Cryptocurrency Venues: Segmentation, Fees, and Tax Policies*

Kerry Back, Oğuzhan Çelebi, Ali Kakhbod and A. Max Reppen

May 2022

Abstract

We study interactions between cryptocurrency trading venues, traders, and taxation in which the venues differ in technology (fast vs. slow). The property distinguishing this market from other markets like equities is the fact that each venue clears trades separately from one another. We show that trading fees rise when the venues are technologically differentiated. Improvements in the slow venue's technology is associated with greater trading volumes, whereas improvements in the fast venue's speed has an ambiguous impact. When the two venues have similar technologies, differentiation has a positive effect on trading volumes. Our welfare analysis suggests that in equilibrium, the tax rate to optimize tax revenue depends only on trader preferences, and it is thus independent of the venue properties and competition. When the government assigns different weights to its revenue, trader welfare, and venue welfare, the aggregate welfare can have a maximum for nontrivial tax rates.

Keywords: Digital economy. Crypto market. Fragmentation.
JEL Classification: C7, D86, G2, L13, L86.

*We would like to thank Kevin Li for helpful comments. Kerry Back and Ali Kakhbod are with the Jesse H. Jones School of Business at Rice University. Oğuzhan Çelebi is with the Department of Economics at MIT. A. Max Reppen is with the Questrom School of Business at Boston University. Emails: Kerry Back (kerry.e.back@rice.edu), Oğuzhan Çelebi (ocelebi@mit.edu), Ali Kakhbod (akakhbod@rice.edu), A. Max Reppen (amreppen@bu.edu).

1 Introduction

Exchanges and venues for cryptocurrency—crypto for short—trading regularly exhibit large deviations in price, both across and within jurisdictions. This price dispersion is considerable, relative to other types of trading, such as equity markets. Part of the reason for such differences between equities and cryptocurrencies is the strong regulation of equity markets. In particular, equity markets are subject to investor protection rules that aim to keep price segmentation low. Notably, the U.S. Securities and Exchange Commission's Rule 611 of Regulation NMS effectively creates a single market on which transactions from all trading venues clear.[1] This mechanism, which contributes to the lower segmentation, is not present in cryptocurrency markets, which are both international and largely unregulated. Instead, crypto assets are traded in multiple venues, most of which clear individually.[2]

In this context, we study the interaction between cryptocurrency traders, trading venues, and a tax entity using a model in which venues with different technologies compete on fees. Venues maximize their profit from trading, which is generated by the demand of traders to maximize their utility of holding the crypto asset. The trading is also subject to a trading tax set by a government, with welfare implications.

The technology of the venue is referred to as *speed*. Speed is here understood 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 impacting the time between the decision to trade and when the asset has changed hands. Given the infamously rapid price movements of crypto—at the time of writing, bitcoin has moved on average more than 4% per day—speed is of the essence.

Investors observe shocks to their marginal utility of holding the asset in an infinite horizon problem where 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 preferences, and traders choose whether to trade in the fast or slow market accordingly. The choice of venue is dynamic, and the traders may update their choice at any time.

[1] "Rule 611, among other things, requires a trading center to establish, maintain, and enforce written policies and procedures reasonably designed to prevent "trade-throughs" – the execution of trades at prices inferior to protected quotations displayed by other trading centers." from `https://www.sec.gov/divisions/marketreg/rule611faq.pdf`

[2] This is evidenced by the price dispersion observed in cryptocurrency markets Makarov and Schoar (2020).

Markets are either *integrated* or *segmented*. In an integrated market, two distinct venues represent different access points to the same market and thus also the same clearing price. In light of investor protection rules that aim to keep the prices in sync, for instance by requiring execution at the best available price in any market, market integration is a stylized fact of equity markets. In contrast, a *segmented* market consists of multiple venues that clear separately. Investor protection rules are designed to make traditional markets more integrated, but these rules do not cover cryptocurrency venues. Indeed, cryptocurrency exchanges clear separately from one another, with a distinct price in each.

This property of a segmented market in which investors may dynamically choose the vendor is, to our knowledge, new to the literature and an appropriate representation of cryptocurrency markets. In principle, the model applies to any such segmented market, of which cryptocurrencies is a natural and timely example to better understand.

Our findings give insight into the competition between cryptocurrency trading venues and their dependence on the trading technology. We find general conditions under which (i) no trading occurs, (ii) the market is segmented, i.e., both venues are active, and (iii) traders do not have sufficient preference for speed to trade in the fast venue at all (Theorem 1). We study the transition between these regimes and how speed segmentation affects welfare, trading volumes, and costs. With these general conditions, we can analyze fee competition among venues and determine the optimal trading fees.

The model also allows us to study trading volumes in cryptocurrency venues and show how they depend on taxes, fees, and the speeds of the venues (Proposition 1). We analyze how venues with different trading speeds compete in fees (Proposition 2) and how their equilibrium trading fees depend on their speeds (Proposition 3). In particular, we show how differentiation affects competition: when the differentiation between two venues decreases (i.e. the slower venue becomes faster or the faster venue becomes slower), the trading fees in both venues decrease, whereas 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.

We then focus on the effect of transaction speed on trading volume. First, a change in the transaction speed of a venue directly affects the instantaneous trading volume in the venue. Second, it affects the fees charged by the venues in equilibrium and, thus, the market structure itself. Both effects are positive for the speed of the slower venue, and total trading volume is increasing in the speed of the slower venue (Proposition

4). Moreover, we show that the effect of an increase in the speed of the fast venue is ambiguous. However, we consider the special case of *full competition*, where the speed difference between the venues is arbitrarily small. We find that whenever the rate of the preference shock is greater than the trading speed, differentiation increases trading volume (Proposition 5). Particularly, if the trading speed is greater than the rate of the preference shock, the effect of differentiation on trading volume depends on the discount factor of the traders. We characterize a threshold such that whenever the discount rate of the traders is below that threshold (i.e. the traders are sufficiently patient), differentiation increases trading volume. These results show that how trading speed affects trading volume depends on the characteristics of the traders in cryptocurrency markets.

Finally, we study the impact of a cryptocurrency trading tax levied by a government. The tax that optimizes tax revenue is independent of the venue speed and asset supply. Instead, it only depends on the preference parameters of the traders (Proposition 7). We also find that the welfare of all participants depends only on the taxes through the ratio of taxes to marginal utility. This means that the effect on the share of welfare of any participants due to a change in marginal utility can be offset by an appropriate tax change. In other words, the welfare distribution exhibits a natural equivalence between marginal utility and taxes (Proposition 6). In a growing and maturing landscape of crypto asset trading, our model provides a valuable understanding of healthy competition and optimal taxation.

1.1 Literature Review

In the years since the publication of the Bitcoin whitepaper Nakamoto (2008), and particularly in recent years, the public and researchers alike have shown great interest in cryptocurrencies. Rigorous studies of its properties range from questions of security and privacy to mining and economic aspects of markets. The former are topics of interest in computer science, and we refer the interested reader to the book by Narayanan et al. (2016) for an introduction.

In finance, Yermack (2015) is one of the first papers that brings academic attention to the field of crypto. A number of recent papers develop models of cryptocurrencies (see, e.g. Abadi and Brunnermeier (2018); Biais et al. (2019); Chiu and Koeppl (2017); Cong and He (2019); Cong, Li and Wang (2021); Cong, He and Li (2021); Lehar and Parlour (2020); Pagnotta (2018); Schilling and Uhlig (2019); Sockin and Xiong (2020); Weber (2016)). These papers have different focuses (e.g., in mining, smart contracts, price dispersion across venues, transaction fees, among others). For example, mining is the process that upholds the integrity of the distributed ledgers of cryptocurrencies. Because miners compete against each other, mining is naturally modeled as a game, cf. e.g. Biais et al. (2019), Lehar and Parlour (2020). Other mining examples include Cong, He and Li (2021) who analyze mining in pools, a form of mining in which some miners cooperate to gain an advantage; and Roşu and Saleh (2021) who analyze the wealth effects in the alternative, and likely successor,[3] system to mining called proof-of-stake. Economic aspects of cryptocurrency networks have been studied in e.g. Easley, O'Hara and Basu (2019) who investigate the emergence and role of transaction fees on the Bitcoin blockchain. Cong and He (2019) show how smart contract features provided by many blockchain systems can lead to welfare improvements in contracting. Lehar and Parlour (2021) compare liquidity provision at a centralized exchange with liquidity provision at an AMM which utilizes a constant product function. Finally, several recent papers document empirical facts related to crypto trading and investments (e.g., Stoffels (2017); Hu, Parlour and Rajan (2019); Borri and Shakhnov (2019); Makarov and Schoar (2020); Liu and Tsyvinski (2021)).

Most cryptocurrency research, such as the above examples, revolves around properties of the currencies and their structures themselves. In contrast, we here focus on the market structure at the level of trading venues, i.e., the entities that facilitate trading of the crypto assets. To the best of our knowledge this paper is the first paper that analyzes market structure properties with heterogeneous crypto trading venues.

The dynamic trading framework of this paper is related to the literature of the integrated (equity) markets (e.g., Kyle (1985*a*,*b*); Glosten and Milgrom (1985); Back (1992); Wang (1994); Back, Cao and Willard (2000)). Similar to these works we also present a fully dynamic trading model, however, in sharp contrast to these important papers, we instead focus on impacts of having multiple venues on trading volume, tax, transaction costs and welfare. It is worth mentioning that the competition feature of our model relates to the recent paper by Pagnotta and Philippon (2018). However, there are several distinct modeling differences: in particular, their model is a subscription model in which each trader initially chooses a venue and stays there forever, however, in our model traders can dynamically move between the venues depending on their preference shocks and how fast they would like to execute their trades, moreover, their model is designed for equity (integrated) markets because venues share one clearing condition.

[3]The Ethereum network has recently made great progress in moving to a proof-of-stake system and is expected to transition during Q2 2022: `https://ethereum.org/en/upgrades/merge/`.

Instead, we solve a model with multiple clearing conditions—one for each venue—as is the case for a segmented (crypto) market.

1.2 Institutional facts about crypto

We briefly account for some properties of cryptocurrencies and their trading that is relevant for this paper. As previously mentioned, price dispersion is widespread in the crypto markets, cf. e.g. Makarov and Schoar (2020). This is caused by and evidence of how trading happens and clears separately in different venues. In addition to this, each venues sets their own fees for trading.[4] That these trading fees are an important source of differentiation and competition is clear from how fees are prominently featured in most trading venue comparisons[5]. Cryptocurrency trading venues are also compared on some other fronts, such as accessibility and convenience; payment features; and customer service.[5] We consider all of these features to be encompassed in our wide notion of speed. Indeed, accessibility reduces the time to trade; venues accepting credit and debit cards avoid the delay associated with wire transfers; and customer service could reduce trading downtime due to difficulties. These facts are evidence of competition among exchanges with different speeds and fees.

The rest of the paper is organized as follows. In Section 2.1, we analyze the decision of the traders for a given market structure (the trading speed and the transaction fees as given) and characterize the trading equilibrium with two venues; Section 2.3 characterizes the trading volume under this trading equilibrium. In Section 2.4, we analyze how cryptocurrency venues compete in transaction fees in order to maximize their profit with trading speeds given. In Section 3, we analyze welfare and optimal tax policies in these markets.

2 Model: Cryptocurrency Venues

We consider a continuous time model with 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; a trader who owns the asset is called a *holder*, and a trader who does

[4]These fees set by trading venues should not be confused with transaction fees for making transactions on the blockchain itself.

[5]This is exemplified in the comparison
https://www.investopedia.com/best-crypto-exchanges-5071855.

not is called a *nonholder*. Each trader has a personalized/intrinsic value $\eta \in [\eta_l, \eta_h]$ representing their desire to holding the asset, which changes over time.

The instantaneous utility of a holder with value η is $u(\eta)$ while a nonholder gets zero instantaneous utility value (regardless of their intrinsic value). The value η changes, independently across the traders, with a Poisson shock with rate γ. Conditional on the arrival of a shock, the new valuation is chosen from $[\eta_l, \eta_h]$ according to the cumulative distribution function $F(\cdot)$ and the probability distribution function $f(\cdot)$.

There are multiple *crypto trading venues* with different trading speeds and transaction fees (i.e., costs to traders). At any given time, a holder decides to sell or hold the asset and a nonholder decides to buy the asset or do nothing. If a trader decides to sell or buy the asset, she then decides which venue to use.

A venue ν, $\nu \in \{s, f\}$, is characterized by its transaction speed σ_ν and transaction fee c_ν. We assume venue f is faster, i.e., $\sigma_f \geq \sigma_s$ and $c_f > c_s$.[6] In addition to the transaction costs, traders also pay an amount $\theta \geq 0$ per trade. The difference between θ and the transaction fees c_s and c_f arises because venues may strategically choose c_s and c_f, whereas θ is fixed (c.f. Section 2.4). We will consider various interpretations of θ in later sections: for example, θ is interpreted as a tax charged by a regulatory body.

2.1 Analysis: Traders' Decisions

We first characterize the behavior of traders as a function of the speed and transaction fees of the venues. Let $\nu \in \{s, f\}$ denote the venue and $m \in \{0, 1\}$ a trader's position, where $m = 0$ denotes a nonholder and $m = 1$ denotes a holder. At each time t, a holder may Sell (S) or Hold (H), and a nonholder may Buy (B) or do Nothing (N). The action set of a trader is (H, S^ν) if $m = 1$ and (N, B^ν) if $m = 0$, where the superscripts denote the venue choices of traders.

We focus on the stationary equilibrium. Let p_ν denote the equilibrium asset price in venue ν. Let $V_{m,\eta}$ denote a trader's expected payoff whose current trading position and value are, respectively, $m \in \{0, 1\}$ and $\eta \in [\eta_l, \eta_h]$. To derive the Hamilton–Jacobi–Bellman (HJB) equations, depending on traders' actions, we define the following func-

[6]Note that $\sigma_f \geq \sigma_s$ is without loss of generality. If $c_f > c_s$ does not hold, then in equilibrium, there is no demand for the slower venue. When we endogenize the transaction fees, the slower venue always chooses a lower transaction fee in order to make positive revenue.

tions that will be used to construct the value function:

$$\begin{aligned}
\rho V_H(\eta) &= \underbrace{u(\eta)}_{\text{flow gain of holding the asset}} + \gamma \underbrace{\left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_H(\eta)\right)}_{\text{net gain of the pref. shock}}, \\
\rho V_S^s(\eta) &= u(\eta) + \gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^s(\eta)\right) + \sigma_s \underbrace{\left(V_{0,\eta} + p_s - c_s - \theta - V_S^s(\eta)\right)}_{\text{net gain of selling in the slow venue}}, \\
\rho V_S^f(\eta) &= u(\eta) + \gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^f(\eta)\right) + \sigma_f \underbrace{\left(V_{0,\eta} + p_f - c_f - \theta - V_S^f(\eta)\right)}_{\text{net gain of selling in the fast venue}}.
\end{aligned} \tag{1}$$

For a trader who currently owns the asset, $V_H(\eta)$ denotes the value of holding it and $V_S^\nu(\eta), \nu \in \{s, f\}$ denotes the value of selling the asset in venue ν.

In a similar fashion, we define

$$\begin{aligned}
\rho V_N(\eta) &= \gamma \underbrace{\left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_N(\eta)\right)}_{\text{net gain of the pref. shock}}, \\
\rho V_B^s(\eta) &= \gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^s(\eta)\right) + \sigma_s \underbrace{\left(V_{1,\eta} - p_s - c_s - \theta - V_B^s(\eta)\right)}_{\text{net gain of buying in the slow venue}}, \\
\rho V_B^f(\eta) &= \gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^f(\eta)\right) + \sigma_f \underbrace{\left(V_{1,\eta} - p_f - c_f - \theta - V_B^f(\eta)\right)}_{\text{net gain of buying in the fast venue}}.
\end{aligned} \tag{2}$$

For a trader who currently does not own the asset, $V_N(\eta)$ denotes his value of doing nothing and $V_S^\nu(\eta), \nu \in \{s, f\}$ the value of buying the asset in venue ν.

Next, to derive the optimal action of a trader, we note that she has three options. If she is a holder, she chooses between holding the asset, selling it in the slower venue, or selling it in the faster venue. If she is a nonholder, she decides between doing nothing, buying the asset in the slow venue, or buying the asset in the fast venue. Recall that $V_{0,\eta}$ and $V_{1,\eta}$ denote the (continuation) values (i.e., the expected payoff) for a trader as a function of her current position m and her value η. Given (1) and (2), we obtain $V_{0,\eta} = \max\{V_N(\eta), V_B^s(\eta), V_B^f(\eta)\}$ and $V_{1,\eta} = \max\{V_H(\eta), V_S^s(\eta), V_S^f(\eta)\}$. Moreover, 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 (e.g., $\eta \in N$ if and only if $V_{0,\eta} = V_N$; $\eta \in B_s$ if and only if $V_{0,\eta} = V_B^s(\eta)$ and so on).

In order to specify the stationary equilibrium, we need to characterize the stationary distribution of values η of holders and nonholders, whose densities we denote as $f_h(\eta)$

and $f_{nh}(\eta)$. We now formally define the stationary equilibrium.

Definition 1. *A stationary equilibrium consists of sets N, B_s, B_f, H, S_s, and S_f; prices p_s and p_f; and f_h and f_{nh} such that:*

- $f_h(\eta) + f_{nh}(\eta) = f(\eta)$
- *Traders behave optimally*
- *Asset market clears:*
$$\int_{\eta_l}^{\eta_h} f_h(\eta) d\eta = Z \tag{3}$$
- *Fast venue clears:*
$$\int_{B_f} f_{nh}(\eta) d\eta = \int_{S_f} f_h(\eta) d\eta \tag{4}$$
- *Slow venue clears:*
$$\int_{B_s} f_{nh}(\eta) d\eta = \int_{S_s} f_h(\eta) d\eta \tag{5}$$

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*, 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 and tax 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 occur 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 theorem statement:

$$g(\sigma_f, \sigma_s, c_s, c_f, \theta) = \frac{\sigma_f(\sigma_s + \gamma + \rho)(c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(c_s + \theta)}{\sigma_f - \sigma_s}. \tag{6}$$

The function g increases whenever the fast venue becomes better for the traders (when σ_f increases or c_f decreases) and decreases whenever the slow venue becomes better for the 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 $u(\eta_h) \leq 2(c_s+\theta)(\gamma+\rho)$, then there is no equilibrium at which a positive measure of traders trade.*

*(ii) [**Market segmentation.**] 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 value 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:

$$(1-Z)F(\eta_1) + ZF(\eta_4) = 1-Z, \tag{7}$$
$$(1-Z)F(\eta_2) + ZF(\eta_3) = 1-Z, \tag{8}$$
$$u(\eta_1) - u(\eta_4) = 2(\gamma+\rho)(c_s+\theta), \tag{9}$$
$$u(\eta_2) - u(\eta_3) = 2g(\sigma_f, \sigma_s, c_s, c_f, \theta). \tag{10}$$

*(iii) [**No segmentation**] If $u(\eta_h) > 2(c_s+\theta)(\gamma+\rho)$ and $u(\eta_h) \leq 2g(\sigma_f, \sigma_s, c_s, c_f, \theta)$, 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

$$(1-Z)F(\eta_1) + ZF(\eta_4) = 1-Z, \tag{11}$$
$$\frac{u(\eta_1) - u(\eta_4)}{\gamma+\rho} = 2(c_s+\theta). \tag{12}$$

The theorem characterizes the cutoffs in terms of $u(\eta_h)$. Figure 1 illustrates a segmented market.

Theorem 1 is the building block for our analysis. Before we proceed with the analysis, we discuss the conditions that give rise to the three main cases. Note that

$$u(\eta_h) = u(\eta_h) - u(0) = (V_H(\eta_h) - V_H(0)) \cdot (\gamma+\rho). \tag{13}$$

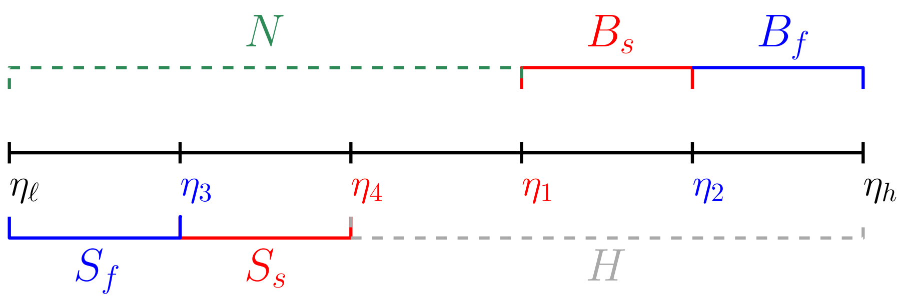

Figure 1: Threshold values in Theorem 1. The symbols S_f and S_s denote types who will sell in the fast and slow venue, respectively; B_f and B_s are those who will buy in the fast and slow venue, respectively. Types in N will not buy, and types in H will hold.

This quantity corresponds to the value of a transaction between traders with values 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 $2(\gamma + \rho)(c_s + \theta) \geq u(\eta_h)$, 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 above condition as

$$2(c_s + \theta) > V_H(\eta_h) - V_H(0). \tag{14}$$

In this form, the equation simply says that the total transaction cost (fees and tax) 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:

$$(\gamma + \rho)(V_H(\eta_h) - V_H(0)) > 2g(\sigma_f, \sigma_s, c_s, c_f, \theta). \tag{15}$$

As $c_f > c_s$, the numerator is always positive and bounded away from zero, 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.

Equation 9 shows the trade-off between trading in the slow venue versus no trading. Rewriting that equation, we obtain $V_H(\eta_1) - V_H(\eta_4) = 2(c_s + \theta)$, which relates the value of a transaction between the closest buyer and seller values of η for traders in the slow venue. Intuitively, the value of a transaction between these two types must be equal to the total transaction cost in equilibrium.

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.

2.2 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., which of the three cases in Theorem 1 is active) and the corresponding cutoff thresholds as defined in Theorem 1.

2.2.1 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 a 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 tax θ is less than 1 (approximately). Once the tax 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 tax is even higher, the no-trade condition in Theorem 2 obtains and no traders engage in any trading. In Figure 2, we see that even the slow venue thresholds hit their corner solutions (i.e., when $\theta \geq 2.3$ in the figure). Thus, the figure showcases all three cases of Theorem 1.

2.2.2 Comparative statics of the market structure

In Figure 2, we illustrated how the market structure changes as θ (the external "tax") increases. In general, the relevant exogenous parameters in Theorem 1 are (in addition

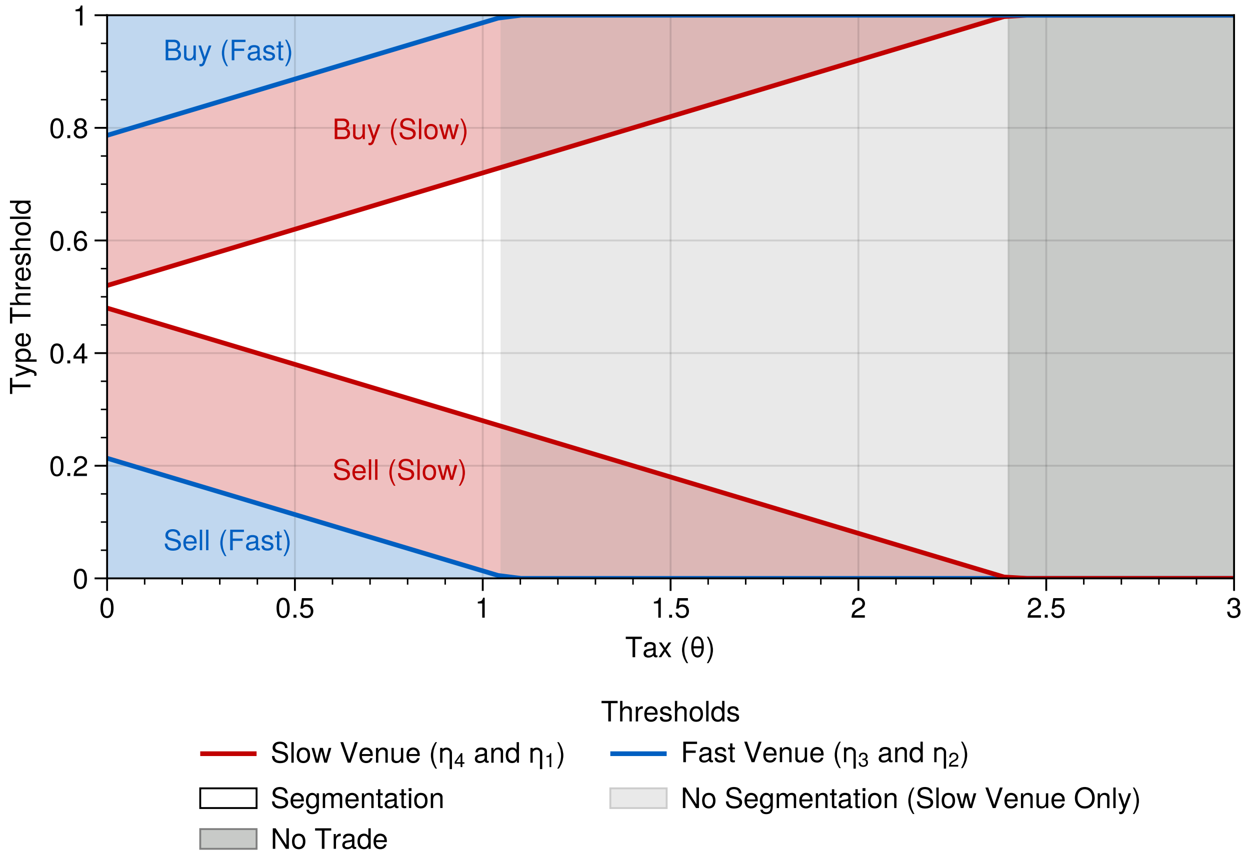

Figure 2: Type thresholds $\eta_1, \ldots, \eta_4$ as in Theorem 1 as a function of the tax parameter θ. Here η is uniformly distributed on $[0, 1]$ and the utility function $u(\eta) = \eta$ is linear. The other parameters are $Z = \frac{1}{2}$; $\rho = \gamma = 0.1$; $c_s = 0.1$ and $\sigma_s = 1$;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

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 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. Note that this results holds even when the speed offered by the fast venue is overwhelmingly large; according to Theorem 1, segmentation only occurs when

$$u(\eta_h) = u(1) > 2g(\sigma_f, \sigma_s, c_s, c_f, \theta). \tag{16}$$

Holding all parameters constant (say, with $\sigma_s = 1$) and taking $\sigma_f \to \infty$, the expression on the right-hand side approaches $c_f - c_s$. Market segmentation does not occur if $c_f - c_s$ is

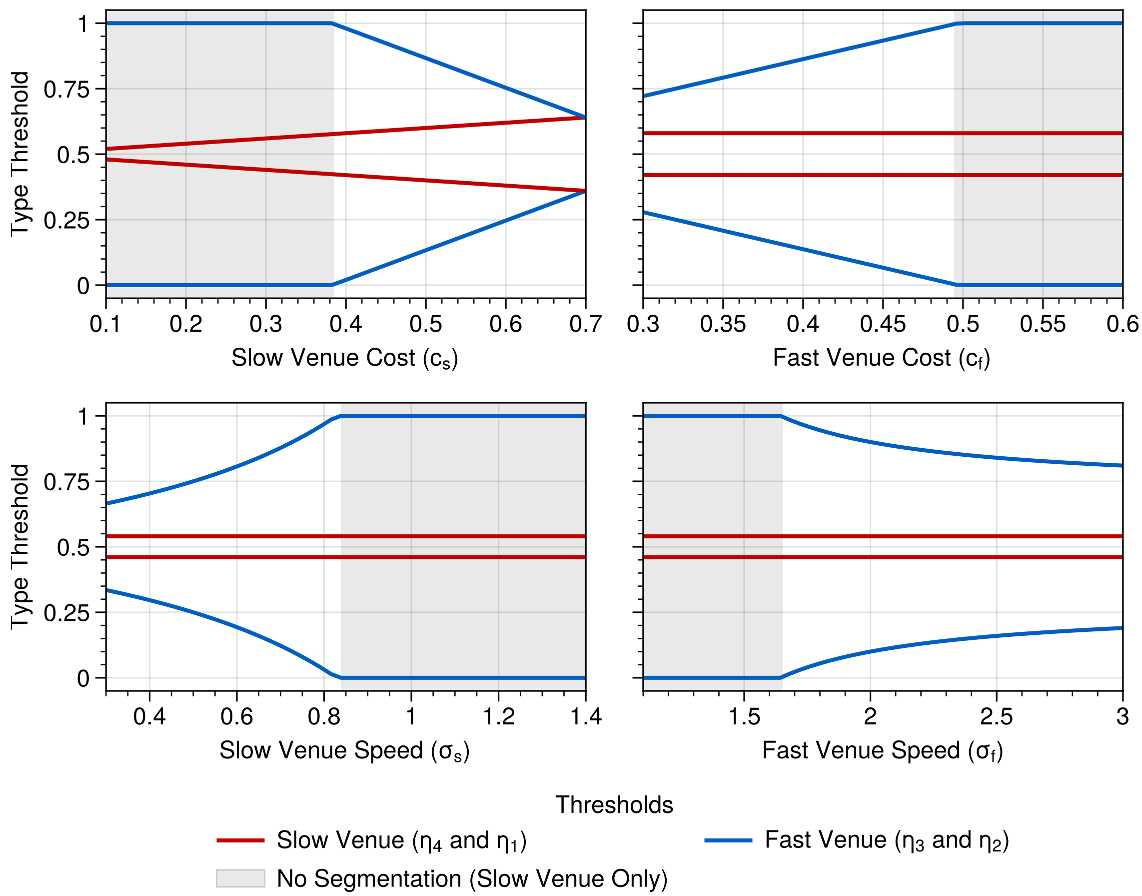

Figure 3: Comparative statics for Theorem 1. The type thresholds (y-axis) are plotted as a function of changing a single parameter at a time (x-axis); the parameters of interest are the transaction fees and speed offered by the slow and fast venues. Figure 2 presents an analogous plot for θ. As usual, the thresholds are ordered: $\eta_3 < \eta_4 < \eta_1 < \eta_2$. Thresholds for the the fast venue are in blue, and those for the slow venue are in red.

large, no matter how large the speed advantage. This finding is illustrated in the top left panel of Figure 3, where $\sigma_f = 10$ and $\sigma_s = 1$. As c_s increases, we see that an increasingly smaller set of types trade in the slow venue, as expected. Once c_s is large enough, the fast venue is competitive enough that the "extreme types" (i.e., holders whose types η are close to zero or nonholders whose types are close to one) leave the slow venue and begin trading in the fast venue. As c_s increases past this threshold, more and more types begin trading in the fast venue, and at the limit $c_s = c_f$, the slow venue disappears and all trading occur in the fast venue. In terms of the thresholds, this corner case corresponds to $\eta_3 = \eta_4$ and $\eta_1 = \eta_2$, as depicted in the right edge of the top-left panel of Figure 3.

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. Intuitively, the reason is that traders who are on the cutoff are actually indifferent 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 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: intuitively, 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

2.3 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:

$$\begin{aligned} m_f(\sigma_s, \sigma_f, c_s, c_f, \theta) &= \int_{\eta_l}^{\eta_3} f_h(\eta) d\eta + \int_{\eta_2}^{\eta_h} f_{nh}(\eta) d\eta \\ &= F(\eta_3) \frac{\gamma Z}{\gamma + \sigma_f} + (1 - F(\eta_2)) \frac{\gamma(1-Z)}{\gamma + \sigma_f} \end{aligned} \tag{17}$$

and

$$\begin{aligned} m_s(\sigma_s, \sigma_f, c_s, c_f, \theta) &= \int_{\eta_3}^{\eta_4} f_h(\eta) d\eta + \int_{\eta_1}^{\eta_2} f_{nh}(\eta) d\eta \\ &= (F(\eta_4) - F(\eta_3)) \frac{\gamma Z}{\gamma + \sigma_s} + (F(\eta_2) - F(\eta_1)) \frac{\gamma(1-Z)}{\gamma + \sigma_s}. \end{aligned} \tag{18}$$

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 prices and speeds.[7]

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. 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. In the next section, we 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.

2.3.1 Illustrations

We illustrate the results of Proposition 1 in Figures 4, 5, and 6. Figure 4 plots the trading volumes in the two venues as the transaction fees in the venues change. The results confirm part of the statement in the proposition. The figure also reveals that (at least for

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

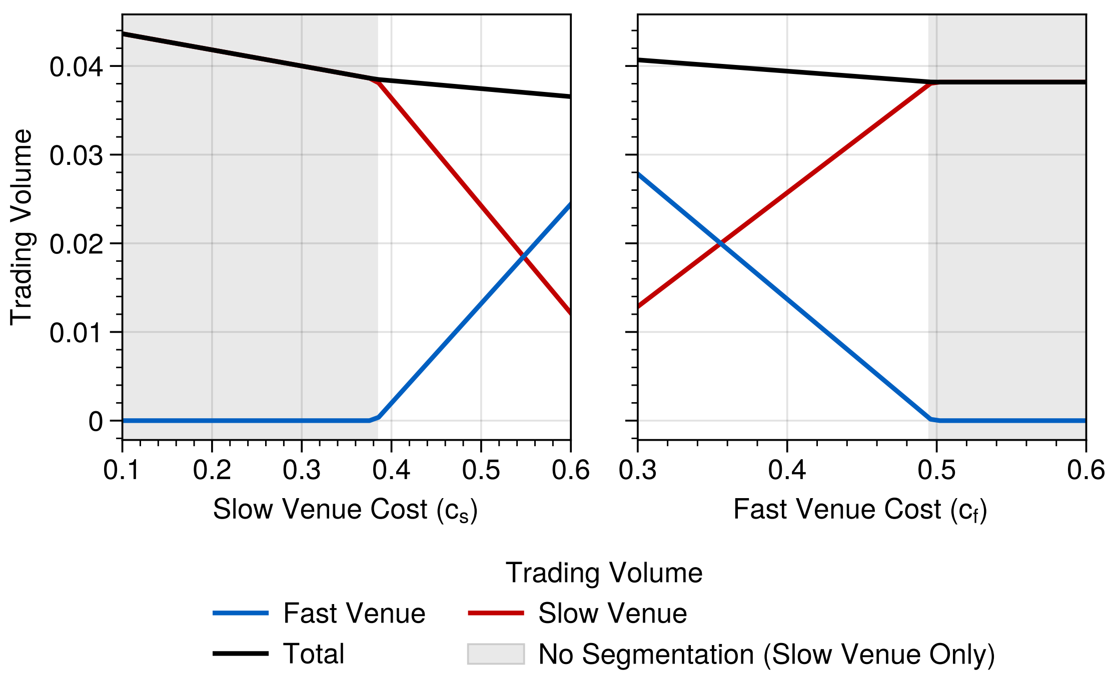

Figure 4: Trading volume in the two venues as a function of transaction fees c_s and c_f with $\eta \sim \text{Unif}[0,1]$ and $u(\eta) = \eta$. The gray regions indicate cost regimes in which only the slow venue is active. The primary driver of the trading volume in either venue is whether the fees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

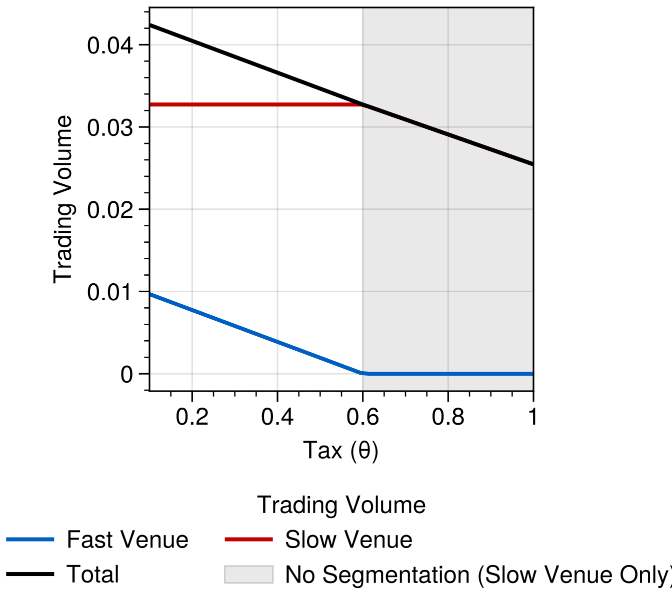

Figure 5: Trading volume as a function of the tax θ with $\eta \sim \text{Unif}[0,1]$ and $u(\eta) = \eta$.. The gray region indicates the threshold above which only the slow venue is active. For our linear model, increasing θ has a *layered* effect: as θ increases, traders from the fast venue switch to the slow venue until the fast venue is inactive. From that point, further increases drive traders away from the slow venue, becoming nontraders.

this set of functional parameters), total trading volume decreases in the slow venue fee c_s; in other words, the increase in trading volume in the fast venue does not compensate for the loss in the slow trading venue.

This result is somewhat surprising: if it were the case that traders simply migrated from the slow to the fast venue as fees goes up, then total trading volume would increase due to the net increase in speed (i.e., recall that trading volume is speed times the measure of traders). We know, however, that this is not the case, as the top left panel of Figure 3 shows that increasing c_s will induce some traders to stop trading. Figure 4 shows that the effect carries over even when taking trading speed into account: total trading volume decreases.

The behavior is similar when we change the fast venue fee. Note, however, that after trading in the fast venue ceases, increasing c_f has no more effect on the market structure, such that $\partial TV / \partial c_f = 0$ for sufficiently large c_f.

Next, consider the effect of increasing the tax θ, as shown in Figure 5; as predicted, it is decreasing. When θ is so low that both venues are active (see Theorem 1), increasing θ will only decrease trading volume in the fast venue while keeping the slow venue trading volume constant. Note that the set of types trading in either venue decreases per

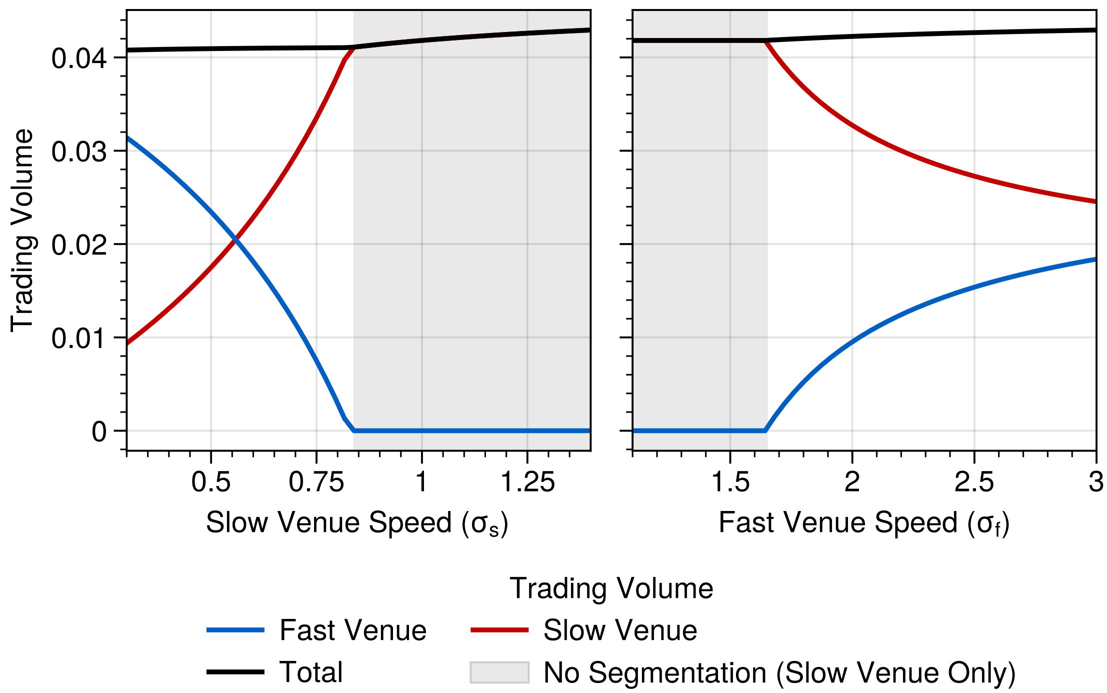

Figure 6: Trading volume as a function of venue speed with $\eta \sim \text{Unif}[0,1]$ and $u(\eta) = \eta$.. The gray region indicates the thresholds beyond which only the slow venue is active.

Figure 2; the interpretation is that the inflow of traders from the fast to the slow venue exactly balances out the flow out of the slow venue (flow being measured by speed times the measure of traders). Only when the tax is high enough and trading ceases in the fast venue does the trading volume in the slow venue begin to decrease, as there is no inflow to make up for the outflow of traders.

The final set of figures (see Figure 6) show the effect of increasing venue speed (σ_s and σ_f). It agrees with the proposition that, e.g., $\partial TV_s / \partial \sigma_s > 0$, and likewise for the other partial derivatives. Note, however, that total trading volume is always increasing in the venue speed.[8] Intuitively, better *technology* (i.e., faster speeds) induces more trading; the results in the next section characterize this behavior in greater detail.

2.4 Fee Competition: Improvement versus Differentiation

Having characterized trading volume, we next analyze the competition between the two venues, starting with transaction fees. The revenues of the fast and slow venues for a

[8]If σ_f is sufficiently small, then increasing σ_f by a small enough amount will not affect the market structure, in which case $\partial TV_s / \partial \sigma_f = \partial TV_f / \partial \sigma_f = 0$.

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 two assumptions, the first to keep the analysis tractable, and the second to guarantee that the tax imposed by government (or external party) is not so high as to prohibit trading *a priori*.

Assumption 2. $u(\eta) = a\eta$*, where we refer to* a *as the* marginal utility*.*

Assumption 3. $a > 2\theta(\gamma + \rho)$

An equilibrium is a set of fees c_s^*, c_f^* such that $c_s^* \in \arg\max_{c_s} R_s(\sigma_f, c_f^*, \sigma_s, c_s)$ $c_f^* \in \arg\max_{c_f} R_f(\sigma_f, c_f, \sigma_s, c_s^*)$. For the rest of the paper, we also assume the following to make the analysis tractable:

Assumption 4. *F is uniform over* $[0, 1]$*.*

The following proposition characterizes the equilibrium prices of two competing venues.

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

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

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

Consequently,

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

Proposition 2 characterizes how the venues compete to attract traders. The slower venue always undercuts the fast venue, as otherwise no trader would choose 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

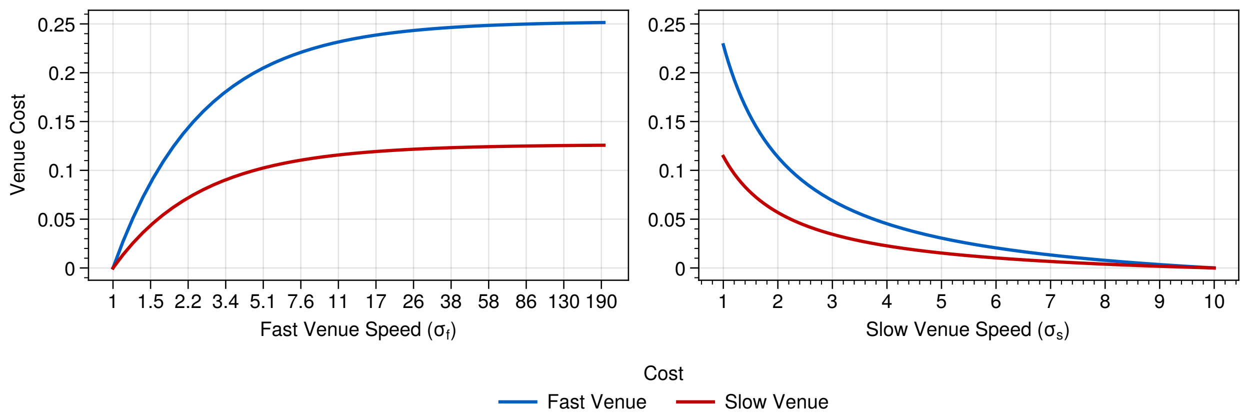

Figure 7: Optimal transaction fees (cost) as a function of venue speeds.

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.

Corollary 1. *Fix parameters $Z \in (0,1)$, $\gamma > 0$, $\rho > 0$, and $\theta \geq 0$. 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.

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.*

This result shows 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.

The results above are illustrated in Figure 7, 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 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 (decreasing) in 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 the transaction fees, making everyone better off. On the other hand, an increase in σ_f increases the transaction fees and makes everyone except the traders with the most extreme valuations worse off.

Lastly, we analyze the effect of speed on trading volume. Our fourth 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, for given choices of traders, it increases the trading speed in slow venue, which increases trading volume (the *improvement channel*). Second, it makes the slow venue more competitive. Competition lowers the trading fees, which also increases trading volume (the *differentiation channel*). A speed increase for the slow venue affects both channels positively by decreasing differentiation and improving the overall trading speed in the market, which leads to increased trading volume.

Figure 8 plots the trading volumes in each venue as a function of σ_s. Note that (in addition to being increasing) trading volume is a concave function of σ_s, such that increasing σ_s has a diminishing effect.

In fact, the two components of the competition channel driving up the trading volume exhibit diminishing returns. We have already seen in the right panel of Figure 7 that increasing σ_s^* has diminishing effects on lowering costs. Moreover, Figure 9 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 exogeneously, 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.

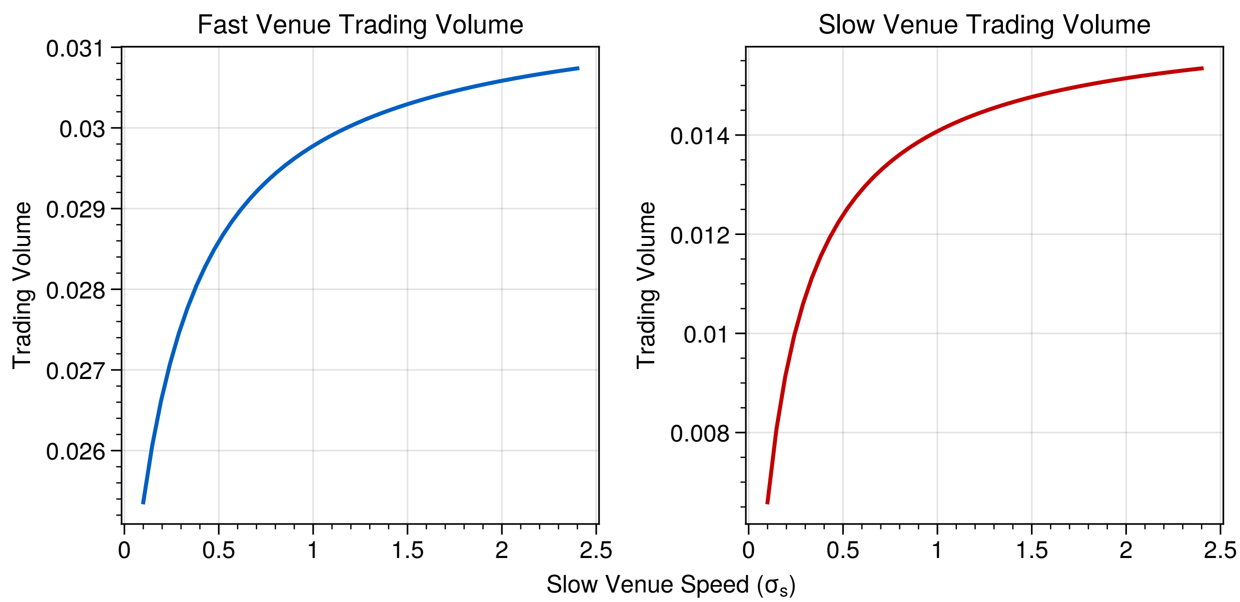

Figure 8: Trading volumes in both venues increase as σ_s increases.

The effect of speed in the fast venue is more complicated. On the one hand, speed increases the transaction rate in the fast venue, which increases trading volume. However, speed also increases differentiation, thus increasing the transaction fees charged in equilibrium. Increased transaction fees reduce the incentives of traders to trade and pay the fee, which reduces trading volume. Thus, the improvement and differentiation channels are working in opposite directions. In general, the effect of σ_f is ambiguous. We consider an instructive special case in which σ_s and σ_f are very close, a scenario we call the *full competition benchmark*. We analyze the effect of a speed increase for the faster venue. The following proposition characterizes the effect of an increase in σ_f when the two venues start from the same speed level.

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* $\gamma \geq \sigma$, *then* $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0$
- *If* $\gamma < \sigma$, *then* $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0$ *if and only if* $\rho < \frac{\gamma^2 + \gamma\sigma}{\sigma - \gamma}$

The proposition shows that, first, if $\sigma < \gamma$, i.e., the trading speed is slow relative to the frequency of the preference shock, a speed increase in the fast venue causes an increase in trading volume. This result shows that the effect of higher transaction speed dominates the effect of higher fees; in other words, the improvement channel dominates

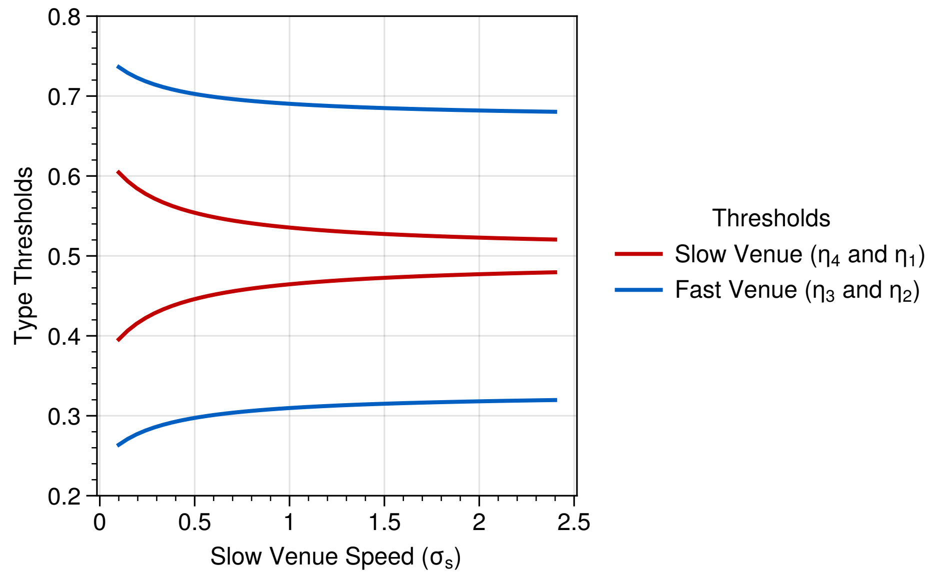

Figure 9: Increasing σ_s has a diminishing impact on type thresholds.

the differentiation channel. The intuition for this result is that when the trading speed is low compared to the rate of the preference shock, improvements in speed are more important for the traders and higher prices causes less distortion in the choices of the traders. Second, if $\gamma < \sigma$, i.e., trading speed is fast relative to the frequency of the preference shock, then the discount factor of the traders becomes important. In particular, the trading volume is still increasing if they are patient relative to the frequency of the preference shock. The reason is that when traders are patient and the preference shock is not very frequent, the traders have a higher value of changing their position, making them less susceptible to the adverse effects of higher fees. As a result, the improvement channel dominates the differentiation channel.

If traders are impatient, then they are discouraged from trading due to rising transaction fees as the benefit of changing their position is lower. In that case, they would avoid paying higher fees and the differentiation channel dominates the improvement channel. These results show the importance of competition when there are multiple venues: the fast venue becomes faster, but trading volume decreases due to decreased competition and higher fees.

The following corollary shows how these forces interact if the trading speed is fast:

Corollary 2. *In the setting of Proposition 5, the following are equivalent:*

1. $\rho \leq \gamma$.

2. $\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0$ *for all sufficiently large* σ.

If the transaction speed is very fast, then the trade-off depends on the comparison between ρ and γ. If $\rho > \gamma$, then there is more benefit from faster transactions as the traders will obtain a higher utility from owning the asset before a preference shock comes. This makes them continue trading under higher fees and increases trading volume. On the other hand, if $\rho < \gamma$, then the effect of increased transaction fees dominates faster trading speed, leading to a decrease in trading volume.

3 Welfare Analysis

Because the financial transactions in this model preserve the financial value in the system, they are purely distributional and do not affect total welfare. The system welfare is thus constituted by only the welfare gained from holders' utility. Formally, this welfare can be separated into three different terms corresponding to the trader welfare, venue welfare, and government welfare:

$$W = \int_0^1 a\eta f_h(\eta)d\eta \overset{\text{formally}}{=} \underbrace{\int_0^1 \eta f_h(\eta)d\eta - \text{costs}}_{\text{trader welfare}} + \underbrace{R_f + R_s}_{\text{venue welfare}} + \underbrace{\theta\, TV}_{\text{government welfare}} . \quad (25)$$

More precisely, with the costs (per unit time) given by $\sigma_\nu(c_\nu + \theta)$ for $\nu \in \{f, s\}$ depending on η, we get from Theorem 1 that the trader welfare is

$$\begin{aligned} W_{\text{trader}} &= \int_0^1 a\eta f_h(\eta)d\eta - \int_{\eta_l}^{\eta_3} \sigma_f(c_f + \theta)f_h(\eta)d\eta - \int_{\eta_2}^{\eta_h} \sigma_f(c_f + \theta)f_{nh}(\eta)d\eta \\ &\quad - \int_{\eta_3}^{\eta_4} \sigma_s(c_s + \theta)f_h(\eta)d\eta - \int_{\eta_1}^{\eta_2} \sigma_s(c_s + \theta)f_{nh}(\eta)d\eta \\ &= W - R_f - R_s - \theta\, TV, \end{aligned}$$

the venue welfare is

$$\begin{aligned} W_{\text{venue}} &= \int_{\eta_l}^{\eta_3} \sigma_f c_f f_h(\eta)d\eta + \int_{\eta_2}^{\eta_h} \sigma_f c_f f_{nh}(\eta)d\eta + \int_{\eta_3}^{\eta_4} \sigma_s c_s f_h(\eta)d\eta + \int_{\eta_1}^{\eta_2} \sigma_s c_s f_{nh}(\eta)d\eta \\ &= c_f\, TV_f + c_s\, TV_s = R_f + R_s, \end{aligned}$$

and the government welfare is

$$\begin{aligned} W_{\text{government}} &= \int_{\eta_l}^{\eta_3} \sigma_f \theta f_h(\eta) d\eta + \int_{\eta_2}^{\eta_h} \sigma_f \theta f_{nh}(\eta) d\eta + \int_{\eta_3}^{\eta_4} \sigma_s \theta f_h(\eta) d\eta + \int_{\eta_1}^{\eta_2} \sigma_s \theta f_{nh}(\eta) d\eta \\ &= \theta\, TV\,. \end{aligned}$$

As the marginal utility a increases, the following results show that the increased welfare from holders' utility is distributed proportionally among market participants if the tax θ is adjusted appropriately. Alternatively, from a welfare distribution perspective, we can say that an increase in the marginal utility a is *equivalent* to a tax decrease in terms of the share of welfare.

Proposition 6. *The normalized welfare quantities W_{trader}/a, W_{venue}/a, $W_{\text{government}}/a$, and W/a depend on a and θ only through θ/a.*

3.1 Tax Optimization

The government tax rate is carefully chosen to maximize welfare. However, the government could consider optimizing total welfare, its own welfare, or a weighted welfare, depending on its objective. As discussed above, because financial transactions only redistribute assets, total welfare is the utility welfare from traders holding the traded asset. Total welfare is hence optimized with minimal frictions, which is accomplished with zero taxes, (see Figure 10). Therefore, we focus on the nontrivial cases of maximizing tax revenue and weighted welfare.

3.1.1 Tax revenue maximization

The government maximizes its tax revenue, which is the tax θ times the trading volume TV. The tax is set as the solution to $\max_\theta W_{\text{government}}$. Interestingly, we see in Proposition 7 that this quantity is independent of σ_f and σ_s, meaning that it is independent of the technology driving the speed of the venues. The optimal tax also does not depend on the asset supply Z. In fact, θ/a depends only on γ and ρ. With the other values fixed, we see from (9) that as ρ and γ increase, the trading boundaries η_1 and η_4 separate. Hence, the traders are less inclined to trade as their impatience grows or their shocks are more frequent. This tendency is natural, as the fees and taxes are paid at the time of trading, but utility is gained over the time the asset is held. Both parameters ρ and γ decrease this utility: the impatience parameter ρ does so directly by discounting

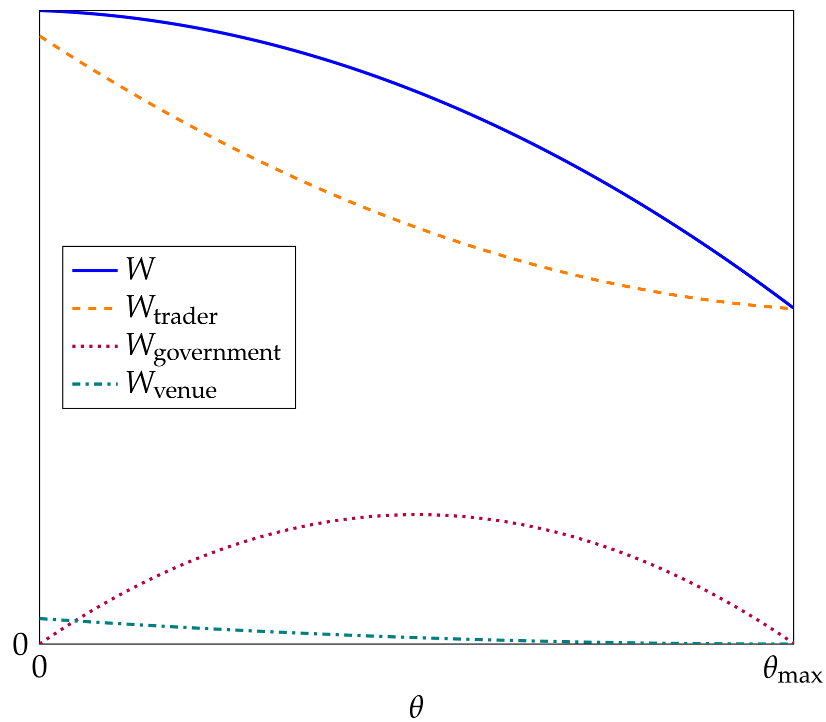

Figure 10: Welfare of market participants as a function of the tax θ. Here $\theta_{\max}$ denotes the maximum tax possible for Assumption 3 to hold.

future utility, whereas γ does so indirectly by increasing the likelihood of having other holding preferences in the future. Consequently, as shown in Proposition 7, if traders are relatively more sensitive to the costs of trading, the tax is lower to encourage taxable trading. This is the same effect observed for venue fees in Proposition 2.

Proposition 7. *The tax rate θ that optimizes the government tax revenue is $\frac{a}{4(\gamma+\rho)}$.*

Note that although the optimal tax rate only depends on γ and ρ, tax revenue also depends on the other parameters. The government, therefore, has an interest in the level of all parameters, even those that do not directly shape the tax policy.

3.1.2 Weighted welfare

The government has an interest in welfare distribution. It takes this into account by optimizing the weighted welfare

$$W_w = w_{\text{trader}} W_{\text{trader}} + w_{\text{venue}} W_{\text{venue}} + w_{\text{government}} W_{\text{government}}.$$

The extreme case of Section 3.1.1 is a special case when $w_{\text{trader}} = w_{\text{venue}} = 0$ and $w_{\text{government}} > 0$. It is in the interest of both traders and venues to have low taxes, as that facilitates trade. Indeed, as long as the weight on traders and venues is high enough, the weighted welfare optimized tax is always zero:

Proposition 8. *If $w_{\text{government}} \leq \min\{w_{\text{trader}}, w_{\text{venue}}\}$, then $\frac{\partial W_w}{\partial \theta} < 0$, which means welfare is optimized at $\theta = 0$.*

The intuition behind Proposition 8 is the same as why W is optimized at zero, but no longer self-evident because W_w is, unlike W, not equal to the holding utility. Moreover, when $w_{\text{government}}$ is not small, the weighted welfare is instead maximized for nontrivial θ.

Proposition 9. *If $w_{\text{government}}$ is sufficiently large relative to w_{trader} and w_{venue}, then W_w is optimized at $\theta > 0$.*

We see in Figure 11 that $w_{\text{government}}$ must not be particularly large for this effect to appear. In the figure, the government weight is only 20% larger than that of the other two. In fact, the effect is still present when the weighting is only 10% but is less visibly apparent; consequently, we plot the case of 20%.

4 Conclusion

We set up a model to study the equilbria and interactions between cryptocurrency trading venues, traders, and taxation in which the venues differ in technology (fast vs. slow). The property distinguishing this market from other markets like equities is the fact that each venue clears trades separately from one another. We first analyze the trading equilibrium and then the equilibrium of fee competition among trading venues. Trading fees rise when the venues are technologically differentiated. Improvements in the slow venue's technology is associated with greater trading volumes, whereas improvements in the fast venue's speed has an ambiguous impact. When the two venues have similar technologies, differentiation has a positive effect on trading volumes.

Our welfare analysis suggests that in equilibrium, the tax rate to optimize tax revenue depends only on trader preferences, and it is thus independent of the venue properties and competition. Because aggregate welfare is generated from holding utility, it is maximized with minimal market frictions, i.e., no taxes. When the government assigns

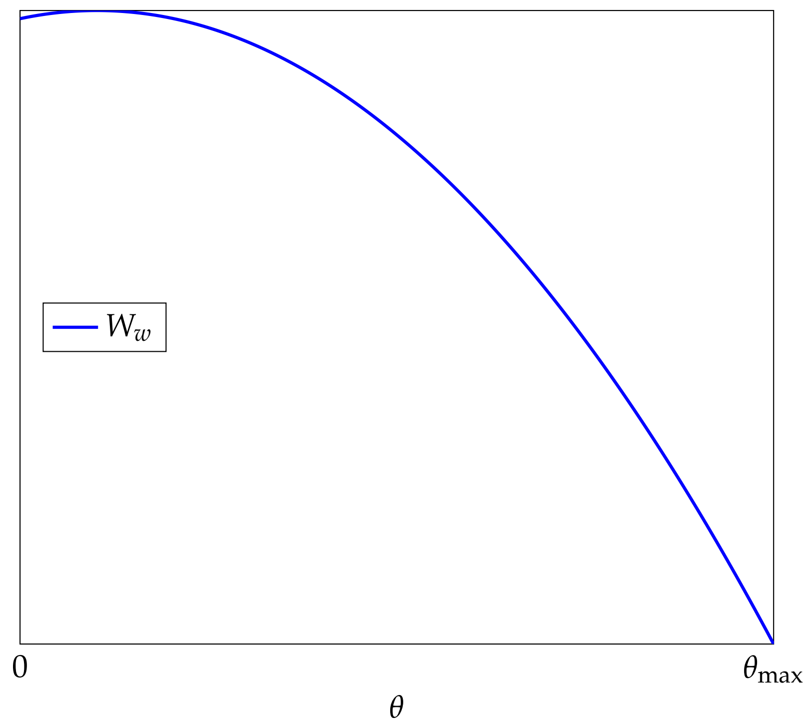

Figure 11: Example of non-trivial taxes when maximizing weighted welfare. Here $w_{\text{government}} = 1.2w_{\text{trader}} = 1.2w_{\text{venue}}$ and $\theta_{\max}$ denotes the maximum tax possible for Assumption 3 to hold.

different weights to its revenue, trader welfare, and venue welfare, the aggregate welfare can have a maximum for nontrivial tax rates.

In light of the recent growth in cryptocurrency trading, these results provide timely insight into the structure of cryptocurrency markets and the effects of taxation. We hope that the speed and technological aspects of venue differentiation can be fruitful to future analysis of DeFi platforms such as decentralized exchanges, of which there are many that differ in technology and speed.

References

Abadi, Joseph, and Markus Brunnermeier. 2018. "Blockchain economics." National Bureau of Economic Research.

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, Christophe Bisiere, Matthieu Bouvard, and Catherine Casamatta. 2019. "The blockchain folk theorem." *Review of Financial Studies*, 32(5): 1662–1715.

Borri, Nicola, and Kirill Shakhnov. 2019. "The cross-section of cryptocurrency returns." *Available at SSRN 3241485*.

Chiu, Jonathan, and Thorsten V Koeppl. 2017. "The economics of cryptocurrencies–bitcoin and beyond." *Available at SSRN 3048124*.

Cong, Lin William, and Zhiguo He. 2019. "Blockchain disruption and smart contracts." *Review of Financial Studies*, 32(5): 1754–1797.

Cong, Lin William, Ye Li, and Neng Wang. 2021. "Tokenomics: Dynamic adoption and valuation." *Review of Financial Studies*, 34(3): 1105–1155.

Cong, Lin William, Zhiguo He, and Jiasun Li. 2021. "Decentralized mining in centralized pools." *The Review of Financial Studies*, 34(3): 1191–1235.

Easley, David, Maureen O'Hara, and Soumya Basu. 2019. "From mining to markets: The evolution of bitcoin transaction fees." *Journal of Financial Economics*, 134(1): 91–109.

Glosten, L. R., and P. Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of Financial Economics*, 14(1): 71–100.

Hu, Albert S, Christine A Parlour, and Uday Rajan. 2019. "Cryptocurrencies: Stylized facts on a new investible instrument." *Financial Management*, 48(4): 1049–1068.

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.

Lehar, Alfred, and Christine A Parlour. 2020. "Miner collusion and the bitcoin protocol." working paper, University of California, Berkeley.

Lehar, Alfred, and Christine A Parlour. 2021. "Decentralized exchanges." working paper, University of California, Berkeley.

Liu, Yukun, and Aleh Tsyvinski. 2021. "Risks and returns of cryptocurrency." *Review of Financial Studies*, 34(6): 2689–2727.

Makarov, Igor, and Antoinette Schoar. 2020. "Trading and arbitrage in cryptocurrency markets." *Journal of Financial Economics*, 135(2): 293–319.

Nakamoto, Satoshi. 2008. "Bitcoin: A peer-to-peer electronic cash system." *Decentralized Business Review*, 21260.

Narayanan, Arvind, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. 2016. *Bitcoin and cryptocurrency technologies: a comprehensive introduction.* Princeton University Press.

Pagnotta, Emiliano. 2018. "Bitcoin as decentralized money: Prices, mining rewards, and network security." *Mining Rewards, and Network Security (October 26, 2018)*.

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

Roşu, Ioanid, and Fahad Saleh. 2021. "Evolution of shares in a proof-of-stake cryptocurrency." *Management Science*, 67(2): 661–672.

Schilling, Linda, and Harald Uhlig. 2019. "Some simple bitcoin economics." *Journal of Monetary Economics*, 106: 16–26.

Sockin, Michael, and Wei Xiong. 2020. "A model of cryptocurrencies." National Bureau of Economic Research.

Stoffels, Jaron. 2017. "Asset pricing of cryptocurrencies and momentum based patterns." *Erasmus School of Economics working paper*.

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

Weber, Warren E. 2016. "A Bitcoin standard: Lessons from the gold standard." Bank of Canada Staff Working Paper.

Yermack, David. 2015. "Is Bitcoin a real currency? An economic appraisal." In *Handbook of digital currency*. 31–43. Elsevier.

Appendix

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 taxes 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 holders and non-holders 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. $p_s - c_s \geq p_f - c_f$,

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

Proof. Assume for a contradiction that $p_s - c_s \geq p_f - c_f$ is not satisfied. Then we have $p_s - c_s < p_f - c_f$. For a contradiction, assume that there exists η such that $V_S^s(\eta) \geq V_H(\eta)$ and $V_S^s(\eta) \geq V_S^f(\eta)$. The former implies that $V_{0,\eta} + p_s - c_s - \theta - V_S^s(\eta) > 0$. Note that the values for $V_S^s(\eta)$ and $V_S^f(\eta)$ are given by following equations:

$$\rho V_S^s(\eta) = u(\eta) + \gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^s(\eta) \right) + \sigma_s \left(V_{0,\eta} + p_s - c_s - \theta - V_S^s(\eta) \right) \tag{26}$$

$$\rho V_S^f(\eta) = u(\eta) + \gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^f(\eta) \right) + \sigma_f \left(V_{0,\eta} + p_f - c_f - \theta - V_S^f(\eta) \right) \tag{27}$$

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

$$\gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^f(\eta) \right) \geq \gamma \left(\mathbf{E}_{\eta'}[V_{1,\eta'}] - V_S^s(\eta) \right).$$

Moreover, $V_S^s(\eta) \geq V_S^f(\eta)$, $\sigma_f > \sigma_s$, $V_{0,\eta} + p_s - c_s - \theta - V_S^s(\eta) > 0$ together with $p_s - c_s < p_f - c_f$ implies that

$$\sigma_f \left(V_{0,\eta} + p_f - c_f - \theta - V_S^f(\eta) \right) > \sigma_s \left(V_{0,\eta} + p_s - c_s - \theta - V_S^s(\eta) \right)$$

Thus, we have $V_S^s(\eta) < V_S^f(\eta)$, which is a contradiction. As a result, there is no η such that $V_S^s(\eta) \geq V_H(\eta)$ and $V_S^s(\eta) \geq V_S^f(\eta)$. 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.[9] 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 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 $V_{1,\eta^*} - p_f - c_f - \theta - V_B^f(\eta^*) = 0$. But as $p_s + c_s < p_f + c_f$ and $V_B^s \leq V_s^f$ we have $V_{1,\eta^*} - p_s - c_s - \theta - V_b^f(\eta^*) > 0$. 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 showed 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 second part is analogous. Assume for a contradiction that $p_s + c_s \leq p_f + c_f$ is not satisfied. Then we have $p_s + c_s > p_f + c_f$. For a contradiction, assume that there exists η such that $V_B^s(\eta) \geq V_N(\eta)$ and $V_B^s(\eta) \geq V_B^f(\eta)$. The former implies that $V_{1,\eta} - p_s - c_s - \theta - V_B^s(\eta) > 0$. Note that the values for $V_B^s(\eta)$ and $V_B^f(\eta)$ are given by following equations:

$$\rho V_B^s(\eta) = \gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^s(\eta) \right) + \sigma_s \left(V_{1,\eta} - p_s - c_s - \theta - V_B^s(\eta) \right) \tag{28}$$

and

$$\rho V_B^f(\eta) = \gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^f(\eta) \right) + \sigma_f \left(V_{1,\eta} - p_f - c_f - \theta - V_B^f(\eta) \right) \tag{29}$$

Looking at term by term, as $V_B^s(\eta) \geq V_B^f(\eta)$, we have $\gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^f(\eta) \right) \geq \gamma \left(\mathbf{E}_{\eta'}[V_{0,\eta'}] - V_B^s(\eta) \right)$. Moreover, $V_B^s(\eta) \geq V_B^f(\eta)$, $\sigma_f > \sigma_s$, $V_{1,\eta} - p_s - c_s - \theta - V_B^s(\eta) > 0$ together with $p_s + c_s > p_f + c_f$ implies that

$$\sigma_f \left(V_{1,\eta} - p_f - c_f - \theta - V_B^f(\eta) \right) > \sigma_s \left(V_{1,\eta} - p_s - c_s - \theta - V_B^s(\eta) \right)$$

Thus, we have $V_B^s(\eta) < V_B^f(\eta)$, which is a contradiction. As a result, There is no η such that $V_B^s(\eta) \geq V_N(\eta)$ and $V_B^s(\eta) \geq V_B^f(\eta)$. This means that there is measure 0 of

[9] The continuity of V_N and V_H in η follows directly from continuity of $u(\eta)$. The continuity of V_S^s, V_S^f, V_B^s and V_B^f follows from continuity of $u(\eta)$, $V_{1,\eta}$ and $V_{0,\eta}$. To show the continuity of $V_{1,\eta}$ and $V_{0,\eta}$ 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 $V_{1,\eta}$ when she owns the asset and $V_{0,\eta}$ when she does not by playing the same strategy. Thus we have $V_{1,\eta'} \geq V_{1,\eta}$ and $V_{0,\eta'} > V_{0,\eta}$. Next, if η plays the equilibrium strategy of η', then $V_{1,\eta'} - V_{1,\eta}$ and $V_{0,\eta'} - V_{0,\eta}$ 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.

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 positive demand in buying 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 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 slow. 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 $V_{0,\eta^*} + p_f - c_f - \theta - V_B^f(\eta^*) = 0$. But as $p_s - c_s > p_f - c_f$ and $V_S^s \leq V_S^f$ we have $V_{0,\eta^*} + p_s - c_s - \theta - V_b^f(\eta^*) > 0$. 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 showed there are no traders that prefer to buy slowly, 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 $V_{1,\eta} = V_S^s(\eta)$ and $V_{0,\eta} = V_B^s(\eta)$. Then by substituting for RHS and summing:

$$(\gamma + \rho)(V_{1,\eta} + V_{0,\eta}) = u(\eta) + \gamma(\mathbf{E}_{\eta'}[V_{1,\eta'}] + \mathbf{E}_{\eta'}[V_{0,\eta'}])) - 2\sigma_s(c_s + \theta) \tag{30}$$

From the optimality of sell slow, we have:

$$(\sigma_s + \gamma + \rho)(\gamma + \rho)[V_S^s(\eta) - V_H] = -\sigma_s(u(\eta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta}]) + \sigma_s(\gamma + \rho)(V_{0,\eta} + p_s - c_s - \theta) \geq 0$$

Rearranging:

$$(\gamma + \rho)V_{0,\eta} \geq u(\eta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}] - (\gamma + \rho)(p_s - c_s - \theta) \tag{31}$$

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

$$(\gamma+\rho)V_{1,\eta} \geq \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}] + (\gamma+\rho)(p_s+c_s+\theta) \tag{32}$$

Summing equations 31 and 32, we obtain:

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

Which contradicts equation 30. Replacing s with f in above proof proves (ii) and (iv). □

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

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

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

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

Which implies:

$$\begin{aligned}
(\sigma_s+\gamma+\rho)(V_{1,\eta} - p_s - c_s - \theta) &> \sigma_s(V_{1,\eta} - p_s - c_s - \theta) + \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}] \\
V_{1,\eta} - p_s - c_s - \theta &> V_B^s(\eta) = V_{0,\eta} \\
V_{1,\eta} &> V_{0,\eta} + p_s + c_s + \theta
\end{aligned} \tag{33}$$

Which is very 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:

$$V_{1,\eta} < V_{0,\eta} + p_f - c_f - \theta \tag{34}$$

Equations 33 and 34 imply $p_f - c_f > p_s + c_s + 2\theta$. Subtracting $2c_s - 2\theta$ from LHS, we obtain $p_f - c_f > p_s - c_s$ which is a contradiction. Switching s with f in the above proof proves (ii).

□

Following lemma is immediate given the previous ones:

Lemma 4.

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

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

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

Proof. As $u(\eta)$ is strictly increasing, we have following:

$$\frac{\partial V_N(\eta)}{\partial \eta} = 0$$

$$\frac{\partial V_H(\eta)}{\partial \eta} > 0$$

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

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

Where first inequality follows from the optimality of $V_{0,\eta}$, second follows from the fact $u(\eta)$ is strictly increasing and third equality follows from $\eta \in B_s$. Similarly, assume $\eta \in B_f$. Take any $\eta' > \eta$. With the exactly same reasoning, we obtain:

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

Hence $\frac{\partial V_{0,\eta}}{\partial \eta} \geq 0$, 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.

$\square$

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.

$\square$

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.

$\square$

The structure follows from Lemmas 6 and 7. The fact that $\eta_1 < \eta_4$ follows from $S_f \cap B_s = \emptyset$ and $S_s \cap B_s = \emptyset$.

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 following two conditions is necessary for any trader to prefer trade in any equilibrium:

$$V_B^s(\eta_h) > V_N \tag{35}$$

$$V_S^s(0) > V_H(0) \tag{36}$$

From 35:

$$\frac{\sigma_s(V_{1,\eta_h} - p_s - c_s - \theta) + \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{(\sigma_s + \gamma + \rho)} > \frac{\gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{(\gamma + \rho)}$$

Which reduces to:

$$(\gamma + \rho)\sigma_s(V_{1,\eta_h} - p_s - c_s - \theta) > \sigma_s \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]$$

Hence:

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

From 36, using $u(0) = 0$:

$$\frac{\sigma_s(V_{0,0} + p_s - c_s - \theta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{(\sigma_s + \gamma + \rho)} > \frac{\gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{(\gamma + \rho)}$$

Hence

$$(\gamma + \rho)(V_N + p_s - c_s - \theta) > (\gamma + \rho)V_H(0) \tag{38}$$

Summing 37 and 38:

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

□

Following lemma characterizes the condition under which 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:

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

After some algebra, we obtain:

$$V_N - V_H(0) > -\frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} \tag{39}$$

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

$$V_H(\eta_h) - V_N > \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f + c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s + c_s + \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} \tag{40}$$

Summing equations 39 and 40, 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.

□

First, we find the distributions for f_h and f_{nh} (we leave the values undefined at the cutoffs $\{\eta_i\}_{i=1,\ldots,4}$ as values in individual points do not matter):

Lemma 10.

$$f_{nh}(\eta) = \begin{cases} f(\eta)\frac{\sigma_f + (1-Z)\gamma}{\sigma_f + \gamma} & \text{if } \eta \in (\eta_l, \eta_3) \\ f(\eta)\frac{\sigma_s + (1-Z)\gamma}{\sigma_s + \gamma} & \text{if } \eta \in (\eta_3, \eta_4) \\ f(\eta)(1-Z) & \text{if } \eta \in (\eta_4, \eta_1) \\ f(\eta)\frac{(1-Z)\gamma}{\sigma_s + \gamma} & \text{if } \eta \in (\eta_1, \eta_2) \\ f(\eta)\frac{(1-Z)\gamma}{\sigma_f + \gamma} & \text{if } \eta \in (\eta_2, \eta_h) \end{cases}, \quad f_h(\eta) = \begin{cases} f(\eta)\frac{Z\gamma}{\sigma_f + \gamma} & \text{if } \eta \in (\eta_l, \eta_3) \\ f(\eta)\frac{Z\gamma}{\sigma_s + \gamma} & \text{if } \eta \in (\eta_3, \eta_4) \\ f(\eta)Z & \text{if } \eta \in (\eta_4, \eta_1) \\ f(\eta)\frac{\gamma Z + \sigma_s}{\sigma_s + \gamma} & \text{if } \eta \in (\eta_1, \eta_2) \\ f(\eta)\frac{\gamma Z + \sigma_f}{\sigma_f + \gamma} & \text{if } \eta \in (\eta_2, \eta_h) \end{cases}$$

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

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

As $f_h(\eta) + f_{nh}(\eta) = f(\eta)$, we have:

$$f_{nh}(\eta) = \frac{\sigma_f + (1-Z)\gamma}{\sigma_f + \gamma} f(\eta)$$

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

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

As $f_h(\eta) + f_{nh}(\eta) = f(\eta)$, we have:

$$f_{nh}(\eta) = \frac{\sigma_s + (1-Z)\gamma}{\sigma_s + \gamma} f(\eta)$$

Let $\eta \in (\eta_4, \eta_1)$. Inflow outflow balance for holders with valuation η requires:

$$\gamma f_h(\eta) = \gamma Z f(\eta) \implies f_h(\eta) = Z f(\eta)$$

Hence $f_{nh}(\eta) = (1-Z) f(\eta)$.

Let $\eta \in (\eta_1, \eta_2)$. Inflow-outflow balance requires:

$$\gamma f_{nh}(\eta) + \sigma_s f_{nh}(\eta) = (1-Z)\gamma f(\eta) \implies f_{nh}(\eta) = f(\eta) \frac{(1-Z)\gamma}{\gamma + \sigma_s}$$

As $f_h(\eta) + f_{nh}(\eta) = f(\eta)$, we have:

$$f_h(\eta) = f(\eta) \frac{\gamma Z + \sigma_s}{\gamma + \sigma_s}$$

Let $\eta \in (\eta_2, \eta_h)$. Inflow-outflow balance requires:

$$\gamma f_{nh}(\eta) + \sigma_f f_{nh}(\eta) = (1-Z)\gamma f(\eta) \implies f_{nh}(\eta) = f(\eta) \frac{(1-Z)\gamma}{\gamma + \sigma_f}$$

As $f_h(\eta) + f_{nh}(\eta) = f(\eta)$, we have:

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

$\square$

We first need find the market clearing conditions.

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

$$(1-Z)F(\eta_1) + ZF(\eta_4) = 1 - Z \tag{41}$$

$$(1-Z)F(\eta_2) + ZF(\eta_3) = 1 - Z \tag{42}$$

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

$$\int_{\eta_2}^{\eta_h} f_{nh}(\eta)d\eta = \int_{\eta_l}^{\eta_3} f_h(\eta)d\eta \tag{43}$$

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

$$\frac{1-Z}{Z} = \frac{F(\eta_3)}{1-F(\eta_2)} \tag{45}$$

$$ZF(\eta_3) + (1-Z)F(\eta_2) = (1-Z). \tag{46}$$

From slow market clearing condition:

$$\int_{\eta_1}^{\eta_2} f_{nh}(\eta)d\eta = \int_{\eta_3}^{\eta_4} f_h(\eta)d\eta \tag{47}$$

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

$$\frac{1-Z}{Z} = \frac{F(\eta_4)-F(\eta_3)}{F(\eta_2)-F(\eta_1)} \tag{49}$$

$$(1-Z)F(\eta_1) + ZF(\eta_4) = (1-Z), \tag{50}$$

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:*

$$\frac{u(\eta_1)-u(\eta_4)}{\gamma+\rho} = 2(c_s+\theta) \tag{51}$$

$$\frac{u(\eta_2)-u(\eta_3)}{\gamma+\rho} = 2\frac{\sigma_f(c_f+\theta)(\sigma_s+\gamma+\rho)-\sigma_s(c_s+\theta)(\sigma_f+\gamma+\rho)}{(\sigma_f-\sigma_s)(\gamma+\rho)} \tag{52}$$

Proof. It observes that

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

In addition as η_1 is the cutoff type for buying slowly and doing nothing, we have $V_B^s(\eta_1) = V_N$[10]. Combining these:

$$V_H(\eta_1) = V_N + p_s + c_s + \theta \tag{53}$$

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

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

Last equality implies

$$V_H(\eta_4) = V_N + p_s - c_s - \theta \tag{54}$$

Using equations 53 and 54, we get $V_H(\eta_1) - V_H(\eta_4) = 2(c_s + \theta)$. Hence:

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

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

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

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

Rearranging:

$$(\sigma_f - \sigma_s)(u(\eta_3) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]) = (\sigma_f - \sigma_s)(\gamma + \rho)V_N \\ + \sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta) \tag{56}$$

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

$$\begin{aligned} V_N &= \frac{u(\eta_3)}{\gamma + \rho} + \frac{\gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{\gamma + \rho} - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} \\ &= V_H(\eta_3) - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} \end{aligned} \tag{57}$$

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

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

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

$$V_H(\eta_2) - \frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f + c_f + \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s + c_s + \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} = V_N \tag{58}$$

Solving 57 and 58 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. □

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 41, 42, 51 and 52 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 54, we obtain the characterization of p_s:

$$\begin{aligned}
p_s &= V_h(\eta_4) - V_N + c_s + \theta \\
&= \frac{u(\eta_4) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{\gamma + \rho} + c_s + \theta - \frac{\gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{\gamma + \rho} \\
&= \frac{u(\eta_4)}{\gamma + \rho} + c_s + \theta + \frac{\gamma}{\gamma + \rho}\Big(\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}]\Big)
\end{aligned} \tag{59}$$

From equation 57:

$$\frac{\sigma_f(\sigma_s + \gamma + \rho)(p_f - c_f - \theta) - \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta)}{(\sigma_f - \sigma_s)(\gamma + \rho)} = \frac{u(\eta_3)}{\gamma + \rho} + \frac{\gamma}{\gamma + \rho}\Big(\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}]\Big) \tag{60}$$

Following lemma shows that 59 and 60 yields a unique price vector and completes the characterization of equilibrium.

Lemma 13. *Equations 59 and 60 characterize a unique price vector.*

Proof. We only need to show that (59) and (60) lead to unique solutions for p_s and p_f. In particular, we will show that the quantity $\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}]$ does not depend on either p_s nor p_f; once this is done, p_s is directly pinned down by (59), and after substituting into (60), 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 $\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}]$ only depends on the endogenous thresholds η_1, η_2, η_3 and η_4. First recall that

$$\begin{aligned}
V_S^f(\eta) &= \frac{u(\eta) + \sigma_f(V_N + p_f - c_f - \theta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{\sigma_f + \rho + \gamma} \\
V_S^s(\eta) &= \frac{u(\eta) + \sigma_s(V_N + p_s - c_s - \theta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{\sigma_s + \rho + \gamma} \\
V_B^f(\eta) &= \frac{\sigma_f(V_H(\eta) - p_f - c_f - \theta) + \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{\sigma_f + \rho + \gamma} \\
V_B^s(\eta) &= \frac{\sigma_s(V_H(\eta) - p_s - c_s - \theta) + \gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{\sigma_s + \rho + \gamma} \\
V_H(\eta) &= \frac{u(\eta) + \gamma \mathbf{E}_{\eta'}[V_{1,\eta'}]}{\gamma + \rho} \\
V_N &= \frac{\gamma \mathbf{E}_{\eta'}[V_{0,\eta'}]}{\gamma + \rho}.
\end{aligned} \tag{61}$$

Next we derive $\mathbf{E}_{\eta'}[V_{0,\eta'}]$ and $\mathbf{E}_{\eta'}[V_{1,\eta'}]$ in closed forms.

- **Computing $\mathbf{E}_{\eta'}[V_{0,\eta'}]$.** Using (61) we have

$$\frac{dV_{0,\eta}}{d\eta} = \mathbf{1}_{\{\eta\in[\eta_l,\eta_1]\}} \times 0 + \mathbf{1}_{\{\eta\in[\eta_1,\eta_2]\}} \frac{\sigma_s u'(\eta)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} + \mathbf{1}_{\{\eta\in[\eta_2,\eta_h]\}} \frac{\sigma_f u'(\eta)}{(\sigma_f+\rho+\gamma)(\rho+\gamma)}. \tag{62}$$

  Integrating (62) implies

$$\begin{aligned} V_{0,\eta} &= \mathbf{1}_{\{\eta\in[\eta_l,\eta_1]\}} V_N + \mathbf{1}_{\{\eta\in[\eta_1,\eta_2]\}} \left[V_N + \int_{\eta_1}^{\eta} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi\right] \\ &\quad + \mathbf{1}_{\{\eta\in[\eta_2,\eta_h]\}} \left[V_N + \int_{\eta_1}^{\eta_2} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi + \int_{\eta_2}^{\eta} \frac{\sigma_f u'(\xi)}{(\sigma_f+\rho+\gamma)(\rho+\gamma)} d\xi\right]. \end{aligned} \tag{63}$$

  Finally, taking an expectation from (63) gives

$$\begin{aligned} \mathbf{E}_{\eta'}[V_{0,\eta'}] &= V_N + \int_{\eta_1}^{\eta_2} \left(\int_{\eta_1}^{\tilde\eta} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi\right) dF(\tilde\eta) \\ &\quad + \left(\int_{\eta_1}^{\eta_2} \frac{\sigma_s u'(\xi)}{(\sigma_s+\rho+\gamma)(\rho+\gamma)} d\xi\right) \Big(1 - F(\eta_2)\Big) \\ &\quad + \int_{\eta_2}^{\eta_h} \left(\int_{\eta_2}^{\tilde\eta} \frac{\sigma_f u'(\xi)}{(\sigma_f+\rho+\gamma)(\rho+\gamma)} d\xi\right) dF(\tilde\eta). \end{aligned} \tag{64}$$

  Thus, (64) shows that $\mathbf{E}_{\eta'}[V_{0,\eta'}]$ only depends on the endogenous thresholds η_1 and η_2.

- **Computing $\mathbf{E}_{\eta'}[V_{1,\eta'}]$.** Using (61) we have

$$\frac{dV_{1,\eta}}{d\eta} = \mathbf{1}_{\{\eta\in[\eta_l,\eta_3]\}} \frac{u'(\eta)}{\sigma_f+\rho+\gamma} + \mathbf{1}_{\{\eta\in[\eta_3,\eta_4]\}} \frac{u'(\eta)}{\sigma_s+\rho+\gamma} + \mathbf{1}_{\{\eta\in[\eta_4,\eta_h]\}} \frac{u'(\eta)}{\rho+\gamma}. \tag{65}$$

Then, integrating gives

$$
\begin{aligned}
V_{1,\eta} = & \mathbf{1}_{\{\eta\in[\eta_l,\eta_3]\}} \int_{\eta_l}^{\eta} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi \\
& + \mathbf{1}_{\{\eta\in[\eta_3,\eta_4]\}} \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi + \int_{\eta_3}^{\eta} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) \\
& + \mathbf{1}_{\{\eta\in[\eta_4,\eta_h]\}} \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi + \int_{\eta_3}^{\eta_4} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi + \int_{\eta_4}^{\eta} \frac{u'(\xi)}{\gamma+\rho} d\xi \right). \quad (66)
\end{aligned}
$$

Therefore, taking an expectation from (66) gives

$$
\begin{aligned}
\mathbf{E}_{\eta'}[V_{1,\eta'}] = & \int_{\eta_l}^{\eta_3} \left(\int_{\eta_l}^{\tilde\eta} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi \right) dF(\tilde\eta) + \left(\int_{\eta_l}^{\eta_3} \frac{u'(\xi)}{\sigma_f+\rho+\gamma} d\xi \right) \Big(1-F(\eta_3)\Big) \\
& + \int_{\eta_3}^{\eta_4} \left(\int_{\eta_3}^{\tilde\eta} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) dF(\tilde\eta) + \Big(1-F(\eta_4)\Big) \left(\int_{\eta_3}^{\eta_4} \frac{u'(\xi)}{\sigma_s+\rho+\gamma} d\xi \right) \\
& + \int_{\eta_4}^{\eta_h} \left(\int_{\eta_4}^{\tilde\eta} \frac{u'(\xi)}{\gamma+\rho} d\xi \right) dF(\tilde\eta) \quad (67)
\end{aligned}
$$

Thus, (67) shows that $\mathbf{E}_{\eta'}[V_{1,\eta'}]$ only depends on the endogenous thresholds η_3 and η_4. Together, (64) and (67) finish the proof that the difference $\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}]$ does not depend on p_s nor p_f.

To finish the proof of the lemma, note that equations (59) and (60) yield unique solutions for p_s and p_f. Since $\mathbf{E}_{\eta'}[V_{1,\eta'}] - \mathbf{E}_{\eta'}[V_{0,\eta'}] > 0$, all terms in (59) are positive, and hence $p_s > 0$. Since we have assumed that $u(\cdot) \geq 0$, it follows that p_f is also positive, which finishes the lemma. □

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:

$$(1 - Z)F(\eta_4) + ZF(\eta_1) = 1 - Z \tag{68}$$

$$\frac{u(\eta_1) - u(\eta_4)}{\gamma + \rho} = 2(c_s + \theta) \tag{69}$$

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

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

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, fast venue clearing condition cannot hold.). To see that, there is no demand for selling in the fast venue if $V_S^s(0) - V_S^f(0) \geq 0$. This corresponds to:

$$(V_H(0) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta) - (p_f - c_f - \theta)\sigma_f(\sigma_s + \gamma + \rho) \geq 0$$

Which is equivalent to:

$$\begin{aligned} L = {} & (V_H(0) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s - c_s - \theta) + (c_f + \theta)\sigma_f(\sigma_s + \gamma + \rho) \\ & \geq p_f(\sigma_f(\sigma_s + \gamma + \rho)) \end{aligned} \tag{70}$$

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

$$\begin{aligned} U = {} & (V_H(\eta_h) - V_N)(\sigma_f - \sigma_s)(\gamma + \rho) + \sigma_s(\sigma_f + \gamma + \rho)(p_s + c_s + \theta) - (c_f + \theta)\sigma_f(\sigma_s + \gamma + \rho) \\ & \leq p_f(\sigma_f(\sigma_s + \gamma + \rho)) \end{aligned} \tag{71}$$

Note that there exists a p_f such that there is no demand in selling or buying fast if $L > U$. 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 8, 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 $\frac{\partial u(\eta_3)}{\partial x} < 0$.[11] 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. $\square$

Following lemma shows all the comparative statics of the model:

Lemma 15. *We have 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. $\frac{\partial \eta_1}{\partial \theta} < 0$, $\frac{\partial \eta_4}{\partial \theta} > 0$, $\frac{\partial \eta_2}{\partial \theta} > 0$, $\frac{\partial \eta_3}{\partial \theta} < 0$

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

Proof. **Part 1**

First two equations are trivial as σ_s does not appear in equations that determine η_1 and η_4. For the last two:

$$\frac{\partial(u(\eta_2) - u(\eta_3))}{\partial \sigma_s} = 2\frac{(c_f - c_s)\sigma_f(\sigma_f + \gamma + \rho)}{(\sigma_f - \sigma_s)^2} > 0$$

Result follows from lemma 14.

Part 2

First two equations are trivial as σ_s does not appear in equations that determine η_1 and η_4. We have:

$$\frac{\partial(u(\eta_2) - u(\eta_3))}{\partial \sigma_f} = -2\frac{(c_f - c_s)\sigma_f(\sigma_f + \gamma + \rho)}{(\sigma_f - \sigma_s)^2} < 0$$

Result follows from lemma 14.

Parts 3 and 4

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

$$\frac{\partial(u(\eta_1) - u(\eta_4))}{\partial \gamma} = 2c_s > 0$$

Result follows from lemma 14. Proof for ρ is exactly same.

Part 5

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

$$\frac{\partial(u(\eta_1) - u(\eta_4))}{\partial c_s} = 2(\gamma + \rho) > 0$$

Result follows from lemma 14.

Part 6

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

$$\frac{\partial(u(\eta_1)-u(\eta_4))}{\partial c_f}=0$$

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 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 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 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 same.

Part 7 of 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:

$$\frac{\partial R_s}{\partial c_s}=\frac{4\gamma\sigma_f\sigma_s(\gamma+\sigma_s+\rho)(1-Z)Z}{a(\gamma+\sigma_s)(\sigma_f-\sigma_s)}(c_f-2c_s) \tag{72}$$

$$\frac{\partial R_f}{\partial c_f}=\frac{2\gamma\sigma_f(Z-1)Z\Big(a(\sigma_s-\sigma_f)-2c_s\sigma_s(\gamma+\sigma_f+\rho)+4c_f\sigma_f(\gamma+\sigma_s+\rho)+2(\sigma_f-\sigma_s)(\gamma+\rho)\theta\Big)}{a(\gamma+\sigma_f)(\sigma_f-\sigma_s)} \tag{73}$$

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 $c_s = 0$:

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

which is positive by Assumption 3. 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 (72) to zero, we find

$$c_s^*(c_f) = \frac{c_f}{2}. \tag{74}$$

Solving for the root of (73) and using (74) 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 3. 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 3. 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 3 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 3,

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

Proof of Proposition 5

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

The fraction is positive, so

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

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

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

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

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

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

$$\lim_{\epsilon \to 0} \frac{\partial TV}{\partial \epsilon} > 0 \iff \rho < \frac{\gamma^2 + \gamma\sigma}{\sigma - \gamma}$$

Proof of Proposition 6

First, we observe from Proposition 2 that c_f/a and c_s/a depend only on θ/a. Therefore, after dividing (9) and (10) by a, the same holds for η_1, η_2, η_3, and η_4. That

$$W/a = \int_0^1 \eta f_h(\eta) d\eta$$

depends only on θ/a follows from f_h only depending on the thresholds $\{\eta_i\}_{i=1,\ldots,4}$.

Next, consider the trading volumes m_f and m_s in (17) and (18). Because they also only depend on θ and a through the thresholds $\{\eta_i\}_{i=1,\ldots,4}$, their dependence is also on θ/a. Hence, the scaled revenue, $c_v\, TV_v\, /a$ depends only on θ/a because c_v/a does. The quantity W_{venue}/a is the sum of two such term.

We have already concluded that TV depends only on θ/a. It is therefore a direct consequence that the scaled tax revenue $W_{\text{government}}/a = \theta\, TV\, /a$ also only depends on θ/a.

Finally, because $W_{\text{trader}} = W - W_{\text{venue}} - W_{\text{government}}$, the claim follows from the above arguments.

Proof of Proposition 7

We differentiate $W_{\text{government}}$ with respect to θ:

$$\frac{\partial W_{\text{government}}}{\partial \theta} = -\frac{2\gamma\sigma_f(3\sigma_f\sigma_s + \gamma(2\sigma_f + \sigma_s))(\gamma + \sigma_s + \rho)(4(\gamma + \rho)\theta - a)(1 - Z)Z}{a(\gamma + \sigma_f)(\gamma + \sigma_s)((\gamma + \rho)(4\sigma_f - \sigma_s) + 3\sigma_f\sigma_s)}.$$

We see that the first order condition $\frac{W_{\text{government}}}{\partial \theta} = 0$ is satisfied if and only if $\theta = \frac{a}{4(\gamma+\rho)}$. Because the coefficient in front of θ is negative, we directly verify that this indeed corresponds to a maximum.

Proof of Proposition 8

Denote by W' the derivative $\frac{\partial W}{\partial \theta}$, and similarly for other functions. We first observe that

$$W' < 0.$$

Indeed,

$$
\begin{aligned}
W' = &-\frac{2\sigma_f(\gamma+\rho)(\gamma+\sigma_s+\rho)(1-Z)Z}{a(\gamma+\sigma_f)(\gamma+\sigma_s)((\gamma+\rho)(-4\sigma_f+\sigma_s)-3\sigma_f\sigma_s)^2} \\
&\times \Big(a(\sigma_f-\sigma_s)(\gamma^2(4\sigma_f+\sigma_s)+3\sigma_f\sigma_s\rho+\gamma(5\sigma_f\sigma_s+4\sigma_f\rho+\sigma_s\rho)) \\
&\quad +2\sigma_f(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s)(\gamma+\rho)(\gamma+\sigma_s+\rho)\theta\Big) < 0.
\end{aligned}
$$

Next, we show that the derivatives W'_{trader} and W'_{venue} are negative:

$$
W'_{\text{venue}} = -\frac{4\gamma\sigma_f(\sigma_f-\sigma_s)(5\sigma_f\sigma_s+\gamma(4\sigma_f+\sigma_s))(\gamma+\rho)(\gamma+\sigma_s+\rho)(a-2(\gamma+\rho)\theta)(1-Z)Z}{a(\gamma+\sigma_f)(\gamma+\sigma_s)(-3\sigma_f\sigma_s+\gamma(-4\sigma_f+\sigma_s)-4\sigma_f\rho+\sigma_s\rho)^2},
$$

which we directly verify to be negative. For the traders,

$$
\begin{aligned}
W'_{\text{trader}} = &-\frac{2\sigma_f(\gamma+\sigma_s+\rho)(1-Z)Z}{a(\gamma+\sigma_f)(\gamma+\sigma_s)((\gamma+\rho)(-4\sigma_f+\sigma_s)-3\sigma_f\sigma_s)^2} \\
&\times \Big(a\big[\gamma\sigma_f(\gamma+\sigma_s)(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s)+\gamma(8\gamma\sigma_f^2+2\sigma_f(\gamma+5\sigma_f)\sigma_s-(\gamma+\sigma_f)\sigma_s^2)\rho \\
&\qquad +(\sigma_f-\sigma_s)(3\sigma_f\sigma_s+\gamma(4\sigma_f+\sigma_s))\rho^2\big] \\
&\quad +2\sigma_f(\gamma+\rho)\big[-\gamma(\gamma+\sigma_s)(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s)+\sigma_s(9\sigma_f\sigma_s+\gamma(8\sigma_f+\sigma_s))\rho \\
&\qquad +(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s)\rho^2\big]\theta\Big),
\end{aligned}
$$

which only has one negative term in the parenthesis. Comparing the terms on lines 2 and 4,

$$
a\gamma\sigma_f(\gamma+\sigma_s)(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s) > 2\sigma_f(\gamma+\rho)\gamma(\gamma+\sigma_s)(4\gamma\sigma_f+5\gamma\sigma_s+9\sigma_f\sigma_s)\theta,
$$

by Assumption 3, and we conclude W'_{trader} is also negative.

Suppose $W'_w \geq 0$. Then

$$
\begin{aligned}
0 &\leq w_{\text{government}}W'_{\text{government}} + w_{\text{trader}}W'_{\text{trader}} + w_{\text{venue}}W'_{\text{venue}} \\
&\leq w_{\text{government}}(W'_{\text{government}} + W'_{\text{trader}} + W'_{\text{venue}}) \\
&= w_{\text{government}}W' < 0,
\end{aligned}
$$

where we used that W'_{trader} and W'_{venue} are negative. This is obviously a contradiction, so we conclude that $W'_w < 0$.

Proof of Proposition 9

For instance at the tax optimizing $\theta = \frac{a}{4(\gamma+\rho)} > 0$ it holds that $W_{\text{government}} > 0$. Because $W_{\text{government}}|_{\theta=0} = 0$, $W_w|_{\theta=0}$ is independent of $w_{\text{government}}$. Hence,

$$\lim_{w_{\text{government}} \to \infty} W_w|_{\theta=\frac{a}{4(\gamma+\rho)}} = \infty > W_w|_{\theta=0},$$

which concludes the proof.