NBER WORKING PAPER SERIES

COMPETING ON SPEED

Emiliano Pagnotta
Thomas Philippon

Competing on Speed
Emiliano Pagnotta and Thomas Philippon
NBER Working Paper No. 17652
December 2011, Revised December 2011
JEL No. G12,L13,L15

ABSTRACT

Two forces have reshaped global securities markets in the last decade: Exchanges operate at much faster speeds and the trading landscape has become more fragmented. In order to analyze the positive and normative implications of these evolutions, we study a framework that captures (i) exchanges' incentives to invest in faster trading technologies and (ii) investors' trading and participation decisions. Our model predicts that regulation that protect prices will lead to fragmentation and faster trading speed. Asset prices decrease when there is intermediation competition and are further depressed by price protection. Endogenizing speed can also change the slope of asset demand curves. On normative side, we find that for a given number of exchanges, faster trading is in general socially desirable. Similarly, for a given trading speed, competition among exchange increases participation and welfare. However, when speed is endogenous, competition between exchanges is not necessarily desirable. In particular, speed can be inefficiently high. Our model sheds light on important features of the experience of European and U.S. markets since the implementation of Reg. NMS, and provides some guidance for optimal regulations.

Emiliano Pagnotta
New York University
Stern School of Business
44 West 4th Street, Suite 9-81
New York, NY 10012
emiliano.pagnotta@stern.nyu.edu

Thomas Philippon
New York University
Stern School of Business
44 West 4th Street, Suite 9-190
New York, NY 10012-1126
and NBER
tphilipp@stern.nyu.edu

"In this high-tech stock market, Direct Edge and the other exchanges are sprinting for advantage. All the exchanges have pushed down their latencies [...] Almost each week, it seems, one exchange or another claims a new record [...] The exchanges have gone warp speed because traders have demanded it. Even mainstream banks and old-fashioned mutual funds have embraced the change". *The New York Times*, January 1st 2011, The New Speed of Money, Reshaping Markets

The securities exchange industry has been been deeply transformed over the past decade. In particular, the speed at which investors trade has increased a lot, and stock trading, particularly in the U.S., have become significantly more fragmented. The consequences of these transformations are the subject of heated debates in academic and policy circles. In this paper we provide a framework for the joint analysis of trading speed, trading regulations, and market fragmentation.

Let us consider trading speed first. Major market centers around the world have made costly investments in fast computerized trading platforms to reduce communication latencies. This process has gone beyond stock exchanges to include derivatives and currencies, and it has accelerated during the second half of the 2000s. Figure 1 illustrate this trend. It displays the reduction in execution times of small orders on NYSE and NASDAQ. What is driving this race for speed? In the human-driven trading era, higher execution speeds helped reduce moral hazard with floor brokers, but this is less relevant in the current environment. In this paper, we emphasize investor heterogeneity and vertical differentiation. We argue that market centers seeking to attract order flows have an incentive to relax price competition by differentiating along the speed dimension.

The second major feature of the new trading landscape is fragmentation, illustrated by Figure 2 for Europe and the U.S. The top panel (Europe) shows that traditional markets

2

Figure 1: Speed of Executing small orders (seconds)

Source: Angel, Harris and Spatt (2010)

such as the London Stock Exchange have lost market share to faster entrants such as Chi-X. The bottom panel (US) shows an even more dramatic evolution: the fraction of NYSE-listed stocks traded at the NYSE has decreased from 80% in 2004 to just over 20% in 2009. Most of the lost trading volume has been captured by new entrants (e.g. Direct Edge and BATS).

Market regulators were not passive witnesses of this process. In the U.S., policy makers have encouraged fragmentation to reduce the market power of exchanges and other intermediaries (prominently with Reg. NMS, which we discuss below).[1] The effects are tangible: big cap stocks that previously traded in one or two exchanges can now be traded in near fifty venues (including internalization. pools and dark venues). But, does fragmentation achieve the desired goal? Should it be fostered in the first place? We argue that the answer to this question depends on the nature of innovations in intermediation services (e.g., execution delays). This dimension, however, is largely absent from the fragmentation literature.

---

[1]For example, the SEC stated in 2010: "mandating the consolidation of order flow in a single venue would create a monopoly and thereby lose the important benefits of competition among markets. The benefits of such competition include incentives for trading centers to create new products, provide high quality trading services that meet the needs of investors, and keep trading fees low."

Figure 2: Market Fragmentation
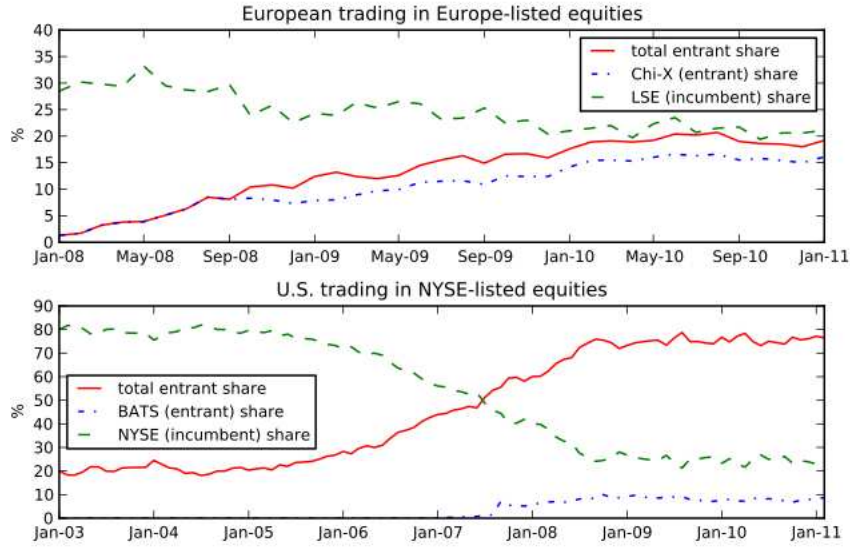
source: Barclays Capital Equity Research and Federation of European Securities Exchanges

Source: Menkveld (2011)

Analyzing these issues is difficult because it requires modeling four separate components: (i) why and how investors value trading speed; (ii) how differences in speed affect competition among trading venues and the affiliation choices of investors; (iii) how trading regulations affect (i) and (ii); and (iv) how these choices affect investment in speed and equilibrium fragmentation. These requirements explain our modeling choices and the structure of our paper, which is depicted in Figure 3.

Figure 3: Timing and Structure of the Model
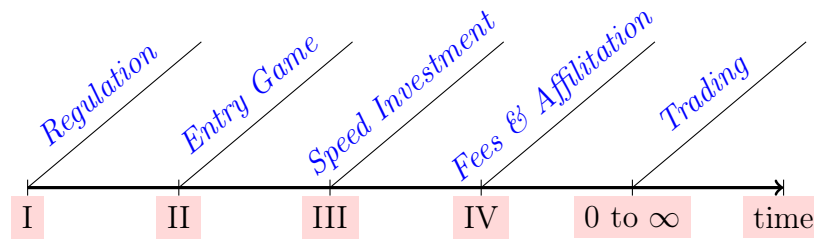
Our first task is thus to provide explicit micro-foundations for how investors value speed in financial markets. We consider a dynamic infinite horizon model where heterogenous investors buy and a sell a single security. Ex-post gains from trade arise from shocks to the marginal utility (or marginal cost) of holding the asset.[2] High-marginal-utility investors are natural buyers, while low-marginal-utility investors are natural sellers of the asset. In this model, speed allows investors to realize a larger fraction of the ex-post gains from trade (Prop 1).

Our second task is to analyze the allocation of investors across trading venues. To do so we model ex-ante heterogeneity among venues and investors. Investors differ (ex-ante) by the volatility of their private value process, i.e. by how much their marginal utility can be expected to fluctuate once they start trading. Since gains from trade are larger for investors with high expected volatility, these investors attach a higher value to speed. Venues differ in their trading speeds and compete in prices. We characterize the equilibrium with one venue (monopoly), and the equilibrium with two venues and different speeds (differentiated duopoly). Competition leads to lower fees and higher participation by investors. Faster venues charge a higher price and attract speed-sensitive investors. The first contribution of our paper is to characterize the pricing decisions and equilibrium profits of trading venues and the participation and affiliation choices of investors (Prop 2).

Our third task is to analyze the impact of trading regulations. In the U.S., the trade-through rule provided by Rule 611 of Regulation NMS essentially requires that any venue execute its trades at the National Best Bid and Offer, thereby consolidating prices from a

---

[2]As is well understood in the literature, these shocks can capture liquidity demand (i.e., a need for cash), financing costs, hedging demand, or any other personal use of the asset, including specific arbitrage opportunities. See Duffie et al. (2007) for a discussion. The important point is that these shocks affect the private value of the asset, not its common value. Therefore they generate gains from trade. These gains from trade are a required building block in any trading model.

5

scattered trading map. We propose a stylized analysis of this regulation by considering two polar cases. In one case, which we refer to as "free segmentation" any venue can refuse to execute the trades of investors from the other venue. The venues are effectively segmented, and trades occur at different prices. The other case corresponds to "price protection". We find that "price protection" acts as a subsidy for the relatively slow market. At the trading stage, investors in the slow venue enjoy being able to trade with investors from the fast venue. Anticipating this, they are more willing to join the slow venue under "price protection" than under "free segmentation". An important contribution of our paper is to analyze how trading regulations affect ex-ante competition among exchanges (Prop 2 and 3).

When we endogenize the speed and the market structure, we find that price protection encourages entry. Without price protection, when venues form prices separately, there is a greater tendency towards consolidation, even in the absence of liquidity externalities (e.g. Pagano (1989)). In addition we show that fragmentation leads to more investment in trading technologies and faster trading speed. Putting these various pieces together, our model provides a consistent interpretation of the U.S. experience in recent years: after the implementation of Reg NMS, new market centers proliferated and trading speed increased rapidly (Prop 3, 4 and 5).

Modeling entry and speed choices can fundamentally change the prediction of the model regarding assets prices. With endogenous speed, we show that asset prices *increase* with asset supply. In addition, competition among exchanges tends to lower asset prices (Prop 6 and 7).

Finally, we analyze the welfare implications of entry, speed and affiliation choices. As a benchmark we characterize the efficient outcome under the constraint that venues break even. Somewhat surprisingly, we find that, even in the absence of liquidity externalities and

6

fixed entry costs, a planner would choose to operate only one venue. Our model then allows us to ask several question: When does competition increase welfare? When does investment in trading speed increase welfare?

We find that the market outcome is generally inefficient. In the monopoly case, participation is always too low and allowing for endogenous speed always improves welfare. The resulting speed might be higher or lower than the one chosen by the planner. In the frictional finance literature, it is often stated that higher asset prices are socially desirable. By analyzing explicitly the welfare impact of trading in secondary markets, we highlight that observed prices are not a sufficient statistic for welfare. For instance, prices can be inefficiently high due to limited investor participation.

In the duopoly case, both entry and speed can be inefficient. On the entry side, there is the usual trade-off between price competition and product diversity on the one hand, and business stealing on the other. Excessive entry is possible when entry costs are relatively high. Regarding speed choices, we find a fairly clear and intuitive condition: allowing venues to compete on speed improves welfare if the default speed is relatively low, but decreases welfare once the default speed reaches a certain threshold.

Regarding price protection, we find that it always has a first order negative impact on prices, but its impact on welfare depends crucially on entry. When protection increases entry, it has a first order positive impact on welfare. When it does not increase entry, it has a small negative impact on welfare.

<u>Discussion of the literature.</u>

Theory analyses of fragmentation include Mendelson (1987), Pagano (1989), and Madhavan (1995). Biais (1993), Glosten (1994), Hendershott and Mendelson (2000) and Parlour and Seppi (2003) study inter market competition under different trading rules. Other dimensions of

competition between exchanges have been addressed. Santos and Scheinkman (2001) study competition in margin requirements, while Foucault and Parlour (2004) study competition in listing fees. Recent empirical analyses include Amihud et al. (2003) and O'Hara and Ye (2011). Investors' participation and welfare has recently been study by Huang and Wang (2010).

Amihud and Mendelson (1986) have pioneered the analysis of the effect of liquidity on asset prices. The literature of trading with search frictions was fostered by Duffie et al. (2005). Differently from models in the Duffie et al. (2005) tradition, higher asset price is not necessarily socially desirable. Our trading model is closest to Lagos and Rocheteau (2009). Weill (2007) used a related framework to analyze exchanges. Vayanos and Wang (2007) study concentration of liquidity across assets. Our model contributes to this literature studying trading across strategic venues, thus endogenizing the trading environment. Jarrow and Protter (2011) argue that, unlike standard arbitrageurs, high frequency traders may exacerbate mispricing.

Gabszewicz and Thisse (1979), Shaked and Sutton (1982) and Shaked and Sutton (1983) have developed the theory of vertically differentiated oligopolies. Differently from these papers, we endogenize the value of "quality" (trading delays) through a micro-founded trading game. The result that, through endogenous speed, equilibrium asset prices can increase with the supply of assets is, to the best of our knowledge, new to the literature. The economic intuition is related to Acemoglu (1998) in the context of labor economics and directed technological change.

The rest of the paper is organized as follows. Section 1 presents our benchmark trading model and we derive the value functions of investors. Section 2 analyzes competition among trading venues with and without price protection. Section 3 analyzes trading venues' entry

decisions and investment in speed. Section 4 focuses on equilibrium asset prices. Section 5 characterizes the solutions to the Planner's problem and studies the efficiency of the market equilibrium. Section 6 contains a numerical analysis of the model, and Section 7 concludes.

1 Trading Model

We present our trading model in the case of one market. This section provides explicit micro-foundations for how investors value speed in financial markets. The key result of this section is a characterization of value functions as function of speed and investors characteristics.

1.1 Preferences and Technology

We start by describing the main building blocks of our model: investors' preferences and trading technology. Preferences need to incorporate heterogeneity to create gains from trade as well as interesting participation decisions among exchanges. The trading technology must capture the role of speed in financial markets.

Time is continuous and we fix a probability space. The model has a continuum of heterogeneous investors, two goods and one asset. The measure of investors is normalized to one and their preferences are quasi-linear. The numeraire good (cash) has a constant marginal utility normalized to one, and can be freely invested at the constant rate of return r. The asset is in fixed supply $\bar{a}$, which is also the endowment of each investor. One unit of asset pays a constant dividend equal to 1 of a perishable non-tradable good. The flow utility that an investor derives from holding a_t units of the asset at time t is

$$u_{\sigma,\epsilon_t}(a_t) = (\mu + \sigma\epsilon_t)\, a_t,$$

where (σ, ϵ_t) denotes the type of the investor. This type is defined by a fixed component σ and a time varying (random) component ϵ_t. The fixed component $\sigma \in [0, \bar{\sigma})$ is known at time 0 and distributed according to the cumulative distribution G. The type $\epsilon_t \in [-1, +1]$ changes randomly over time. The times when a change can occur are distributed exponentially with parameter γ.[3] Conditional on a change, ϵ is *iid* with mean zero and probabilities $\phi(+1) = \phi(-1) = \frac{1}{2}$. Finally, we restrict asset holdings to $a_t \in \{0, 1\}$.

The focus of our paper is on the trading technology for the asset. For clarity, we describe here the case where all investors trade at the same speed (later we will endogenize speed choices and consider markets with different speeds). The market where investors trade the asset is characterized by the constant contact rate ρ. Conditional on being in contact, the market is walrasian and clears at the price p_t.[4] Any investor in contact with the market at time t can trade at the price p_t. Investors who are not in contact simply keep their holdings constant.

Our assumptions about technology and preferences imply that the value function of a class-σ investor with current valuation $\epsilon(t)$ and current asset holdings a at time t is

$$V_{\sigma,\epsilon_t}(a, t) = \mathbb{E}_t \left[\int_t^T e^{-r(s-t)} u_{\sigma,\epsilon_s}(a) ds + e^{-r(T-t)} \left(V_{\sigma,\epsilon_T}(a_T, \tau) - p_T(a_T - a) \right) \right] \qquad (1)$$

where the realization of the random type at time $s > t$ is $\epsilon(s)$ and T denotes the next time the investor makes contact with the market. Expectations are defined over the random

---

[3]As explained in the introduction, the ϵ shocks capture time varying liquidity demand, financing costs, hedging demand, or specific investment opportunities. For instance, an corporate investor might need to sell its financial assets to finance a real investment. A household might do the same for the purchase of a durable good or a house. The parameter σ then simply measures the size of these shocks. If we think of delegated management, the effective shock is the sum of the shocks affecting all the investors in the fund. In this example, heterogeneity comes from the fact that different funds cater to different clienteles.

[4]It would be straightforward to add bargaining with market makers and bid-ask spreads, but this would not bring new insights compared to Duffie et al. (2005) and Lagos and Rocheteau (2009). For simplicity we therefore assume competitive trading conditional on being in contact with the market.

variables T and $\epsilon(s)$ and are conditional on the current type $\epsilon(t)$ and the speed of the market ρ.

1.2 Trading Equilibrium

We will show that the asset price remains constant during the trading game. The value functions are thus time-independent, and equation (1) becomes simply

$$rV_{\sigma\epsilon}(a) = u_{\sigma,\epsilon}(a) + \gamma \sum_{\epsilon'} \phi_{\epsilon'} \left[V_{\sigma\epsilon'}(a) - V_{\sigma\epsilon}(a)\right] + \rho \left[V_{\sigma\epsilon}(a_{\sigma,\epsilon}^*) - V_{\sigma\epsilon}(a) - p(a_{\sigma,\epsilon}^* - a)\right] \quad (2)$$

Following Lagos and Rocheteau (2009), we define the adjusted holding utility as

$$\bar{u}(a; \sigma, \epsilon) \equiv \frac{(r + \rho)\, u_{\sigma,\epsilon}(a) + \gamma \mathbb{E}\left[u_{\sigma,\epsilon'}(a) \mid \epsilon\right]}{r + \rho + \gamma}$$

Lagos and Rocheteau (2009) (Lemma 1) show that $\bar{u}$ is the object that investors seek to maximize when deciding how much to trade. Note that since ϵ is i.i.d. with mean 0, we have for any a and any ϵ

$$\mathbb{E}\left[u_{\sigma,\epsilon'}(a) \mid \epsilon\right] = \mu a$$

This expected utility over ϵ' does not depend on σ or ϵ. This implies that $\bar{u}(a; \sigma, \epsilon) = \left(\mu + \sigma\epsilon\frac{r+\rho}{r+\rho+\gamma}\right) a$.

Recall that G was the ex-ante distribution of permanent types. Let $\tilde{G}(\sigma)$ be the number of traders of type less than σ in the market. If all potential investors join the market we simply have $\tilde{G} = G$. In the generic case, however, we have $\tilde{G} \leq G$ since some investors do not participate. Indeed, we shall see that in the multiple venues model, the distribution $\tilde{G}$

11

is typically discontinuous. We therefore present our results without putting any restriction on the function $\tilde{G}$.

Lemma 1. *An equilibrium with constant price p is characterized by the demand functions*

$$a^* (p; \sigma, \epsilon) = \arg\max_a \bar{u} (a; \sigma, \epsilon) - rp. \tag{3}$$

and the market clearing condition

$$\int_\sigma \sum_\epsilon \phi_\epsilon a^* (p; \sigma, \epsilon) \, d\tilde{G} (\sigma) = \bar{a} \tilde{G} (\bar{\sigma}), \tag{4}$$

Proof. See Proposition 1 in Lagos and Rocheteau (2009). The proposition only needs to be adapted to take into account heterogeneity in σ. $\square$

There is a clear symmetry around $\bar{a} = 1/2$ since half the investors are of trading type $\epsilon = +1$ and half are of trading type $\epsilon = -1$. It is therefore sufficient to analyze a market where $\bar{a} \leq 1/2$. In this case, supply is short and low types always sell their entire holdings when they contact the market. Moreover, there is a marginal type $\hat{\sigma}$ who is indifferent between buying and not buying when $\epsilon = 1$. This marginal type is defined by

$$\hat{\sigma} (p, \rho) \equiv \left(1 + \frac{\gamma}{r + \rho}\right) (rp - \mu). \tag{5}$$

The demand functions are therefore $a^* = 0$ when $\epsilon = -1$ or when $\sigma < \hat{\sigma}$; and $a^* = 1$ when $\epsilon = +1$ and $\sigma \geq \hat{\sigma}$.

We can use these demand curves to rewrite the market clearing condition. All negative trading types $\epsilon = -1$ want to hold $a = 0$ and they represent is half of the traders. The

12

trading types $\epsilon = +1$ want to hold one unit if $\sigma > \hat{\sigma}$ and nothing if $\sigma < \hat{\sigma}$. The demand for the asset is $1/2 \times (\tilde{G}(\bar{\sigma}) - \tilde{G}(\hat{\sigma}))$. The ex-ante supply of the asset (per capita) is $\bar{a}$. The market clearing condition is therefore

$$\frac{\tilde{G}(\bar{\sigma}) - \tilde{G}(\hat{\sigma})}{2} = \bar{a}\tilde{G}(\bar{\sigma}). \tag{6}$$

Notice that the asset holdings of types $\sigma < \hat{\sigma}$ are non-stationary since they never purchase the asset. A type $\sigma < \hat{\sigma}$ sells its holding $\bar{a}$ on the first contact with the market and never holds the asset again. Over time, the assets move from the low-σ types to the high-σ types, and then keep circulating among high types in response to ϵ-shocks and trading opportunities. It is easy to see that the price remains constant along the transition path. The gross supply of assets is always $\rho\bar{a}\tilde{G}(\bar{\sigma})$. The gross demand from high types is always $\rho\left(\tilde{G}(\bar{\sigma}) - \tilde{G}(\hat{\sigma})\right)/2$. From (6), the market always clears.

We can now characterize the steady state distribution among types $\sigma > \hat{\sigma}$. Let $\alpha_{\sigma,\epsilon}(a)$ be the share of class-σ investors with trading type ϵ currently holding a units of asset. Consider first a type $(\epsilon = +1, a = 1)$. This type is satisfied with its current holding and does not trade even if it contacts the market. Outflows come only from changes of ϵ from $+1$ to -1, which happens with intensity $\gamma/2$. There are two sources of inflows: types $(\epsilon = -1, a = 1)$ who switch to $\epsilon = 1$, and types $(\epsilon = +1, a = 0)$ who purchase one unit when they contact the market. In steady state, outflows must equal inflows:

$$\frac{\gamma}{2}\alpha_{\sigma,+}(1) = \frac{\gamma}{2}\alpha_{\sigma,-}(1) + \rho\alpha_{\sigma,+}(0). \tag{7}$$

13

Dynamics for types $(\epsilon = -1, a = 0)$ are similar:

$$\frac{\gamma}{2}\alpha_{\sigma,-}(0) = \rho\alpha_{\sigma,-}(1) + \frac{\gamma}{2}\alpha_{\sigma,+}(0) \tag{8}$$

Finally for types $(\epsilon = +1, a = 0)$ and $(\epsilon = -1, a = 1)$ trade creates outflows so we have

$$\left(\frac{\gamma}{2} + \rho\right)\alpha_{\sigma,+}(0) = \frac{\gamma}{2}\alpha_{\sigma,-}(0) \tag{9}$$

$$\left(\frac{\gamma}{2} + \rho\right)\alpha_{\sigma,-}(1) = \frac{\gamma}{2}\alpha_{\sigma,+}(1) \tag{10}$$

Finally, the shares must add up to one, therefore

$$\sum_{\epsilon=\pm,a=0,1} \alpha_{\sigma,\epsilon}(a) = 1 \tag{11}$$

We summarize our results in the following Lemma

Lemma 2. *The trading equilibrium is characterized by the price p and marginal type $\hat{\sigma}$ defined in (5) and (6). The transition dynamics are as follows. The price remains constant while asset holdings shift from low σ-types to high σ-types. Low types $(\sigma < \hat{\sigma})$ sell their initial holdings $\bar{a}$ and never purchase the asset again. High types $\sigma \geq \hat{\sigma}$ buy when $\epsilon = 1$ and sell when $\epsilon = -1$. The distribution of holdings among high σ-types converges to the steady state distribution of well-allocated assets $\alpha_{\sigma,+}(1) = \alpha_{\sigma,-}(0) = \frac{1}{4}\frac{2\rho+\gamma}{\gamma+\rho}$, and mis-allocated assets $\alpha_{\sigma,+}(0) = \alpha_{\sigma,-}(1) = \frac{1}{4}\frac{\gamma}{\gamma+\rho}$. These allocation converge the Walrasian allocation when $\rho \to \infty$.*

Proof. To see the steady state allocations, add (7) and (10) to get $\alpha_{\sigma,-}(1) = \alpha_{\sigma,+}(0)$. This immediately implies $\alpha_{\sigma,-}(0) = \alpha_{\sigma,+}(1)$. Using (7), we obtain $\alpha_{\sigma,+}(1) = \left(1 + 2\frac{\rho}{\gamma}\right)\alpha_{\sigma,-}(1)$. We can then solve for the shares of each type $\alpha_{\sigma,+}(1) = \frac{1}{4}\frac{\gamma+2\rho}{\gamma+\rho}$; and $\alpha_{\sigma,+}(0) = \frac{1}{4}\frac{\gamma}{\gamma+\rho}$.

14

Notice also that the market clearing condition among high types is simply $\alpha_{\sigma,+}(1) + \alpha_{\sigma,-}(1) = 1/2$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

1.3 Value Functions

Our goal is to analyze the provision of speed in financial markets. We therefore need to estimate the value that investors attach to trading in each market. We do it in two steps. We first compute the steady state value functions for investors who keep on trading. We later compute the ex-ante values taking into account the transition dynamics.

Consider the steady state value functions for any type $\sigma > \hat{\sigma}$. They solve the following system. For the types holding the assets, we have

$$
rV_{\sigma,+}(1) \;=\; \mu + \sigma + \frac{\gamma}{2}\left[V_{\sigma,-}(1) - V_{\sigma,+}(1)\right] \tag{12}
$$

$$
rV_{\sigma,-}(1) \;=\; \mu - \sigma + \frac{\gamma}{2}\left[V_{\sigma,+}(1) - V_{\sigma,-}(1)\right] + \rho\left(p + V_{\sigma,-}(0) - V_{\sigma,-}(1)\right) \tag{13}
$$

For the types not holding the assets, we have

$$
rV_{\sigma,-}(0) = \frac{\gamma}{2}\left[V_{\sigma,+}(0) - V_{\sigma,-}(0)\right] \tag{14}
$$

$$
rV_{\sigma,+}(0) = \frac{\gamma}{2}\left[V_{\sigma,-}(0) - V_{\sigma,+}(0)\right] + \rho\left(V_{\sigma,+}(1) - V_{\sigma,+}(0) - p\right) \tag{15}
$$

Define $H \equiv V_{\sigma,+,1} - V_{\sigma,+,0}$ and $L \equiv V_{\sigma,-,1} - V_{\sigma,-,0}$. Then, taking differences of the above equations we get

$$
rL \;=\; \mu - \sigma + \frac{\gamma}{2}(H - L) + \rho(p - L)
$$

$$
rH \;=\; \mu + \sigma - \frac{\gamma}{2}(H - L) - \rho(H - p)
$$

15

Note that the asset price p is pinned down by the marginal (minimum type in each market). For now we keep it as a (market specific) parameter. We can then solve $r(H - L) = 2\sigma - (\gamma + \rho)(H - L)$ and obtain the gains from trade for type σ in market ρ:

$$H - L = \frac{2\sigma}{r + \gamma + \rho}.$$

Note that these gains from trade do not depend on the equilibrium price. Hence they do not depend on the allocation of types to the market. They only depend on the market speed ρ and on the individual type σ. Using the gains from trade $H - L$, we can reconstruct the functions L and H and finally for the initial value functions. The no-trade outside option of any investor is

$$W_{out} = \frac{\mu\bar{a}}{r}. \tag{16}$$

The following proposition characterizes the ex-ante value functions, taking into account the transition dynamics leading up to the steady state allocations.

Proposition 1. *The ex-ante value W for type σ of participating in a market with speed ρ and price p is the sum of the value of ownership and the value of trading:*

$$W(\sigma, \hat{\sigma}, s) - W_{out} = \frac{s\bar{a}\hat{\sigma}}{r} + \frac{s}{2r}\max(0; \sigma - \hat{\sigma}), \tag{17}$$

where effective speed s defined by

$$s(\rho) \equiv \frac{\rho}{r + \gamma + \rho}, \tag{18}$$

and the marginal type $\hat{\sigma}(p, \rho)$, defined in (5), is increasing in p and decreasing in ρ.

16

Proof. See Appendix. □

The intuition is that W is made of two parts. The value of ownership is $\frac{\mu\bar{a}+s\bar{a}\hat{\sigma}}{r}$. It is independent of σ. It is the value that can be achieved by all types $\sigma < \hat{\sigma}$ with the "sell-and-leave" strategy. The second part $\frac{s}{2r}\max(0;\sigma-\hat{\sigma})$ is the value of trading repeatedly, and it depends on the type σ. This part of the value function is super-modular in (s,σ).

2   Competition and Affiliation

In this section we analyze competition among a given set of trading venues, and the result-ing allocation of investors across these venues. We characterize the pricing decisions and equilibrium profits of trading venues and the affiliations choices of investors. Importantly, we analyze how price protection in the trading game affects these equilibrium outcomes. In other words, we analyze how trading regulations affect the ex-ante competition among exchanges. In this section, we take the set of venues as given. In the next section we will endogenize entry and speed.

In all cases, we start with mass one of investors, and aggregate supply $\bar{a}$. G is the ex-ante distribution of types. The participation decision of type σ is described by

$$\mathcal{P} : [0,\bar{\sigma}] \longrightarrow \{0,1,2\},$$

where $\mathcal{P}(\sigma) = 0$ means staying out, 1 means joining market 1, and 2 joining market 2. Trading venues compete a la Bertrand. If an investor joins venue i, it pays a membership fee q_i and is then allowed to use the trading venue (staying out costs nothing, so formally $q_0 = 0$ and $W = W_{out}$). Recall that we denote by $\tilde{G}_i(.)$ the distribution of types who join market

17

i, so $\tilde{G}_i(\bar{\sigma})$ is the total number of investors who join market i. This is the key equilibrium object. Let us know formally define an equilibrium of the affiliation game.

Definition 1. An equilibrium of the affiliation game is a set of participation decisions $\mathcal{P}$ by investors, and pricing decisions q by trading venues such that:

- Participation decisions are optimal: for all σ and all i, $\mathcal{P}(\sigma) = i$ implies $W(\sigma, \hat{\sigma}_i, s) - q_i \geq W(\sigma, \hat{\sigma}_j, s) - q_j$ for all $j \neq i$; reciprocally when $W(\sigma, \hat{\sigma}_i, s) - q_i > W(\sigma, \hat{\sigma}_j, s) - q_j$ for all $j \neq i$ then we must have $\mathcal{P}(\sigma) = i$.

- Venues maximize profits: $q_i = \arg\max q_i \tilde{G}_i(\bar{\sigma})$;

- The investor market clears: $\sum_{i=0,1,2} \tilde{G}_i(\sigma) = G(\sigma)$ for all $\sigma \in [0, \bar{\sigma}]$;

- Subsequent asset prices and marginal types satisfy (5) and (6).

In the remaining of this section, we consider several versions of the affiliation game: with one or two venues, and with or without trading regulations.

2.1 One Speed

With one speed, the marginal trading type must be indifferent between joining the market and not joining the market. So we must have $W(\hat{\sigma}, \hat{\sigma}, s) - W_{out} = q$ and therefore

$$q = \frac{s\bar{a}\hat{\sigma}}{r}. \tag{19}$$

All types below $\hat{\sigma}$ are indifferent between joining and staying out. Let δ be the mass of investors that join, sell and leave.[5] Market clearing requires

$$\delta = \left(\frac{1}{2\bar{a}} - 1\right)(1 - G(\hat{\sigma}))$$

This condition holds at an interior solution as long as $\delta < G(\hat{\sigma})$, or in other words as long as

$$\frac{G(\hat{\sigma})}{1 - G(\hat{\sigma})} > \frac{1}{2\bar{a}} - 1$$

In the remaining of the paper, we assume that either $\bar{a}$ is close enough to $1/2$ or that there is a sufficient mass of low type investors to ensure the existence of interior solutions.

Total profits for the exchange are $\pi = q(1 - G(\hat{\sigma}) + \delta)$ which we can write using market clearing as

$$\pi = q\frac{1 - G(\hat{\sigma})}{2\bar{a}}.$$

Notice that if $\bar{a} = 1/2$ we get $\delta = 0$, the simplest case to analyze. When $\bar{a}$ is less than $1/2$, we simply need to remember that δ investors sell and become inactive. The equilibrium is depicted in the top panel of Figure 4.

---

[5]There can also be a corner solution with full participation, characterized by the market clearing condition $G(\sigma_{\min}^{\mathsf{T}}) = 1 - 2\bar{a}$. All investor pay the participation fee $q_{\min}$, which is also the total profit of the trading venue. Then $G(\sigma_{\min}^{\mathsf{T}})$ sell and drop out, while the remaining $1 - G(\sigma_{\min}^{\mathsf{T}})$ trade in the market with a supply per capita of $1/2$. The participation condition is simply $\hat{V} - q \geq \mu\frac{\bar{a}}{r}$. There is full participation as long as $q \leq q_{\min} = \frac{s}{r}\bar{a}\sigma_{\min}^{\mathsf{T}}$.

Consolidated Market (monopoly)

A consolidated market center with exogenous speed s behaves like a classic monopolist. We index this market structure by m. Using (19), the program of the monopoly is

$$\max_q q \frac{1 - G(\hat{\sigma})}{2\bar{a}}$$

The FOC for profit maximization is[6]

$$1 - G(\hat{\sigma}_m) = g(\hat{\sigma}_m) \hat{\sigma}_m \tag{20}$$

This is a standard result. The monopoly restricts participation to maximize its profits. Note that the choice of $\hat{\sigma}_m$ is independent of the speed in the market. The fee q_m increases one to one with s.

Fragmented Markets (Bertrand duopoly)

In the fragmented case, exchanges compete in fees a la Bertrand. In equilibrium, fees and profits are both zero. All investors participate and the distribution of investors across trading venues is immaterial. The solution is

$$q_{Bertrand} = 0.$$

In the presence of fixed costs, this would not be an equilibrium. Without differentiation by speed, there is a natural monopoly.

---

[6]We can check concavity of $\pi = \sigma(1 - G(\sigma))$: $\pi' = 1 - G - g\sigma$ and $\pi" = -2g - \sigma g' < 0 \Leftrightarrow 2g + \sigma g' > 0$. Many distribution functions, e.g. uniform, satisfy this constraint over the relevant range.

Figure 4: Investors' Market Affiliation Choice

Affiliation Choices with One Market

Affiliation Choices with Free Segmentation

Affiliation Choices with Protected Prices

21
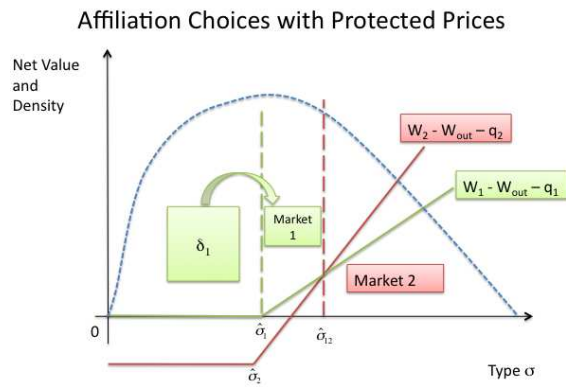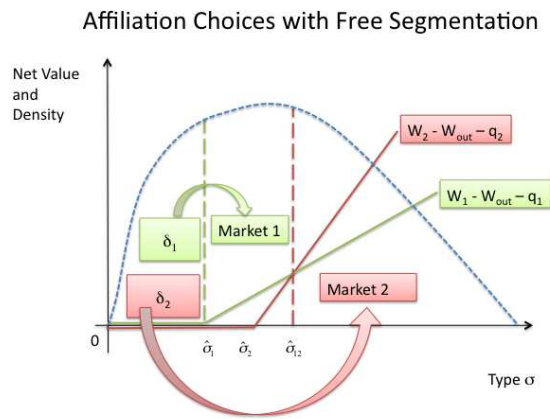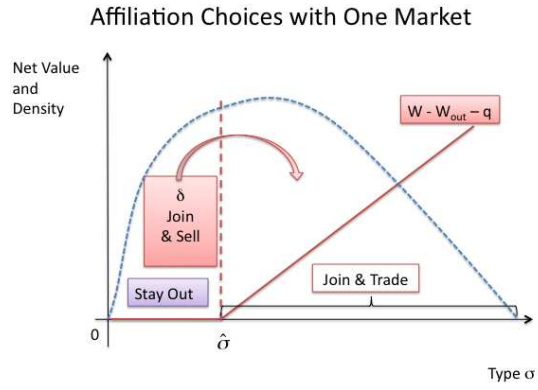
2.2 Segmented Venues

Formally, suppose there are two venues, 1 and 2, with speeds ρ_1 and ρ_2 and participation fees q_1 and q_2. We define venue 2 as the fast market, so $\rho_2 > \rho_1$. A critical issue is the segmentation of trades and the possibility of having different prices. We consider two types of regulations.

Definition 2. We say that there is *segmentation* if a venue refuses to execute trades coming from investors of another venue. Otherwise, we say that there is *price protection*.[7]

Under free segmentation, an investor joins a market and cannot trade with an investor in the other market. The trading venues are effectively segmented and equilibrium asset prices can be different. Under price protection, assets prices must be the same in both venues.

Consider first the case where there is free segmentation. Prices can then be different in the two venues because exchange 2 can refuse to execute the trade of an investor from exchange 1. The key issue is to understand the affiliation choices of investors. We proceed by backward induction. Investors anticipate that each market will be characterized by its speed and its price, which together define the marginal trading type $\hat{\sigma}$. Investors can then estimate their value functions W defined in (17). The net value from joining market $i = 1, 2$ is $W(\sigma, \hat{\sigma}_i, s_i) - W_{out} - q_i$. These value functions are depicted in the middle panel of Figure 4.

Let $\hat{\sigma}_1$ be the marginal type who is indifferent between joining market 1 and staying out. It must satisfy equation (19), therefore we have

$$q_1 = \frac{\bar{a} s_1 \hat{\sigma}_1}{r} \tag{21}$$

---

[7]This is our simple way to capture access and trade-through rules in SEC's Reg. NMS

It is useful to keep in mind that the value functions are not super modular for low types. In addition, we know that each market must attract a mass δ of types who join and sell their assets. Because these types must be indifferent between joining and staying out, we must have $W(\hat{\sigma}_i, \hat{\sigma}_i, s_i) - W_{out} - q_i = 0$ in both markets. Otherwise, all the low types would strictly prefer one market to another. The above condition guarantees this for market 1. For market 2, we must also have

$$q_2 = \frac{\bar{a} s_2 \hat{\sigma}_2}{r} \qquad (22)$$

Notice an important point here: $\hat{\sigma}_2$ is defined as the marginal trader in market 2, i.e., the type who would be indifferent between trading repeatedly and dropping out after selling. It is clear from Figure 4 (and we prove it below) that $\hat{\sigma}_2$ does not in fact join market 2. Rather, $\hat{\sigma}_2$ joins market 1.

With two markets, we must define a new marginal type, $\hat{\sigma}_{12}$, who is indifferent between joining market 1 and market 2. By definition, this type must be such that $W(\hat{\sigma}_{12}, \hat{\sigma}_2, s_2) - q_2 = W(\hat{\sigma}_{12}, \hat{\sigma}_1, s_1) - q_1$. This implies $\frac{s_1 \bar{a} \hat{\sigma}_1}{r} + \frac{s_1}{2r}(\hat{\sigma}_{12} - \hat{\sigma}_1) - q_1 = \frac{s_2 \bar{a} \hat{\sigma}_2}{r} + \frac{s_2}{2r}(\hat{\sigma}_{12} - \hat{\sigma}_2) - q_2$, and therefore using (21) and (22):

$$\hat{\sigma}_{12} = \frac{r}{\bar{a}} \frac{q_2 - q_1}{s_2 - s_1} \qquad (23)$$

Note that $\hat{\sigma}_1 < \hat{\sigma}_2 < \hat{\sigma}_{12}$. The set of types who join market 2 cannot be an interval. It is made of all the types above $\hat{\sigma}_{12}$ and some types below $\hat{\sigma}_1$. The affiliation is depicted in the middle panel of Figure 4.

Market clearing in market 2 requires $(1 - G(\hat{\sigma}_{12}) + \delta_2)\bar{a} = \frac{1 - G(\hat{\sigma}_{12})}{2}$. Total profits for the fast exchange under free segmentation are $\pi_2^{seg} = q_2(1 - G(\hat{\sigma}_{12}) + \delta_2) = q_2 \frac{1 - G(\hat{\sigma}_{12})}{2\bar{a}}$. Market clearing for the slow exchange requires $(G(\hat{\sigma}_{12}) - G(\hat{\sigma}_1) + \delta_1)\bar{a} = \frac{G(\hat{\sigma}_{12}) - G(\hat{\sigma}_1)}{2}$. Total profits

23

for the slow exchange are $\pi_1^{seg} = q_1 \frac{G(\hat{\sigma}_{12}) - G(\hat{\sigma}_1)}{2\bar{a}}$. The affiliation of investors to markets 1 and 2 are given by the marginal types 19 and 23. Exchanges 1 and 2 solve simultaneously

$$\max_{q_1} \pi_1^{seg} = \frac{q_1}{2\bar{a}} \left(G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) \right) \tag{24}$$

$$\max_{q_2} \pi_2^{seg} = \frac{q_2}{2\bar{a}} \left(1 - G\left(\hat{\sigma}_{12}\right) \right)$$

Taking first order conditions from the previous system, we obtain the following Lemma:

Lemma 3. *Under free segmentation the allocation* $(\hat{\sigma}_1^{seg}, \hat{\sigma}_{12}^{seg})$ *among trading venues solves the following system:*

$$1 - G\left(\hat{\sigma}_{12}\right) = g\left(\hat{\sigma}_{12}\right) \left(\hat{\sigma}_{12} + \hat{\sigma}_1 \frac{s_1}{s_2 - s_1} \right), \tag{25}$$

$$G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) = \left(g\left(\hat{\sigma}_1\right) + \frac{s_1}{s_2 - s_1} g\left(\hat{\sigma}_{12}\right) \right) \hat{\sigma}_1. \tag{26}$$

2.3   Protected Prices

Now consider the case where is there is a single price, but two venues with different speeds. The asset price is p in both markets. Market 1 is still characterized by the indifference condition (21) for the marginal type $\hat{\sigma}_1$. However this condition does not hold for market 2, because low types can join market 1 and then sell their assets to investors in market 2. Instead, we have the condition that the asset price is the same in both markets. From (5) this implies the constraint

$$\left(1 + \frac{\gamma}{r + \rho_1} \right) \hat{\sigma}_2 = \left(1 + \frac{\gamma}{r + \rho_2} \right) \hat{\sigma}_1 \tag{27}$$

24

This means that $\hat{\sigma}_2 < \hat{\sigma}_1$. The indifference condition for $\hat{\sigma}_{12}$ is still $W(\hat{\sigma}_{12}, \hat{\sigma}_2, s_2) - q_2 = W(\hat{\sigma}_{12}, \hat{\sigma}_1, s_1) - q_1$. We show in the Appendix that this leads to

$$\hat{\sigma}_{12} = \frac{2r}{s_2 - s_1}\left(q_2 - \frac{z}{2\bar{a}}q_1\right), \tag{28}$$

where

$$z \equiv 1 - \frac{1 + \frac{r}{\rho_1}}{1 + \frac{r}{\rho_2}}(1 - 2\bar{a}).$$

The structure of the value functions is still as depicted in the bottom panel of Figure 4. There is now only one market clearing condition. As a result, the sell and drop traders join market 1 where they can sell at a higher price because they can sell to investors in market 2. We then have $\delta_2 = 0$ and the market clearing condition is

$$(1 - G(\hat{\sigma}_1) + \delta_1)\bar{a} = \frac{1 - G(\hat{\sigma}_1)}{2}$$

The following Lemma summarizes the protected price equilibrium

Lemma 4. *Under price protection the allocation* $\left(\hat{\sigma}_1^{prot}, \hat{\sigma}_{12}^{prot}\right)$ *among trading venues solves the following system:*

$$1 - G(\hat{\sigma}_{12}) = g(\hat{\sigma}_{12})\left(\hat{\sigma}_{12} + z\frac{s_1}{s_2 - s_1}\hat{\sigma}_1\right)$$

$$G(\hat{\sigma}_{12}) - \frac{G(\hat{\sigma}_1)}{2\bar{a}} = \left(\frac{g(\hat{\sigma}_1)}{2\bar{a}} + z\frac{s_1}{s_2 - s_1}g(\hat{\sigma}_{12})\right)\hat{\sigma}_1 + 1 - \frac{1}{2\bar{a}}$$

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Price protection has two consequences. It increases the profits of the slower exchange, and it decreases price competition and participation for given speeds and given exchanges.

We can now compare the outcome of the various market structures. To derive analytical results, we assume that the ex-ante distribution of types G is exponential

Assumption A1: $G(\sigma) = 1 - e^{-\frac{\sigma}{\nu}}$ We can now state the following proposition:

Proposition 2. *Competition among exchanges increases participation. With or without price protection, participation in the fast venue is higher than total participation with a monopoly, i.e. $\hat{\sigma}_{12} < \hat{\sigma}_m$. Total participation is even higher since $\hat{\sigma}_1 < \hat{\sigma}_{12}$. Under A1 price protection increases the profits of the slow venue and decreases total active participation, i.e. $\pi_1^{prot} \geq \pi_1^{seg}$ and $\hat{\sigma}_1^{prot} \geq \hat{\sigma}_1^{seg}$. Price protection does not affect the fee $q_2 = \frac{\nu}{2r}(s_2 - s_1)$, and it has an ambiguous impact on participation in the fast venue.*

Proof. See Appendix. ☐

The intuition for the first half of the proposition is simply that price competition increases participation. A result that is perhaps less obvious is that participation in just the fast venue is already higher than total participation with a monopoly. The intuition for the second half of the proposition is as follows. Price protection is a subsidy to the slow market because its investors are allowed to sell their assets to investors in the fast market. This creates a larger demand for the slow market. When considering its profits $q_1(1 - G(\hat{\sigma}_1) + \delta_1)$, the presence of this demand makes it more attractive for the slow market to increase its price. This is why $\hat{\sigma}_1^{prot} \geq \hat{\sigma}_1^{seg}$. Protection also soften the price elasticity of the marginal type $\hat{\sigma}_{12}$, which again is good for the slow venue. Thus profits of the slow venue increase under protection for two reason: more demand, and less price elasticity.[8]

The impact on participation in the high speed market is small in practice, and positive

---

[8]We check numerically the robustness of the result $\pi_1^{prot} \geq \pi_1^{seg}$ to alternative assumptions about the underlying distribution of σ in Section 6.

Table 3.1: Entry Game

Markets $1 \downarrow$ and $2 \rightarrow$	In	Out
In	$(\pi_1^{\intercal} - \kappa, \pi_2^{\intercal} - \kappa)$	$(\pi_1^m - \kappa, 0)$
Out	$(0, \pi_2^m - \kappa)$	$(0, 0)$

for the parameter values that we consider, as discussed in the appendix: we typically find $\hat{\sigma}_{12}^{prot} \leq \hat{\sigma}_{12}^{seg}$.

Proposition 2 plays an important role in our paper. The results regarding profits are important to understand the impact of price protection on entry and therefore on the equilibrium market structure. The results regarding participation are important to understand the welfare implications of various regulations. We explore these issue in the next section.

3   Endogenous Speed and Entry

In this section we complete the description of the equilibrium market structure by analyzing the entry decisions of trading venues, as well as their optimal investment in speed.

3.1   Price Protection and Entry

We analyze in this section the relationship between trading regulation and entry for given speeds. There are two potential entrants, with speeds s_1 and s_2, with the convention that $s_1 < s_2$. The entry cost κ is the same for both exchanges. Market i's net profit is then given by $\pi_i^{\intercal} - \kappa$ where $\intercal \in \{seg; prot\}$ denotes trading regulations. For a given speed, asset supply $\overline{a} \leq 1/2$ and regulatory framework, the profit functions π are as in Section 2. A given venue i finds it optimal to enter whenever net profits are non negative.

We model entry as a simultaneous game. The payoffs of the entry game are in table

3.1. From our previous analysis, we know the following: (i) for a given trading regulation $\intercal$, $\pi_1^\intercal < \pi_2^\intercal$ simply because 2 is faster and (ii) $\pi_1^{seg} < \pi_1^{prot}$ from Proposition 2. Consequently,

Proposition 3. *Price protection at the trading stage helps sustain entry at the initial stage.*

As shown in Figure 5, price protection expands the ex-ante number of markets for economies with intermediate entry costs (between π_1^{seg} and π_1^{prot}). The expected level of fragmentation hence depends on price regulation.

Figure 5: Entry Cost, Regulation and Equilibrium Fragmentation
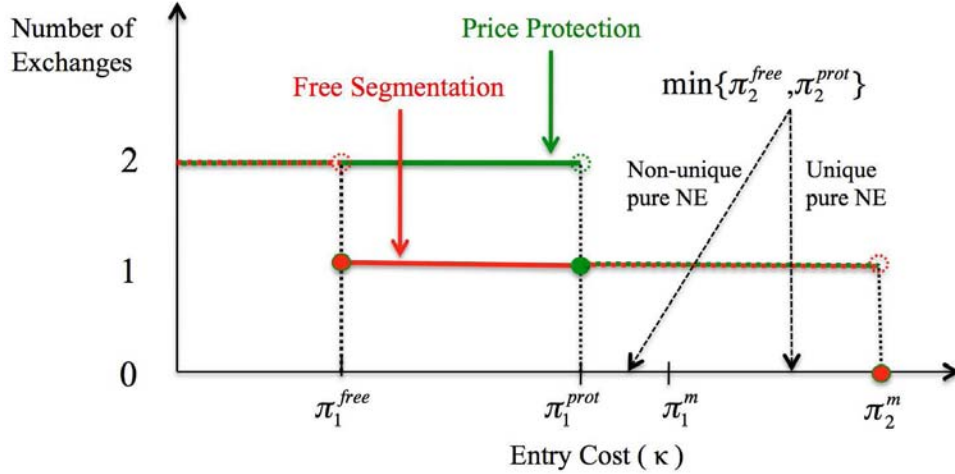
The graph shows the equilibrium number of exchanges, as a function of entry costs κ. Price protection affects the equilibrium number of exchanges that enter the market when entry costs are between the expected profits of the slow venue under segmentation, π_1^{seg}, and under price protection, π_1^{prot}. When there are two Nash equilibriums, the outcomes are that either the fast or slow venue decides to enter, and the other venue stays out.

Depending on parameter values, the entry game may have more than one Nash equilibrium in pure strategies. To simplify our presentation, we assume hereafter that our economies satisfy the inequality $\pi_1^m < \min\left\{\pi_2^{seg}, \pi_2^{prot}\right\}$. Thus, only the fast exchange enters whenever $\kappa > \pi_1^{prot}$. We characterize the cases with multiple equilibria in the proof of Proposition 3 in the Appendix.

28

3.2 Speed Choices

In this section we analyze speed choices taking the number of active markets as given. For simplicity, we concentrate here on the case where $\bar{a} = 1/2$. In this limiting case, trading regulation does not affect markets' profit functions and thus trading regulations become immaterial. When convenient, we assume the following cost to derive analytical results.

Assumption A2. *The cost of achieving contact rate ρ is given by $c \max\{\rho - \underline{\rho}; 0\}$, where $c > 0$ is the constant marginal cost of speed beyond the default level $\underline{\rho}$.*

Under A2, the total cost of entering and reaching the effective speed s is

$$C(s) = c \max\left\{(r + \gamma)\frac{s}{1-s} - \underline{\rho}; 0\right\} \tag{29}$$

These costs are convex in effective speed s. We analyze first the case of a monopolist.

Consolidated Market

Given the monopolist speed, denoted s_M, the marginal type is such that

$$q_M = \frac{s_M}{2r}\hat{\sigma}_M \tag{30}$$

The program of the monopolist then is

$$\max_{q,s} q\left(1 - G(\hat{\sigma})\right) - C(s)$$

We can now characterize the consolidated market equilibrium.

Proposition 4. *Monopoly*. *The equilibrium with consolidated markets and endogenous speed has the following properties: (i) Participation is the same as with exogenous speed: $\hat{\sigma}_M = \hat{\sigma}_m$; (ii) Effective speed is given by*

$$2rC'(s_M) = (1 - G(\hat{\sigma}_M))\,\hat{\sigma}_M; \tag{31}$$

and (iii) Under A1-A2 optimal effective speed is given by

$$s_M = 1 - (2rc(\gamma + r)e)^{1/2}\,\nu^{-1/2} \tag{32}$$

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

The monopolist determines market participation based on the distribution of investors' types only. Note than in any interior solution, optimal speed does not depend on the default speed level. Naturally, investments in speed increase with investors heterogeneity ν. When the distribution of permanent types G has fatter right tails, the average investor gains from trade increase. Interestingly, the contact rate ρ_M is concave in the frequency of preference shocks γ: it first increases with (γ) and then decreases, and has a global maximum at $\gamma = \frac{\nu}{8cer} - r$. On the one hand, when the frequency of preference shocks increase, investors want to reallocate their assets more frequently, which increases demand for speed. The marginal value of each trade decreases though since the desired holding period shrinks. Since speed is costly, there is a maximum speed that can be supported in any market equilibrium.

Fragmented Market

When trading is fragmented, exchanges have an incentive to differentiate their intermediation services by offering different speeds since Bertrand competition with fixed speed drives profits down to zero. We simplify the analysis of this case by assuming that market 1's speed is exogenously given $(s_1 = \frac{\rho}{r+\gamma+\underline{\rho}})$ while market 2 chooses an effective speed s_2 which cost is $C(s_2)$. After market 2's speed is chosen, there is simultaneous affiliation fee competition as in section 2.

In the speed choice stage, market 2 solves

$$\max_{s_2} (1 - G(\hat{\sigma}_{12})) q_2 - C(s_2)$$

The following proposition characterizes the equilibrium.

Proposition 5. *Duopoly.* *The equilibrium with fragmented markets and endogenous speed has the following properties:(i) Participation is determined by the marginal types $\hat{\sigma}_1$ and $\hat{\sigma}_{12}$ as in Lemma 4; (ii) Participation in the fast venue alone is higher than participation in the monopolist case; (iii) Speed in market 2 is determined by*

$$2rC'(s_2) = (1 - G(\hat{\sigma}_{12})) \left\{ \hat{\sigma}_{12} + s_1 \frac{\partial \hat{\sigma}_1}{\partial s_2} \right\} ; \tag{33}$$

and (iv) Under A1, the duopoly chooses a higher speed than the monopoly.

Proof. See Appendix. □

The incentives of exchanges to differentiate their services thus increases trading speed. The intuition is as follows. There are two forces at play: scale and differentiation. On the one hand, a monopolist earns higher profits and mechanically wants to invest more in speed.

31

In the limit of Bertrand competition, profits are zero irrespective of speed and there is no incentive to invest in speed. On the other hand, the incentive to differentiate pushes towards higher speed in a duopoly. We study the welfare consequences in Section 5.

4   Asset Prices

Regulations, market structure, speed and affiliation choices affect asset prices through the endogenous determination of the marginal participating types and the liquidity frictions.

4.1   Volume Weighted Prices

From equation (5), we know that the equilibrium asset price is given by

$$p_i = \frac{\mu}{r} + \frac{\hat{\sigma}_i}{r} \left(\frac{r + \gamma s_i}{r + \gamma} \right) \tag{34}$$

Under price protection, $\hat{\sigma}^{prot}$ is given by Lemma 4. Under free segmentation, there are two prices. The asset price in venue i is as in (34), where $(\hat{\sigma}_1^{seg}, \hat{\sigma}_2^{seg})$ are given by (21) and (22). Let us define the instantaneous transaction rates τ_i by

$$\tau_1 = \frac{\gamma}{4} \frac{(\gamma + r) s_1}{(\gamma + r s_1)} \left(G(\hat{\sigma}_{12}) - G(\hat{\sigma}_1) \right)$$
$$\tau_2 = \frac{\gamma}{4} \frac{(\gamma + r) s_2}{(\gamma + r s_2)} \left(1 - G(\hat{\sigma}_{12}) \right)$$

where $\hat{\sigma}_{12}$ is given by 23. Then we have the following Lemma

Lemma 5. *Under free segmentation, The volume-weighted average price $\overline{p}^{seg}$ (VWAP) is*

given by

$$\overline{p}^{free} = \left(\frac{\tau_1}{\tau_1 + \tau_2}\right) p_1^{free} + \left(\frac{\tau_2}{\tau_1 + \tau_2}\right) p_2^{free} \tag{35}$$

In what follows, when we talk about 'the' price in the free duopoly case, we mean the VWAP.

4.2   Asset Supply and Asset Price

We now study the link between the supply of the asset $\bar{a}$ and its price. Under A1-2, we can use Proposition 4 to find the equilibrium price under monopoly

$$p_M = \frac{\mu}{r} + \frac{\nu}{r} \frac{1}{r+\gamma} \left(r + \gamma \left(1 - \left(\frac{r}{\bar{a}} ce\left(\gamma + r\right)\right)^{1/2} \nu^{-1/2} \right) \right) \tag{36}$$

In a frictionless Walrasian market, i.e., $s = 1$ and $\kappa = 0$ (see Section 6 for details.), the equilibrium price under A1 p_w is given by

$$p_w = \frac{1}{r} \left[\mu - \nu \log\left(2\bar{a}\right)\right] \tag{37}$$

The closed-form expression (36) yields interesting insights. With a single venue, the equilibrium asset price decreases with the frequency of the temporary shocks (γ) and with the marginal cost of technology c, and increases with the degree of investor heterogeneity (ν). There is also an interesting relationship between the asset price and the asset supply (see Figure 6), which we summarize in the following proposition:

Proposition 6. *Asset Supply.* *The Walrasian price decreases with the asset supply. For given speed, the monopoly and free duopoly prices are independent of asset supply. With endogenous speed, the prices increases with the asset supply.*

Figure 6: Equilibrium Asset Price and Asset Supply
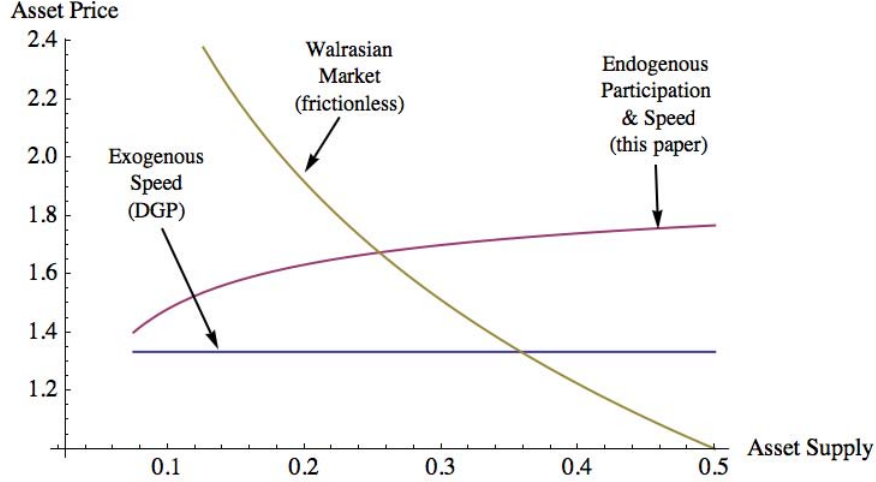
The graph shows the equilibrium asset price, as a function of asset supply $\overline{a}$. The asset price is scaled by (μ/r). "Walrasian market" corresponds to the equilibrium price in a frictionless market where entry and speed costs are zero. "Exogenous (Endogenous) Speed" corresponds to the resulting price where there is a single venue, and investor participation and trading speeds are exogenous (endogenous). The graph is created under assumptions A1-2 and the following parameters: $\gamma = 10$, $r = 5\%/252$, $\nu = \exp(1)$, $\underline{\rho} = 5$, $c = 5$, $\mu = 1$.

Proposition 6 is an immediate consequence of our previous results. In a Walrasian market, the equilibrium price depends entirely on the marginal trading type which decreases in $\overline{a}$.

F

The marginal type under monopoly is given by (20) and is therefore independent of $\bar{a}$. For the segmented duopoly we know from Lemma 3 that the marginal types do not depend on $\bar{a}$, and therefore neither does the price at a given speed. This explain the flat horizontal line in Figure 6. When speed is endogenous, however, we know from Proposition 4 that the optimal speed increases with total asset supply. This will then lead to an increasing price.

4.3 Market Structure and Asset Price

We now compare asset prices across market structures. The link is not obvious because there are two effects. On the one hand, the monopoly restricts entry and chooses a high marginal type, thereby increasing the equilibrium price. On the other hand, speed increases prices, and competition increases speed. In equilibrium, however, the first effect dominates, and we obtain the following proposition.

Proposition 7. *Market Structure.* *Under A1, the asset price – with or without endogenous speed – is higher under monopoly than under competition, .*

Proof. See Appendix. □

The message of Proposition 7 is that, independently of price protection, competition between venues decreases the equilibrium asset price level.[9] This is consistent with Amihud et al. (2003) who provide evidence that trading consolidation increases asset prices.

5 Welfare and Efficient Solution

5.1 Welfare Functions

We study the welfare gains of a given market equilibrium with respect to the no trade benchmark:

$$\mathcal{W} \equiv \underbrace{\sum_i \int_\sigma (W(\sigma, \hat{\sigma}_i, s_i) - W_{out}) dG(\sigma)}_{\text{Partic. gains \& Allocation efficiency}} - \underbrace{\sum_i (\kappa + C(s_i))}_{\text{Entry+Speed Investment}}$$

---

[9]Note that for the comparison of the monopoly with price protection A1 is not required.

Table 5.1: Cases of analysis		
	Consolidated Market	Free Competition
No Speed Choice	$\mathcal{W}_m$	$\mathcal{W}_{Bertrand}$
Endogenous Speed	$\mathcal{W}_M$	$\mathcal{W}_{comp}$

The following Lemma characterizes the welfare functions.

Lemma 6. *Social welfare in a single market equals*

$$\mathcal{W} = \frac{s}{2r} \int_{\hat{\sigma}}^{\bar{\sigma}} \sigma dG(\sigma) - C(s) - \kappa$$

With two trading venues, social welfare is

$$\mathcal{W} = \frac{s_1}{2r} \int_{\hat{\sigma}_1}^{\hat{\sigma}_{12}} \sigma dG(\sigma) + \frac{s_2}{2r} \int_{\hat{\sigma}_{12}}^{\bar{\sigma}} \sigma dG(\sigma) - \sum_{i=1,2} C(s_i) - 2\kappa. \tag{38}$$

Proof. See Appendix. □

To simplify the exposition in this section we consider only the case $\bar{a} = 1/2$ where price regulation is immaterial (we denote social welfare in this case $\mathcal{W}_{comp}$ but it is the same as $\mathcal{W}_\top$, $\top \in \{seg, prot\}$). We analyze the welfare consequences of price protection in Section 6. In the remaining of this section we want to compare in this section the social gains of different market organizations. We assume that every single venue equilibrium of the entry game involves speed investment[10]. Table 5.1 summarizes the relevant cases.

---

[10]In the proof of Proposition 3 we characterize the cases in which the outcome of the entry game has a fixed-speed monopolist

One Speed

As a benchmark, we discuss in the context of our paper the social gains of market organization when investments that improve trading speeds are not available. This is the case considered in the existing literature. Welfare in the monopoly case is given by

$$\mathcal{W}_m = \frac{s}{2r} \int_{\hat{\sigma}_m}^{\overline{\sigma}} \sigma dG\left(\sigma\right)$$

In the fragmented case, exchanges compete in fees a la Bertrand. In equilibrium, fees and profits are both zero. All investors participate and the distribution of investors across trading venues is immaterial. Social welfare in this case is

$$\mathcal{W}_{Bertrand} = \frac{s}{2r} E\left(\sigma\right)$$

For any given effective speed s, welfare is higher than under monopoly. This is the classic case for inter market competition when liquidity externalities are moderate (Economides (1996)).

5.2   Welfare, Speed, and Competition

Does speed increase welfare?

We discuss here the welfare consequences of the advent of innovations in technologies that permits faster trading.

Proposition 8. *When trading is consolidated, social welfare is always higher with endogenous speed. With fragmented markets, under A1, there exists a unique default speed $\underline{s}_0$ such that welfare increases with endogenous speed if and only if $\underline{s} < \underline{s}_0$.*

Proof. See Appendix. □

The intuition for Proposition 8 is as follows. First with a monopoly, we know from Proposition 4 that speed does not affect market participation. However, the monopolist has an incentive to invest in speed to extract higher fees from investors with types $\sigma > \hat{\sigma}_M$. Because the monopolist bears the investment costs entirely, there are not negative externalities. Consequently, taxing technology investments is never optimal in this environment.

Under the duopoly, speed allows venues to differentiate and relax Bertrand competition. Whether social welfare increases with technology investments depends thus on the trade-off between investor participation levels and trading efficiency. When the default effective speed is low, the gains from trading efficiency are large and dominate the negative impact on participation. The opposite happens when the default effective speed is high. In this case, taxing technology investments can increase welfare.

Does competition increase welfare?

In this section we endogenize entry and ask whether market competition increases welfare. Section 3.2 shows that competition affects investor participation and speed investments. In turn, the outcome of the entry game in section 3.1 determines the number of active venues. The net social gains of competition are given by $\mathcal{W}_{comp} - \mathcal{W}_M - \kappa$. Let $\overline{\kappa}$ be the entry cost value that makes these gains equal to zero (see Figure 7). The social benefits of higher speeds and higher participation may in principle be offset by inefficient cost duplication. We can establish the following.

Proposition 9. *Under A1, consolidation increases welfare only when entry costs satisfy $\overline{\kappa} < \kappa \leq \pi_1^{comp}$. Otherwise, fragmentation always (weakly) increases welfare. When the marginal cost of technology approaches zero, $\overline{\kappa}$ approaches $\frac{\nu}{2r}(1 - 1/e)$).*

Proof. See Appendix.　　　　　　　　　　　　　　　　　　　　　　　　　　□

The intuition is the usual one from the IO literature. The benefits of higher speeds and higher participation are offset by inefficient cost duplication in economies with intermediate entry costs. Notice from 24that the slow venue profits can be expressed as

$$\pi_1^{comp} = \frac{s_1}{2r} \hat{\sigma}_1 \left(G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) \right)$$

Naturally, π_1^{comp} approaches zero when the default speed speed $\underline{s}$ is close to zero. When $\underline{s}$ approaches one, on the other hand, differentiation becomes increasingly difficult for the exchanges and equilibrium profits approach zero (the Bertrand outcome with $\hat{\sigma}_1 = 0$). Positive values of π_1^{comp} occur for intermediate $\underline{s}$.

In the proof of Proposition 9 we show that, when the marginal cost of technology approaches zero, $\overline{\kappa}$ approaches $\frac{\nu}{2r}(1 - 1/e))$. This limiting value is intuitive. The term $(1 - 1/e))$ represents the participation loss under the monopoly when $\underline{s} \to 1$. The term $\frac{\nu}{2r}$ represents the ex-ante participation value of investors of types $\sigma < \nu$ that decide not to participate under monopoly (using 17 we have $W\left(\sigma, \nu, 1\right) - W_{out} = \frac{\nu}{2r}$ when $\overline{a} = 1/2$).

5.3　Constrained Efficiency

How would a planner organize trading in our economy? With fixed setup costs and no marginal cost, the unconstrained solution is obviously to have one fast exchange with full participation financed by lump-sum taxes on all agents. This, however, is not a very relevant benchmark. Following a long tradition in public finance, we therefore solve a constrained planner's program where we rule out external subsidies. In other words, we require trading venues to (at least) break even.

Figure 7: Competition, Entry Costs and Welfare
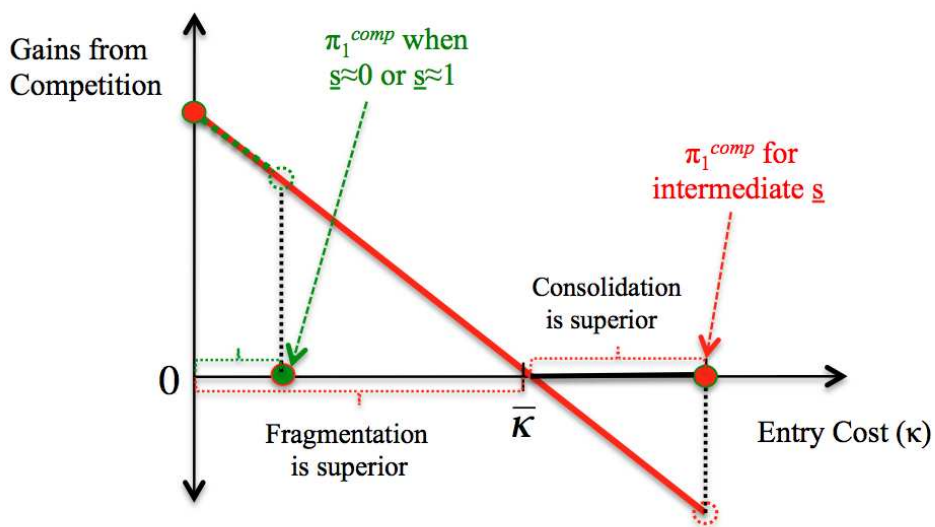
The graph shows the welfare gains from competition $\mathcal{W}_{comp} - \mathcal{W}_M - \kappa$, as a function of entry costs κ. Gains are zero for entry cost $\overline{\kappa}$. When π_1^{comp} is low relative to $\overline{\kappa}$, competition is never socially sub-optimal. When π_1^{comp} is high relative to $\overline{\kappa}$, there is a range of entry costs $[\overline{\kappa}, \pi_1^{comp}]$ where consolidation is socially preferred. This occurs for intermediate values of default speed $\underline{s}$.

The planner faces the same cost structure: a set-up cost κ for each venue, the default effective speed $\underline{s}$ is available at no cost, and a speed $s_2 > \underline{s}$ is available at cost $C(s_2)$. In the constrained program one might expect the planner to create two trading venues. This would allow the planner to relax the break-even constraints by charging a high price for the fast venue, while maintaining participation in the slower, but cheaper, venue. Surprisingly, however, we find that the planner chooses not to do so:

Proposition 10. *The planner operates a single venue.*

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

The planner chooses a single venue for investors even when there are no concerns of cost duplication (the result holds for $\kappa = 0$). This result extends to the case where trading venues can be consolidated or when the planner can choose s_1. The intuition is that the marginal dollar raised by increasing the price and lowering participation in the slow venue is always more valuable than the direct welfare loss. The planner therefore increases $\hat{\sigma}_1$ until it is equal to $\hat{\sigma}_{12}$.

In the remaining of this section, we therefore focus, without loss of generality, on the case where the planner operates a single venue. Its program is simply

$$\max_{s,q} \frac{s}{2r} \int_{\hat{\sigma}}^{\overline{\sigma}} \sigma dG(\sigma) - C(s), \tag{39}$$

$$\text{s.t. } q(1 - G(\hat{\sigma})) \geq C(s) \tag{40}$$

where the marginal type $\hat{\sigma}$ is defined as in (30). The following proposition compares the constrained efficient solution to the market outcome.

41

Proposition 11. *In the constrained efficient allocation, participation is always higher, but trading speed can be higher or lower, than under monopoly.*

Proof. See Appendix. □

We provide in the proof of Proposition 11 an example where $s_M > s*$. Intuitively, the monopolist chooses an inefficiently high speed when the distribution of permanent types has a fat right tail. In this case, the monopolist targets investors with high private gains from trade, offering a high-speed-high-price package. The planner might prefer to include the "middle class" of investors even if that means lower speed because of the break even constraint.

We know from Proposition 5 that the fast venue chooses a higher speed than the monopolist. Thus, it follows from Proposition 11 that the fast venue can also choose a sub-optimally high speed. In the numerical analysis, we show that equilibrium participation in the duopoly game can be inefficiently low or inefficiently high.

6   Numerical Analysis

In this section we illustrate the implications of the model numerically under A1 and A2. We compare market outcomes to a Walrasian market, which represents a frictionless competitive market with $\kappa_w = 0$ and $c_w = 0$ (which implies $\rho_w = \infty$ and $s_w = 1$). The Walrasian outcomes are as follows.

Lemma 7. *The Walrasian equilibrium outcomes are as follows: (i) investor participation equals $2\overline{a}$, (ii) $p_w = \frac{1}{r}\left[\mu - \nu\log(2\overline{a})\right]$, (iii) the instantaneous transaction rate equals $\tau_w = \frac{\overline{a}\gamma}{2}$, (iv) $\mathcal{W}_w = \frac{\overline{a}\nu}{r}(1 - \log(2\overline{a}))$.*

Table 6.1: Baseline Parameters

γ	$\underline{\rho}$	c	ν	$\bar{a}$	μ	r
608	1070	0.0002	e	0.45	1	$0.05/252$

Proof. See Appendix. □

6.1 Parametrization

We select parameters for a numerical illustration of the implications of the model. for a representative stock with a daily volume of $100,000$ shares and mean trade size of $1,000$ shares. Table 6.1 contains the baseline parameter values. We match trading delays to the prevalent ones around the implementation of decimalization in US equity markets, so that the default speed $\underline{\rho}$ is relatively low given current standards (average round trip trade is approx. 20 seconds). These values are consistent with the ones reported by Angel et al. (2010). The model implied daily preference shock rate is then 608. This is a high value for most individual investors. We think of our investors as either investment buy-side institutions or intermediary firms representing a group of end users. We analyze a range of values of c that generate endogenous speeds consistent with more recent standards (figures 8 and 9).

6.2 Investor Participation

As discussed in sections 5, participation with consolidated trading is always inefficiently low. The equilibrium analysis also shows that participation can be inefficiently high in a fragmented environment. Intuitively, when the marginal cost of technology increases, the competition outcome increasingly resembles the fragmented equilibrium without speed choice. We can observe in figure 8 that when c is high enough, total participation with speed choice crosses the optimal level from below. Analogously, participation under duopoly

is inefficiently low when the cost of speed is low. A similar argument can be made when default speed is very high. In this case, the scope for markets differentiation is very limited and total participation with speed choice approaches the maximum sustainable level $2\overline{a}$. Participation is also inefficiently high when γ and v are high.

6.3 Asset Prices

Figure 10 displays the effect of parameter changes under different market organizations. As discussed in section 4, a first general observation is that competition in intermediation decreases asset price levels.

Second, we find that asset prices are further depressed under price protection (i.e. $p^{prot} < \overline{p}^{seg}$). This is due to the fact that, for all possible WVAP prices, the marginal type is lower under protection and the equilibrium price liquidity adjustment $\frac{r+\gamma}{r+\gamma\underline{s}}$ is higher. An empirical implication of our analysis is then that, fixing a given asset characteristics, its price will be higher in consolidated trading economies (most Asian countries). Moreover, the price will be lower in markets that enforce price protection. All else equal (i.e., for similar assets) our model predicts that prices will be higher in Japan or China than in Europe, and in Europe than in the US[11].

A third important observation is that, in general, equilibrium asset prices bear a very weak relationship with social welfare (compare welfare and prices in figures 8 and 10).

The comparative statics show that, as expected, $\frac{\partial p}{\partial c} < 0$ and $\frac{\partial p}{\partial \nu} > 0$. However, while $\frac{\partial p^{seg}}{\partial \gamma}, \frac{\partial p^{prot}}{\partial \gamma} > 0$ and $\frac{\partial p^{seg}}{\partial \rho_1}, \frac{\partial p^{prot}}{\partial \rho_1} < 0$, the asset price is essentially not affected by the frequency of preference shocks and default speed levels when trading is consolidated (see sections 3 and

---

[11]Japan and China are cases where there is little or no fragmentation. Trading in fragmented in Europe, but there is not price protection like in the US.

4).

6.4   Is price protection socially desirable?

We analyze in this case effects on price protection when $\overline{a} < 1/2$. Price protection not only affects investor participation and speed choices but, as shown in Proposition 3, it can also affect the number of equilibrium exchanges. The latter effect can be dominant in economies with "intermediate" set-up costs. When a given trading regulation $\tau \in \{seg, prot\}$ is set at time I, the net social effects can be computed by analyzing its effect on the entry game and the posterior speed-fee competition and affiliation decisions. Let

$$\mathcal{G}_{prot}\left(\kappa\right) \equiv \mathcal{W}_{prot}^{net}\left(\kappa\right) - \mathcal{W}_{free}^{net}\left(\kappa\right)$$

denote the net social gains of price protection. In section 3.1 we showed that regulation may affect the equilibrium number of active venues. We have the following.

$$\mathcal{G}_{prot}\left(\kappa\right) = \begin{cases} \mathcal{W}_{prot} - \mathcal{W}_{free} & \text{if} & \kappa \leq \pi_1^{free} \\ \mathcal{W}_{prot} - \mathcal{W}_M - \kappa & \text{if} & \pi_1^{free} < \kappa \leq \pi_1^{prot} \\ 0 & \text{if} & \kappa > \pi_1^{prot} \end{cases}$$

We found that, when price protection affects entry, it has a first order positive effect on Welfare. This is due to more participation and more allocation efficiency when the entry outcome changes from a single to a dual venue equilibrium. However, in economies with low entry costs, price protection does not affects entry and has a small negative effect on welfare. This is due mainly to a loss in total market participation that also reduces transaction volumes (see figure 9).

7 Concluding Remarks

We have provided a positive and normative analysis of trading speed in financial markets. On the positive side, our model provides an explanation for the joint evolution of trading regulations, fragmentation, and speed. On the normative side, our model clarifies the circumstances under which competition, fragmentation and speed improve or reduce welfare. Our approach to liquidity is distinct from the usual liquidity externality based on increasing returns in the number of traders.

The most important caveat to our analysis is that our model ignores asymmetric information. It is important to point out, however, that our approach is the logical first step, since the trade-offs and economic forces that we have identified would be present in any model, with or without asymmetric information. In particular, speed-sensitive gains from trade are required to think about investment in speed. With free entry, if the average investor does not care about speed, then there would be no investment in speed. Nothing prevents the formation of a relatively slow and cheap exchange. If uninformed traders choose to join fast exchanges, it must be that they value speed. Otherwise they would all join the slow exchange, depriving the fast exchange from liquidity. The idea that speed is provided exclusively to satisfy a fraction of informed traders is therefore inconsistent with free entry. What information would do, then, would be to change the value of speed for some investors. It is possible that some participants might use speed to take advantage of other investors (e.g. Jovanovic and Menkveld (2010)). It is also possible that speed could allow uninformed traders to hedge. Hence, we certainly do not claim that asymmetric information is irrelevant, but we do claim that the building blocks of our model are required to analyze speed, fragmentation, and welfare with or without asymmetric information.

References

Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089. 8

Amihud, Y., Lauterbach, B., and Mendelson, H. (2003). The Value of Trading Consolidation : Evidence from the Exercise of Warrants. *Journal of Financial and Quantitative Analysis*, 38(4):829 –846. 8, 35

Amihud, Y. and Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–249. 8

Angel, J., Harris, L., and Spatt, C. S. (2010). Equity Trading in the 21st Century. 43

Biais, B. (1993). Price Formation and Equilibrium Liquidity in Fragmented and Centralized Markets. *Journal of Finance*, 48(1):157–185. 7

Duffie, D., Garleanu, N., and Pedersen, L. H. (2005). Over-the-counter Markets. *Econometrica*, 73(6):1815–1847. 8, 10

Duffie, D., Garleanu, N., and Pedersen, L. H. (2007). Valuation in Over-the-Counter Markets. *Review of Financial Studies*, 20(6):1865–1900. 5

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization*, 14(6):673–699. 37

Foucault, T. and Parlour, C. A. (2004). Competition for Listings. *Rand Journal of Economics*, 35(2):329–355. 8

Gabszewicz, J. and Thisse, J.-F. (1979). Price Competition, Quality and Income Disparities. *Journal of Economic Theory*, 20:340–359. 8

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *Journal of Finance*, 49(4):1127–1161. 7

Hendershott, T. and Mendelson, H. (2000). Crossing Networks and Dealer Markets: Competition and Performance. *Journal of Finance*, 55(5):2071–2115. 7

Huang, J. and Wang, J. (2010). Market liquidity, asset prices, and welfare. *Journal of Financial Economics*, 95(1):107–127. 8

Jarrow, R. A. and Protter, P. (2011). A Dysfunctional Role of High Frequency Trading in Electronic Markets. 8

Jovanovic, B. and Menkveld, A. J. (2010). Middlemen in Electronic Markets. 46

Lagos, R. and Rocheteau, G. (2009). Liquidity in Asset Markets With Search Frictions. *Econometrica*, 77(2):403–426. 8, 10, 11, 12

Madhavan, A. (1995). Consolidation, Fragmentation and the Disclosure of Trading Information. *Review of Financial Studies.* 7

Mendelson, H. (1987). Consolidation, fragmentation, and market performance. *Journal of Financial and Quantitative Analysis*, 22:187–207. 7

Menkveld, A. J. (2011). Electronic Trading and Market Structure. In *UK Government Driver Review, DR16.* 4

O'Hara, M. and Ye, M. (2011). Is market fragmentation harming market quality? *Journal of Financial Economics.* 8

Pagano, M. (1989). Trading volume and Asset Liquidity. *Quarterly Journal of Economics.* 6, 7

Parlour, C. A. and Seppi, D. J. (2003). Liquidity-based competition for order flow. *Review of Financial Studies*, 16(2):329–355. 7

Santos, T. and Scheinkman, J. A. (2001). Competition Among Exchanges. *Quarterly Journal of Economics*, 116(3):225–1061. 8

Shaked, A. and Sutton, J. (1982). Relaxing Price Competition Through Product Differentiation. *Review of Economic Studies*, 49(1):3–13. 8

Shaked, A. and Sutton, J. (1983). Natural Oligopolies. *Econometrica*, 51(5):1469–1483. 8

Vayanos, D. and Wang, T. (2007). Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory*, 136(1):66–104. 8

Weill, P.-O. (2007). Leaning Against the Wind. *Review of Economic Studies*, 74(October 1987):1329–1354. 8

Appendix

Proof of Proposition 1

For steady state values we have

$$(r + \rho) L = \mu - \sigma + \frac{\gamma}{2} (H - L) + \rho p$$

$$(r + \rho) H = \mu + \sigma - \frac{\gamma}{2} (H - L) + \rho p$$

and

$$\left(r + \frac{\gamma}{2} \right) V_{\sigma,-} (0) = \frac{\gamma}{2} V_{\sigma,+} (0)$$

$$\left(r + \frac{\gamma}{2} \right) V_{\sigma,+} (1) = \mu + \sigma + \frac{\gamma}{2} V_{\sigma,-} (1)$$

$$\left(r + \frac{\gamma}{2} \right) V_{\sigma,+} (0) = \frac{\gamma}{2} V_{\sigma,-} (0) + \rho (H - p)$$

$$\left(r + \frac{\gamma}{2} \right) V_{\sigma,-} (1) = \mu - \sigma + \frac{\gamma}{2} V_{\sigma,+} (1) + \rho (p - L)$$

Let us now compute the ex-ante value functions. Let us first consider types $\sigma < \hat{\sigma}$. They join the market to sell at price p. Averaging over types $\epsilon = \pm 1$, we get the ex-ante value function $\hat{V}$ that solves the Bellman equation

$$r \hat{V} = \mu \bar{a} + \rho \left(p \bar{a} - \hat{V} \right) \Longrightarrow \hat{V} = \frac{\mu + \rho p}{r + \rho} \bar{a}$$

Since $\mu + \rho p = \frac{\mu}{r} (r + \rho) + \rho \left(p - \frac{\mu}{r} \right)$ we can rewrite

$$\hat{V} = \frac{\mu}{r} \bar{a} + \frac{\rho}{r + \rho} \left(p - \frac{\mu}{r} \right) \bar{a}$$

From the definition of $\hat{\sigma}$ we also now that

$$p - \frac{\mu}{r} = \frac{1}{r} \frac{r + \rho}{r + \rho + \gamma} \hat{\sigma}$$

therefore

$$\hat{V} = \frac{\mu \bar{a}}{r} + \frac{\rho}{r + \rho + \gamma} \frac{\bar{a}}{r} \hat{\sigma}$$

since $s(\rho) \equiv \frac{\rho}{r + \gamma + \rho}$, we get finally

$$\hat{V} = \frac{\mu \bar{a}}{r} + s \frac{\bar{a}}{r} \hat{\sigma}$$

the marginal type $\hat{\sigma}(p, \rho)$ is defined in (5), is increasing in p and decreasing in ρ. The key point is that $\hat{V}$ does not depend on the type σ, but only on the price and speed of the market.

Let us now consider the steady state types, $\sigma > \hat{\sigma}$. They also start with endowment $\bar{a}$ and they are going to jump directly to their steady state value functions as soon as they can trade

$$rW_{\sigma,+} = \bar{a}(\mu + \sigma) + \frac{\gamma}{2}(W_{\sigma,-} - W_{\sigma,+}) + \rho(V_{\sigma,+}(1) - W_{\sigma,+} - p(1 - \bar{a}))$$

$$rW_{\sigma,-} = \bar{a}(\mu - \sigma) + \frac{\gamma}{2}(W_{\sigma,+} - W_{\sigma,-}) + \rho(V_{\sigma,-}(0) - W_{\sigma,-} + p\bar{a})$$

and the ex-ante value is

$$W(\sigma, \rho) \equiv \frac{W_{\sigma,+} + W_{\sigma,-}}{2}$$

Adding up the two lines, we get

$$(r + \rho)W(\sigma) = (\mu + \rho p)\bar{a} + \rho \frac{V_{\sigma,+}(1) + V_{\sigma,-}(0) - p}{2}$$

Note the nice intuition

$$W(\sigma, \rho) = \hat{V} + \frac{\rho}{r + \rho} \frac{V_{\sigma,+}(1) + V_{\sigma,-}(0) - p}{2}$$

This gives a nice continuous link between the low and high types around $\hat{\sigma}$. Of course, we are going to find that the second term on the RHS is zero for $\sigma = \hat{\sigma}$. Therefore we need to compute

$$X \equiv V_{\sigma,+}(1) + V_{\sigma,-}(0).$$

Note that X is the sum of the value of the efficient allocation. Using the system in the previous sections, let us define

$$S \equiv \sum_{\epsilon, a} V_{\sigma,\epsilon}(a)$$

It is easy to see that

$$rS = 2\mu + \rho\left(H - L\right) = 2\left(\mu + \rho\frac{\sigma}{r + \gamma + \rho}\right)$$

For X we get

$$\left(r + \frac{\gamma}{2}\right)X = \mu + \sigma + \frac{\gamma}{2}\left(V_{\sigma,+}\left(0\right) + V_{\sigma,-}\left(1\right)\right) = \mu + \sigma + \frac{\gamma}{2}\left(S - X\right)$$

Therefore

$$\left(r + \gamma\right)X = \mu + \sigma + \frac{\gamma}{2}S$$
$$= \mu + \sigma + \frac{\gamma}{r}\left(\mu + \rho\frac{\sigma}{r + \gamma + \rho}\right)$$
$$= \frac{\mu}{r}\left(r + \gamma\right) + \frac{\sigma}{r}\left(r + \gamma\right)\frac{r + \rho}{r + \gamma + \rho}$$

which we can simplify to

$$X = \frac{\mu}{r} + \frac{\sigma}{r}\frac{r + \rho}{r + \gamma + \rho}$$

Therefore, we have, when $\sigma > \hat{\sigma}$, we have

$$W\left(\sigma, \rho\right) = \hat{V} + \frac{1}{2}\frac{\rho}{r + \gamma + \rho}\frac{\sigma - \hat{\sigma}}{r}$$

QED.

Proof of Lemma 4

First notice that $W\left(\hat{\sigma}_{12}, \hat{\sigma}_{2}, s_{2}\right) - q_{2} = W\left(\hat{\sigma}_{12}, \hat{\sigma}_{1}, s_{1}\right) - q_{1}$ can be written as:

$$\frac{s_{2}\bar{a}\hat{\sigma}_{2}}{r} + \frac{s_{2}}{2r}\left(\hat{\sigma}_{12} - \hat{\sigma}_{2}\right) - q_{2} = \frac{s_{1}\bar{a}\hat{\sigma}_{1}}{r} + \frac{s_{1}}{2r}\left(\hat{\sigma}_{12} - \hat{\sigma}_{1}\right) - q_{1}.$$

Since $q_1 = \frac{\bar{a}s_1\hat{\sigma}_1}{r}$, we get $\frac{s_2 - s_1}{2r}\hat{\sigma}_{12} = q_2 - \frac{\bar{a}s_2\hat{\sigma}_2}{r} + \frac{s_2\hat{\sigma}_2 - s_1\hat{\sigma}_1}{2r}$. Using $\hat{\sigma}_2 = m\hat{\sigma}_1$, we get $\frac{s_2 - s_1}{2r}\hat{\sigma}_{12} = q_2 - q_1\left(\frac{1}{2\bar{a}} - \frac{s_2}{s_1}m\left(\frac{1}{2\bar{a}} - 1\right)\right)$ where $m \equiv \frac{1 + \frac{\gamma}{r+\rho_2}}{1 + \frac{\gamma}{r+\rho_1}}$. Since $\frac{s_2}{s_1}m = \frac{\rho_2}{\rho_1}\frac{r+\rho_1}{r+\rho_2}$, we get

$$\hat{\sigma}_{12} = \frac{2r}{s_2 - s_1}\left(q_2 - \frac{z}{2\bar{a}}q_1\right)$$

where

$$z \equiv 1 - \frac{1 + \frac{r}{\rho_1}}{1 + \frac{r}{\rho_2}}(1 - 2\bar{a}).$$

Note that $z \leq 1$, $z \approx 1$ when $a \approx 0.5$, and $z \approx 2a$ when r/ρ is small (the realistic case). The profits of market 1 are

$$\pi_1^{prot} = q_1\left(G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) + \delta_1\right)$$

We know introduce notations to simplify the equation

$$\alpha \equiv 2\bar{a}$$

$$k \equiv \frac{s_1}{s_2 - s_1}$$

In the protected price equilibrium, firms therefore maximize

$$\max_{q_1} \pi_1^{prot} = \frac{q_1}{\alpha}\left(1 - \alpha + \alpha G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right)\right)$$

$$\max_{q_2} \pi_2^{prot} = q_2\left(1 - G\left(\hat{\sigma}_{12}\right)\right)$$

The conditions $\frac{\partial \pi_1^{prot}}{\partial q_1} = 0$ and $\frac{\partial \pi_2^{prot}}{\partial q_2} = 0$ lead to

$$1 - G\left(\hat{\sigma}_{12}\right) = g\left(\hat{\sigma}_{12}\right)\left(\hat{\sigma}_{12} + zk\hat{\sigma}_1\right)$$

$$1 - \alpha + \alpha G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) = \left(g\left(\hat{\sigma}_1\right) + \alpha zkg\left(\hat{\sigma}_{12}\right)\right)\hat{\sigma}_1$$

QED

Proof of Proposition 2

Define

$$\nu\left(\hat{\sigma}\right) \equiv \frac{1 - G\left(\hat{\sigma}\right)}{g\left(\hat{\sigma}\right)}$$

Let us compare the three market structures. The monopoly allocation $\hat{\sigma}_m$ is the solution to

$$\hat{\sigma}_m = \nu\left(\hat{\sigma}_m\right)$$

Rearranging the first order conditions, the free segmentation allocation $\left(\hat{\sigma}_1^{seg}, \hat{\sigma}_{12}^{seg}\right)$ is the solution to

$$
\begin{aligned}
\hat{\sigma}_{12} &= \nu\left(\hat{\sigma}_{12}\right) - k\hat{\sigma}_1 \\
\hat{\sigma}_1\left(\frac{g\left(\hat{\sigma}_1\right)}{g\left(\hat{\sigma}_{12}\right)} + k\right) &= \frac{g\left(\hat{\sigma}_1\right)}{g\left(\hat{\sigma}_{12}\right)}\nu\left(\hat{\sigma}_1\right) - \nu\left(\hat{\sigma}_{12}\right)
\end{aligned}
$$

The price protection allocation $\left(\hat{\sigma}_1^{prot}, \hat{\sigma}_{12}^{prot}\right)$ is the solution to

$$
\begin{aligned}
\hat{\sigma}_{12} &= \nu\left(\hat{\sigma}_{12}\right) - z\left(\alpha\right)k\hat{\sigma}_1 \\
\hat{\sigma}_1\left(\frac{g\left(\hat{\sigma}_1\right)}{g\left(\hat{\sigma}_{12}\right)} + \alpha z\left(\alpha\right)k\right) &= \frac{g\left(\hat{\sigma}_1\right)}{g\left(\hat{\sigma}_{12}\right)}\nu\left(\hat{\sigma}_1\right) - \alpha\nu\left(\hat{\sigma}_{12}\right)
\end{aligned}
$$

where we highlight in red the differences to help the comparison. Notice first $\hat{\sigma}_{12} < \hat{\sigma}_m$ irrespective of whether prices are free or protected.

Exponential Distribution

Under A1, we have $G\left(\sigma\right) = 1 - e^{-\sigma/\nu}$ and therefore $\nu\left(\hat{\sigma}\right) = \nu$ and the system is

$$
\begin{aligned}
\frac{\hat{\sigma}_{12}}{\nu} &= 1 - zk\frac{\hat{\sigma}_1}{\nu} \\
\frac{\hat{\sigma}_1}{\nu}\left(e^{\frac{\hat{\sigma}_{12} - \hat{\sigma}_1}{\nu}} + \alpha zk\right) &= e^{\frac{\hat{\sigma}_{12} - \hat{\sigma}_1}{\nu}} - \alpha
\end{aligned}
$$

It is convenient to defined $\Delta \equiv (\hat{\sigma}_{12} - \hat{\sigma}_1)/\nu$ and $x \equiv \frac{\hat{\sigma}_1}{\nu}$, so that we can write the system in (x, Δ):

$$(1 + zk)\,x = 1 - \Delta \tag{41}$$

$$e^\Delta - \alpha = (e^\Delta + \alpha zk)\,x \tag{42}$$

Impact of protection on $\hat{\sigma}_1$

The second equation of the system is

$$1 - x = \frac{\alpha(1 + zk)}{e^\Delta + \alpha zk}$$

This leads to a schedule x increasing in Δ. The issue is how it changes with α. We study the function on the RHS, namely: $\log\left(\frac{\alpha(1+zk)}{e^\Delta+\alpha zk}\right) = \log(\alpha) + \log(1 + zk) - \log(e^\Delta + \alpha zk)$. Taking the derivative w.r.t. α

$$\frac{1}{\alpha} + \frac{kz'}{1 + zk} - \frac{\alpha kz' + kz}{e^\Delta + \alpha kz} = \frac{1}{\alpha} - \frac{1}{\alpha + \frac{e^\Delta}{kz}} + kz'\left(\frac{1}{1 + kz} - \frac{1}{\frac{e^\Delta}{\alpha} + kz}\right)$$

since $\frac{e^\Delta}{\alpha} > 1$ we have $\frac{1}{1+kz} - \frac{1}{\frac{e^\Delta}{\alpha}+kz} > 0$. Similarly $\frac{1}{\alpha} - \frac{1}{\alpha + \frac{e^\Delta}{kz}} > 0$. So $\frac{\alpha(1+kz)}{e^\Delta+\alpha zk}$ is increasing in α. Therefore the equilibrium condition $e^\Delta - \alpha = (e^\Delta + \alpha kz)\,x$ implies a schedule x increasing in Δ and decreasing in α. The first equilibrium condition $(1 + zk)\,x = 1 - \Delta$ gives a schedule x decreasing in Δ and decreasing in α. Straightforward analysis then shows that x must be decreasing in α. The free price structure corresponds to $\alpha = 1$, while the protected price structure corresponds to $\alpha = 2a < 1$. Therefore, since $\hat{\sigma}_1 = \nu x$, $\hat{\sigma}_1$ must be higher under price protection.

Impact of protection on $\hat{\sigma}_{12}$

The analysis of $\hat{\sigma}_{12}$ is ambiguous. It is clear that when $k \to 0$ we have $\hat{\sigma}_{12} \to \nu$, which is the monopoly solution. Define $y = \frac{\hat{\sigma}_{12}}{\nu} = x + \Delta$, and get the system

$$(1 + kz)\,y = 1 + kz\Delta$$

$$1 - y = kz\frac{e^\Delta - \alpha}{e^\Delta + \alpha kz}$$

The first curve is y increasing in Δ and decreasing in α. The second curve can be written gives $y = 1 - kz + \frac{kz\alpha(kz+1)}{e^\Delta + \alpha kz}$, which shows y decreasing in Δ. With respect to α, however, it is not clear. In the realistic case where $\frac{r}{\rho_1}$ is small, we have $z(\alpha) = \alpha$ so

$$(1 + k\alpha)\, y = 1 + k\alpha\Delta$$

$$1 - y = k\alpha \frac{e^\Delta - \alpha}{e^\Delta + k\alpha^2}$$

We study the case where α is close to one. The free price solution is

$$(1 + k)\,\bar{y} = 1 + \bar{\Delta} k$$

$$1 - \bar{y} = k \frac{e^{\bar{\Delta}} - 1}{e^{\bar{\Delta}} + k}$$

and we look for small deviations: $\alpha = 1 - \epsilon$, $\Delta = \bar{\Delta} + \hat{\Delta}$, $y = \bar{y} + \hat{y}$. The first equation is simply

$$(1 + k)\,\hat{y} - k\bar{y}\epsilon = k\left(\hat{\Delta} - \bar{\Delta}\epsilon\right)$$

$$(1 + k)\,\hat{y} = k\hat{\Delta} + k\left(\bar{y} - \bar{\Delta}\right)\epsilon$$

The second one gives

$$1 - \bar{y} - \hat{y} = \frac{k}{e^{\bar{\Delta}} + k}\left(e^{\bar{\Delta}} - 1 + \hat{\Delta}e^{\bar{\Delta}} + \left(2 - e^{\bar{\Delta}}\right)\epsilon - \frac{e^{\bar{\Delta}} - 1}{e^{\bar{\Delta}} + k}\left(e^{\bar{\Delta}}\hat{\Delta} - 2k\epsilon\right)\right)$$

$$-\left(e^{\bar{\Delta}} + k\right)^2 \hat{y} = ke^{\bar{\Delta}}\left((1 + k)\,\hat{\Delta} + \left(2 - e^{\bar{\Delta}} + k\right)\epsilon\right)$$

From the first schedule we get $k\hat{\Delta} = (1 + k)\,\hat{y} - k\left(\bar{y} - \bar{\Delta}\right)\epsilon$. The second schedule then becomes

$$-\left(\left(e^{\bar{\Delta}} + k\right)^2 + e^{\bar{\Delta}}(1 + k)^2\right)\hat{y} = ke^{\bar{\Delta}}\left(2 + k - e^{\bar{\Delta}} - (1 + k)\left(\bar{y} - \bar{\Delta}\right)\right)\epsilon$$

The evolution of y therefore depends on the sign of $\chi = 2 + k - e^{\bar{\Delta}} - (1 + k)\left(\bar{y} - \bar{\Delta}\right)$. From

the equilibrium condition at $\alpha = 1$, we get $\bar{y} = \frac{1+\bar{\Delta}k}{1+k}$, and the Δ under free prices solves

$$\left(\bar{\Delta} + k\right) e^{\bar{\Delta}} = 1 + k \left(2 - \bar{\Delta}\right)$$

In the special case $k = 0$, we get $\bar{y} = 1$ and $\bar{\Delta}e^{\bar{\Delta}} = 1$ implies $\bar{\Delta} = 0.5671$ then $\chi = 1 - e^{\bar{\Delta}} + \bar{\Delta} = -0.1961 < 0$. In this case $\hat{y}$ increases with ϵ: σ_{12} is higher under price protection. However, as long as k is not too small ($k > 0.185$), we have $2 + k - e^{\bar{\Delta}} - (1 + k)\left(\bar{y} - \bar{\Delta}\right) > 0$ and $\hat{y}$ decreasing with ϵ: σ_{12} is lower, and participation in the fast market is higher under price protection.

Comparing Profits

It is convenient to define a system that nests price protection and free competition as special cases. Fist, define the scaled controls

$$t_1 \equiv \frac{2r}{\alpha s_1}q_1,$$
$$t_2 \equiv \frac{2r}{s_1}q_2.$$

Next the scaled profits by $F_i \equiv \frac{2r}{s_1}\pi_i$. With these notations, the profit functions are

$$F_1\left(t_1, t_2, \alpha\right) = t_1 \left(1 - \alpha + \alpha G\left(\hat{\sigma}_{12}\right) - G\left(t_1\right)\right)$$
$$F_2\left(t_1, t_2, \alpha\right) = t_2 \left(1 - G\left(\hat{\sigma}_{12}\right)\right)$$

and we have

$$\hat{\sigma}_{12} = k \left(t_2 - z\left(\alpha\right)t_1\right)$$
$$\hat{\sigma}_1 = t_1$$

The general system is the one with protected prices with $\alpha < 1$ and $z\left(\alpha\right) = 1 - \frac{1 + \frac{r}{\rho_1}}{1 + \frac{r}{\rho_2}}\left(1 - \alpha\right)$. The free segmentation case corresponds to $\alpha = 1$ and $z = 1$. We can always return to the system in σ using $t_2 = \frac{\hat{\sigma}_{12}}{k} + z\hat{\sigma}_1$ and $t_1 = \hat{\sigma}_1$.

56

Let us now derive the FOCs. Using $\frac{\partial \pi_1^{prot}}{\partial t_1} = 0$ and $\frac{\partial \pi_2^{prot}}{\partial t_2} = 0$ we get

$$1 - \alpha + \alpha G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right) = t_1 \left(\alpha z\left(\alpha\right) kg\left(\hat{\sigma}_{12}\right) + g\left(\hat{\sigma}_1\right)\right)$$
$$1 - G\left(\hat{\sigma}_{12}\right) = t_2 kg\left(\hat{\sigma}_{12}\right)$$

With exponential distributions we have that t_2 is constant: $t_2 = \frac{\nu}{k}$. Note that this implies $q_2 \frac{2r}{s_1} = \frac{\nu}{k}$ so $q_2 = \frac{\nu}{2r}\left(s_2 - s_1\right)$. The fees of the fast venue are proportional to the difference in effective speed.

To understand the impact of price protection of profits, take the total differential

$$\frac{dF_1}{d\alpha} = \frac{\partial F_1}{\partial t_1}\frac{dt_1}{d\alpha} + \frac{\partial F_1}{\partial t_2}\frac{dt_2}{d\alpha} + \frac{\partial F_1}{\partial \alpha}$$

Optimality implies $\frac{\partial F_1}{\partial t_1} = 0$, and we have just seen that $\frac{dt_2}{d\alpha} = 0$. Therefore $\frac{dF_1}{d\alpha} = \frac{\partial F_1}{\partial \alpha}$ and

$$\frac{\partial F_1}{\partial \alpha} = t_1 \left(-1 + G\left(\hat{\sigma}_{12}\right) + \alpha g\left(\hat{\sigma}_{12}\right)\frac{\partial \hat{\sigma}_{12}}{\partial \alpha}\right) = t_1 \left(-1 + G\left(\hat{\sigma}_{12}\right) - \alpha g\left(\hat{\sigma}_{12}\right) kt_1 z'\left(\alpha\right)\right)$$

Since $z'\left(\alpha\right) > 0$, we see that $\frac{\partial F_1}{\partial \alpha} < 0$: price protection increases the profits of the slow market. The economic intuition is simple. The term $-1 + G\left(\hat{\sigma}_{12}\right)$ corresponds to the "sell and leave" investors who come to the slow venue under protection. The term with z' corresponds to the softer price effect on the marginal type $\hat{\sigma}_{12}$. QED.

Proof of Proposition 3

The relationship between entry costs κ and profits determines the number of active venues in equilibrium. Let $\overline{\pi}_i \equiv \max\left\{\pi_i^{prot}, \pi_i^{seg}\right\}$ and $\underline{\pi}_i \equiv \min\left\{\pi_i^{prot}, \pi_i^{seg}\right\}$. We analyze below the existence of NE in pure strategies of the normal-form game shown in figure 5.

- *Two-venues equilibriums.* Suppose $\kappa \le \overline{\pi}_1$, By Proposition 2, we have that $\underline{\pi}_1 = \pi_1^{seg}$. It is immediate then that entry is always optimal for the slow venue when $\kappa \le \pi_1^{seg}$ and that, for any $\pi_1^{seg} < \kappa \le \pi_1^{prot}$, we have $\pi_1^{seg} - \kappa < 0$ and $\pi_1^{prot} - \kappa \ge 0$. A duopoly is never sustainable whenever $\kappa > \pi_1^{prot}$.

- *Single-venue equilibriums.* Suppose $\pi_1^{prot} < \kappa \le \pi_2^m$.

- Case 1: $\pi_2^m \geq \kappa > \pi_1^m$. The only NE has the slow venue out and the fast venue entering, with payoff π_2^m.

- Case 2: $\overline{\pi}_1 \leq \kappa \leq \underline{\pi}_2$. In this case there is a single NE where only the fast venue enters.

- Case 3: $\overline{\pi}_2 < \kappa < \pi_1^m$. There are two NE where only one venue enters, either the slow of fast one.

- Case 4: $\underline{\pi}_2 < \kappa \leq \min\{\overline{\pi}_2, \pi_1^m\}$. When $\pi_2^{\mathsf{T}} = \overline{\pi}_2$, there is a single NE where only the fast venue enters. When $\pi_2^{\mathsf{T}} = \underline{\pi}_2$, there are two NE where only one venue enters, either the slow of fast one.

- *No-entry equilibrium.* Whenever $\kappa > \pi_2^m$ the only NE has both venues out. QED

Proof of Proposition 4

Part (i) is straightforward. The interior solution FOC for speed is

$$-g\left(\hat{\sigma}_M\right)\frac{\partial \hat{\sigma}_M}{\partial s_M}q = C'\left(s_M\right) \tag{43}$$

Combining 43 with the FOC for q, and using 30 to compute the partial derivative delivers (ii). Using A2 we have that the LHS of 31 is given by

$$2rC'\left(s_M\right) = \frac{2rc\left(\gamma + r\right)}{\left(1 - s\right)^2}$$

Using A1 we have that the RHS of 31 equals ν/e. Combining these expressions yields 32. Using 18 in 32 we have that the optimal speed ρ_M is given by

$$\rho_M = \frac{\left(\gamma + r\right)^{1/2}\left(\nu^{1/2} - \left(2rce\left(\gamma + r\right)\right)^{1/2}\right)}{\left(2rce\right)^{1/2}}$$

When c is sufficiently low, it is easy to see that the value of $\frac{\partial \rho_M}{\partial \gamma}$ is positive (negative) for small (large) values of γ and achieves a maximum at $\frac{\nu}{8cer} - r$. We also have

$$\frac{\partial^2 \rho_M}{\partial \gamma^2} = -\frac{\nu^{1/2}}{4\left(2rce\right)^{1/2}\left(\gamma + r\right)^{3/2}}$$

which is negative for all $\gamma > 0$. QED

Proof of Proposition 5

Parts (i) and (ii) are straightforward. To prove part (iii) we proceed in 3 steps.

Step 1: Necessary condition

Under A1 and with $\alpha = 1$, we have

$$\hat{\sigma}_{12} = \nu - \frac{s_1}{s_2 - s_1}\hat{\sigma}_1$$

and

$$q_2 = \frac{\nu}{2r}\left(s_2 - s_1\right)$$

The profits of the fast venue are $\pi_2 = q_2\left(1 - G\left(\hat{\sigma}_{12}\right)\right)$ therefore

$$\pi_2 = \frac{\nu}{2r}\left(s_2 - s_1\right)\left(1 - G\left(\hat{\sigma}_{12}\right)\right)$$

Note that this system is equivalent to the monopoly case when $s_1 = 0$. The FOC for speed is

$$2rC'\left(s_2\right) = \nu\left(1 - G\left(\hat{\sigma}_{12}\right)\right) - \nu\left(s_2 - s_1\right)g\left(\hat{\sigma}_{12}\right)\frac{\partial\hat{\sigma}_{12}}{\partial s_2} \tag{44}$$

The consolidated solution is $2rC'\left(\overline{s}_2\right) = \nu e^{-1}$. With two active venues we have

$$\frac{\partial\hat{\sigma}_{12}}{\partial s_2} = \frac{k}{s_2 - s_1}\hat{\sigma}_1 - k\frac{\partial\hat{\sigma}_1}{\partial s_2}$$

59

Then,

$$2rC'\left(s_2\right) = \nu\left(1 - G\left(\hat{\sigma}_{12}\right)\right) - \nu g\left(\hat{\sigma}_{12}\right)\left[k\hat{\sigma}_1 - s_1\frac{\partial\hat{\sigma}_1}{\partial s_2}\right]$$

$$= e^{-\frac{\hat{\sigma}_{12}}{\nu}}\left(\nu - \left[k\hat{\sigma}_1 - s_1\frac{\partial\hat{\sigma}_1}{\partial s_2}\right]\right)$$

Using $x \equiv \frac{\hat{\sigma}_1}{\nu}, \Delta \equiv \frac{\hat{\sigma}_{12} - \hat{\sigma}_1}{\nu}$

$$2rC'\left(s_2\right) = \nu e^{kx-1}\left(1 - kx + s_1\frac{\partial x}{\partial s_2}\right) \tag{45}$$

Since C' is an increasing function, market 2 chooses a higher speed whenever the RHS of 45 is greater than νe^{-1}. That is,

$$e^{kx}\left(1 - kx + s_1\frac{\partial x}{\partial s_2}\right) - 1 > 0 \tag{46}$$

Step 2: Finding $\frac{\partial x}{\partial s_2}$

Differentiating the system 41-42 we have

$$\left(1 + k\right)dx + d\Delta - \frac{k}{\left(s_2 - s_1\right)}ds_2 = 0$$

$$\left(e^{\Delta} + k\right)dx + e^{\Delta}\left(x - 1\right)d\Delta - \frac{k}{\left(s_2 - s_1\right)}ds_2 = 0$$

After appropriate substitutions we get

$$\frac{\partial x}{\partial s_2} = \frac{k}{s_2 - s_1}\left(\frac{x\left(1 + e^{\Delta}\left(1 - x\right)\right)}{e^{\Delta}\left(2 + k - x\left(1 + k\right)\right) + k}\right)$$

Re-arranging,

$$s_1\frac{\partial x}{\partial s_2} = \frac{k^2 x\left(1 + e^{\Delta}\left(1 - x\right)\right)}{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right)} \tag{47}$$

Step 3: Verifying inequality 46

Substituting 47 in 46 let

$$S\left(k\right) \equiv e^{kx}\left(1 - kx + \frac{k^2 x\left(1 + e^{\Delta}\left(1 - x\right)\right)}{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right)}\right) - 1 \tag{48}$$

Re-arranging we have

$$S(k) = e^{kx}\left(\frac{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right) - kxe^{\Delta}\left(1 + \Delta - kx\right)}{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right)}\right) - 1 \tag{49}$$

To satisfy the inequality we need $S\left(k\right) > 0$ for all $k > 0$ and $S\left(0\right) = 0$ (corresponding to the monopolist case where $s_1 = 0$). Let $x\left(k\right)$ and $\Delta\left(k\right)$ denote the solutions to the system 41-42 for a given $k \geq 0$. Since $x\left(k\right)$ and $\Delta\left(k\right)$ are continuous functions, $S(k)$ is continuous. Using 41-42 one can see that

$$\lim_{k\to\infty} x\left(k\right) = 0 \tag{50}$$
$$\lim_{k\to\infty} \Delta\left(k\right) = \underline{\Delta}$$

where $\underline{\Delta}$ is defined by $e^{\underline{\Delta}} + \underline{\Delta} = 2$. Notice that $\lim_{k\to\infty} x\left(k\right) k = 1 - \underline{\Delta}$. Similarly,

$$\lim_{k\to 0} x\left(k\right) = 1 - \overline{\Delta} \tag{51}$$
$$\lim_{k\to 0} \Delta\left(k\right) = \overline{\Delta}$$

where where $\overline{\Delta}$ is defined by $e^{\overline{\Delta}}\overline{\Delta} = 1$. Taking limits of 48 we find

$$\lim_{k\to 0} S\left(k\right) = e^0 - 1 = 0$$
$$\lim_{k\to\infty} S\left(k\right) = e^{1-\underline{\Delta}} - 1 > 0$$

A sufficient condition for $S(k) > 0$ for all $k > 0$ is to show that the term between brackets in 49 is greater than one. This is the case whenever

$$e^{\Delta}\left(1 + \Delta + k\right) + k + e^{\Delta}k\left[\left(1 - x\right) + \left(xk\right)^2 - x\Delta\right] > e^{\Delta}\left(1 + \Delta + k\right) + k \tag{52}$$

61

Note from 41 that $1 - x = kx + \Delta$. Then,

$$(1 - x) + (xk)^2 - x\Delta = kx + \Delta(1 - x) + (xk)^2 > 0$$

We conclude that $S(k) > 0$ for all $k > 0$. QED

Proof of Lemmas and Propositions of Section 4

The market clearing condition 6 depends on the marginal trading type in equation 5. Using 5 to solve for p, and using 18 we obtain 34. From Lemma 4 we know that $\hat{\sigma}_1^{prot}$ clears the asset market when prices are protected. The marginal type is then given by

$$\hat{\sigma}^{prot} = \left(1 + \frac{\gamma}{r + \rho_1}\right)(rp - \mu)$$

In this case, the marginal type joins market 1 with speed ρ_1. Using 18 and re-arranging yields (ii).

To find the VWAP price in the free segmentation case, we need the steady-state instantaneous transaction rate τ_i. In the case of a single venue, this rate is given by the product between the mass of active investors that want to trade, and the market speed ρ_i. The mass of active investors in market 1 is given by $G\left(\hat{\sigma}_{12}^{seg}\right) - G\left(\hat{\sigma}_1^{seg}\right)$, and a fraction $\alpha_{\sigma,+}(0) = \alpha_{\sigma,-}(1)$ of them want to trade in the steady-state. Using **??** we then have

$$\tau_1 = \frac{\gamma}{4(\gamma + \rho_1)} G\left(\hat{\sigma}_{12}^{free}\right) - G\left(\hat{\sigma}_1^{free}\right) \rho_1$$

Using using 18 delivers **??**. The rate τ_2 can be found analogously. In steady state both relative transaction volumes $\frac{\tau_i}{\tau_i + \tau_j}$ and prices are constant, which explains 35.

Expression 36 is found by replacing 32 in 34 and using $\hat{\sigma}_M = \nu$ under A1. The frictionless competitive is derived in Lemma 6.

Proof of Proposition 7

To prove $p_M > \overline{p}^{seg}$, it is sufficient to show that $p_M \geq p_2^{seg}$. The latter requires

$$\hat{\sigma}_M \frac{r + \gamma s_M}{r + \gamma} \geq \hat{\sigma}_2 \frac{r + \gamma s_2}{r + \gamma} \tag{53}$$

$$\frac{r + \gamma s_M}{r + \gamma s_2} \geq \frac{\hat{\sigma}_2}{\hat{\sigma}_M} = \frac{\hat{\sigma}_2}{\nu}$$

where we used A1's implication $\hat{\sigma}_M = \nu$, and skipped super-indexes for notation simplicity. Section 2 shows that $\hat{\sigma}_1 < \hat{\sigma}_2 < \hat{\sigma}_{12}$. Using $x \equiv \frac{\hat{\sigma}_1}{\nu}, \Delta \equiv \frac{\hat{\sigma}_{12} - \hat{\sigma}_1}{\nu}$, we then have $x + \Delta > \frac{\hat{\sigma}_2}{\nu}$. It is thus sufficient to prove that

$$\frac{r + \gamma s_M}{r + \gamma s_2} - (x + \Delta) \geq 0$$

Using $k = s_1 / (s_2 - s_1)$ we can define the following function

$$U(k) \equiv \frac{r + \gamma s_M k}{r + \gamma (1 + k) s_1} - (x(k) + \Delta(k))$$

To prove 53 it is sufficient to show that $U(k) \geq 0$ for all $k \geq 0$. Note first that $k \to 0$ iff $s_1 \to 0$. Then,

$$\lim_{k \to 0} U(k) = 1 - \left(1 - \overline{\Delta} + \overline{\Delta}\right) = 0$$

where we used 51. By 50, we have $\lim_{k \to \infty} U(k) > 0$. To show that $U'(k)$ is positive for all k we need that

$$s_M (r + \gamma(1 + k)s_1) > s_1 (r + \gamma s_M k)$$

which holds true since $s_M > s_1$. QED

Proof of Lemma 6

The welfare formula reflects the joint welfare of four groups: exchange owners, dropout investors, active traders in the slow market and active traders in the fast market. Transfers from investors to exchange owners do not represent net social gains and hence are not reflected in 38. When dropout investors join market i, their before-fees gains are independent

of their types and equal to

$$W\left(\sigma, \hat{\sigma}_i, s_i\right) - W_{out} = \frac{1}{r} s_i \bar{a} \hat{\sigma}_i$$

The total mass of these investors equals $\left(\frac{1}{2\bar{a}} - 1\right)\left(1 - G\left(\hat{\sigma}_1\right)\right)$, and under free competition a mass equal to $\left(\frac{1}{2\bar{a}} - 1\right)\left(1 - G\left(\hat{\sigma}_{12}^{seg}\right)\right)$ joins the fast market. Thus, the social gains for this group in the free and protected cases are given by

$$\frac{\bar{a}}{r}\left(\frac{1}{2\bar{a}} - 1\right)\left[\left(G\left(\hat{\sigma}_{12}\right) - G\left(\hat{\sigma}_1\right)\right) s_1 \hat{\sigma}_1^{free} + \left(1 - G\left(\hat{\sigma}_{12}\right)\right) s_2^{free} \hat{\sigma}_{12}^{free}\right] \tag{54}$$

$$\frac{\bar{a}}{r}\left(\frac{1}{2\bar{a}} - 1\right)\left[\left(1 - G\left(\hat{\sigma}_1\right)\right) s_1 \hat{\sigma}_1^{prot}\right] \tag{55}$$

Using Proposition 1, the welfare of active investors in the slow market is given by

$$\int_{\hat{\sigma}_1^R}^{\hat{\sigma}_{12}^R} \left[\frac{s_1}{r}\left(\bar{a} - \frac{1}{2}\right)\hat{\sigma}_1^R + \frac{s_1}{2r}\sigma\right] dG\left(\sigma\right) \tag{56}$$

The welfare of active investors in the fast market under free and protected prices are given by

$$\int_{\hat{\sigma}_{12}^{free}}^{\bar{\sigma}} \left[\frac{s_2^{free}}{r}\left(\bar{a} - \frac{1}{2}\right)\hat{\sigma}_2^{free} + \frac{s_2^{free}}{2r}\sigma\right] dG\left(\sigma\right) \tag{57}$$

$$\int_{\hat{\sigma}_{12}^{prot}}^{\bar{\sigma}} \left[\frac{s_1}{r}\left(\bar{a} - \frac{1}{2}\right)\hat{\sigma}_1^{prot} + \frac{s_2^{prot}}{2r}\sigma\right] dG\left(\sigma\right) \tag{58}$$

Adding up 54, 56 and 57 yields gross social welfare under free competition. Similarly, adding up 55, 56 and 58 yields gross social welfare under price protection. Expression ?? is obtained by subtracting speed investment costs. The single speed market equation is a particular case. QED

Proof of Proposition 8

<u>Step 1: Finding net social value of speed investment.</u> Endogenous speed choice increases social welfare anytime that $\mathcal{W}_{seg} - \mathcal{W}_{Bertrand} > 0$. Under A1 we then require that

$$(s_2 - s_1) \int_{\hat{\sigma}_{12}^{free}}^{\bar{\sigma}} \sigma e^{\frac{-\sigma}{\nu}} d\sigma - s_1 \int_{\hat{\sigma}_1^{free}}^{\bar{\sigma}} \sigma e^{\frac{-\sigma}{\nu}} d\sigma - 2r\nu C(s_2) > 0$$

Computing the integrals, we have

$$(s_2 - s_1) \left(\hat{\sigma}_{12}^{free} + \nu \right) e^{\frac{-\hat{\sigma}_{12}^{free}}{\nu}} - s_1 \left(\hat{\sigma}_1^{free} + \nu \right) e^{\frac{-\sigma}{\nu}} - \nu - 2r\nu C(s_2) > 0$$

Dividing by νs_1 and using $x \equiv \frac{\hat{\sigma}_1}{\nu}, \Delta \equiv \frac{\hat{\sigma}_{12} - \hat{\sigma}_1}{\nu}, k \equiv s_1 / (s_2 - s_1)$, we can write the LHS of the above inequality as follows

$$R(k) \equiv \frac{1}{k} (\Delta + x + 1) e^{-(\Delta + x)} - (x + 1) e^{-x} - 1 - 2rC(s_2) \tag{59}$$

The net social value of speed investments is positive anytime that $R(k) > 0$.

<u>Step 2: $R'(k) < 0$</u>

Differentiating 59 we have that $R'(k) < 0$ iff

$$-e^{-\Delta} \left[(\Delta + x)(1 + \Delta' + x') + 1 \right] + xx' < 0$$

Re-arranging

$$(\Delta + x)(\Delta' + 1) + (1 + \Delta x') > xx' \left(e^{\Delta} - 1 \right)$$

Differentiating 41-42 we find that

$$x'(k) = \frac{-x \left(e^{\Delta}(1 - x) + 1 \right)}{e^{\Delta}(1 + \Delta) + k(1 + e^{\Delta})} \tag{60}$$

$$\Delta'(k) = \frac{-x \left(e^{\Delta} - 1 \right)}{e^{\Delta}(1 + \Delta) + k(1 + e^{\Delta})} \tag{61}$$

Thus, $x(k)$ and $\Delta(k)$ are decreasing functions. The RHS is thus negative. The sign of the

LHS depends on the expressions $(\Delta' + 1)$ and $(1 + \Delta x')$. Using 60 and 61 we have

$$\Delta' + 1 = \frac{e^{\Delta}\left(1 + \Delta + k - x\right) + k + x}{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right)} > 0$$

and

$$1 + \Delta x' = \frac{e^{\Delta}\left(1 + k + \Delta\left(1 - x(1 - x)\right)\right) + k - \Delta x}{e^{\Delta}\left(1 + \Delta\right) + k\left(1 + e^{\Delta}\right)} > 0$$

Thus, $R'(k) < 0$

Step 3. Verifying the claim

Note that the inequality is always satisfied when $s_1 \to 0$ ($k \to 0$) since $R(k) \to \infty$. When $s_1 \to 1$, for any solution $s_2 > s_1$ where the fast venue is active, we know from the proof of proposition 5 that $\lim_{k \to \infty} x(k) = 0$. Then,

$$\lim_{k \to \infty} R(k) = -2 - 2rC\left(s_2\right) < 0$$

Consequently, we found that $\lim_{k \to \infty} R(k) < 0$ and $\lim_{k \to 0} R(k) \to \infty$. Since R is a continuous function, by the intermediate value theorem there is a number $\overline{k} > 0$ such that $R(\overline{k}) = 0$. Since R is monotonically decreasing, $\overline{k}$ is unique. The, for any $s_2 > s_1$, $\overline{s}_1$ is given by $\overline{s}_1 = \frac{s_2 \overline{k}}{1 + \overline{k}}$. QED

Proof of Proposition 9

Using Lemma 6, the gains of competition are given by

$$\mathcal{W}_{comp} - \mathcal{W}_M - \kappa = \frac{s_1}{2r} \int_{\hat{\sigma}_1}^{\hat{\sigma}_{12}} \sigma dG\left(\sigma\right) + \frac{s_2}{2r} \int_{\hat{\sigma}_{12}}^{\overline{\sigma}} \sigma dG\left(\sigma\right) - \frac{s_M}{2r} \int_{\hat{\sigma}_M}^{\overline{\sigma}} \sigma dG\left(\sigma\right)$$
$$- C\left(s_2\right) + C\left(s_M\right) - \kappa$$

Using A1, computing the integrals, and re-arranging, we have that the gains from competition are equal to zero if and only if

$$\frac{s_1}{2r}e^{-\frac{\hat{\sigma}_1}{\nu}}\left(\hat{\sigma}_1+\nu\right)+\frac{(s_2-s_1)}{2r}e^{-\frac{\hat{\sigma}_{12}}{\nu}}\left(\hat{\sigma}_{12}+\nu\right)-\frac{\nu s_M}{re}= \tag{62}$$
$$\left(\overline{\kappa}+C\left(s_2\right)-C\left(s_M\right)\right)$$

which yields $\overline{\kappa}$. Note that using 24 under A1 we have

$$\pi_1^{comp}=\frac{s_1}{2r}\hat{\sigma}_1\left(e^{-\frac{\hat{\sigma}_1}{\nu}}-e^{-\frac{\hat{\sigma}_{12}}{\nu}}\right)$$
$$\pi_2^{comp}=\frac{1}{2r}e^{-\frac{\hat{\sigma}_{12}}{\nu}}\left(\hat{\sigma}_1 s_1+\hat{\sigma}_{12}\left(s_2-s_1\right)\right)-C(s_2)$$
$$\pi_M=\frac{\nu s_M}{2re}-C\left(s_M\right)$$

Re-arranging 62, and using the profit functions above, we obtain the following expression

$$\pi_1^{comp}-\overline{\kappa}=-\pi_2^{comp}-\frac{\nu}{\hat{\sigma}_1}\pi_1^{comp}+\pi_M-\frac{\nu}{2r}s_2 e^{-\frac{\hat{\sigma}_{12}}{\nu}} \tag{63}$$

We from the proof of 5 that when $\underline{s}\to 1$, $\hat{\sigma}_1\to 0$ and $\hat{\sigma}_{12}\to\nu\underline{\Delta}$. Thus, π_1^{comp} and π_2^{comp} converge to zero and we have from 63 that $\overline{\kappa}$ approaches $\frac{\nu}{2r}(1-1/e)$. QED

Proof of Proposition 10

In general, its objective function is

$$\max_{s_2,q_1,q_2}\frac{s}{2r}\int_{\sigma_1}^{\sigma_{12}}\sigma dG\left(\sigma\right)+\frac{s_2}{2r}\int_{\sigma_{12}}^{\overline{\sigma}}\sigma dG\left(\sigma\right)-C\left(s_2\right)$$

and the marginal types are given by 21 and 23, so we have

$$q_1=s_1\frac{\sigma_1}{2r},$$
$$q_2=\left(s_2-s_1\right)\frac{\sigma_{12}}{2r}+q_1.$$

The break-even constraint is $q_2\left(1-G\left(\hat{\sigma}_{12}\right)\right)\geq C\left(s_2\right)$, so the Lagrangian (scaled by $2r$) is

$$\mathcal{L} = \underline{s} \int_{\sigma_1}^{\overline{\sigma}} \sigma dG(\sigma) + (s - \underline{s}) \int_{\sigma_{12}}^{\overline{\sigma}} \sigma dG(\sigma) - 2rC(s) + \lambda \left\{ ((s - \underline{s}) \sigma_{12} + \underline{s}\sigma_1)(1 - G(\sigma_{12})) - 2rC(s) \right\}$$

and the FOCs for affiliations are

$$\sigma_1^* g(\sigma_1^*) = \lambda,$$

$$\sigma_{12}^* g(\sigma_{12}^*) = \frac{\lambda}{1 + \lambda} \left(1 - G(\sigma_{12}^*) - \frac{\underline{s}}{s - \underline{s}} g(\sigma_{12}^*) \sigma_1^* \right).$$

Optimal speed satisfies

$$2r \frac{\partial C}{\partial \rho}(s^*) = \frac{1}{1 + \lambda} \int_{\sigma_{12}^*}^{\overline{\sigma}} \sigma dG(\sigma) + \frac{\lambda}{1 + \lambda} (1 - G(\sigma_{12}^*)) \sigma_{12}^*,$$

and the break-even constraint is simply $2rC(s^*) = (1 - G(\sigma_{12}))((s - \underline{s})\sigma_{12}^* + \underline{s}\sigma_1^*)$. From the first two FOCs it is immediate that $\sigma_1^* > \sigma_{12}^*$, which is inconsistent with our assumption that market 1 is active. We conclude that there must be a single venue.

This result can be extended to the case where the planner operates the two venues with one budget constraint. In this case, the constraint is $(G(\hat{\sigma}_{12}) - G(\hat{\sigma}_1)) q_1 + (1 - G(\hat{\sigma}_{12})) q_2 > C(s_2)$ and the Lagrangian is

$$\mathcal{L} = \underline{s} \int_{\sigma_1}^{\overline{\sigma}} \sigma dG(\sigma) + (s - \underline{s}) \int_{\sigma_{12}}^{\overline{\sigma}} \sigma dG(\sigma) - 2rC(s) + \lambda \left((1 - G(\sigma_1)) \underline{s}\sigma_1 + (1 - G(\sigma_{12}))(s - \underline{s}) \sigma_{12} - 2rC(s) \right)$$

and the FOCs for affiliations are

$$1 - G(\sigma_1^*) = g(\sigma_1^*) \frac{1 + \lambda}{\lambda} \sigma_1^*$$

$$1 - G(\sigma_{12}^*) = g(\sigma_2^*) \frac{1 + \lambda}{\lambda} \sigma_{12}^*$$

Optimal speed satisfies the same equation as before. In this case, we see that $\sigma_1^* = \sigma_{12}^*$, market 1 is still inactive.

Proof of Proposition 11

With one venue, the Lagrangian is

$$\mathcal{L} = s \int_{\hat{\sigma}}^{\overline{\sigma}} \sigma dG(\sigma) - 2rC(s) + \lambda \left(s\hat{\sigma} \left(1 - G(\hat{\sigma}) \right) - 2rC(s) \right)$$

From the previous section, it is immediate that

$$1 - G(\sigma^*) = g(\sigma^*) \frac{1 + \lambda}{\lambda} \sigma_1^*$$

Since the monopoly solution is $\frac{1-G(\sigma_M)}{g(\sigma_M)} = \sigma_M$, it is clear that $\sigma_M > \sigma^*$.

Regarding speed, the planner chooses

$$2r \frac{\partial C}{\partial \rho}(s^*) = \frac{1}{1 + \lambda} \int_{\sigma^*}^{\overline{\sigma}} \sigma dG(\sigma) + \frac{\lambda}{1 + \lambda} \left(1 - G(\sigma^*) \right) \sigma^*,$$

while the monopoly chooses $2r\frac{\partial C}{\partial s}(s_M) = (1 - G(\sigma_M)) \sigma_M$. If $\lambda = 0$, it is clear that $s^* > s_M$, as expected. However, when the break-even constraint binds, the comparison is ambiguous. We now provide a simple example to show that it is indeed possible for the monopoly to over-invest in speed.

Counter-example

Consider a binary distribution. High $\sigma^H = \bar{\sigma}$ with population share n. Low sigma $\sigma^L = \alpha\bar{\sigma}$ with $\alpha < 1$ and population share $1 - n$. Cost function $2rC = \frac{c}{2}s^2$. The marginal price is $q^i = \rho\sigma^i$. The monopoly has two choices:

- Set price to $\rho\alpha\bar{\sigma}$, get everyone to participate, then $\pi = \rho\alpha\bar{\sigma} - c(s)$.

- Set high price $\rho\bar{\sigma}$, only high types participate, then $\pi = \rho n\bar{\sigma} - c(s)$.

The monopoly chooses high speed low participation if and only if $n > \alpha$. The speed choice is $\max(n, \alpha) \bar{\sigma}/c$.

The Planner has two main choices. If all participate $W = \rho\bar{\sigma}((1 - n)\alpha + n) - c(s)$. Then it depends on whether the break-even constraint binds. If it does not, then the planner chooses a higher speed than any monopoly: $s^* = \frac{\bar{\sigma}((1-n)\alpha+n)}{c}$. The break-even constraint

69

binds if $s^* \alpha \bar{\sigma} < c(s^*)$, which is equivalent to $cs > 2\alpha\bar{\sigma} \Leftrightarrow (1 - n)\alpha + n > 2\alpha \Leftrightarrow \alpha < n(1 - \alpha)$. The planner can still choose full participation, but at limit price $c(s) = s\alpha\bar{\sigma} \Leftrightarrow s = \frac{2\alpha\bar{\sigma}}{c}$. Then welfare is $W = s\bar{\sigma}n(1 - \alpha) = \frac{2}{c}(\bar{\sigma})^2 n\alpha(1 - \alpha)$.

The other choice for the planner is that only high type participate. This is same program as monopoly. Speed choice is $n\bar{\sigma}/c$. Welfare is $\frac{1}{2c}(n\bar{\sigma})^2$. The Planner chooses low speed high participation iff $\frac{2}{c}(\bar{\sigma})^2 n\alpha(1 - \alpha) > \frac{1}{2c}(n\bar{\sigma})^2$ or $4\alpha(1 - \alpha) > n$.

To summarize, for the planner to choose lower speed than monopoly, we need: (i) $n > \alpha$ so monopoly goes for high speed low participation; (ii) $4\alpha(1 - \alpha) > n$ so planner chooses high participation; (iii) $\alpha < n(1 - \alpha)$ so break-even violated; and (iv) $n\bar{\sigma}/c > \frac{2\alpha\bar{\sigma}}{c} \Leftrightarrow n > 2\alpha$ so monopoly speed indeed higher. It is easy to see that (i) is not binding. So we have the three following conditions

1. $4\alpha(1 - \alpha) > n$

2. $\alpha < \frac{n}{1+n}$

3. $n > 2\alpha$

Take $n = 1/4$ then we need $\alpha < 1/8$ for third, second is not binding, and it is easy to find a solution for the first. QED.

Proof of Lemma 7

In a frictionless competitive market we have maximum investor participation. Thus, the marginal type is given by

$$\frac{G(\hat{\sigma}_w)}{1 - G(\hat{\sigma}_w)} = \frac{1}{2\bar{a}} - 1$$

Using A1 we obtain $\hat{\sigma}_w = -\nu \log(2\bar{a})$, which combined with $s_w = 1$ and 34 yields $p_w = \frac{1}{r}[\mu - \nu \log(2\bar{a})]$. With $s_w = 1$ the instantaneous transaction rate ?? becomes

$$\tau_w = \frac{\gamma}{4}(1 - G(\hat{\sigma}_w)) = \frac{\gamma}{4}\left(e^{-\frac{\hat{\sigma}_w}{\nu}}\right) = \frac{\gamma\bar{a}}{2}$$

By Lemma 6, social welfare is given by

$$
\begin{aligned}
\mathcal{W}_w &= \frac{\overline{a}}{r} \int_{\hat{\sigma}_w}^{\overline{\sigma}} \sigma dG\left(\sigma\right) \\
&= \frac{\overline{a}}{r\nu} \int_{\hat{\sigma}_w}^{\infty} \sigma e^{-\frac{\hat{\sigma}_w}{\nu}} d\sigma = \frac{\overline{a}}{r}\nu\left(1 - \log\left(2\overline{a}\right)\right)
\end{aligned}
$$

Note that when $\overline{a} = 1/2$, Walrasian social welfare is simply given by $\frac{\nu}{2r}$. QED.

Figure 8: Equilibrium Outcomes and Technology cost

Equilibrium outcomes when asset supply $\overline{a}$ equals 0.5. In this case price regulation is immaterial. The labels are as follows. "Consolidated" correspond to a single monopolist venue. "Two-venues" correspond to a speed-differentiated duopoly. "Planner" correspond to the constrained-efficient solution in section 5. In all cases entry costs κ are zero. Parameter choices are described in section 6.
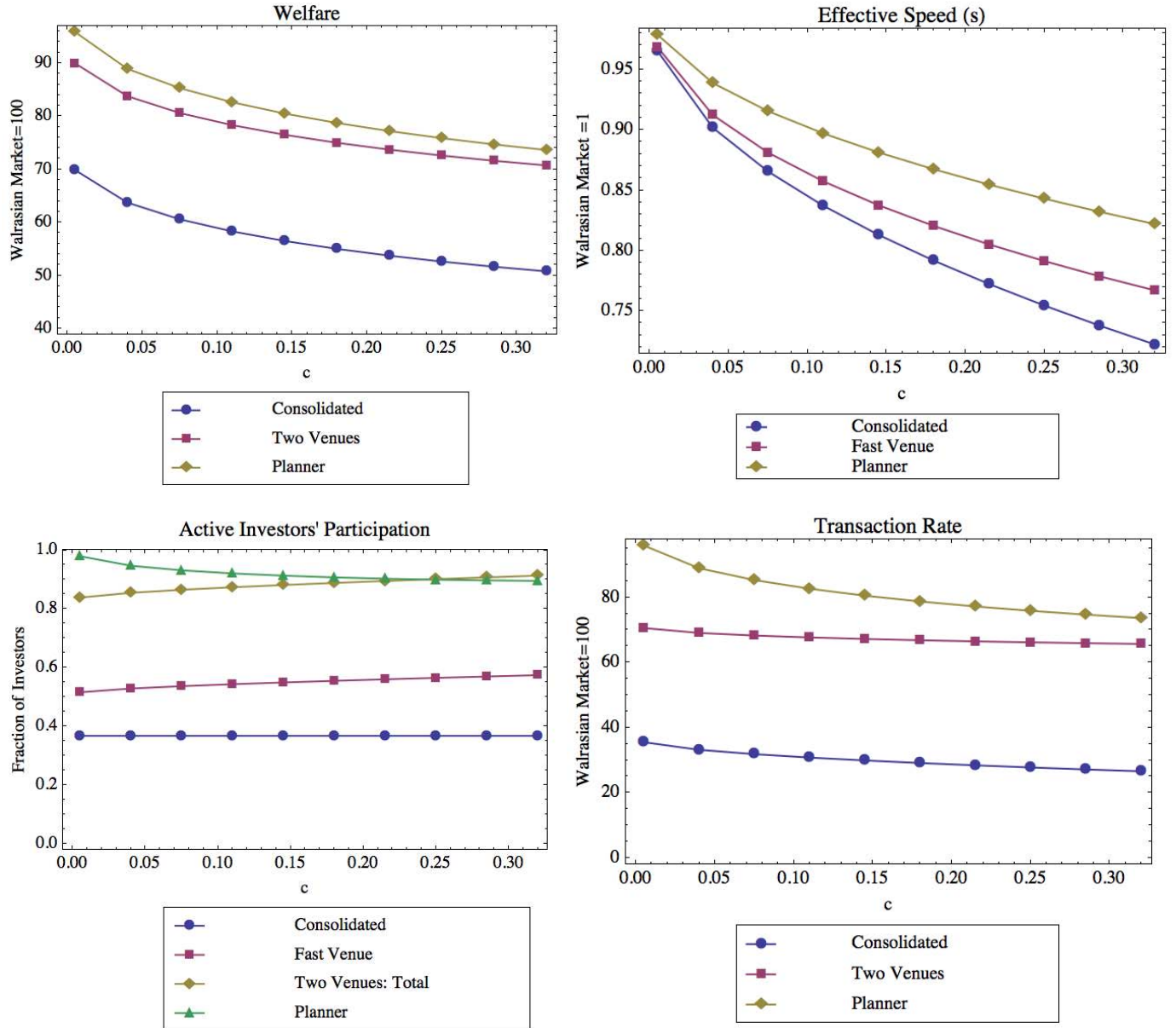
Figure 9: Equilibrium Outcomes, Technology Cost and Price Regulation

Equilibrium outcomes when asset supply $\overline{a}$ equals 0.45. In this case price regulation matters. The labels are as follows. "Free Competition" corresponds to the case of segmented markets, with two asset prices. "Price Protection" corresponds to the case with a single asset price in all venues. "Free: Total" denotes total investor participation under free competition. In all cases entry costs κ are zero. Parameter choices are described in section 6.
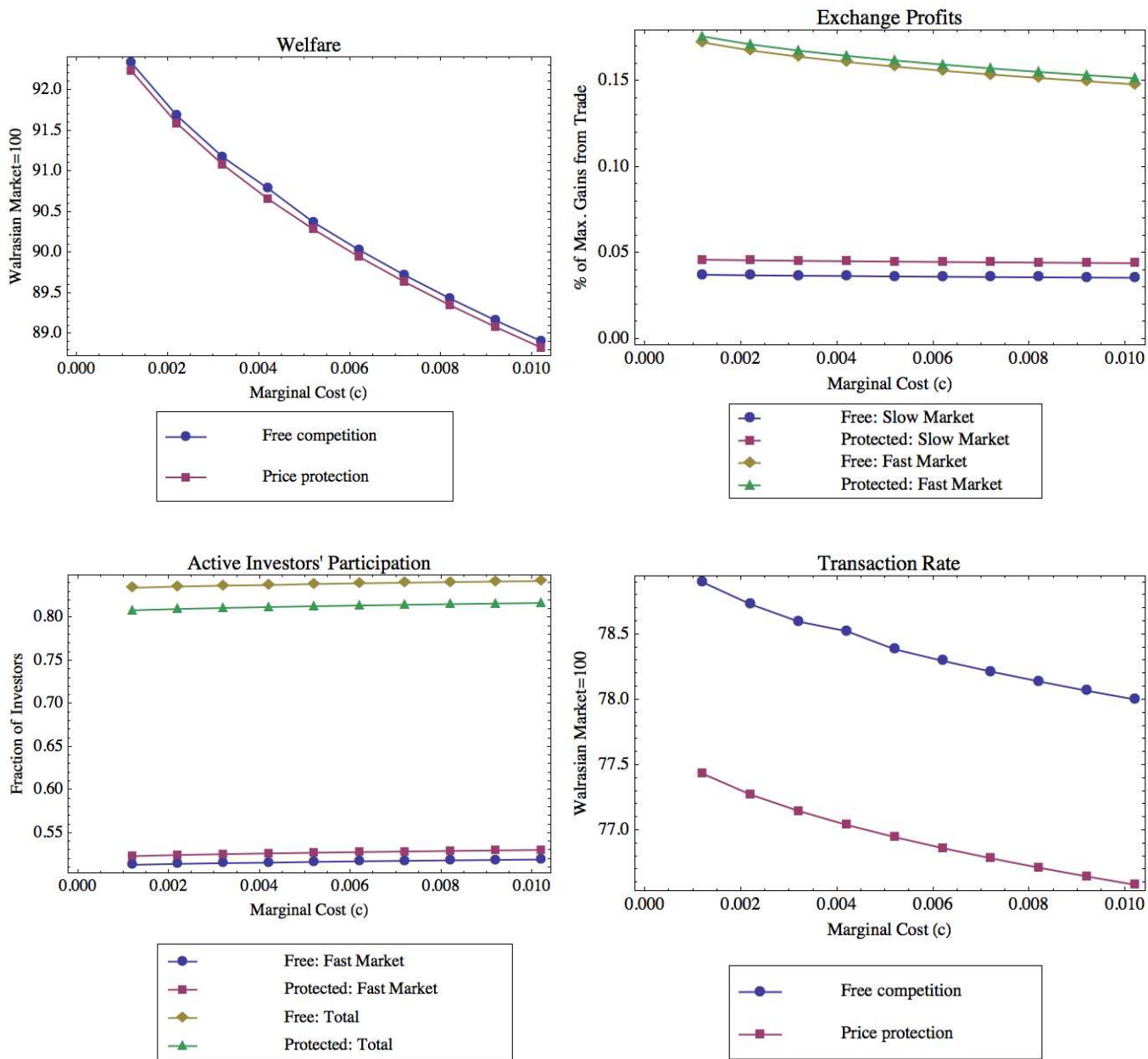
Figure 10: Equilibrium Asset Prices

Equilibrium asset prices when asset supply $\overline{a}$ equals 0.45. In this case price regulation matters. The labels are as follows. "Consolidated" correspond to a single monopolist venue. "Free: VWAP" corresponds to the volume-weighted average price under free segmentation. "Price Protection" denotes the unique asset price under price protection. "Planner" corresponds to the equilibrium price in the constrained-efficient solution in section 5. Parameter choices are described in section 6.