, there is positive demand in buying slow, which is a contradiction as the slow venue clearing condition cannot hold in that case. Note that $p_s + c_s > p_f + c_f$ and $c_s < c_f$ implies $p_s - c_s > p_f - c_f$.

In any equilibrium with positive trade, there is a type η that prefers selling fast or slow to holding. If type η prefers selling slow, i.e., $V_B^S(\eta) > \max\{V_S^s(\eta), V_N(\eta)\}$, the continuity of V_N, V_S^s and V_S^f in η implies that there is a positive measure of types around η that prefer to buy slow. However, this will be a contradiction to the slow venue clearing condition and cannot happen under any equilibrium.

Next, assume for a contradiction there are no types that prefer to sell slowly. Then each type either prefers holding or selling fast. Let η^* denote the type that is indifferent between selling fast and holding, i.e. $V_H(\eta^*) = V_S^f(\eta^*)$. This means that $V_{no,\eta^*} + p_f - c_f - \theta - V_B^f(\eta^*) = 0$. But as $p_s - c_s > p_f - c_f$ and $V_S^s \leqslant V_S^f$ we have $V_{no,\eta^*} + p_s - c_s - \theta - V_b^f(\eta^*) > 0$. This implies that $V_S^s(\eta^*) > V_S^f(\eta^*)$. But then, there is a positive measure of traders that prefer to sell slowly. As we have shown there are no traders that prefer to buy slowly, the slow venue clearing condition cannot hold and this is a contradiction to the assertion that that we have an equilibrium.

□

We start by proving some lemmas:

Lemma 2. *(i)$S_s \subseteq N \cup B_f$, (ii) $S_f \subseteq N \cup B_s$, (iii) $B_s \subseteq H \cup S_f$, (iv) $B_f \subseteq H \cup S_s$*

Proof. To prove (i) and (iii), assume for a contradiction there exist $\eta \in S_s \cap B_s$. Then we have $V_{o,\eta} = V_S^s(\eta)$ and $V_{no,\eta} = V_B^s(\eta)$. Then by substituting for RHS and summing:

$$(\gamma + \rho)(V_{o,\eta} + V_{no,\eta}) = u(\eta) + \gamma(\mathbb{E}_{\eta'}[V_{o,\eta'}] + \mathbb{E}_{\eta'}[V_{no,\eta'}])) - 2\sigma_s(c_s + \theta). \tag{29}$$

From the optimality of selling slow, we have:

$$\begin{aligned}(\sigma_s + \gamma + \rho)(\gamma + \rho)[V_S^s(\eta) - V_H] \\ &= -\sigma_s(u(\eta) + \gamma \mathbb{E}_{\eta'}[V_{o,\eta}]) + \sigma_s(\gamma + \rho)(V_{no,\eta} + p_s - c_s - \theta) \\ &\geqslant 0.\end{aligned}$$

Rearranging implies:

$$(\gamma + \rho)V_{no,\eta} \geqslant u(\eta) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}] - (\gamma + \rho)(p_s - c_s - \theta). \tag{30}$$

From the optimality of buying slow, using $V_B^s(\eta) - V_N(\eta) > 0$ we obtain:

$$(\gamma + \rho)V_{o,\eta} \geqslant \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}] + (\gamma + \rho)(p_s + c_s + \theta). \tag{31}$$

Summing equations 30 and 31, we obtain:

$$(\gamma + \rho)(V_{0,\eta} + V_{1,\eta}) \geqslant u(\eta) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}] + \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}] + 2(\gamma + \rho)(c_s + \theta).$$

Which contradicts equation 29. Replacing s with f in the above proof proves (ii) and (iv).

$\square$

Next Lemma shows that fast sellers do nothing after selling and slow buyers hold after buying.

Lemma 3. *(i)* $S_f \cap B_s = \varnothing$, *(ii)* $S_s \cap B_f = \varnothing$

Proof. To prove (i), assume for a contradiction $\eta \in S_f \cap B_s$. Then as $\eta \in B_s$, we have:

$$\begin{aligned} V_B^s(\eta) - V_N(\eta) &> 0 \\ (\gamma + \rho)\sigma_s(V_{1,\eta} - p_s - c_s - \theta) + (\gamma + \rho)\gamma \mathbb{E}_{\eta'}[V_{0,\eta'}] - \gamma(\sigma_s + \gamma + \rho)\mathbb{E}_{\eta'}[V_{0,\eta'}] &> 0 \\ (\gamma + \rho)\sigma_s(V_{1,\eta} - p_s - c_s - \theta) - \sigma_s \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}] &> 0,\end{aligned}$$

which implies:

$$
\begin{aligned}
(\sigma_s+\gamma+\rho)(V_{1,\eta}-p_s-c_s-\theta) &> \sigma_s(V_{1,\eta}-p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{0,\eta'}] \\
V_{1,\eta}-p_s-c_s-\theta &> V_B^s(\eta)=V_{0,\eta} \\
V_{1,\eta} &> V_{0,\eta}+p_s+c_s+\theta. \qquad (32)
\end{aligned}
$$

This is intuitive as the individual prefers the continuation value while holding the asset and paying p_s to the value of not holding the asset. From $\eta \in S_f$, we have:

$$V_S^f(\eta)-V_H(\eta)>0.$$

Similar calculations as above yield:

$$V_{1,\eta}<V_{0,\eta}+p_f-c_f-\theta. \qquad (33)$$

Equations 32 and 33 imply $p_f-c_f>p_s+c_s+2\theta$. Subtracting $2c_s-2\theta$ from the left-hand side, we obtain $p_f-c_f>p_s-c_s$ which is a contradiction. Switching s with f in the above proof proves (ii). $\square$

The following lemma is immediate given the previous ones:

Lemma 4.

$$
\begin{aligned}
V_S^f(\eta) &= \frac{u(\eta)+\sigma_f(V_N+p_f-c_f-\theta)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_f+\gamma+\rho)}, \\
V_S^s(\eta) &= \frac{u(\eta)+\sigma_s(V_N+p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s+\gamma+\rho)}, \\
V_B^f(\eta) &= \frac{\sigma_f(V_H(\eta)-p_f-c_f-\theta)+\gamma\mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_f+\gamma+\rho)}, \\
V_B^s(\eta) &= \frac{\sigma_s(V_H(\eta)-p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_s+\gamma+\rho)}.
\end{aligned}
$$

The following lemma helps us prove the structure of the speed choices.

Lemma 5. *$\frac{\partial V_{1,\eta}}{\partial\eta}>0$ and $\frac{\partial V_{0,\eta}}{\partial\eta}\geqslant 0$.*

Proof. As $u(\eta)$ is strictly increasing, we have the following: $\frac{\partial V_N(\eta)}{\partial\eta}=0$, and $\frac{\partial V_H(\eta)}{\partial\eta}>0$.

Next, assume $\eta \in B_s$. Take any $\eta' > \eta$. We have:

$$V_{0,\eta'} \geqslant V_B^s(\eta') > V_B^s(\eta) = V_{0,\eta},$$

where the first inequality follows from the optimality of $V_{no,\eta}$, the second follows from the fact $u(\eta)$ is strictly increasing, and the third equality follows from $\eta \in B_s$. Similarly, assume $\eta \in B_f$. Take any $\eta' > \eta$. With the exact same reasoning, we obtain:

$$V_{0,\eta'} \geqslant V_B^f(\eta') > V_B^f(\eta) = V_{0,\eta}.$$

Hence, $\frac{\partial V_{0,\eta}}{\partial \eta} \geqslant 0$, which proves the second claim. Given this, it is immediate to conclude $V_H(\eta), V_S^s(\eta)$ and $V_S^f(\eta)$ are all strictly increasing in η as $u(\eta)$ is strictly increasing. Thus, $V_{1,\eta}$, which is their maximum, is strictly increasing, $\frac{\partial V_{1,\eta}}{\partial \eta} > 0$, which proves the first claim. □

Lemma 6. *There exist cutoffs η_1 and η_2 such that $N = [\eta_l, \eta_1]$, $B_s = [\eta_1, \eta_2]$ and $B_f = [\eta_2, \eta_h]$*

Proof. Let $\eta_1 = \sup N$ and $\eta_2 = \inf B_f$. Notice that given Lemma 5, the differences: $V_B^f(\eta) - V_B^s(\eta)$, $V_B^s(\eta) - V_N(\eta)$ and $V_B^f(\eta) - V_N(\eta)$ are all strictly increasing in η. Then, if $\eta \in B_f$, we have $V_B^f(\eta) > V_B^s(\eta)$ and $V_B^f(\eta) > V_N(\eta)$. Then as above differences are increasing in η, $\eta' > \eta$ implies $V_B^f(\eta') > V_B^s(\eta')$ and $V_B^f(\eta') > V_N(\eta')$. Hence, $\eta' \in B_f$, which proves that $B_f = [\eta_2, \eta_h]$.

Similarly, if $\eta \in N$, then $V_N(\eta) > V_B^s(\eta)$ and $V_N(\eta) > V_B^f(\eta)$. Then as above differences are increasing in η, $\eta' < \eta$ implies $V_N(\eta') > V_B^s(\eta')$ and $V_N(\eta') > V_B^f(\eta')$. Hence, $\eta' \in N$, which proves that $N = [\eta_l, \eta_1]$. The fact that $B_s = [\eta_1, \eta_2]$ follows immediately.

□

Lemma 7. *There exists cutoffs η_3 and η_4 such that $S_f = [\eta_l, \eta_3]$, $S_s = [\eta_3, \eta_4]$ and $H = [\eta_4, \eta_h]$*

Proof. Let $\eta_3 = \sup\{\eta \in S_f\}$ and $\eta_4 = \inf\{\eta \in H\}$. Notice that the differences $V_H(\eta) - V_S^s(\eta)$, $V_H(\eta) - V_S^f(\eta)$, $V_S^s(\eta) - V_S^f(\eta)$ are all strictly increasing in η. Then, if $\eta \in H$, we have $V_H(\eta) > V_S^s(\eta)$ and $V_H(\eta) > V_S^f(\eta)$. Then as above differences are increasing in η, $\eta' > \eta$ implies $V_H(\eta') > V_S^s(\eta')$ and $V_H(\eta') > V_S^f(\eta)$. Hence, $\eta' \in H$, which proves that $H = [\eta_4, \eta_h]$. Similarly, if $\eta \in S_s^f$, then $V_S^f(\eta) > V_H(\eta)$ and $V_S^f(\eta) > V_S^s(\eta)$. Then, as above differences are increasing in η, $\eta' < \eta$ implies $V_S^f(\eta') > V_H(\eta')$ and $V_S^f(\eta') > V_S^s(\eta')$. Hence, $\eta' \in S_f$, which proves that $S_f = [\eta_l, \eta_3]$. The fact that $S_s = [\eta_3, \eta_4]$ then follows.

□

The structure follows from Lemmas 6 and 7. The fact that $\eta_1 < \eta_4$ follows from $S_f \cap B_s = \varnothing$ and $S_s \cap B_s = \varnothing$. The following lemma proves part (i) of the proposition.

Lemma 8. *If $u(\eta_h) > 2(c_s + \theta)(\gamma + \rho)$, then there is no trade.*

Proof. See that the following two conditions are necessary for any trader to prefer trade in any equilibrium:

$$V_B^s(\eta_h) > V_N, \tag{34}$$

$$V_S^s(0) > V_H(0). \tag{35}$$

From 34 we have:

$$\frac{\sigma_s(V_{o,\eta_h} - p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_s + \gamma + \rho)} > \frac{\gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\gamma + \rho)},$$

which reduces to $(\gamma + \rho)\sigma_s(V_{o,\eta_h} - p_s - c_s - \theta) > \sigma_s \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]$. Therefore,

$$(\gamma + \rho)(V_H(\eta_h) - p_s - c_s - \theta) > (\gamma + \rho)V_N. \tag{36}$$

From 35, using $u(0) = 0$, we have $\frac{\sigma_s(V_{no,0} + p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s + \gamma + \rho)} > \frac{\gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\gamma + \rho)}$. Hence,

$$(\gamma + \rho)(V_N + p_s - c_s - \theta) > (\gamma + \rho)V_H(0). \tag{37}$$

Summing 36 and 37:

$$\begin{aligned} V_H(\eta_h) + V_N - 2(c_s + \theta) &> V_N + V_H(0), \\ V_H(\eta_h) - V_H(0) &> 2(c_s + \theta), \\ u(\eta_h) &> 2(c_s + \theta)(\gamma + \rho). \end{aligned}$$

$\square$

The following lemma characterizes the condition under which the fast venue is active in any equilibrium:

Lemma 9. *There is positive trade in fast venue only if*

$$u(\eta_h) > 2\frac{\sigma_f(\sigma_s+\gamma+\rho)(c_f+\theta)-\sigma_s(\sigma_f+\gamma+\rho)(c_s+\theta)}{\sigma_f-\sigma_s}.$$

Proof. Given the structure of cutoffs, one necessary condition for positive trade in the fast venue is $V_S^s(0) < V_S^f(0)$. As $u(0)=0$, this is equivalent to:

$$\frac{\sigma_s(V_N+p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s+\gamma+\rho)} < \frac{\sigma_f(V_N+p_f-c_f-\theta)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_f+\gamma+\rho)}.$$

After some algebra, we obtain:

$$V_N - V_H(0) > -\frac{\sigma_f(\sigma_s+\gamma+\rho)(p_f-c_f-\theta)-\sigma_s(\sigma_f+\gamma+\rho)(p_s-c_s-\theta)}{(\sigma_f-\sigma_s)(\gamma+\rho)}. \tag{38}$$

Another necessary condition is $V_B^s(\eta_h) < V_B^f(\eta_h)$. Doing similar algebra as above, we obtain:

$$V_H(\eta_h) - V_N > \frac{\sigma_f(\sigma_s+\gamma+\rho)(p_f+c_f+\theta)-\sigma_s(\sigma_f+\gamma+\rho)(p_s+c_s+\theta)}{(\sigma_f-\sigma_s)(\gamma+\rho)}. \tag{39}$$

Summing equations 38 and 39, we obtain

$$u(\eta_h) > 2\frac{\sigma_f(\sigma_s+\gamma+\rho)(c_f+\theta)-\sigma_s(\sigma_f+\gamma+\rho)(c_s+\theta)}{\sigma_f-\sigma_s},$$

which, after observing that the conditions are together sufficient, completes the proof. $\square$

First, we find the distributions for f_o and f_{no} (we leave the values undefined at the cutoffs $\{\eta_i\}_{i=1,\ldots,4}$ as values in individual points do not matter):

Lemma 10.

$$f_{no}(\eta)=\begin{cases} f(\eta)\frac{\sigma_f+(1-Z)\gamma}{\sigma_f+\gamma} & \textit{if } \eta\in(\eta_l,\eta_3)\\ f(\eta)\frac{\sigma_s+(1-Z)\gamma}{\sigma_s+\gamma} & \textit{if } \eta\in(\eta_3,\eta_4)\\ f(\eta)(1-Z) & \textit{if } \eta\in(\eta_4,\eta_1)\\ f(\eta)\frac{(1-Z)\gamma}{\sigma_s+\gamma} & \textit{if } \eta\in(\eta_1,\eta_2)\\ f(\eta)\frac{(1-Z)\gamma}{\sigma_f+\gamma} & \textit{if } \eta\in(\eta_2,\eta_h)\end{cases},\quad f_o(\eta)=\begin{cases} f(\eta)\frac{Z\gamma}{\sigma_f+\gamma} & \textit{if } \eta\in(\eta_l,\eta_3)\\ f(\eta)\frac{Z\gamma}{\sigma_s+\gamma} & \textit{if } \eta\in(\eta_3,\eta_4)\\ f(\eta)Z & \textit{if } \eta\in(\eta_4,\eta_1)\\ f(\eta)\frac{\gamma Z+\sigma_s}{\sigma_s+\gamma} & \textit{if } \eta\in(\eta_1,\eta_2)\\ f(\eta)\frac{\gamma Z+\sigma_f}{\sigma_f+\gamma} & \textit{if } \eta\in(\eta_2,\eta_h)\end{cases}$$

Proof. Let $\eta\in(\eta_l,\eta_3)$. The outflow of asset owners with valuation η is $\gamma f_o(\eta)+\sigma_f f_o(\eta)$, while the inflow is $\gamma Z f(\eta)$, hence we have:

$$\gamma f_o(\eta)+\sigma_f f_o(\eta)=\gamma Z f(\eta)\implies f_o(\eta)=\frac{\gamma Z}{\sigma_f+\gamma}f(\eta).$$

As $f_o(\eta)+f_{no}(\eta)=f(\eta)$, we have:

$$f_{no}(\eta)=\frac{\sigma_f+(1-Z)\gamma}{\sigma_f+\gamma}f(\eta).$$

Let $\eta\in(\eta_3,\eta_4)$. Inflow-outflow balance requires:

$$\gamma f_o(\eta)+\sigma_s f_o(\eta)=\gamma Z f(\eta)\implies f_o(\eta)=\frac{\gamma Z}{\sigma_s+\gamma}f(\eta).$$

As $f_o(\eta)+f_{no}(\eta)=f(\eta)$, we have

$$f_{no}(\eta)=\frac{\sigma_s+(1-Z)\gamma}{\sigma_s+\gamma}f(\eta).$$

Let $\eta\in(\eta_4,\eta_1)$. The inflow-outflow balance for owners with valuation η requires:

$$\gamma f_o(\eta)=\gamma Z f(\eta)\implies f_o(\eta)=Zf(\eta).$$

Therefore, $f_{no}(\eta)=(1-Z)f(\eta)$. Let $\eta\in(\eta_1,\eta_2)$. The inflow-outflow balance requires:

$$\gamma f_{no}(\eta)+\sigma_s f_{no}(\eta)=(1-Z)\gamma f(\eta)\implies f_{no}(\eta)=f(\eta)\frac{(1-Z)\gamma}{\gamma+\sigma_s}.$$

As $f_o(\eta) + f_{no}(\eta) = f(\eta)$, we have $f_o(\eta) = f(\eta)\frac{\gamma Z+\sigma_s}{\gamma+\sigma_s}$. Let $\eta \in (\eta_2, \eta_h)$. The inflow-outflow balance requires

$$\gamma f_{no}(\eta) + \sigma_f f_{no}(\eta) = (1-Z)\gamma f(\eta) \implies f_{no}(\eta) = f(\eta)\frac{(1-Z)\gamma}{\gamma+\sigma_f}.$$

As $f_o(\eta) + f_{no}(\eta) = f(\eta)$, we have $f_o(\eta) = f(\eta)\frac{\gamma Z+\sigma_f}{\gamma+\sigma_f}$. □

We first need to find the market clearing conditions.

Lemma 11. *In any equilibrium, asset market clearing conditions imply:*

$$(1-Z)F(\eta_1) + ZF(\eta_4) = 1-Z, \tag{40}$$

$$(1-Z)F(\eta_2) + ZF(\eta_3) = 1-Z. \tag{41}$$

Proof. We have two market clearing conditions: one for the slow venue and one for the fast venue. Fast venue clearing condition:

$$\int_{\eta_2}^{\eta_h} f_{no}(\eta)d\eta = \int_{\eta_l}^{\eta_3} f_o(\eta)d\eta \tag{42}$$

$$(1-F(\eta_2))\frac{\gamma(1-Z)}{\sigma_f+\gamma} = F(\eta_3)\frac{\gamma Z}{\sigma_f+\gamma} \tag{43}$$

$$\frac{1-Z}{Z} = \frac{F(\eta_3)}{1-F(\eta_2)} \tag{44}$$

$$ZF(\eta_3) + (1-Z)F(\eta_2) = (1-Z). \tag{45}$$

From slow market clearing condition:

$$\int_{\eta_1}^{\eta_2} f_{no}(\eta)d\eta = \int_{\eta_3}^{\eta_4} f_o(\eta)d\eta \tag{46}$$

$$\Big(F(\eta_2) - F(\eta_1)\Big)\frac{\gamma(1-Z)}{\sigma_s+\gamma} = \Big(F(\eta_4) - F(\eta_3)\Big)\frac{\gamma Z}{\sigma_s+\gamma} \tag{47}$$

$$\frac{1-Z}{Z} = \frac{F(\eta_4)-F(\eta_3)}{F(\eta_2)-F(\eta_1)} \tag{48}$$

$$(1-Z)F(\eta_1) + ZF(\eta_4) = (1-Z), \tag{49}$$

where the last line is obtained by plugging in fast market clearing equality. □

Lemma 12. *If* $u(\eta_h) > 2\frac{\sigma_f(\sigma_s+\gamma+\rho)(c_f+\theta)-\sigma_s(\sigma_f+\gamma+\rho)(c_s+\theta)}{\sigma_f-\sigma_s}$*, then in any equilibrium there is positive trade in both venues. The equilibrium cutoffs are given by:*

$$\frac{u(\eta_1)-u(\eta_4)}{\gamma+\rho} = 2(c_s+\theta), \tag{50}$$

$$\frac{u(\eta_2)-u(\eta_3)}{\gamma+\rho} = 2\frac{\sigma_f(c_f+\theta)(\sigma_s+\gamma+\rho)-\sigma_s(c_s+\theta)(\sigma_f+\gamma+\rho)}{(\sigma_f-\sigma_s)(\gamma+\rho)}. \tag{51}$$

Proof. It observes that

$$\begin{aligned} V_B^s(\eta_1) &= \frac{\sigma_s(V_H(\eta_1)-p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_s+\gamma+\rho)}, \\ &= \frac{\sigma_s(V_H(\eta_1)-p_s-c_s-\theta)+(\gamma+\rho)V_N}{(\sigma_s+\gamma+\rho)}, \\ &= \frac{\sigma_s(V_H(\eta_1)-V_N-p_s-c_s-\theta)}{(\sigma_s+\gamma+\rho)}+V_N. \end{aligned}$$

In addition, as η_1 is the cutoff type for buying slowly and doing nothing, we have $V_B^s(\eta_1) = V_N$[7]. Combining these:

$$V_H(\eta_1) = V_N + p_s + c_s + \theta. \tag{52}$$

Similarly, as η_4 is the cutoff type for selling slowly and holding the asset, we have $V_H(\eta_4) = V_S^s(\eta_4)$. Hence,

$$\begin{aligned} V_H(\eta_4) &= \frac{u(\eta_4)+\sigma_s(V_N+p_s-c_s-\theta)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s+\gamma+\rho)}, \\ &= \frac{u(\eta_4)+\gamma\mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s+\gamma+\rho)}+\frac{\sigma_s(V_N+p_s-c_s-\theta)}{(\sigma_s+\gamma+\rho)}, \\ &= \frac{(\gamma+\rho)V_H(\eta_4)}{(\sigma_s+\gamma+\rho)}+\frac{\sigma_s(V_N+p_s-c_s-\theta)}{(\sigma_s+\gamma+\rho)}. \end{aligned}$$

T last equality implies

$$V_H(\eta_4) = V_N + p_s - c_s - \theta. \tag{53}$$

[7]Notice that $V_N(\eta) = V_N$ for any η

Using equations 52 and 53, we get $V_H(\eta_1) - V_H(\eta_4) = 2(c_s + \theta)$. Therefore,

$$V_H(\eta_1) - V_H(\eta_4) = \frac{u(\eta_1) - u(\eta_4)}{\gamma + \rho} = 2(c_s + \theta).$$

Now we use $V_S^s(\eta_3) = V_S^f(\eta_3)$, so

$$\frac{u(\eta_3) + \sigma_s(V_N + p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s + \gamma + \rho)} = \frac{u(\eta_3) + \sigma_f(V_N + p_f - c_f - \theta) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_f + \gamma + \rho)}. \tag{54}$$

Rearranging implies:

$$(\sigma_f - \sigma_s)(u(\eta_3) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]) = (\sigma_f - \sigma_s)(\gamma + \rho)V_N \\ + \sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta). \tag{55}$$

Dividing by $(\sigma_f - \sigma_s)(\gamma + \rho)$:

$$\begin{aligned} V_N &= \frac{u(\eta_3)}{\gamma + \rho} + \frac{\gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{\gamma + \rho} - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} \\ &= V_H(\eta_3) - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)}. \end{aligned} \tag{56}$$

Similarly, using $V_B^f(\eta_2) = V_B^s(\eta_2)$, we obtain:

$$\frac{\sigma_f(V_H(\eta_2) - p_f - c_f - \theta) + \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_f + \gamma + \rho)} = \frac{\sigma_s(V_H(\eta_2) - p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_s + \gamma + \rho)}.$$

Rearranging and dividing by $(\sigma_f - \sigma_s)(\gamma + \rho)$:

$$V_H(\eta_2) - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f + c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s + c_s + \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} = V_N. \tag{57}$$

Solving 56 and 57 gives

$$\frac{u(\eta_2) - u(\eta_3)}{\gamma + \rho} = 2\frac{\sigma_f(c_f + \theta)(\sigma_s + \gamma + \rho) - \sigma_s(c_s + \theta)(\sigma_f + \gamma + \rho)}{(\sigma_f - \sigma_s)(\gamma + \rho)}.$$

Note that if $u(\eta_h) > 2\frac{\sigma_f(\sigma_s+\gamma+\rho)(c_f+\theta) - \sigma_s(\sigma_f+\gamma+\rho)(c_s+\theta)}{\sigma_f - \sigma_s}$, then there exists $\eta_2 < \eta_h$

and $\eta_3 > 0$ such that equation above holds. Moreover, any type $\eta < \eta_3$ and $\eta > \eta_2$ strictly prefers fast venue to slow venue and there is positive trade in the fast venue. $\square$

Next, note that $u(\eta_h) > 2\frac{\sigma_f(\sigma_s+\gamma+\rho)(c_f+\theta)-\sigma_s(\sigma_f+\gamma+\rho)(c_s+\theta)}{\sigma_f-\sigma_s}$ implies $u(\eta_h) > 2(c_s + \theta)(\gamma+\rho)$ whenever $c_f > c_s$, which is assumed. Thus, whenever the former inequality holds, η_1, η_2, η_3 and η_4 that solves equations 40, 41, 50 and 51 constitutes equilibrium cutoffs where both venues are active. To finish the proof of part (ii) of theorem 1 we characterize the equilibrium prices in case (ii). From equation 53, we obtain the characterization of p_s:

$$\begin{aligned}
p_s &= V_h(\eta_4) - V_N + c_s + \theta, \\
&= \frac{u(\eta_4) + \gamma \mathbb{E}_{\eta'}[V_{1,\eta'}]}{\gamma+\rho} + c_s + \theta - \frac{\gamma \mathbb{E}_{\eta'}[V_{0,\eta'}]}{\gamma+\rho}, \\
&= \frac{u(\eta_4)}{\gamma+\rho} + c_s + \theta + \frac{\gamma}{\gamma+\rho}\Big(\mathbb{E}_{\eta'}[V_{1,\eta'}] - \mathbb{E}_{\eta'}[V_{0,\eta'}]\Big).
\end{aligned} \tag{58}$$

From equation 56:

$$\begin{aligned}
&\frac{\sigma_f(\sigma_s+\gamma+\rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f+\gamma+\rho)(p_s - c_s - \theta)}{(\sigma_f-\sigma_s)(\gamma+\rho)} \\
&\qquad\qquad = \frac{u(\eta_3)}{\gamma+\rho} + \frac{\gamma}{\gamma+\rho}\Big(\mathbb{E}_{\eta'}[V_{1,\eta'}] - \mathbb{E}_{\eta'}[V_{0,\eta'}]\Big).
\end{aligned} \tag{59}$$

The following lemma shows that 58 and 59 yields a unique price vector and completes the characterization of equilibrium.

Lemma 13. *Equations 58 and 59 characterize a unique price vector.*

Proof. We only need to show that (58) and (59) lead to unique solutions for p_s and p_f. In particular, we will show that the quantity $\mathbb{E}_{\eta'}[V_{o,\eta'}] - \mathbb{E}_{\eta'}[V_{no,\eta'}]$ does not depend on either p_s nor p_f; once this is done, p_s is directly pinned down by (58), and after substituting into (59), p_f is also uniquely determined. To finish up, we will need to show that the numbers p_s and p_f recovered this way are positive: this is done in the last step of the proof.

To begin, in the following two steps we establish that $\mathbb{E}_{\eta'}[V_{o,\eta'}] - \mathbb{E}_{\eta'}[V_{no,\eta'}]$

only depends on the endogenous thresholds η_1, η_2, η_3 and η_4. First recall that

$$
\begin{aligned}
V_S^f(\eta) &= \frac{u(\eta) + \sigma_f(V_N + p_f - c_f - \theta) + \gamma \mathbb{E}_{\eta'}[V_{o,\eta'}]}{\sigma_f + \rho + \gamma}, \\
V_S^s(\eta) &= \frac{u(\eta) + \sigma_s(V_N + p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{o,\eta'}]}{\sigma_s + \rho + \gamma}, \\
V_B^f(\eta) &= \frac{\sigma_f(V_H(\eta) - p_f - c_f - \theta) + \gamma \mathbb{E}_{\eta'}[V_{no,\eta'}]}{\sigma_f + \rho + \gamma}, \\
V_B^s(\eta) &= \frac{\sigma_s(V_H(\eta) - p_s - c_s - \theta) + \gamma \mathbb{E}_{\eta'}[V_{no,\eta'}]}{\sigma_s + \rho + \gamma}, \\
V_H(\eta) &= \frac{u(\eta) + \gamma \mathbb{E}_{\eta'}[V_{o,\eta'}]}{\gamma + \rho}, \\
V_N &= \frac{\gamma \mathbb{E}_{\eta'}[V_{no,\eta'}]}{\gamma + \rho}.
\end{aligned} \tag{60}
$$

Next we derive $\mathbb{E}_{\eta'}[V_{no,\eta'}]$ and $\mathbb{E}_{\eta'}[V_{o,\eta'}]$ in closed forms.

- **Computing $\mathbb{E}_{\eta'}[V_{no,\eta'}]$.** Using (60) we have

$$
\begin{aligned}
\frac{dV_{no,\eta}}{d\eta} = \mathbf{1}_{\{\eta \in [\eta_l, \eta_1]\}} \times 0 + \mathbf{1}_{\{\eta \in [\eta_1, \eta_2]\}} & \frac{\sigma_s u'(\eta)}{(\sigma_s + \rho + \gamma)(\rho + \gamma)} \\
+ \mathbf{1}_{\{\eta \in [\eta_2, \eta_h]\}} & \frac{\sigma_f u'(\eta)}{(\sigma_f + \rho + \gamma)(\rho + \gamma)}.
\end{aligned} \tag{61}
$$

  Integrating (61) implies

$$
\begin{aligned}
V_{no,\eta} = \mathbf{1}_{\{\eta \in [\eta_l, \eta_1]\}} V_N + \mathbf{1}_{\{\eta \in [\eta_1, \eta_2]\}} & \left[V_N + \int_{\eta_1}^{\eta} \frac{\sigma_s u'(\xi)}{(\sigma_s + \rho + \gamma)(\rho + \gamma)} d\xi \right] \\
+ \mathbf{1}_{\{\eta \in [\eta_2, \eta_h]\}} & \left[V_N + \int_{\eta_1}^{\eta_2} \frac{\sigma_s u'(\xi)}{(\sigma_s + \rho + \gamma)(\rho + \gamma)} d\xi + \int_{\eta_2}^{\eta} \frac{\sigma_f u'(\xi)}{(\sigma_f + \rho + \gamma)(\rho + \gamma)} d\xi \right].
\end{aligned} \tag{62}
$$

Finally, taking an expectation from (62) gives

$$\mathbb{E}_{\eta'}[V_{no,\eta'}] = V_N + \int_{\eta_1}^{\eta_2} \left(\int_{\eta_1}^{\tilde{\eta}} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi \right) dF(\tilde{\eta}) + \left(\int_{\eta_1}^{\eta_2} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi \right) \Big(1 - F(\eta_2)\Big) + \int_{\eta_2}^{\eta_h} \left(\int_{\eta_2}^{\tilde{\eta}} \frac{\sigma_f u'(\xi)}{(\sigma_f+\rho+\gamma)(\rho+\gamma)} d\xi \right) dF(\tilde{\eta}). \tag{63}$$

Thus, (63) shows that $\mathbb{E}_{\eta'}[V_{no,\eta'}]$ only depends on the endogenous thresholds η_1 and η_2.

- **Computing $\mathbb{E}_{\eta'}[V_{o,\eta'}]$.** Using (60) we have

$$\frac{dV_{o,\eta}}{d\eta} = \mathbf{1}_{\{\eta\in[\eta_l,\eta_3]\}} \frac{u'(\eta)}{\sigma_f+\rho+\gamma} + \mathbf{1}_{\{\eta\in[\eta_3,\eta_4]\}} \frac{u'(\eta)}{\sigma_s+\rho+\gamma} + \mathbf{1}_{\{\eta\in[\eta_4,\eta_h]\}} \frac{u'(\eta)}{\rho+\gamma}. \tag{64}$$

Then, integrating gives

$$V_{o,\eta} = \mathbf{1}_{\{\eta\in[\eta_l,\eta_3]\}} \int_{\eta_l}^{\eta} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi + \mathbf{1}_{\{\eta\in[\eta_3,\eta_4]\}} \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi + \int_{\eta_3}^{\eta} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) + \mathbf{1}_{\{\eta\in[\eta_4,\eta_h]\}} \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi + \int_{\eta_3}^{\eta_4} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi + \int_{\eta_4}^{\eta} \frac{u'(\xi)}{\gamma+\rho} d\xi \right). \tag{65}$$

Therefore, taking an expectation from (65) gives

$$\mathbb{E}_{\eta'}[V_{o,\eta'}] = \int_{\eta_l}^{\eta_3} \left(\int_{\eta_l}^{\tilde{\eta}} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi \right) dF(\tilde{\eta}) + \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi \right) \Big(1 - F(\eta_3)\Big) + \int_{\eta_3}^{\eta_4} \left(\int_{\eta_3}^{\tilde{\eta}} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) dF(\tilde{\eta}) + \Big(1 - F(\eta_4)\Big) \left(\int_{\eta_3}^{\eta_4} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) + \int_{\eta_4}^{\eta_h} \left(\int_{\eta_4}^{\tilde{\eta}} \frac{u'(\xi)}{\gamma+\rho} d\xi \right) dF(\tilde{\eta}). \tag{66}$$

Thus, (66) shows that $\mathbb{E}_{\eta'}[V_{o,\eta'}]$ only depends on the endogenous thresholds η_3 and η_4. Together, (63) and (66) finish the proof that the difference $\mathbb{E}_{\eta'}[V_{o,\eta'}] - \mathbb{E}_{\eta'}[V_{no,\eta'}]$ does not depend on p_s nor p_f.

To finish the proof of the lemma, note that equations (58) and (59) yield unique solutions for p_s and p_f. Since $\mathbb{E}_{\eta'}[V_{o,\eta'}] - \mathbb{E}_{\eta'}[V_{no,\eta'}] > 0$, all terms in (58) are positive, and hence $p_s > 0$. Since we have assumed that $u(\cdot) \geqslant 0$, it follows that p_f is also positive, which finishes the lemma. $\square$

To prove part (iii), assume $u(\eta_h) > 2(c_s + \theta)(\gamma + \rho)$ and

$$u(\eta_h) < 2\frac{\sigma_f(\sigma_s + \gamma + \rho)(c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(c_s + \theta)}{\sigma_f - \sigma_s}.$$

From Lemma 9, we know that there cannot be any trade in the fast venue, i.e. $\eta_3 = 0$ and $\eta_2 = \eta_h$. Lemmas 10 and 11 still hold, and given $u(\eta_h) > 2(c_s + \theta)(\gamma + \rho)$, following the same steps in lemma 12, we see that the cut-offs η_1 and η_4 are uniquely pinned down by:

$$(1 - Z)F(\eta_4) + ZF(\eta_1) = 1 - Z, \tag{67}$$

$$\frac{u(\eta_1) - u(\eta_4)}{\gamma + \rho} = 2(c_s + \theta). \tag{68}$$

As in the earlier case, the equilibrium price in the slow venue is given by:

$$p_s = \frac{u(\eta_4)}{\gamma + \rho} + c_s + \theta + \frac{\gamma}{\gamma + \rho}\Big(\mathbb{E}_{\eta'}[V_{1,\eta'}] - \mathbb{E}_{\eta'}[V_{0,\eta'}]\Big).$$

To show that this is indeed an equilibrium, we need to find a price p_f such that there is no demand for trade in the fast venue (otherwise, the fast venue clearing condition cannot hold.). To see that, there is no demand for selling in the fast venue if $V_S^s(0) - V_S^f(0) \geqslant 0$. This corresponds to:

$$\begin{aligned}&(V_H(0) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)\\&\qquad - (p_f - c_f - \theta)\sigma_f(\sigma_s + \gamma + \rho)\\&\geqslant 0,\end{aligned}$$

which is equivalent to:

$$
\begin{aligned}
L = &(V_H(0) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta) \\
&+ (c_f + \theta)\sigma_f(\sigma_s + \gamma + \rho) \\
\geqslant\ & p_f(\sigma_f(\sigma_s + \gamma + \rho)).
\end{aligned}
$$

Similarly, there is no demand for buying in the fast venue if $V_B^s(\eta_h) - V_B^f(\eta_h) \geqslant 0$. This is equivalent to:

$$
\begin{aligned}
U = &(V_H(\eta_h) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s + c_s + \theta) \\
&- (c_f + \theta)\sigma_f(\sigma_s + \gamma + \rho) \\
\leqslant\ & p_f(\sigma_f(\sigma_s + \gamma + \rho)).
\end{aligned}
$$

Note that there exists a p_f such that there is no demand in selling or buying fast if $L > U$. The following condition is sufficient to have $L > U$:

$$
u(\eta_h) < 2\frac{\sigma_f(\sigma_s + \gamma + \rho)(c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(c_s + \theta)}{\sigma_f - \sigma_s}.
$$

This condition holds under the assumptions of (iii), thus $L > U$. Hence, under any $p_f \in (U, L)$, there is no demand for fast venue and p_f fast venue clearing condition is satisfied, finishing the characterization of the equilibrium and Theorem 1.

Proof of Proposition 1

First, we prove a short lemma:

Lemma 14. *Let x be any variable. Then if $\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial x} > (<)0$, then $\frac{\partial \eta_2}{\partial x} > (<)0$ and $\frac{\partial \eta_3}{\partial x} < (>)0$. Similarly, if $\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial x} > (<)0$, then $\frac{\partial \eta_1}{\partial x} > (<)0$ and $\frac{\partial \eta_4}{\partial x} < (>)0$.*

Proof. We prove this for the first case, rest is similar. From equation 9, we see that if $\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial x} > 0$, this is only possible under $\frac{\partial u(\eta_2)}{\partial x} > 0$ and $\frac{\partial u(\eta_3)}{\partial x} < 0$.[8] Then $\frac{\partial \eta_2}{\partial x} > 0$ and $\frac{\partial \eta_3}{\partial x} < 0$ follows from the fact that u is strictly increasing. □

[8] It is clear that one of these must hold. To see why both are necessary, see that equation 9 requires cut-offs to move in opposite direction.

The following lemma shows all the comparative statics of the model:

Lemma 15. *We have the following comparative statics:*

1. $\frac{\partial \eta_1}{\partial \sigma_s} = \frac{\partial \eta_4}{\partial \sigma_s} = 0$, $\frac{\partial \eta_2}{\partial \sigma_s} > 0$, $\frac{\partial \eta_3}{\partial \sigma_s} < 0$
2. $\frac{\partial \eta_4}{\partial \sigma_f} = 0$, $\frac{\partial \eta_1}{\partial \sigma_f} = 0$, $\frac{\partial \eta_2}{\partial \sigma_f} < 0$, $\frac{\partial \eta_3}{\partial \sigma_f} > 0$
3. $\frac{\partial \eta_1}{\partial \gamma} > 0$, $\frac{\partial \eta_4}{\partial \gamma} < 0$, $\frac{\partial \eta_2}{\partial \gamma} > 0$, $\frac{\partial \eta_3}{\partial \gamma} < 0$
4. $\frac{\partial \eta_1}{\partial \rho} > 0$, $\frac{\partial \eta_4}{\partial \rho} < 0$, $\frac{\partial \eta_2}{\partial \rho} > 0$, $\frac{\partial \eta_3}{\partial \rho} < 0$
5. $\frac{\partial \eta_1}{\partial c_s} < 0$, $\frac{\partial \eta_4}{\partial c_s} > 0$, $\frac{\partial \eta_2}{\partial c_s} < 0$, $\frac{\partial \eta_3}{\partial c_s} > 0$
6. $\frac{\partial \eta_1}{\partial c_f} = 0$, $\frac{\partial \eta_4}{\partial c_f} = 0$, $\frac{\partial \eta_2}{\partial c_f} > 0$, $\frac{\partial \eta_3}{\partial c_f} < 0$
7. $\frac{\partial \eta_1}{\partial \theta} < 0$, $\frac{\partial \eta_4}{\partial \theta} > 0$, $\frac{\partial \eta_2}{\partial \theta} > 0$, $\frac{\partial \eta_3}{\partial \theta} < 0$

Proof. **Part 1.**

The first two equations are trivial as σ_s does not appear in equations that determine η_1 and η_4. For the last two: $\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial \sigma_s} = 2\frac{(c_f-c_s)\sigma_f(\sigma_f+\gamma+\rho)}{(\sigma_f-\sigma_s)^2} > 0$, the result follows from lemma 14.

Part 2.

The first two equations are trivial as σ_s does not appear in equations that determine η_1 and η_4. We have: $\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial \sigma_f} = -2\frac{(c_f-c_s)\sigma_f(\sigma_f+\gamma+\rho)}{(\sigma_f-\sigma_s)^2} < 0$, the result follows from lemma 14.

Parts 3 and 4.

$$\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial \gamma} = 2\frac{c_f\sigma_f - c_s\sigma_s}{\sigma_f - \sigma_s} > 0,$$
$$\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial \gamma} = 2c_s > 0,$$

the result follows from lemma 14. Proof for ρ is exactly the same.

Part 5.

$$\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial c_s} = -2\frac{\sigma_s(\sigma_f+\gamma+\rho)}{(\sigma_f-\sigma_s)(\gamma+\rho)} < 0,$$
$$\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial c_s} = 2(\gamma+\rho) > 0,$$

the result follows from lemma 14.

Part 6.

$$\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial c_f} = \frac{\sigma_f(\sigma_s+\gamma+\rho)}{(\sigma_f-\sigma_s)(\gamma+\rho)} > 0,$$
$$\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial c_f} = 0,$$

the result follows from lemma 14.

Part 7.

$$\frac{\partial(u(\eta_2)-u(\eta_3))}{\partial \theta} = 2 > 0,$$
$$\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial \theta} = 2 > 0.$$

Note that $\frac{\partial \eta_2}{\partial c_s} < 0$ and $\frac{\partial \eta_3}{\partial c_s} > 0$ implies m_f is increasing in c_s. Then TV_f is increasing in c_s. The proof for σ_f is exactly the same. $\frac{\partial \eta_2}{\partial c_f} > 0$ and $\frac{\partial \eta_3}{\partial c_f} < 0$ implies m_f is decreasing in c_f. Then TV_f is decreasing in c_f. The proof for σ_s and θ is exactly the same. Note that $\frac{\partial \eta_1}{\partial c_f} < 0$ and $\frac{\partial \eta_4}{\partial c_f} > 0$ implies m_s is increasing in c_f. Then TV_s is increasing in c_f. The proof for σ_s is exactly the same. $\frac{\partial \eta_1}{\partial c_s} > 0$ and $\frac{\partial \eta_4}{\partial c_s} < 0$ implies m_s is decreasing in c_s. Then TV_s is decreasing in c_s. The proof for σ_f is exactly the same. Part 7 of the above Lemma implies that increasing θ results in some types switching from fast venue to slow venue and some types switching from slow venue to no trade. Thus, trading volume is decreasing in θ. □

Proof of Proposition 2

Derivatives of the revenue in fast and slow venues are:

$$\frac{\partial R_s}{\partial c_s} = \frac{4\gamma\sigma_f\sigma_s(\gamma+\sigma_s+\rho)(1-Z)Z}{a(\gamma+\sigma_s)(\sigma_f-\sigma_s)}(c_f - 2c_s), \tag{69}$$

and

$$\frac{\partial R_f}{\partial c_f} = \frac{2\gamma\sigma_f(Z-1)Z\Big(a(\sigma_s-\sigma_f) - 2c_s\sigma_s(\gamma+\sigma_f+\rho) + 4c_f\sigma_f(\gamma+\sigma_s+\rho) + 2(\sigma_f-\sigma_s)(\gamma+\rho)\theta\Big)}{a(\gamma+\sigma_f)(\sigma_f-\sigma_s)}. \tag{70}$$

First, we see that $\frac{\partial^2 R_s}{\partial c_s^2} = \frac{8\gamma\sigma_f\sigma_s(\gamma+\sigma_s+\rho)Z}{a(\sigma_f-\sigma_s)(\gamma+\sigma_s)}(Z-1) < 0$ and $\frac{\partial^2 R_f}{\partial c_f^2} = \frac{8\gamma\sigma_f^2(\gamma+\sigma_s+\rho)Z}{a(\gamma+\sigma_f)(\sigma_f-\sigma_s)}(Z-1) < 0$, so the fee competition game has unique interior optimum. Moreover, the derivative of R_f evaluated at $c_s = 0$:

$$\left.\frac{\partial R_f}{\partial c_f}\right|_{c_s=0, c_f=0} = \frac{2\gamma\sigma_f(a-2(\gamma+\rho)\theta)(1-Z)Z}{a(\gamma+\sigma_f)},$$

which is positive by Assumption 4. Thus, whenever $c_s = 0$, $c_f > 0$. But $\frac{\partial R_s}{\partial c_s} > 0$ at $c_s = 0$ and $c_f > 0$, so there cannot be any equilibrium where $c_s = 0$, and c_s must be interior in any equilibrium. Setting (69) to zero, we find

$$c_s^*(c_f) = \frac{c_f}{2}. \tag{71}$$

Solving for the root of (70) and using (71) yields

$$c_f^*(\sigma_f, \sigma_s) = (a - 2\theta(\gamma+\rho))\frac{\sigma_f - \sigma_s}{(4\sigma_f - \sigma_s)(\gamma+\rho) + 3\sigma_f\sigma_s}.$$

Because the revenues converge to zero when prices are high, the unique equilibrium is given by c_s^* and c_f^*.

Proof of Proposition 3

Taking the derivative of equilibrium fee with respect to σ_f:

$$\frac{\partial c_f^*}{\partial \sigma_f} = \frac{3\sigma_s(\sigma_s + \gamma + \rho)(a - 2(\gamma + \rho)\theta)}{((\gamma + \rho - 3\sigma_f)\sigma_s - 4(\gamma + \rho)\sigma_f)^2} > 0,$$

by Assumption 4. Clearly, $\frac{\partial c_s^*}{\partial \sigma_f} > 0$ as $c_s^* = \frac{c_f^*}{2}$. To prove the second part:

$$\frac{\partial c_f^*}{\partial \sigma_s} = -\frac{3\sigma_f(\sigma_f + \gamma + \rho)(a - 2(\gamma + \rho)\theta)}{((\gamma + \rho - 3\sigma_f)\sigma_s - 4(\gamma + \rho)\sigma_f)^2} < 0,$$

again by Assumption 4. And also again $\frac{\partial c_s^*}{\partial \sigma_s} < 0$ as $c_s^* = \frac{c_f^*}{2}$.

Proof of Proposition 4

Directly differentiating,

$$\frac{\partial TV_s}{\partial \sigma_s} = \frac{2\gamma\sigma_f(a - 2(\gamma + \rho)\theta)(1 - Z)Z}{a(\gamma + \sigma_s)^2(-3\sigma_f\sigma_s + \gamma(-4\sigma_f + \sigma_s) - 4\sigma_f\rho + \sigma_s\rho)^2} \times \left(4\gamma^3\sigma_f + \sigma_s^2\rho(\sigma_f + \rho) + 8\gamma^2\sigma_f(\sigma_s + \rho) + \gamma\left(\sigma_s^2\rho + 4\sigma_f(\sigma_s + \rho)^2\right)\right).$$

Assumption 4 guarantees that $a - 2(\gamma + \rho)\theta > 0$, so the first term is positive. The second term is also positive, so $\frac{\partial TV_s}{\partial \sigma_s} > 0$. Similarly, by Assumption 4,

$$\frac{\partial TV_f}{\partial \sigma_s} = \frac{4\gamma\sigma_f^2(\gamma + \rho)(\gamma + \sigma_f + \rho)(a - 2(\gamma + \rho)\theta)(1 - Z)Z}{a(\gamma + \sigma_f)\left((\gamma + \rho)(4\sigma_f - \sigma_s) + 3\sigma_f\sigma_s\right)^2} > 0.$$

Proof of Proposition 5

$$\frac{\partial TV}{\partial \sigma_f} = -\frac{2\gamma(\gamma+\sigma_s+\rho)(a-2(\gamma+\rho)\theta)(1-Z)Z}{a(\gamma+\sigma_f)^2(\gamma+\sigma_s)(\gamma+\rho)^2\Big((4\sigma_f-\sigma_s)+3\sigma_f\sigma_s\Big)^2}$$
$$\times\Bigg(\gamma^2(\gamma+\rho)(-8\sigma_f^2+4\sigma_f\sigma_s+\sigma_s^2)+\gamma^2 6\sigma_f\sigma_s(-2\sigma_f+\sigma_s)$$
$$+3\sigma_f^2\sigma_s^2\rho-3\gamma\sigma_f\sigma_s(-2\sigma_s\rho+\sigma_f(\sigma_s+2\rho))\Bigg).$$

The fraction is positive, so

$$\text{sign}\left(\frac{\partial TV}{\partial \sigma_f}\right) = -\text{sign}\Bigg(\gamma^2(\gamma+\rho)(-8\sigma_f^2+4\sigma_f\sigma_s+\sigma_s^2)+\gamma^2 6\sigma_f\sigma_s(-2\sigma_f+\sigma_s)$$
$$+3\sigma_f^2\sigma_s^2\rho-3\gamma\sigma_f\sigma_s(-2\sigma_s\rho+\sigma_f(\sigma_s+2\rho))\Bigg).$$

Let $\sigma_f = \sigma + \epsilon$ and $\sigma_s = \sigma$. Simplifying and factoring out constants, yields

$$\text{sign}\left(\frac{\partial TV}{\partial \epsilon}\right) = \text{sign}\left(\frac{\partial TV}{\partial \sigma_f}\right) = \text{sign}\Big(\gamma^2+\gamma\sigma+\rho(\gamma-\sigma)+\mathcal{O}(\epsilon)\Big).$$

Letting $\epsilon \to 0$ and rearranging, we get

$$\text{sign}\left(\lim_{\epsilon\to 0}\frac{\partial TV}{\partial \epsilon}\right) = \text{sign}\left(\lim_{\epsilon\to 0}\frac{\partial TV}{\partial \sigma_f}\right) = \text{sign}\Big(\gamma(\gamma+\sigma)+\rho(\gamma-\sigma))\Big).$$

If $\gamma > \sigma$, then clearly $\lim_{\epsilon\to 0}\frac{\partial TV}{\partial \epsilon} > 0$. If $\gamma < \sigma$, then

$$\lim_{\epsilon\to 0}\frac{\partial TV}{\partial \epsilon} > 0 \iff \sigma < \frac{\gamma(\rho+\gamma)}{\rho-\gamma}.$$

Proof of Proposition 6

$$
\begin{aligned}
\frac{\partial W}{\partial \sigma_f} = &-\frac{(\gamma+\sigma_s+\rho)(1-Z)Z}{(\gamma+\sigma_f)^2(\gamma+\sigma_s)(4\gamma\sigma_f-\gamma\sigma_s+3\sigma_f\sigma_s+4\sigma_f\rho-\sigma_s\rho)^3} \\
&\times\Big(\gamma^4(24\sigma_f^3-18\sigma_f^2\sigma_s+5\sigma_f\sigma_s^2+\sigma_s^3)+3\sigma_f^2\sigma_s^2(-\sigma_f+\sigma_s)\rho^2 \\
&+2\gamma^3(\sigma_s^3\rho+\sigma_f\sigma_s^2(4\sigma_s+5\rho)+3\sigma_f^3(9\sigma_s+8\rho)-\sigma_f^2\sigma_s(13\sigma_s+18\rho)) \\
&+2\gamma\sigma_f\sigma_s(3\sigma_s^2\rho^2-\sigma_f\sigma_s\rho(\sigma_s+7\rho)+\sigma_f^2(6\sigma_s^2+13\sigma_s\rho+10\rho^2)) \\
&+\gamma^2(\sigma_s^3\rho^2+\sigma_f\sigma_s^2\rho(14\sigma_s+5\rho)-\sigma_f^2\sigma_s(5\sigma_s^2+40\sigma_s\rho+18\rho^2) \\
&+\sigma_f^3(41\sigma_s^2+74\sigma_s\rho+24\rho^2))\Big).
\end{aligned}
$$

Let $\sigma_f = \sigma + \epsilon$ and $\sigma_s = \sigma$. Simplifying and factoring out constants, yields

$$
\frac{\partial W}{\partial \sigma_f} = \left(\frac{4(1-Z)Z\gamma}{9(\gamma+\sigma)^2} + \mathcal{O}(\epsilon)\right). \tag{72}
$$

Letting $\epsilon \to 0$ and rearranging, we see that $\lim_{\epsilon\to 0}\frac{\partial W}{\partial \sigma_f} > 0$.

Proof of Proposition 7

$$
\begin{aligned}
\frac{\partial W_{\text{trader}}}{\partial \sigma_f} = &-\frac{(\gamma+\sigma_s+\rho)(1-Z)Z}{(\gamma+\sigma_f)^2(\gamma+\sigma_s)(4\gamma\sigma_f-\gamma\sigma_s+3\sigma_f\sigma_s+4\sigma_f\rho-\sigma_s\rho)^3} \\
&\times\Big(\gamma(\gamma\sigma_f(\gamma+\sigma_s)(4\gamma\sigma_f-(5\gamma+\sigma_f)\sigma_s)(3\sigma_f\sigma_s+\gamma(2\sigma_f+\sigma_s)) \\
&+\gamma(32\gamma^2\sigma_f^3+2\gamma\sigma_f^2(-12\gamma+25\sigma_f)\sigma_s+4\sigma_f^2(-9\gamma+4\sigma_f)\sigma_s^2+(\gamma^2+4\gamma\sigma_f-7\sigma_f^2)\sigma_s^3)\rho \\
&+(24\gamma^2\sigma_f^3+2\gamma\sigma_f^2(-9\gamma+10\sigma_f)\sigma_s+(\gamma-3\sigma_f)\sigma_f(5\gamma+\sigma_f)\sigma_s^2+(\gamma^2+6\gamma\sigma_f+3\sigma_f^2)\sigma_s^3)\rho^2\Big).
\end{aligned}
$$

Let $\sigma_f = \sigma + \epsilon$ and $\sigma_s = \sigma$. Simplifying and factoring out constants, yields

$$
\frac{\partial W_{\text{trader}}}{\partial \sigma_f} = \left(\frac{(1-Z)Z\gamma(4\rho-\gamma-\sigma)}{9(\gamma+\sigma)^2(\gamma+\sigma_s+\rho)} + \mathcal{O}(\epsilon)\right). \tag{73}
$$

Letting $\epsilon \to 0$ and rearranging, we see that $\lim_{\epsilon\to 0}\frac{\partial W_{\text{trader}}}{\partial \sigma_f} > 0$ if and only if $\sigma < 4\rho - \gamma$.