Exchange Competition, Entry, and Welfare*

Giovanni Cespa† and Xavier Vives‡

June 2019

Abstract

We integrate a market microstructure model with an exchange competition model with entry in which exchanges supply technological services that enhance market participation, and have market power. We find that technological services can be strategic substitutes or complements in platform competition. Free entry of platforms delivers a superior outcome in terms of liquidity and (generally) welfare compared to the case of an unregulated monopoly. Controlling entry or, even better, platform fees may serve as an instrument to limit market power, further increasing welfare. The market may deliver excessive or insufficient entry.

Keywords: Market fragmentation, welfare, endogenous market structure, platform competition, Cournot with free entry, industrial organization of exchanges.
JEL Classification Numbers: G10, G12, G14

*We thank Fabio Braga, Eric Budish, Thierry Foucault, Jiasun Li, Albert Menkveld, Duane Seppi, René Stultz, Felix Suntheim, and seminar participants at ICEF (Moscow), Universidad Carlos III (Madrid), Imperial College (London), Chicago Fed, University of Chicago, Northwestern, the 2018 EFA meeting (Warsaw), and the 2019 AFA meeting (Atlanta) for useful comments and suggestions. Cespa acknowledges financial support from the Bank of England (grant no. RDC151391). Vives acknowledges the financial support of the Ministry of Economy and Competitiveness - ECO2015-63711-P (MINECO/FEDER/UE) and of the Department of Economy and Knowledge of the Generalitat de Catalunya (Ref. 2017 SGR 1244). Orestis Vravosinos provided excellent research assistance.

†Cass Business School–City, University of London, and CEPR. 106, Bunhill Row, London EC1Y 8TZ, UK. e-mail: `giovanni.cespa@gmail.com`

‡IESE Business School, Avinguda Pearson, 21 08034 Barcelona, Spain.

"We are now living in a much different world, where many are questioning whether the pendulum has swung too far and we have too many venues, creating unnecessary complexity and costs for investors." Mary Jo White, Economic Club of New York, June 2014.

"The cost of market data and exchange access has been a cause of debate and concern for the industry for many years, and those concerns have grown as these costs have risen dramatically in the last several years [...] Exchanges also have been able to charge more for the data center connections [...] since they control access at the locations where the data is produced." Brad Katsuyama, U.S. House of Representatives Committee on Financial Services, June 2017.

1 Introduction

Over the past two decades, governments and regulators moved to foster competition among trading venues, leading to an increase in market fragmentation. However, there is now a concern that the entry of new platforms may have been excessive, and that exchanges exercise too much market power in the provision of technological services. In this paper we show that the move from monopoly to competition has increased liquidity and the welfare of market participants but that the market does not deliver a (constrained) efficient outcome. We characterize how structural and conduct regulation of exchanges has the potential to improve welfare.

The profit orientation of exchanges, when they converted into publicly listed companies, led to regulatory intervention both in the US (Reg NMS in 2005) and the EU (Mifid in 2007), to stem their market power in setting fees. Regulation, together with the removal of barriers to international capital flows and technological developments, led in turn to an increase in fragmentation and competition among trading platforms. Incumbent exchanges such as the NYSE reacted to increased competition by upgrading technology (e.g, creating, NYSE Arca), or merging with other exchanges (e.g., the NYSE merged with Archipelago in 2005 and with Euronext in 2007).[1]

As a result, the trading landscape has changed dramatically. On the one hand, large-cap stocks nowadays commonly trade in multiple venues, a fact that has led to an inexorable decline in incumbents' market shares, giving rise to a "cross-sectional" dimension

[1]See Foucault et al. (2013), Chapter 1.

of market fragmentation (see Figure 1). The automation of the trading process has also spurred fragmentation along a "time-series" dimension, in that some liquidity providers' market participation is limited (Duffie (2010), SEC (2010)), endogenous (Anand and Venkataraman (2015)), or impaired because of the existence of limits to the access of reliable and timely market information (Ding et al. (2014)).[2] On the other hand, trading fees have declined to competitive levels (see, e.g., Foucault et al. (2013), Menkveld (2016), and Budish et al. (2017)), and exchanges have steered their business models towards the provision of technological services (e.g., proprietary data, and co-location space).[3]

Such a paradigm shift has raised a number of concerns highlighted by the fact that even though there are 13 lit stock venues in the US (and 30 alternative ones), 12 of them, which account for two-thirds of daily trading, are controlled by three major players: Intercontinental Exchange, Nasdaq, and CBOE.[4] Indeed, market participants allege that exchanges exercise market power in the provision of technological services.[5] Additionally, regulators and policy makers such as the SEC and the antitrust authorities have also expressed concern about the existence of potential monopoly restrictions or excess entry.[6]

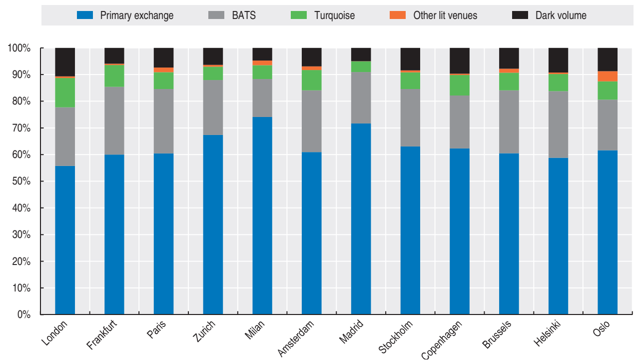
[2]Limited market participation of liquidity providers also arises because of shortages of arbitrage capital (Duffie (2010)) and/or traders' inattention or monitoring costs (Abel et al. (2013)).

[3]Increasing competition in trading services has squeezed the profit margins exchanges drew from traditional activities, leading them to gear their business model towards the provision of technological services (Cantillon and Yin (2011)). There is abundant evidence testifying to such a paradigmatic shift. For example, according to the Financial Times, "After a company-wide review Ms Friedman [Nasdaq CEO] has determined the future lies in technology, data and analytics, which collectively accounted for about 35 per cent of net sales in the first half of this year." (see, "Nasdaq's future lies in tech, data and analytics, says Nasdaq CEO" *Financial Times,* October 2017). Additionally, according to Tabb Group, in the US, exchange data, access, and technology revenues have increased by approximately 62% from 2010 to 2015 (Tabb Group, 2016).

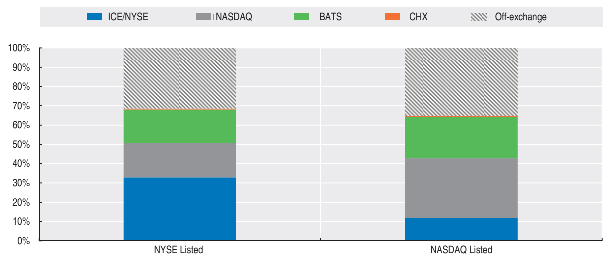
[4]See FT January 8, 2019, where it is also reported that large brokers and banks are setting MEMX a competing exchange to lower costs of trading.

[5]"'Information wants to be free,' the technology activist Stewart Brand once said. 'Information also wants to be expensive.' That is proving true on Wall Street, where stock exchanges–in particular the New York Stock Exchange and Nasdaq–both publicly traded and for-profit, stand accused by rivals and some users of unfairly increasing the price of market data." (*Business Insider,* November 2016). In December 2016 Chicago-based Wolverine Trading LLC stated to the SEC that its total costs related to NYSE equities market data had more than tripled from 2008 to 2016 ("This is a monopoly.")

[6]Responding to a NYSE request to change the fees it charges for premium connectivity services, the SEC in November 2016 stated: "The Commission is concerned that the Exchange has not supported its argument that there are viable alternatives for Users inside the data center in lieu of obtaining such information from the Exchange. The Commission seeks comment on whether Users do have viable alternatives to paying the Exchange a connectivity fee for the NYSE Premium Data Products." The SEC statement echoes industry concerns "'We are pleased that the Commission will be subjecting this incremental fee application to review,' Doug Cifu, the CEO of electronic trading firm Virtu [...] 'As we have repeatedly said we think exchange market data and connectivity fees have 'jumped the shark' as an excessive cost burden on the industry.'" (*Business Insider,* November 2016.) In October 2018, the



(a)



(b)

Figure 1: Market shares among trading venues in Europe (Panel (a)), and the US (Panel (b)). Source: OECD Business and Finance Outlook 2016.

The questions we want to address in this paper are the following: what is the character of platform competition in the supply of technological services? What is the impact of platform competition on the overall quality of the market and on the end users of trading services? If the market outcome is suboptimal, which regulatory tools are more effective? Entry controls (merger policy), or fee regulation?

We assess the consequences for market quality and the welfare of market participants of different exchanges' entry regimes and pricing policies in a context of limited market participation. To this end we propose a stylized framework that captures the above dimensions of market fragmentation and competition among trading venues, integrating a simple two-period, market microstructure model à la Grossman and Miller (1988), with one of platform competition with entry, featuring a finite number of exchanges competing to attract dealers' orders.

The microstructure model defines the *liquidity determination* stage of the game.

SEC did not approve the fee increase for data sought by the NYSE and Nasdaq. See also Okuliar (2014) on whether US competition authorities should intervene more in financial exchange consolidation.

There, two classes of risk averse dealers provide liquidity to two cohorts of rational liquidity traders, who sequentially enter the market. Depending on the structure of the market, at each round traders can submit their orders only to an "established" venue, or also to one of the competing venues. Dealers in the first class are endowed with a technology enabling them to act at both rounds, absorbing the orders of both liquidity traders' cohorts, and are therefore called 'full' (FD); those in the second class can only act in the first round, and are called 'standard' (SD). The possibility to trade in the two rounds captures in a simple way both the limited market participation of standard dealers, and FDs' ability to take advantage of short term return predictability. We assume that there is a best price rule ensuring that the second period price is identical across all the competing trading platforms. This is the case in the US where the combination of the Unlisted Trading Privilege (which allows a security listed on any exchange to be traded by other exchanges), and Regulation National Market System (RegNMS) protection against "trade-throughs," implies that, despite fragmentation, there virtually exists a unique price for each security.[7] We also assume that trading fees are set at the competitive level by the exchanges.[8]

The platform competition model features a finite number of exchanges which, upon incurring a fixed entry cost, offer "technological services" to the full dealers which allow them to trade in the second round. A standard dealer becomes full by paying a fee that reflects the incremental payoff he earns by operating in the second round.[9] This defines an inverse demand for technological capacity; upon entry, each exchange incurs a constant marginal cost to produce a unit of technological service capacity, receiving the corresponding fee from the attracted full dealers. This defines a Cournot game with free entry which represents the *technological capacity determination* stage of the game.

We now describe in more detail the main features of the model and findings. Due to their ability to trade in both rounds, full dealers exhibit a higher risk bearing capacity

[7]Price protection rules were introduced to compensate for the potential adverse effects of price fragmentation when the entry of new platforms was encouraged to limit market power of incumbents. In particular, RegNMS requires market centers to route orders at the top of the book to the trading platform that posts the best price, and exchanges to provide accessible electronic data about their price quotations. The aim is to enforce price priority in all markets. However, for large orders execution pricing may not be the same in all exchanges except if traders have in place cross-exchange order-routing technology. In Europe there is no order protection rule similar to RegNMS. Foucault and Menkveld (2008) show empirically the existence of trade-thoroughs in Amsterdam and London markets. Hendershott and Jones (2005) find that in the US price protection rules improve market quality.

[8]We abstract therefore from competition for order flow issues (see Foucault et al. (2013) for an excellent survey of the topic).

[9]Actually, FDs may have to invest on their own also on items such as speed technology. In our model we will abstract from such investments.

compared to standard dealers. As a consequence, an increase in their mass improves market liquidity. This has two countervailing effects on the welfare of market participants. On the one hand, it lowers the cost of trading and leads traders to hedge more aggressively, increasing their welfare. On the other hand, it hurts standard dealers who face a heightened competitive pressure, and experience a welfare reduction. As liquidity demand augments for both dealers' classes, however, SDs effectively receive a *smaller* share of a *larger* pie. This mitigates the negative impact of increased competition on dealers, implying that on balance the increased liquidity benefits in the aggregate first period market participants other than FDs. In turn, this contributes to make *gross* welfare (i.e., the weighted sum of all market participants' welfare) increasing in the proportion of full dealers, implying that liquidity becomes a measurable welfare indicator.

An important feature of the platform competition stage of the model is that dealers' demand for technological services is log-convex for a wide range of deep parameters. Intuitively, when the proportion of full dealers in the market is small, the margin from acquiring the technology to participate in the second round of trade is way larger than in the polar case when the market is almost exclusively populated by full dealers. Thus, an increase in the proportion of full dealers yields a price reduction which becomes increasingly smaller. We show that this has important implications for the nature of exchange competition. In particular, when two platforms are in the market and their marginal costs are small, strategic complementarities in the supply of technological services arise. Hence, a shock that lowers technology costs can prompt a strong response in technological capacity. Furthermore, log-convexity of the demand function can lead a monopoly platform to step up its technological capacity in the face of an entrant. This magnifies the positive impact of an increase in the number of competing platforms on the aggregate technological service capacity. Given that at equilibrium the latter matches the proportion of full dealers, this in turn amplifies the positive liquidity and welfare impact of heightened platform competition.

An insight of our analysis is that technological services can be viewed as an essential intermediate input in the "production" of market liquidity. This warrants a welfare analysis of the impact of platform competition, which is the subject of the last part of the paper. There, we use our setup to compare the market solution arising with no platform competition (monopoly), and with entry (Cournot free entry), with three different planner solutions which vary depending on the restrictions faced by the planner. An unrestricted planner attains the First Best by choosing the number of competing exchanges as well as the industry technological service fee; a planner who can only regulate the technological service fee but not entry, achieves the Behavioral Second Best; finally, if the planner is unable to affect the way in which exchanges compete but can set the number of exchanges who can enter the market, she achieves the Structural Second Best solution (restricted or unrestricted, depending on whether the planner regulates entry making sure that platforms break even or not).

Insulated from competition, a monopolistic exchange seeks to restrict the supply of technological services to increase the fees it extracts from FDs.[10] Thus, the market at a free entry Cournot equilibrium delivers a superior outcome in terms of liquidity and (generally) welfare. However, compared to the case in which the regulator can control entry, the market solution can feature excessive or insufficient entry. Indeed, in the absence of regulation, an exchange makes its entry decision without internalizing the profit reduction it imposes on its competitors. This "profitability depression" effect is conducive to excessive entry.[11] As new platform entry spurs liquidity, however, it also has a positive "liquidity creation" effect which can offset the profitability depression effect, and lead to insufficient entry. Entry regulation is however typically inferior compared to the alternative of regulating the technological service fee charged by a monopolistic exchange. This is because in this case the planner generally minimizes the setup cost borne by the industry and can more effectively limit market power by forcing the monopolistic exchange to charge the lowest possible technological service fee that is compatible with a break-even condition.

Overall, our analysis suggests that fee regulation achieves the highest and cheapest provision of technological services. However, this form of intervention is subject to rent-seeking efforts by market participants, which indicates that entry regulation can at times work as a simpler alternative instrument.[12] Indeed, spurring entry achieves two objectives. First, it works as a corrective against exchanges' market power in the provision of technological services; additionally, by creating competitive pressure, it achieves the objective of keeping exchanges' trading fees in check.

In the last part of the paper, we use our model to investigate the effects of platforms'

[10]In a similar vein, Cespa and Foucault (2014) find that a monopolistic exchange finds it profitable to restrict the access to price data, to increase the fee it extracts from market participants.

[11]This effect is similar to the "business stealing" effect highlighted by the Industrial Organization literature (see, e.g., Mankiw and Whinston (1986)). Note, however, that business stealing refers to the depressing impact that a firm entry has on its competitors' output. In our context, this effect is not warranted: due to strategic complementarity, heightened competitive pressure can lead an exchange to respond by installing more capacity.

[12]The evidence presented in footnote 6 suggests that regulators' ability to weigh on the technological fee-setting process is far from perfect.

technological capacity decisions on liquidity provision. As usual in models with risk averse dealers, a reduction in risk tolerance reduces liquidity. However, insofar as it penalizes comparatively more SDs than FDs, the same can also boost the demand for technological services, potentially leading to an increase in its supply, which, in some of our simulations is strong enough to produce a liquidity improvement.

Our paper is related to the literature on the welfare effects of platform competition, and investment in technological capacity. Pagnotta and Philippon (2018), consider a framework where trading needs arise from shocks to traders' marginal utilities from asset holding, yielding a preference for different trading speeds. In their model, venues vertically differentiate by speed, with faster venues attracting more speed sensitive investors and charging higher fees. This relaxes price competition, and the market outcome is inefficient. The entry welfare tension in their case is between business stealing and quality (speed) diversity, like in the models of Gabszewicz and Thisse (1979) and Shaked and Sutton (1982). In this paper, as argued above, the welfare tension arises instead from the profitability depression and liquidity creation effects associated with entry.[13] Biais et al. (2015) study the welfare implications of investment in the acquisition of High Frequency Trading (HFT) technology. In their model HFTs have a superior ability to match orders, and possess superior information compared to human (slow) traders. They find excessive incentives to invest in HFT technology, which, in view of the negative externality generated by HFT, can be welfare reducing. Budish et al. (2015) argue that HFT thrives in the continuous limit order book, which is however a flawed market structure since it generates a socially wasteful arms' race to respond faster to (symmetrically observed) public signals. The authors advocate a switch to Frequent Batch Auctions (FBA) instead of a continuous market. Budish et al. (2017), introduce exchange competition in Budish et al. (2015) and analyze whether exchanges have enough incentives to implement the technology required to run FBA. Also building on Budish et al. (2015), Baldauf and Mollner (2017) show that heightened exchange competition has two countervailing effects on market liquidity, since it lowers trading fees, but magnifies the opportunities for cross-market arbitrage, increasing adverse selection.

Our paper is also related to the literature on the Industrial Organization of securities' trading. This literature has identified a number of important trade-offs due to competition among trading venues. On the positive side, platform competition exerts a beneficial impact on market quality because it forces a reduction in trading fees (Foucault and Menkveld (2008) and Chao et al. (2017)), and can lead to improvements in margin

[13]Pagnotta and Philippon (2018) also study the market integration impact of RegNMS.

requirements (Santos and Scheinkman (2001)); furthermore, it improves trading technology and increases product differentiation, as testified by the creation of "dark pools." On the negative side, higher competition can lower the "thick" market externalities arising from trading concentration (Chowdhry and Nanda (1991) and Pagano (1989)), and increase adverse selection risk for market participants (Dennert (1993)). We add to this literature, by pointing out that market incentives may be insufficient to warrant a welfare maximizing solution. Indeed, heightened competition can lead to the entry of a suboptimal number of trading venues, because of the conflicting impact of entry on profitability and liquidity.

The rest of the paper is organized as follows. In the next section, we outline the model. We then turn our attention to study the liquidity determination stage of the game. In section 4, we analyze the payoffs of market participants, and the demand and supply of technological services. We then concentrate on the impact of platform competition with free entry, and contrast the welfare and liquidity effects of different regulatory regimes. A final section contains concluding remarks.

2 The model

A single risky asset with liquidation value $v \sim N(0, \tau_v^{-1})$, and a risk-less asset with unit return are exchanged during two trading rounds.

Three classes of traders are in the market. First, a continuum of competitive, risk-averse, "Full Dealers" (denoted by FD) in the interval $(0, \mu)$, who are active at both rounds. Second, competitive, risk-averse "Standard Dealers" (denoted by SD) in the interval $[\mu, 1]$, who instead are active only in the first round. Finally, a unit mass of traders who enter at date 1, taking a position that they hold until liquidation. At date 2, a new cohort of traders (of unit mass) enters the market, and takes a position. The asset is liquidated at date 3. We now illustrate the preferences and orders of the different players.

2.1 Trading venues

The organization of the trading activity depends on the competitive regime among venues. With a monopolistic exchange, both trading rounds take place on the same venue. When platforms are allowed to compete for the provision of technological services, we assume that a best price rule ensures that the price at which orders are executed is the same across all venues: trading can seamlessly occur on each venue at a unique price at both trading rounds. We thus assume away "cross-sectional" frictions, implying that we have a virtual single platform where all exchanges provide identical access to trading, and stock prices are determined by aggregate market clearing.[14]

We model trading venues as platforms that prior to the first trading round (date 0), supply technology which offers market participants the possibility to trade in the second period. For example, it is nowadays common for exchanges to invest in the supply of co-location facilities which they rent out to traders to store their servers and networking equipment close to the matching engine; additionally, platforms invest in technologies that facilitate the distribution of market data feeds. In the past, when trading was centralized in national venues, exchanges invested in real estate and the facilities that allowed dealers and floor traders to participate in the trading process.

At date $t = -1$ trading venues decide whether to enter and if so they incur a fixed cost. Suppose that there are N entrants, that each venue $i = 1, 2, \ldots, N$ produces a technological service capacity μ_i, and that

$$\sum_{i=1}^{N} \mu_i = \mu, \tag{1}$$

so that the proportion of FDs coincides with the total technological service capacity offered by the platforms. Consistently with the evidence discussed in the introduction (see also Menkveld (2016)), we assume that trading fees are set to the competitive level.

2.2 Liquidity providers

A FD has CARA preferences, with risk-tolerance γ, and submits price-contingent orders x_t^{FD}, to maximize the expected utility of his final wealth: $W^{FD} = (v - p_2)x_2^{FD} + (p_2 - p_1)x_1^{FD}$, where p_t denotes the equilibrium price at date $t \in \{1, 2\}$.[15] A SD also has CARA preferences with risk-tolerance γ, but is in the market only in the first period. He thus submits a price-contingent order x_1^{SD} to maximize the expected utility of his wealth $W^{SD} = (v - p_1)x_1^{SD}$. Therefore, FDs as SDs observe p_1 at the first round; furthermore, FDs also observe p_2, so that their information set at the second round is given by $\{p_1, p_2\}$.

[14] Holden and Jacobsen (2014) find that in the US, only 3.3% of all trades take place outside the NBBO. See also Li (2015) for indirect evidence that the single virtual platform assumption is compelling on non-announcement days.

[15] We assume, without loss of generality with CARA preferences, that the non-random endowment of FDs and dealers is zero. Also, as equilibrium strategies will be symmetric, we drop the subindex i.

The inability of a SD to trade in the second period is a way to capture limited market participation in our model. In today's markets, this friction could be due to technological reasons, as in the case of standard dealers with impaired access to a technology that allows trading at high frequency. In the past, two-tiered liquidity provision occurred because only a limited number of market participants could be physically present in the exchange to observe the trading process and react to demand shocks.[16]

2.3 Liquidity demanders

Liquidity traders have CARA preferences, with risk-tolerance γ^L.

In the first period a unit mass of traders enters the market. A trader receives a random endowment of the risky asset u_1 and submits an order x_1^L in the asset that he holds until liquidation.[17] A first period trader posts a market order x_1^L to maximize the expected utility of his profit $\pi_1^L = u_1 v + (v - p_1)x_1^L$:

$$E[-\exp\{-\pi_1^L/\gamma^L\}|u_1]. \tag{2}$$

In period 2, a new unit mass of traders enters the market. A second period trader observes p_1 (and can thus perfectly infer u_1), receives a random endowment of the risky asset u_2, and posts a market order x_2^L to maximize the expected utility of his profit $\pi_2^L = u_2 v + (v - p_2)x_2^L$:

$$E[-\exp\{-\pi_2^L/\gamma^L\}|p_1, u_2]. \tag{3}$$

We assume that $u_t \sim N(0, \tau_u^{-1})$, $\text{Cov}[u_t, v] = \text{Cov}[u_1, u_2] = 0$. To ensure that the payoff functions of the liquidity demanders are well defined (see Section 4.1), we impose

$$(\gamma^L)^2 \tau_u \tau_v > 1, \tag{4}$$

an assumption that is common in the literature (see, e.g., Vayanos and Wang (2012)).

[16]Alternatively, we can think of SD as dealers who only trade during the day, and FD as dealers who, thanks to electronic trading, can supply liquidity around the clock.

[17]Recent research documents the existence of a sizeable proportion of market participants who do not rebalance their positions at every trading round (see Heston et al. (2010), for evidence consistent with this type of behavior at an intra-day horizon).

2.4 Market clearing and prices

Market clearing in periods 1 and 2 is given respectively by $x_1^L + \mu x_1^{FD} + (1-\mu)\, x_1^{SD} = 0$ and $x_2^L + \mu(x_2^{FD} - x_1^{FD}) = 0$. We restrict attention to linear equilibria where

$$p_1 = -\Lambda_1 u_1 \tag{5a}$$

$$p_2 = -\Lambda_2 u_2 + \Lambda_{21} u_1, \tag{5b}$$

where the price impacts of endowment shocks Λ_1, Λ_2, and Λ_{21} are determined in equilibrium. According to (5a) and (5b), at equilibrium, observing p_1 and the sequence $\{p_1, p_2\}$ is informationally equivalent to observing u_1 and the sequence $\{u_1, u_2\}$.

The model thus nests a standard stock market trading model in one of platform competition. Figure 2 displays the timeline of the model.

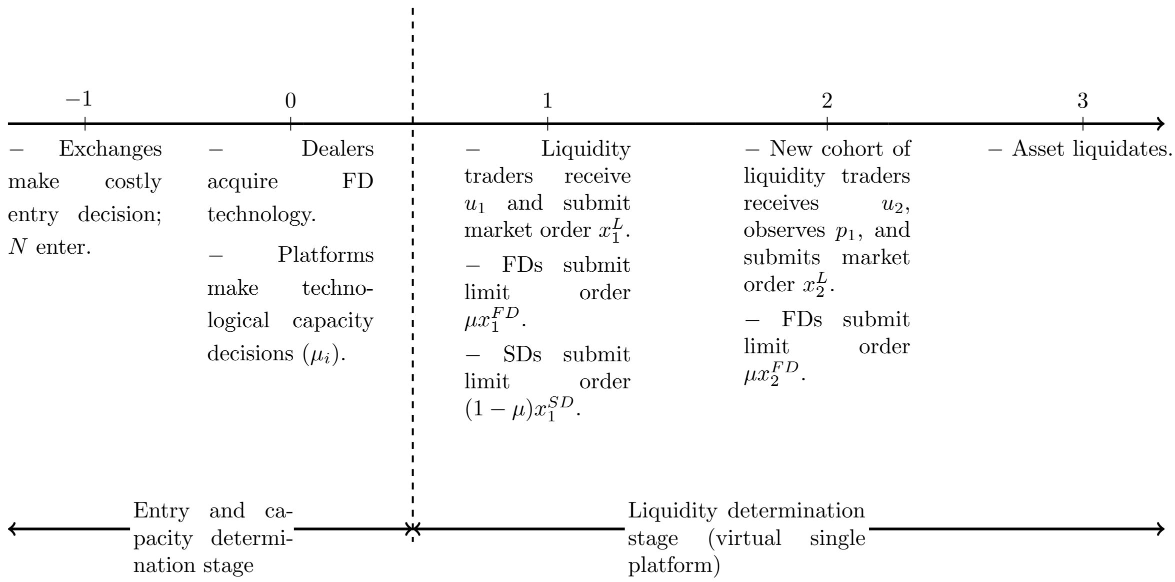


Figure 2: The timeline.

3 Stock market equilibrium

In this section we assume that a positive mass $\mu \in (0,1]$ of FDs is in the market, and present a simple two-period model of liquidity provision à la Grossman and Miller (1988) where dealers only accommodate endowment shocks, but where all traders are expected utility maximizers.

Proposition 1. *For $\mu \in (0,1]$, there exists a unique equilibrium in linear strategies in the stock market, where $x_1^{SD} = -\gamma\tau_v p_1$, $x_1^{FD} = \gamma\tau_u\Lambda_2^{-2}(\Lambda_{21}+\Lambda_1)u_1 - \gamma\tau_v p_1$, $x_2^{FD} = -\gamma\tau_v p_2$, $x_1^L = a_1 u_1$, $x_2^L = a_2 u_2 + b u_1$, and prices are given by* (5a) *and* (5b),

$$\Lambda_1 = \left(1 - \left(1 + a_1 + \mu\gamma\tau_u \frac{\Lambda_{21}+\Lambda_1}{\Lambda_2^2}\right)\right)\frac{1}{\gamma\tau_v} > 0 \tag{6a}$$

$$\Lambda_2 = -\frac{a_2}{\mu\gamma\tau_v} > 0 \tag{6b}$$

$$\Lambda_{21} = -(1 - ((1-\mu)\gamma + \gamma^L)\tau_v\Lambda_1)\Lambda_2 < 0 \tag{6c}$$

$$a_t = \gamma^L\tau_v\Lambda_t - 1 \in (-1,0) \tag{6d}$$

$$b = -\gamma^L\tau_v\Lambda_{21} \in (0,1), \tag{6e}$$

where

$$\Lambda_{21} + \Lambda_1 > 0. \tag{7}$$

The coefficient Λ_t in (5a) and (5b) denotes the period t endowment shock's negative price impact, and is our (inverse) measure of liquidity:

$$\Lambda_t = -\frac{\partial p_t}{\partial u_t}. \tag{8}$$

As we show in the appendix (see (A.3), and (A.14)), a trader's order is given by

$$X_1^L(u_1) = \underbrace{\gamma^L \frac{E[v-p_1|u_1]}{\text{Var}[v-p_1|u_1]}}_{\text{Speculation}} \underbrace{-\, u_1}_{\text{Hedging}}$$

$$X_2^L(u_1,u_2) = \underbrace{\gamma^L \frac{E[v-p_2|u_1,u_2]}{\text{Var}[v-p_2|u_1,u_2]}}_{\text{Speculation}} \underbrace{-\, u_2}_{\text{Hedging}}.$$

According to (6d), a trader speculates and hedges his position to avert the risk of a decline in the endowment value occurring when the return from speculation is low ($a_t \in (-1,0)$). We will refer to $|a_t|$ as the trader's "trading aggressiveness." Additionally, according to (6e), second period traders put a positive weight b on the first period endowment shock. SD and FD provide liquidity, taking the other side of traders' orders. In the first period, standard dealers earn the spread from loading at p_1, and unwinding at the liquidation price. FDs, instead, also speculate on short-term returns. Indeed,

$$x_1^{FD} = \gamma \frac{E[p_2 - p_1|u_1]}{\text{Var}[p_2|u_1]} - \gamma\tau_v p_1.$$

To interpret the above expression, suppose $u_1 > 0$. Then, liquidity traders sell the asset, depressing its price (see (5a)) and, as $E[p_2 - p_1|u_1] = (\Lambda_{21} + \Lambda_1)u_1 > 0$, FDs anticipate a positive short-term return from buying it. When FD unwind their position, the effect of the first period price pressure has not completely disappeared (see (6c)). This induces second period traders to partly absorb FD position, explaining the positive sign of the coefficient b in (6e). Thus, in expectation, FD unload inventory risk from their first period trade to second period liquidity traders.

FDs supply liquidity both by posting a limit order, and a contrarian market order at the equilibrium price, to exploit the predictability of short term returns.[18] In view of this, Λ_1 in (6a) reflects the risk compensation dealers require to hold the portion of u_1 that first period traders hedge and FDs do not absorb via speculation:

$$\Lambda_1 = \left(1 - \left(\underbrace{1 + a_1}_{\text{L1 holding of } u_1} + \underbrace{\mu\gamma\tau_u \frac{\Lambda_{21} + \Lambda_1}{\Lambda_2^2}}_{\text{FD aggregate speculative position}} \right)\right) \frac{1}{\gamma\tau_v}.$$

In the second period, liquidity traders hedge a portion a_2 of their order, which is absorbed by a mass μ of FDs, thereby explaining the expression for Λ_2 in (6b).

Therefore, at both trading rounds, an increase in μ, or in dealers' risk tolerance, increases the risk bearing capacity of the market, leading to a higher liquidity:

Corollary 1. *An increase in the proportion of FDs, or in dealers' risk tolerance increases the liquidity of both trading rounds:* $\partial\Lambda_t/\partial\mu < 0$*, and* $\partial\Lambda_t/\partial\gamma < 0$ *for* $t \in \{1, 2\}$*.*

According to (5b) and (6c), due to FD short term speculation, the first period endowment shock has a persistent impact on equilibrium prices: p_2 reflects the impact of the imbalance FD absorb in the first period, and unwind to second period traders. Indeed, substituting (6c) in (5b), and rearranging yields:

$$p_2 = -\Lambda_2 u_2 + \Lambda_2(\underbrace{(1-\mu)x_1^{SD} + x_1^L}_{=-\mu x_1^{FD}}). \tag{9}$$

[18]This is consistent with the evidence on HFT liquidity supply (Brogaard et al. (2014), and Biais et al. (2015)), and on their ability to predict returns at a short term horizon based on market data (Harris and Saad (2014), and Menkveld (2016)).

Corollary 2. *First period traders hedge the endowment shock more aggressively than second period traders:* $|a_1| > |a_2|$*. Furthermore,* $|a_t|$ *and* b *are increasing in* μ*.*

Comparing dealers' strategies shows that SD in the first period trade with the same intensity as FD in the second period. In view of the fact that in the first period the latter provide additional liquidity by posting contrarian market orders, this implies that

$$\Lambda_1 < \Lambda_2, \tag{10}$$

explaining why traders display a more aggressive hedging behavior in the first period. The second part of the above result reflects the fact that an increase in μ improves liquidity at both dates, but also increases the portion of the first period endowment shock absorbed by FD (see (9)). This, in turn, leads second period liquidity traders to step up their response to u_1.

In view of (6a) and (7), it is easy to see that the price reversion due to FD short term speculation implies

$$\text{Cov}[p_2 - p_1, p_1] = -\Lambda_1(\Lambda_{21} + \Lambda_1)\tau_u^{-1} < 0,$$

so that returns mean revert across trading rounds. A larger FD participation, mitigates price impacts, and attenuates return reversal:

Corollary 3. *An increase in the proportion of FD reduces the mean reversion in the asset returns:* $\partial|\text{Cov}[p_2 - p_1, p_1]|/\partial\mu < 0$.

Summarizing, an increase in μ has two effects: it heightens the risk bearing capacity of the market, and it strengthens the propagation of the first period endowment shock to the second trading round. The first effect makes the market deeper, leading traders to step up their hedging aggressiveness, and lowering the mean reversion in returns.[19] The second effect reinforces second period traders' speculative responsiveness. When all dealers are FDs, liquidity is maximal, and the mean reversion in returns is minimal.

[19] Note that returns would still be negatively correlated even if $\mu \to 0$. This is because, as we show in the appendix (see (A.29) and (A.30)):

$$\Lambda_1 = \frac{1 + (\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v}{(\gamma + \gamma^L + (\gamma + 2\gamma^L)(\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v)\tau_v}$$

$$\Lambda_{21} + \Lambda_1 = \frac{\gamma^L}{\tau_v(\gamma\mu + \gamma^L)(\gamma\mu\tau_u\tau_v(\gamma + 2\gamma^L)(\gamma\mu + \gamma^L) + \gamma + \gamma^L)},$$

both of which are positive when $\mu \to 0$.

Remark 1. *The variance of the first period price is given by* $\text{Var}[p_1] = \Lambda_1^2 \tau_u^{-1}$. *Therefore, a less liquid market increases price volatility.*

Remark 2. *In Appendix B we consider a variation of the liquidity provision model in which we assume that SD enter the market at the second round of the game.*

4 Traders' welfare, technology demand, and exchange equilibrium

In this section we study traders' payoffs, derive demand and supply for technological services, and obtain the platform competition equilibrium.

4.1 Traders' payoffs and the liquidity externality

We measure a trader's payoff with the certainty equivalent of his expected utility:

$$CE^{FD} \equiv -\gamma \ln(-EU^{FD}), \quad CE^{SD} \equiv -\gamma \ln(-EU^{SD}), \quad CE_t^L \equiv -\gamma^L \ln(-EU_t^L), t \in \{1, 2\},$$

where EU^j, $j \in \{SD, FD\}$ and EU_t^L, $t \in \{1, 2\}$ denote respectively the unconditional expected utility of a standard dealer, a full dealer, and a first and second period trader. The following results present explicit expressions for the certainty equivalents.

Proposition 2. *The payoffs of a SD and a FD are given by*

$$CE^{SD} = \frac{\gamma}{2} \ln \left(1 + \frac{\text{Var}[E[v - p_1 | p_1]]}{\text{Var}[v - p_1 | p_1]} \right) \tag{11a}$$

$$CE^{FD} = \frac{\gamma}{2} \Bigg(\ln \left(1 + \frac{\text{Var}[E[v - p_1 | p_1]]}{\text{Var}[v - p_1 | p_1]} + \frac{\text{Var}[E[p_2 - p_1 | p_1]]}{\text{Var}[p_2 - p_1 | p_1]} \right) + \ln \left(1 + \frac{\text{Var}[E[v - p_2 | p_1, p_2]]}{\text{Var}[v - p_2 | p_1, p_2]} \right) \Bigg). \tag{11b}$$

Furthermore:

1. *For all* $\mu \in (0, 1]$, $CE^{FD} > CE^{SD}$.
2. CE^{SD} *and* CE^{FD} *are decreasing in* μ.
3. $\lim_{\mu \to 1} CE^{FD} > \lim_{\mu \to 0} CE^{SD}$.

According to (11a) and (11b), dealers' payoffs reflect the accuracy with which these agents anticipate their strategies' unit profits. A SD only trades in the first period, and the accuracy of his unit profit forecast is given by $\text{Var}[E[v - p_1|p_1]]/\text{Var}[v - p_1|p_1]$ (the ratio of the variance explained by p_1 to the variance unexplained by p_1).

A FD instead trades at both rounds, supplying liquidity to first period traders, as a SD, but also absorbing second period traders' orders, and taking advantage of short-term return predictability. Therefore, his payoff reflects the same components of that of a SD, and also features the accuracy of the unit profit forecast from short term speculation ($\text{Var}[E[p_2 - p_1|p_1]]/\text{Var}[p_2 - p_1|p_1]$), and second period liquidity supply ($\text{Var}[E[v - p_2|p_1, p_2]]/\text{Var}[v - p_2|p_1, p_2]$). In sum, as FD can trade twice, benefiting from more opportunities to speculate and share risk, they enjoy a higher expected utility.

Substituting (6d) and (6e) in (11a) and (11b), and rearranging yields:

$$CE^{SD} = \frac{\gamma}{2} \ln\left(1 + \frac{(1 + a_1)^2}{(\gamma^L)^2 \tau_u \tau_v}\right) \tag{12a}$$

$$CE^{FD} = \frac{\gamma}{2} \left(\ln\left(1 + \frac{(1 + a_1)^2}{(\gamma^L)^2 \tau_u \tau_v} + \left(\frac{1 + a_1}{1 + \mu\gamma\tau_u\tau_v(\mu\gamma + \gamma^L)}\right)^2\right) + \ln\left(1 + \frac{(1 + a_2)^2}{(\gamma^L)^2 \tau_u \tau_v}\right)\right). \tag{12b}$$

An increase in μ has two offsetting effects on the above expressions for dealers' welfare. On the one hand, as it boosts market liquidity, it leads traders to hedge more, increasing dealers' payoffs (Corollaries 1 and 2). On the other hand, as it induces more competition to supply liquidity it lowers them. The latter effect is stronger than the former. Importantly, even in the extreme case in which $\mu = 1$, a FD receives a higher payoff than a SD in the polar case $\mu \approx 0$.

Proposition 3. *The payoffs of first and second period traders are given by*

$$CE_1^L = \frac{\gamma^L}{2} \ln\left(1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} + 2\frac{\text{Cov}[p_1, u_1]}{\gamma^L}\right) \tag{13a}$$

$$CE_2^L = \frac{\gamma^L}{2} \ln\left(1 + \frac{\text{Var}[E[v - p_2|p_1, p_2]]}{\text{Var}[v - p_2|p_1, p_2]} + 2\frac{\text{Cov}[p_2, u_2|p_1]}{\gamma^L} + \frac{\text{Var}[E[v - p_2|p_1]]}{\text{Var}[v]} - \left(\frac{\text{Cov}[p_2, u_1]}{\gamma^L}\right)^2\right). \tag{13b}$$

Furthermore:

1. CE_1^L and CE_2^L are increasing in μ.

2. For all $\mu \in (0,1]$, $CE_1^L > CE_2^L$.

Similarly to SDs, liquidity traders only trade once (either at the first, or at the second round). This explains why their payoffs reflect the precision with which they can anticipate the unit profit from their strategy (see (13a) and (13b)). Differently from SDs, these traders are however exposed to a random endowment shock. As a less liquid market increases hedging costs, it negatively affects their payoff ($\text{Cov}[p_1, u_1] = -\Lambda_1 \tau_u^{-1}$, and $\text{Cov}[p_2, u_2|p_1] = -\Lambda_2 \tau_u^{-1}$). Finally, (13b) shows that a second period trader benefits when the return he can anticipate based on u_1 is very volatile compared to v ($\text{Var}[E[v - p_2|p_1]]/\text{Var}[v]$), since this indicates that he can speculate on the propagated endowment shock at favorable prices. However, a strong speculative activity reinforces the relationship between p_2 and u_1, ($\text{Cov}[p_2, u_1]^2$), leading a trader to hedge little of his endowment shock u_2, and keep a large exposure to the asset risk, thereby reducing his payoff.

Substituting (6d) and (6e) in (13a) and (13b), and rearranging yields:

$$CE_1^L = \frac{\gamma^L}{2} \ln\left(1 + \frac{a_1^2 - 1}{(\gamma^L)^2 \tau_u \tau_v}\right) \tag{14}$$

$$CE_2^L = \frac{\gamma^L}{2} \ln\left(1 + \frac{a_2^2 - 1}{(\gamma^L)^2 \tau_u \tau_v} + \frac{b^2((\gamma^L)^2 \tau_u \tau_v - 1)}{(\gamma^L)^4 \tau_u^2 \tau_v^2}\right). \tag{15}$$

An increase in the proportion of FDs μ makes the market more liquid and leads traders to hedge and speculate more aggressively (Corollary 2), benefiting first period traders (Proposition 3). At the same time, it heightens the competitive pressure faced by SDs, lowering their payoffs (Proposition 2). As liquidity demand augments for both dealers' classes, however, SDs effectively receive a *smaller* share of a *larger* pie. This mitigates the negative impact of increased competition, implying that on balance the positive effect of the increased liquidity prevails:

Corollary 4. *The positive effect of an increase in the proportion of FDs on first period traders' payoffs is stronger than its negative effect on SDs' welfare:*

$$\frac{\partial CE_1^L}{\partial \mu} > -\frac{\partial CE^{SD}}{\partial \mu}, \tag{16}$$

for all $\mu \in (0,1]$.

Aggregating across market participants' welfare yields the following Gross Welfare function:

$$
\begin{aligned}
GW(\mu) &= \mu CE^{FD} + (1-\mu)CE^{SD} + CE_1^L + CE_2^L \\
&= \underbrace{\mu(CE^{FD} - CE^{SD})}_{\text{Surplus earned by FDs}} + \underbrace{CE^{SD} + CE_1^L + CE_2^L}_{\text{Welfare of other market participants}}
\end{aligned} \tag{17}
$$

Corollary 5.

1. *The welfare of market participants other than FDs is increasing in* μ.

2. *Gross welfare is higher at* $\mu = 1$ *than at* $\mu \approx 0$.

The first part of the above result is a direct consequence of Corollary 4: as μ increases, SDs' losses due to heightened competition are more than compensated by traders' gains due to higher liquidity. The second part, follows from Proposition 2 (part 3), and Proposition 3. Note that it rules out the possibility that the payoff decline experienced by FDs as μ increases, leads gross welfare to be higher at $\mu \approx 0$. Therefore, a solution that favors liquidity provision by FDs is also in the interest of all market participants. Finally, we have:

Numerical Result 1. *Numerical simulations show that* $GW(\mu)$ *is monotone in* μ. *Therefore,* $\mu = 1$ *is the unique maximum of the gross welfare function* $GW(\mu)$.

In view of Corollaries 1 and 3, gross welfare is maximal when liquidity (mean reversion in returns) is at its highest (lowest) level.[20] Furthermore, because of monotonicity, the above market quality indicators, become "measurable" welfare indexes.

[20]Numerical simulations where conducted using the following grid: $\gamma, \mu \in \{0.01, 0.02, \ldots, 1\}$, $\tau_u, \tau_v \in \{1, 2, \ldots, 10\}$, and $\gamma^L \in \{1/\sqrt{\tau_u\tau_v} + 0.001, 1/\sqrt{\tau_u\tau_v} + 0.101, \ldots, 1\}$, in order to satisfy (4).

4.2 The demand for technological services

We define the value of becoming a FD as the extra payoff that such a dealer earns compared to a SD. According to (11a) and (11b), this is given by:

$$\begin{aligned}\phi(\mu) &\equiv CE^{FD} - CE^{SD} \qquad (18)\\ &= \frac{\gamma}{2}\Bigg(\underbrace{\ln\left(1+\frac{\text{Var}[E[v-p_1|p_1]]}{\text{Var}[v-p_1|p_1]}+\frac{\text{Var}[E[p_2-p_1|p_1]]}{\text{Var}[p_2-p_1|p_1]}\right)-\ln\left(1+\frac{\text{Var}[E[v-p_1|p_1]]}{\text{Var}[v-p_1|p_1]}\right)}_{\text{Competition}}\\ &\qquad + \underbrace{\ln\left(1+\frac{\text{Var}[E[v-p_2|p_1,p_2]]}{\text{Var}[v-p_2|p_1,p_2]}\right)}_{\text{Liquidity supply}}\Bigg).\end{aligned}$$

FDs rely on two sources of value creation: first, they compete business away from SDs, extracting a larger rent from their trades with first period traders (since they can supply liquidity and speculate on short-term returns); second, they supply liquidity to second period traders.

The function $\phi(\mu)$ can be interpreted as the (inverse) demand for technological services:[21]

Corollary 6. *The inverse demand for technological services $\phi(\mu)$ is decreasing in μ.*

A marginal increase in μ heightens the competition FDs face among themselves, and vis-à-vis SDs. The former effect lowers the payoff of a FD. In the appendix, we show that the same holds also for the latter effect. Thus, an increase in the mass of FDs erodes the rents from competition, implying that $\phi(\mu)$ is decreasing in μ.

Numerical Result 2. *When μ, τ_u, and τ_v are sufficiently large and γ is large above γ^L, $\phi(\mu)$ is log-convex in μ:*

$$\frac{\partial^2 \ln\phi(\mu)}{\partial\mu^2} \geq 0. \qquad (19)$$

In Figure 3 (panel (a)) we plot $\ln(\phi(\mu))$ for a set of parameters yielding log-convexity. When this occurs, the price reduction corresponding to an increase in μ becomes increasingly smaller as μ increases.[22]

[21] As $\phi(\mu)$ reflects the extra margin that FD obtain vis-à-vis D, it formalizes in a simple manner the way in which Lewis (2014) describes Larry Tabb's estimation of traders' demand for the high speed, fiber optic connection that Spread laid down between New York and Chicago in 2009.

[22] We checked log-convexity of the function $\phi(\mu)$, assuming $\tau_u, \tau_v \in \{1, 6, 11\}$, $\gamma, \gamma^L \in \{0.01, 0.02, \ldots, 1\}$, and for $\mu \in \{0.2, 0.4, \ldots, 1\}$. The second derivative of $\ln(\phi(\mu))$ turns negative for μ, τ_u, or τ_v low, and for $\gamma^L > \gamma$ (e.g., this happens when $\tau_u = 1, \tau_v = 6$, $\mu = 0.2$, and $\gamma^L = 0.41$, $\gamma = 0.01$).

Corollary 7. *An increase in γ has two contrasting effects on the inverse demand for technological services $\phi(\mu) = \gamma \ln(EU^{FD}/EU^{SD})^{1/2}$:*

$$\frac{\partial \phi(\mu)}{\partial \gamma} = \underbrace{\frac{1}{2}\ln\left(\frac{EU^{FD}}{EU^{SD}}\right)}_{>0} + \underbrace{\frac{\gamma}{2}\left(\frac{(\partial EU^{FD}/\partial\gamma)EU^{SD} - (\partial EU^{SD}/\partial\gamma)EU^{FD}}{EU^{FD}EU^{SD}}\right)}_{<0} \tag{20}$$

As argued in Corollary 2, as FDs trade twice, they enjoy a larger payoff compared to SDs. An increase in γ leads dealers to trade more aggressively, and for a given expected utility difference, has a positive effect on ϕ. However, a higher risk-tolerance reduces the value of the additional risk-sharing opportunity offered by the second trading round, which has a negative effect on ϕ.

4.3 The supply of technological services and exchange equilibrium

Depending on the industrial organization of exchanges, the supply of technological services is either controlled by a single platform, acting as an "incumbent monopolist," or by $N \geq 2$ venues who compete à la Cournot in technological capacities.

In the former case, the monopolist profit is given by

$$\pi(\mu) = (\phi(\mu) - c)\mu - f, \tag{21}$$

where c and f, respectively denote the and fixed cost of producing a capacity μ. We denote by μ^M the optimal capacity of the monopolist exchange:

$$\mu^M \in \arg\max_{\mu\in(0,1]} \pi(\mu). \tag{22}$$

In the latter case, denoting by μ_i and $\mu_{-i} = \sum_{j\neq i}^{N} \mu_j$, respectively the capacity installed by exchange i and its rivals, and by f and c the fixed and marginal cost incurred by an exchange to enter and produce capacity μ_i, an exchange i's profit function is given by

$$\pi(\mu_i, \mu_{-i}) = (\phi(\mu) - c)\mu_i - f. \tag{23}$$

We define a symmetric Cournot equilibrium as follows:

Definition 1. *A symmetric Cournot equilibrium in technological service capacities is a* *set of capacities $\mu_i^C \in (0,1]$, $i = 1, 2, \dots, N$, such that (i) each μ_i^C maximizes (23), for given capacity choice of other exchanges μ_{-i}^C:*

$$\mu_i^C \in \arg\max_{\mu_i} \pi(\mu_i, \mu_{-i}^C), \tag{24}$$

(ii) $\mu_1^C = \mu_2^C = \cdots = \mu_N^C$, and (iii) $\sum_{i=1}^N \mu_i^C = \mu^C(N)$.

We have the following result:

Proposition 4. *There exists at least one symmetric Cournot equilibrium in technological service capacities and no asymmetric ones.*

Proof. See Amir (2018), Proposition 7, and Vives (1999), Section 4.1. □

Numerical simulations show that the equilibrium is unique and stable.[23]

4.3.1 Strategic complementarity in capacity decisions

With Cournot competition, log-convexity of the inverse demand function implies that the (log of the) revenue of an exchange displays increasing differences in the pair (μ_i, μ_{-i}). Indeed,

$$\ln(\phi(\mu_i, \mu_{-i})\mu_i) = \ln(\phi(\mu_i, \mu_{-i})) + \ln \mu_i,$$

and $\ln(\phi(\mu_i, \mu_{-i}))$ has increasing differences in (μ_i, μ_{-i}) since this is equivalent to ϕ being log-convex.

Thus, with a zero marginal cost, a larger capacity installed by rivals has a negative impact on an exchange profit which decreases in the exchange capacity choice. This leads a platform to respond to an increase in its rivals' capacity choice by increasing the capacity it installs (in this situation a Cournot oligopoly is a game of strategic complements, see e.g., Amir (2018), Proposition 3). This is because when FDs demand is log-convex, the intensive margin effect of a capacity increase is more than offset by the corresponding extensive margin effect. Hence, a platform's decision to step up capacity in the face of rivals' capacity increase, induces a mild price decline that is more than compensated by

[23] In our setup, a sufficient condition for stability (Section 4.3 in Vives (1999)) is that the elasticity of the slope of the FDs inverse demand function is bounded by the number of platforms (plus one):

$$\mathcal{E}|_{\mu=\mu^C(N)} \equiv -\mu \frac{\phi''(\mu)}{\phi'(\mu)}\bigg|_{\mu=\mu^C(N)} < 1 + N.$$

the exchange increase in market share, allowing the platform to boost its revenue (and cut its losses). By continuity, when the marginal cost is sufficiently small, log-convexity of $\phi(\mu)$ can make an exchange best response

$$BR(\mu_{-i}) = \arg\max_{\mu_i}\{\pi(\mu_i, \mu_{-i})|\mu_i \in (0, 1]\}, \tag{25}$$

increasing in its rivals' choices (see Figure 3, panel (b)).[24]

Numerical Result 3. *When $N = 2$, strategic complementarities in capacity decisions can arise for some range of exchanges' best response (see (25)).*

For example, assuming a low value for the marginal cost ($c = 0.0002$), the model easily displays strategic complementarities (see Figure 3, panel (b)).

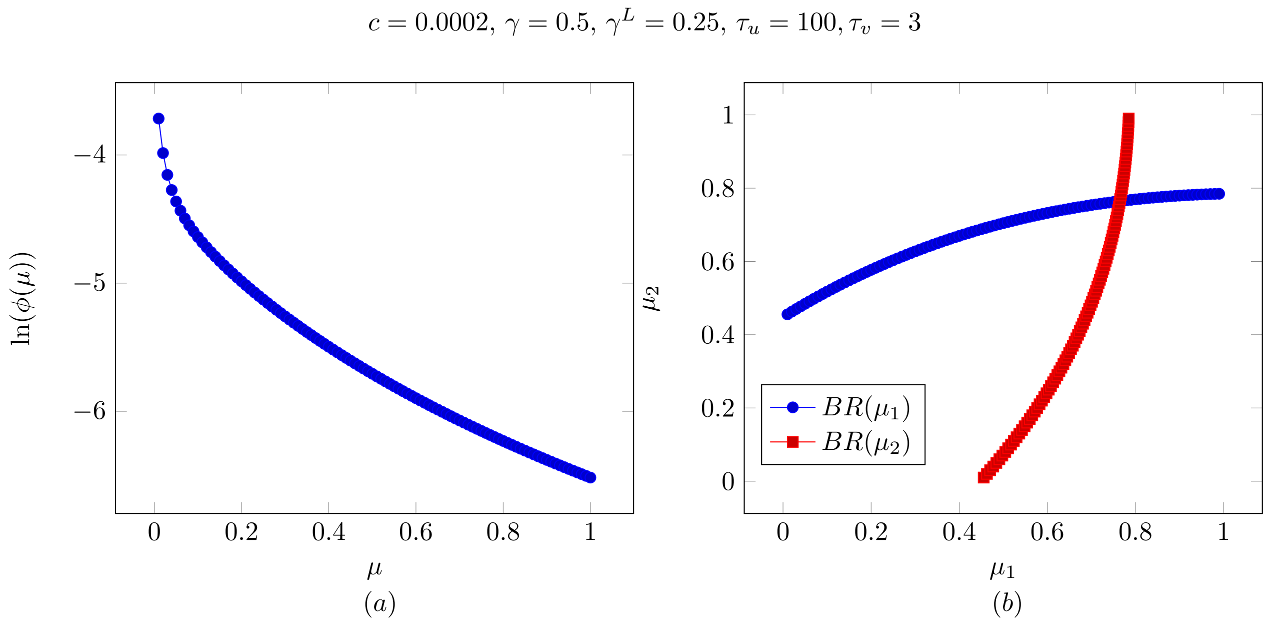


Figure 3: Log-convexity of the demand function (Panel (a)), and strategic complementarities in platforms' capacity decisions (Panel (b)).

For $N > 2$ (when $c > 0$, albeit small) we find instead that an exchange's best response is downward sloping. At a symmetric Cournot equilibrium, we have:

$$\left.\frac{\partial BR_i(\mu_{-i})}{\partial \mu_{-i}}\right|_{\mu=\mu^C(N)} = -\left.\frac{\phi''(\mu)(\mu/N) + \phi'(\mu)}{\phi''(\mu)(\mu/N) + 2\phi'(\mu)}\right|_{\mu=\mu^C(N)}. \tag{26}$$

[24] Parameter values are consistent with Leland (1992).

As N increases, the platform's marginal gain in market share from a capacity increase shrinks (the weight of the positive effect due to demand convexity in (26) declines), yielding a negatively sloped best response.

4.3.2 Comparative statics with respect to N

At a stable Cournot equilibrium, standard comparative statics results apply (see, e.g., Section 4.3 in Vives (1999)). In particular, an increase in the number of exchanges leads to an increase in the aggregate technological service capacity, and a decrease in each exchange profit:

$$\frac{\partial \mu^C(N)}{\partial N} \geq 0 \tag{27a}$$

$$\left.\frac{\partial \pi_i(\mu)}{\partial N}\right|_{\mu=\mu^C(N)} \leq 0. \tag{27b}$$

If the number of competing platforms is exogenously determined, condition (27a) implies that spurring competition in the intermediation industry has positive effects in terms of liquidity and gross welfare (Proposition 1 and Numerical Result 1):

Corollary 8. *At a stable Cournot equilibrium, an exogenous increase in the number of competing exchanges has a positive impact on liquidity and gross welfare:* $\partial\Lambda_t/\partial N < 0$, $\partial GW/\partial N > 0$.

Degryse et al. (2015) study 52 Dutch stocks in 2006-2009 (listed on Euronext Amsterdam and trading on Chi-X, Deutsche Börse, Turquoise, BATS, Nasadaq OMX and SIX Swiss Exchange) and find a positive relationship between market fragmentation (in terms of a lower Herfindhal index, higher dispersion of trading volume across exchanges) and the *consolidated liquidity* of the stock. Foucault and Menkveld (2008) also find that consolidated liquidity increased when in 2004 the LSE launched EuroSETS, a new limit order market to allow Dutch brokers to trade stocks listed on Euronext (Amsterdam).

Upward sloping best responses can lead a platform to respond to a heightened competitive pressure, with an increase in installed capacity, strengthening the aggregate effect in (27a), and the resulting impact this has on liquidity and gross welfare.[25] To illustrate this effect, in Figure 4 we use the same parameters of Figure 3 (panel (b)), and study the impact of an increase in competition. Panel (a) in the figure shows that platforms step

[25]The necessary and sufficient condition for an increase in N to lead to an increase in individual capacity is that $N < \mathcal{E}|_{\mu=\mu^C} < 1 + N$ (see Section 4.3 in Vives (1999)).

up their individual capacity, with a positive effect on liquidity (panels (b) and (c)), and welfare (panel (d)).

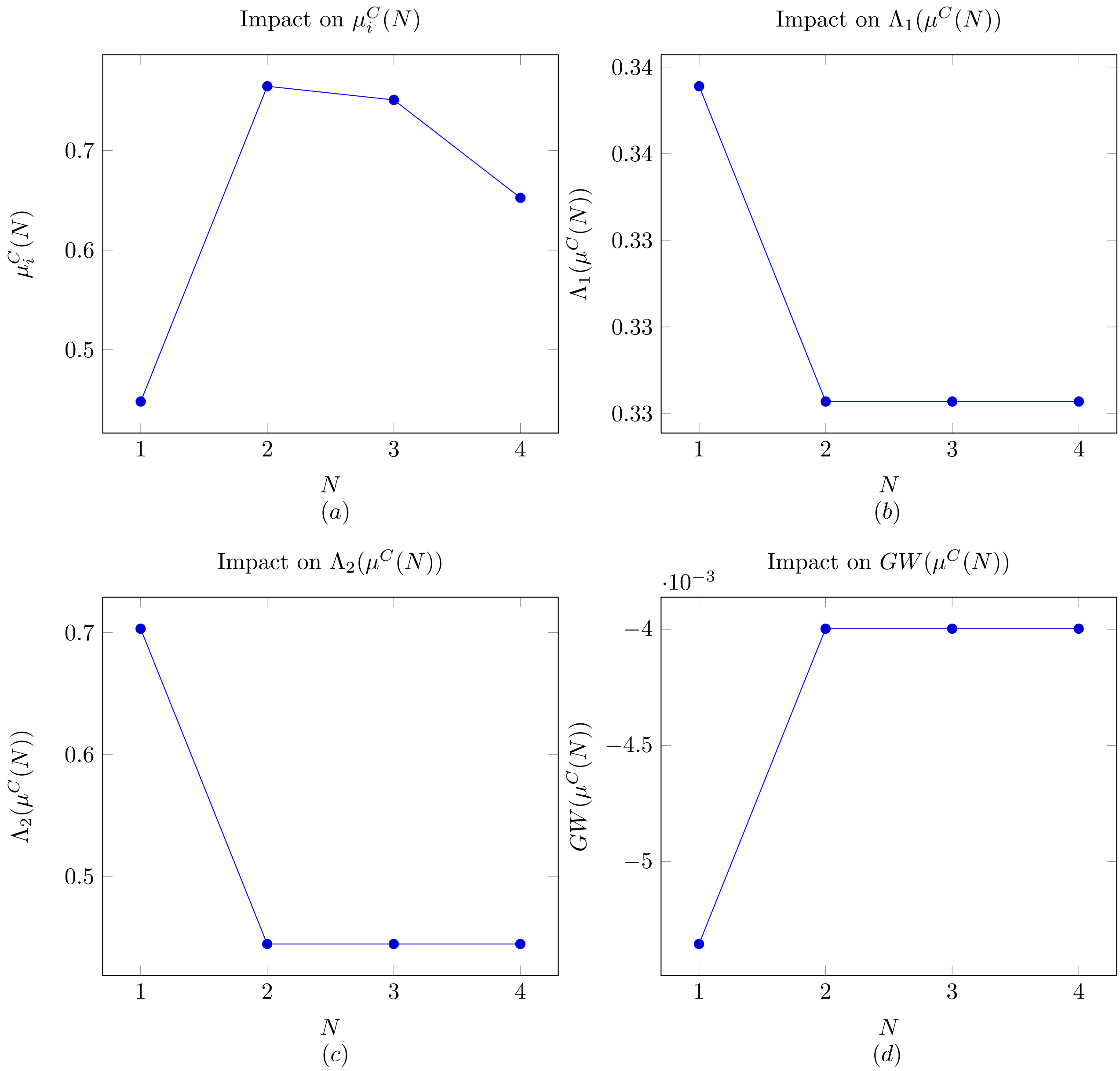


Figure 4: Effect of entry on each platform capacity decisions (panel (a)), liquidity (panels (b) and (c)), and gross welfare (panel (d)) (parameter values as in Figure 3).

5 Endogenous platform entry and welfare

In this section we endogenize platform entry, and study its welfare implications.[26] Assuming that platforms' technological capacities are identical ($\mu = N\mu_i$), a social planner who takes into account the costs incurred by the exchanges faces the following objective function:

$$\begin{aligned} \mathcal{P}(\mu, N) &\equiv GW(\mu) - c\mu - fN \\ &= \pi(\mu_i)N + \psi(\mu). \end{aligned} \tag{28}$$

Expression (28) is the sum of two components. The first component reflects the profit generated by competing platforms, who siphon out FDs surplus, and incur the costs associated with running the exchanges:

$$\pi(\mu_i)N = ((\phi(\mu) - c)\mu_i - f)N = \underbrace{\phi(\mu)\mu}_{(CE^{FD} - CE^{SD})\mu} - c\mu - fN,$$

implying that FDs surplus only contributes indirectly to the planner's function, via platforms' total profit. The second component in (28) reflects the welfare of other market participants:

$$\psi(\mu) = CE^{SD} + CE_1^L + CE_2^L,$$

and highlights the welfare effect of technological capacity choices via the liquidity externality.[27]

We consider five possible outcomes:

1. Cournot with free entry (CFE). In this case, we look for a symmetric Cournot equilibrium in μ, as in Definition 1, and impose the free entry constraint:

$$(\phi(\mu^C(N)) - c)\frac{\mu^C(N)}{N} \geq f > (\phi(\mu^C(N+1)) - c)\frac{\mu^C(N+1)}{N+1}, \tag{29}$$

[26] For example, according to the UK Competition Commission (2011), a platform entry fixed cost covers initial outlays to acquire the matching engine, the necessary IT architecture to operate the exchange, the contractual arrangements with connectivity partners that provide data centers to host and operate the exchange technology, and the skilled personnel needed to operate the business. The Commission estimated that in 2011 this roughly corresponded to £10-£20 million.

[27] Even incumbent exchanges may have to incur an entry cost to supply liquidity in the second round. For example, faced with increasing competition from alternative trading venues, in 2009 LSE decided to absorb Turquoise, a platform set up about a year before by nine of the world's largest banks. (See "LSE buys Turquoise share trading platform," *Financial Times,* December 2009).

which pins down N. We denote by μ^{CFE}, and N^{CFE} the pair that solves the Cournot case. Note that, given Proposition 4 and (27b), a unique Cournot equilibrium with free entry obtains in our setup if (27b) holds and for a given N the equilibrium is unique.

2. Structural Second Best (STR). In this case we posit that the planner can determine the number of exchanges that operate in the market. As exchanges compete à la Cournot in technological capacities, we thus look for a solution to the following problem:

$$\max_{N\geq 1} \mathcal{P}(\mu^C(N), N) \text{ s .t. } \mu^C(N) \text{ is a Cournot equilibrium with } \pi_i^C(N) \geq 0, \tag{30}$$

and denote by μ^{STR}, and N^{STR} the pair that solves (30).

3. Unrestricted Structural Second Best (USTR). In this case we relax the non-negativity constraint in (30), thereby assuming that the planner can make side-payments to exchanges if they do not break-even. Thus, we look for a solution to the following problem:

$$\max_{N\geq 1} \mathcal{P}(\mu^C(N), N) \text{ s .t. } \mu^C(N) \text{ is a Cournot equilibrium,} \tag{31}$$

and denote by μ^{USTR}, and N^{USTR} the pair that solves (31).

4. Behavioral Second Best (BEH). In this case, we let the planner set the fee that exchanges charge to FDs, assuming free entry of platforms. Because of Corollary 6, $\phi(\mu)$ is invertible in μ, implying that setting the fee is equivalent to choosing the aggregate technological capacity μ. Thus, we look for a solution to the following problem:

$$\max_{\mu\in(0,1]} \mathcal{P}(\mu, N) \text{ s.t. } (\phi(\mu) - c)\frac{\mu}{N} \geq f \geq (\phi(\mu) - c)\frac{\mu}{N+1}, \tag{32}$$

and denote by μ^{BEH} and N^{BEH} the pair that solves (32).[28]

5. First Best (FB). In this case, we assume that the planner can regulate the market choosing the fee and the number of competing platforms:

$$\max_{\mu\in(0,1],N\geq 1} \mathcal{P}(\mu, N). \tag{33}$$

We denote by μ^{FB} and N^{FB} the pair that solves (33).

[28]We assume for simplicity that if the second inequality holds with equality, then only N firms enter.

We contrast the above four cases with the "Unregulated Monopoly" outcome (M) defined in Section 4.3.

We make the maintained assumption that both the monopoly profit and $\mathcal{P}(\mu, 1)$ are single-peaked in μ,[29] with maximum monopoly profit being positive. Our first set of results compares the FB and BEH solutions with the monopoly one.

Proposition 5. *1. $N^{FB} = 1$; if the monopoly solution is interior, $\mu^M \in (0,1)$, then $\mu^{FB} > \mu^M$ and $\Lambda_t(\mu^M) > \Lambda_t(\mu^{FB})$; otherwise, $\mu^{FB} = \mu^M$ and $\Lambda_t(\mu^M) = \Lambda_t(\mu^{FB})$.*

2. $\mu^{BEH} \geq \mu^M$ and $\Lambda_t(\mu^M) \geq \Lambda_t(\mu^{BEH})$ always; if at μ^{FB} the monopoly profit is negative, then $N^{BEH} = 1$ and $\mu^{FB} > \mu^{BEH} > \mu^M$ and $\Lambda_t(\mu^M) > \Lambda_t(\mu^{BEH})$.

At the FB the planner minimizes entry costs by letting a single exchange satisfy the industry demand for technological services. Furthermore, it limits the monopolist market power by imposing a fee that is lower than the market solution. A similar logic underpins the BEH solution. More in detail, suppose that at μ^{FB} the monopoly profit is negative. As for a given (aggregate) μ, the profit of an exchange is decreasing in N, for given μ the maximum profit obtains when $N = 1$. We have then that $N^{BEH} = 1$. Furthermore, given that $\mathcal{P}$ is single peaked in μ, it is optimal for μ^{BEH} to be set as large as possible so that monopoly profits are zero. The solution is then $\mu^M < \mu^{BEH} \leq \mu^{FB}$.[30]

Regulating the fee can however be complicated, as our discussion in the introduction suggests. With this in mind, we now focus on the case in which the planner cannot set the technological service fee, but can decide on the number of competing exchanges. In the absence of regulation, a Cournot equilibrium with free entry arises (see (29)). We thus compare this outcome to the Structural Second Best, in both the unrestricted and restricted cases. Evaluating the first order condition of the planner at $N = N^{CFE}$

[29]This condition is satisfied in all of our simulations (see Table 2).

[30]We have numerically verified the above sufficient condition for $N^{BEH} = 1$, and found in our simulations that it is always satisfied. In the reverse order of actions model, in some cases $\pi^M(\mu^{BEH}) > 0$, but the planner still sets $N^{BEH} = 1$. See Table 2 for details.

(ignoring the integer constraint) yields:

$$\left.\frac{\partial \mathcal{P}(\mu^C(N),N)}{\partial N}\right|_{N=N^{CFE}} = \left.\underbrace{\pi_i(\mu^C(N),N)}_{=0}\right|_{N=N^{CFE}} + N^{CFE} \left.\underbrace{\frac{\partial \pi_i(\mu^C(N),N)}{\partial N}}_{\text{Profitability depression } <0}\right|_{N=N^{CFE}} + \left.\underbrace{\psi'(\mu)\frac{\partial \mu^C(N)}{\partial N}}_{\text{Liquidity creation } >0}\right|_{N=N^{CFE}}. \tag{34}$$

According to (34), if the Cournot equilibrium is stable, platform entry has two countervailing welfare effects.[31] The first one is a "profitability depression" effect, and captures the profit decline associated with the demand reduction faced by each platform as a result of entry. This effect is conducive to excessive entry, as each competing exchange does not internalize the negative impact of its entry decision on competitors' profits. The second one is a "liquidity creation" effect and is instead peculiar to a financial market setup in which end users benefit from the possibility to hedge endowment shocks. This effect reflects the welfare creation of an increase in N via the liquidity externality (recall that at a stable equilibrium an increase in N increases $\mu^C(N)$, which has a positive effect on liquidity), and is conducive to insufficient entry since each exchange does not internalize the positive impact of its entry decision on other market participants' payoffs.

These effects differ from the standard ones arising in a Cournot equilibrium with free entry (Mankiw and Whinston (1986)). Liquidity creation relates to the increase in consumer surplus that comes about with an increase in the number of firms because of the quasicompetitiveness property of regular equilibria (that is, total output increasing with the number of firms, see section 4.3 in Vives (1999)). In the Cournot case it so happens that with business stealing (i.e., with individual output decreasing in the number of firms), the profitability depressing effect of entry always dominates, inducing excessive entry (except for the integer problem, insufficient entry can occur by at most one firm). A similar result obtains in our setup, when we compare N^{CFE} with N^{STR}; however, when N^{CFE} is stacked against N^{USTR}, this conclusion does not necessarily hold.

More in detail, N^{CFE} is the the largest N so that platforms break even at a Cournot

[31] This is because at a stable equilibrium (27a) and (27b) hold.

equilibrium. At the STR solution, platforms break even too, but the planner internalizes the profitability depression effect of entry. Thus, we have that

$$N^{CFE} \geq N^{STR}.$$

Conversely, removing the break even constraint, the planner achieves the Unrestricted STR, and depending on which of the effects outlined above prevails, both excessive or insufficient entry can occur:

Proposition 6. *When the planner regulates entry, for stable Cournot equilibria:*

1. $N^{CFE} \geq N^{STR}$, $\mu^{CFE} \geq \mu^{STR}$, *and* $\Lambda_t(\mu^{CFE}) \leq \Lambda_t(\mu^{STR})$.

2. *When the profitability depression effect is stronger than the liquidity creation effect,* $N^{CFE} \geq N^{USTR}$, $\mu^{CFE} \geq \mu^{USTR}$, *and* $\Lambda_t(\mu^{CFE}) \leq \Lambda_t(\mu^{USTR})$. *Otherwise, the opposite inequalities hold.*

3. *The technological capacity at CFE is higher than at STR, which is in turn higher than at M:* $\mu^{CFE} \geq \mu^{STR} \geq \mu^{M}$. *The technological capacity at USTR is higher than at M:* $\mu^{USTR} \geq \mu^{M}$. *Therefore,*

$$\Lambda_t(\mu^M) \geq \Lambda_t(\mu^{STR}) \geq \Lambda_t(\mu^{CFE}), \text{ and } \Lambda_t(\mu^M) \geq \Lambda_t(\mu^{USTR}). \tag{35}$$

4. *Welfare ranking:* $\mathcal{P}^{USTR} \geq \mathcal{P}^{STR} \geq \mathcal{P}^{CFE}$.

The first two items in the proposition reflect our previous discussion. Item 3 shows that while the technological capacity offered at the CFE is higher than at the STR (a natural consequence of excessive entry with respect to the STR benchmark), the comparison with the USTR is inconclusive. Indeed, as explained above, in this case entry can be insufficient, implying that the planner may push entry beyond the break-even level, subsidising the loss-making platforms. Thus, while liquidity maximization is generally at odds with welfare maximization in the STR case, the two may be aligned in the USTR case. Finally, as at the USTR the non-negativity constraint of the exchanges profit is relaxed, $\mathcal{P}^{USTR} \geq \mathcal{P}^{STR}$ must hold.

To verify the possibility of excessive or insufficient entry compared to the USTR, we run two sets of numerical simulations. In the first set, as in Figure 3 we assume standard risk aversion ($\gamma = 0.5$, $\gamma^L = 0.25$), a 10% annual volatility for the endowment shock, and consider a "high" and a "low" payoff volatility scenario (respectively, $\tau_v = 3$, which

which corresponds to a 60% annual volatility for the liquidation value, and $\tau_v = 25$ which corresponds to a 20% annual volatility). Platform costs are set to $f \in \{1 \times 10^{-6}, 2 \times 10^{-6}, \ldots, 31 \times 10^{-6}\}$, and $c = 0.002$.[32] In the second set, we assume lower values for risk aversion ($\gamma = 25$, $\gamma^L = 12$) which are consistent with the literature on price pressure, and recent results on the structural estimation of risk aversion based on insurance market data,[33] and set $\tau_v = \tau_u = 0.1$ (corresponding to a 316% annual volatility for both the endowment shock and the liquidation value), $f \in \{1 \times 10^{-2}, 2 \times 10^{-2}, \ldots, 31 \times 10^{-2}\}$, and $c = 2$. For both sets of simulations, we solve for the technological capacity and the number of platforms, in both the "Original" and "Reverse" order of actions cases, (respectively, OO and RO), and perform robustness analysis (see Tables 1 and 2).

Numerical Result 4. *The results of our numerical simulations are as follows:*

1. *With standard risk aversion values:*
   (a) *With high payoff volatility, entry is excessive: $N^{CFE} > N^{USTR}$, and $\mu^{CFE} > \mu^{USTR}$.*
   (b) *With low payoff volatility, when the marginal cost of technological capacity is low, and for sufficiently large values of the entry cost, entry is insufficient: $N^{CFE} < N^{USTR}$ and $\mu^{CFE} < \mu^{USTR}$.*
2. *With low risk aversion values, for sufficiently large values of the entry cost, entry is insufficient.*

Furthermore, at all solutions N and μ are decreasing in f.[34]

Figure 5 illustrates the output of two simulations in which insufficient entry occurs (when $c = 0.005$, a case we do not display, insufficient entry disappears). Insufficient entry implies that platforms enjoy stronger market power compared to a social planner objective. This situation appears to be in line with the complaints raised by many market participants, as we argue in the introduction, but also with the view of some regulators.[35]

[32] Analyzing the US market, Jones (2018) argues that barriers to entry to the intermediation industry are very low, a consideration that is corroborated by the current state of the market, where 13 cash equity exchanges compete with over 30 ATS. This suggests that entry cost must be low.

[33] See respectively Hendershott and Menkveld (2014), and Cohen and Einav (2007).

[34] Assuming $\gamma = 0.25 < \gamma^L = 0.5$ yields qualitatively similar results in the high volatility case, whereas in the low volatility case insufficient entry disappears.

[35] "[...] For example, one exchange, EDGX, has raised the price on its standard 10GB connection five times since 2010–in total, leaving the price of the connection seven times higher than it was in that year." *Unfair Exchange: The State of America's Stock Markets*, speech of Commissioner Robert J. Jackson Jr., George Mason University, September 2018.

The next result provides a welfare comparison of the different outcomes under the assumption that monopoly profit is negative at the First-Best solution, which implies that at the BEH solution profits are exactly zero (see Lemma 2 in the Appendix).

Proposition 7. *Comparing solutions when $\pi^M(\mu^{FB}) < 0$:*

1. $\mu^{FB} > \mu^{BEH} > \mu^{CFE} \geq \mu^{STR} \geq \mu^M$. *Therefore,*

$$\Lambda_t(\mu^{FB}) < \Lambda_t(\mu^{BEH}) < \Lambda_t(\mu^{CFE}) \leq \Lambda_t(\mu^{STR}) \leq \Lambda_t(\mu^M).$$

2. *The number of exchanges entering the market with Cournot free entry or with entry regulation is no lower than with fee regulation ($N^{BEH} = 1$).*

3. *Welfare comparison:*

$$\mathcal{P}^{FB} > \mathcal{P}^{BEH} > \mathcal{P}^{STR} \geq \max\left\{\mathcal{P}^{CFE}, \mathcal{P}^M\right\} \geq \min\left\{\mathcal{P}^{CFE}, \mathcal{P}^M\right\}, \tag{36a}$$

$$\mathcal{P}^{FB} \geq \mathcal{P}^{USTR} \geq \mathcal{P}^{STR}, \tag{36b}$$

where if the FB solution is interior, then $\mathcal{P}^{FB} > \mathcal{P}^{USTR}$.

According to the proposition, when a planner picking the FB would push an exchange monopolist to supply technological services at a loss, then in a second best world fee regulation is optimal. Indeed, in these conditions, the planner sets μ^{BEH} such that only one platform enters ($N^{BEH} = 1$) and, by Proposition 5, such platform sets the level of technological service capacity which makes the market more liquid than with the monopoly solution. Furthermore, such a solution also implies a level of technological service capacity that is larger than at the CFE.[36] This together with $\mu^{FB} > \mu^{BEH}$—so we are in the increasing (in μ) part of the planner's objective function, $\mu^{CFE} \geq \mu^{STR}$, and $N^{STR} \geq 1 = N^{BEH}$ yields that $P^{BEH} > P^{STR}$.

The results under the assumption of Proposition 7 imply that, if unregulated, the monopoly outcome yields lower liquidity compared to any other alternative. Furthermore, in our simulations, the planner's objective function evaluated at μ^M is always the lowest

[36]To see why notice that μ^{CFE} cannot be higher than μ^{BEH}, as at μ^{BEH} one firm makes zero profit; thus, given single-peakedness of monopoly profit, if there is either one or more firms in the CFE with $\mu^{CFE} > \mu^{BEH}$, profits will be negative. Similarly, it cannot be $\mu^{CFE} = \mu^{BEH}$ because if $N^{CFE} = 1$ then $\mu^{BEH} = \mu^{CFE} = \mu^M$, and by assumption the monopoly profit is positive; if, instead, $N^{CFE} > 1$, then more than one firm shares the revenue that one firm has in the BEH solution, so that its profit must be negative.

compared to the other five alternatives. Thus, both from a liquidity, and welfare point of view the monopoly solution is the worst possible.

5.1 A permanent shock to dealers' risk tolerance

We conclude this section using our model to study the effect of a permanent shock to liquidity providers' risk tolerance. While a lower γ reduces market liquidity (Corollary 1), the latter is at the same time positively affected by an increase in the proportion of FDs (Corollary 1). Thus, as a lower γ leads to a positive shift in the demand for technological services (Corollary 7), it can have an indirect *positive* effect on liquidity. Indeed, in simulations we find that not only can this indirect effect attenuate the direct negative effect of a lower γ on market liquidity, but it can in some cases even overturn it, with a lower γ inducing higher market liquidity overall. We refer the reader to Appendix C for a more detailed analysis.

6 Concluding remarks

We nest a two-period market microstructure model into one of exchange platform competition where trading venues compete à la Cournot in technological services allowing (full) dealers the ability to supply liquidity at both trading rounds to liquidity traders. We show that full dealers have a higher risk bearing capacity compared to those who can only trade in the first round. This implies that as their mass increases, market liquidity and traders' welfare improve. At equilibrium, the mass of full dealers matches the industry technological service capacity. Since at a stable Cournot equilibrium a heightened competition increases industry capacity, this implies that traders' welfare increases in the number of trading venues. We use the model to analyze the welfare effects of different entry regimes. A monopolistic exchange exploits its market power, and under supplies technological services, thereby negatively affecting liquidity and welfare. Allowing competition among trading platforms is beneficial for market quality and (generally) for welfare. However, the market outcome can overprovide or underprovide technological capacity with the corresponding effects on liquidity. If the regulator cannot make transfers to platforms, then entry is never insufficient and the market never underprovides capacity when the benchmark is regulated entry. If, on the other hand, side payments are possible, depending on parameter values entry can also be insufficient. Fee regulation is often superior to entry regulation. Typically, the regulator limits market power by setting a fee low enough so that only one platform can survive and provide a larger (and cheaper) capacity than the market outcome. Both fee and entry regulation are subject to high informational requirements and to lobbying efforts. The choice between them has to weigh the respective costs and benefits.

Our results suggest that exchanges' technological capacity decisions can be an important *driver* of market liquidity, adding to the usual, demand-based factors highlighted by the market microstructure literature (e.g., arbitrage capital, risk bearing capacity of the market). An example is the fact that when a decrease in dealers' risk tolerance increases the demand for technological services, it can prompt a capacity increase which leads in turn to an increase in the mass of FD, attenuating or offsetting the negative direct impact on liquidity. This can provide an explanation for the contrasting liquidity findings of post-crisis regulations aimed at reducing investment banks' trading activities.[37] From this point of view, any argument about market liquidity should also be anchored to the framework in which exchanges interact, and the type of regulatory intervention of the policy maker. Furthermore, we show the limits of the view that aligns liquidity to welfare. Indeed, when excessive entry obtains, even though the market is more liquid, a social planner that internalizes the welfare of exchanges as well as that of market participants, chooses to restrict competition, in this way reducing market liquidity.

Our modelling has integrated industrial organization and market microstructure methods taking technological services as homogeneous. An extension of our approach is to consider that exchanges offer differentiated capacities and introduce asymmetries among exchanges. Differentiation could be both in terms of quality (e.g., speed of connection) and horizontal attributes (e.g., lit vs. dark venues).[38]

[37]Reviewing the literature on the market liquidity impact of post crisis regulations such as the Volcker Rule, an SEC report finds that while dealers in the corporate bond markets have, in aggregate, reduced their capital commitment since the 2007 peak, liquidity measures such as trading activity and average transaction costs have remained flat (see *Access to Capital and Market Liquidity*, SEC report to the US Congress, August 2017).

[38]This would also allow to more directly contrast our results with the differentiated approach of Pagnotta and Philippon (2018).

References

Abel, A. B., J. C. Eberly, and S. Panageas (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica 81*(4), 1455–1481.

Amir, R. (2018). On the Cournot and Bertrand oligopolies and the theory of supermodular games: A survey. *Handbook of Game Theory and Industrial Organization 1*, 40–65.

Anand, A. and K. Venkataraman (2015). Market conditions, fragility and the economics of market making. *Journal of Financial Economics (Forthcoming).*

Baldauf, M. and J. Mollner (2017). Trading in fragmented markets. *Working Paper*.

Biais, B., F. Declerck, and S. Moinas (2015). Who supplies liquidity, how and when? *Working paper*.

Biais, B., T. Foucault, and S. Moinas (2015). Equilibrium fast trading. *Journal of Financial Economics 116*, 292–313.

Brogaard, J., T. Hendershott, and R. Riordan (2014). High frequency trading and price discovery. *Review of Financial Studies 27*, 2267–2306.

Budish, E., P. Cramton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics 139*, 1547–1621.

Budish, E., R. Lee, and J. Shim (2017). Will the market fix the market? a theory of stock exchange competition and innovation. *AFA, Luncheon Talk*.

Cantillon, E. and P.-L. Yin (2011). Competition between exchanges: A research agenda. *International Journal of Industrial Organization 29*(3), 329–336.

Cespa, G. and T. Foucault (2014). Sale of price information by exchanges: Does it promote price discovery? *Management Science 60*(1), 148–165.

Chao, Y., C. Yao, and M. Ye (2017). Why discrete pricing fragments the us stock exchanges and disperses their fee structures. *Working Paper*.

Chowdhry, B. and V. Nanda (1991). Multimarket trading and market liquidity. *Review of Financial Studies 4*(3), 483–511.

Cohen, A. and L. Einav (2007). Estimating risk preferences from deductible choice. *American Economic Review 97*(3), 745–788.

Degryse, H., F. de Jong, and V. van Kervel (2015). The impact of dark trading and visible fragmentation on market quality. *Review of Finance 19*, 1587–1622.

Dennert, J. (1993). Price competition between market makers. *Review of Economic Studies* (60), 735–751.

Ding, S., J. Hanna, and T. Hendershott (2014). How slow is the NBBO? a comparison with direct exchange feeds. *Financial Review* (49), 313–332.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *Journal of Finance 65*(4), 1237–1267.

Foucault, T. and A. J. Menkveld (2008). Competition for order flow and smart order routing systems. *The Journal of Finance 63*(1), 119–158.

Foucault, T., M. Pagano, and A. Röell (2013). *Market Liquidity*. Oxford University Press.

Gabszewicz, J. and J. Thisse (1979). Price competition, quality and income disparities. *Journal of Economic Theory 20*(3), 340–359.

Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *Journal of Finance 43*(3), 617–37.

Harris, J. H. and M. Saad (2014). The sound of silence. *The Financial Review 49*, 203–230.

Hendershott, T. and C. M. Jones (2005). Trade-through prohibitions and market quality. *Journal of Financial Markets* (8), 1–23.

Hendershott, T. and A. J. Menkveld (2014). Price pressures. *Journal of Financial Economics 114*(3), 405 – 423.

Heston, S. L., R. A. Korajczyk, and R. Sadka (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance 65*(4), 1369–1407.

Holden, C. W. and S. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance 69*(4), 1747–1785.

Jones, C. M. (2018). Understanding the market for U.S. equity market data. *Working paper*.

Leland, H. E. (1992). Insider trading: should it be prohibited? *Journal of Political Economy 100*(4), 859–887.

Lewis, M. (2014). *Flash Boys.* London: Allen Lane.

Li, J. (2015). Slow price adjustment to public news in after-hours trading. *Working paper*.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. *RAND Journal of Economics 17*(1), 48–58.

Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics (Forthcoming).*

Okuliar, A. P. (2014). Financial exchange consolidation and antitrust: Is there a need for more intervention? *Antitrust 28*(2), 66–73.

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

Pagnotta, E. and T. Philippon (2018). Competing on speed. *Econometrica (Forthcoming).*

Santos, T. and J. A. Scheinkman (2001). Competition among exchanges. *The Quarterly Journal of Economics 116*(3), 1027–1061.

SEC (2010). Concept release on equity market structure. *Federal Register*.

Shaked, A. and J. Sutton (1982). Relaxing price competition through product differentiation. *Review of Economic Studies 49*(1), 3–13.

UK Competition Commission (2011, November). *BATS Global Markets, Inc/Chi-X Europe Limited, merger inquiry.* UK Competition Commission.

Vayanos, D. and J. Wang (2012). Liquidity and asset returns under asymmetric information and imperfect competition. *Review of Financial Studies 25*(5), 1339–1365.

Vives, X. (1999). *Oligopoly Pricing: old ideas and new tools.* MIT Press.

Vives, X. (2008). *Information and learning in markets: the impact of market microstructure.* Princeton University Press.

A Appendix

The following is a standard result (see, e.g., Vives (2008), Technical Appendix, pp. 382–383) that allows us to compute the unconditional expected utility of market participants.

Lemma 1. *Let the n-dimensional random vector $z \sim N(0, \Sigma)$, and $w = c + b'z + z'Az$, where $c \in \mathbb{R}$, $b \in \mathbb{R}^n$, and A is a $n \times n$ matrix. If the matrix $\Sigma^{-1} + 2\rho A$ is positive definite, and $\rho > 0$, then*

$$E[-\exp\{-\rho w\}] = -|I + 2\rho\Sigma A|^{-1/2} \exp\{-\rho(c - \rho b'(\Sigma + 2\rho A)^{-1} b)\}.$$

Proof of Proposition 1

We start by assuming that at a linear equilibrium prices are given by

$$p_2 = -\Lambda_2 u_2 + \Lambda_{21} u_1 \tag{A.1a}$$

$$p_1 = -\Lambda_1 u_1, \tag{A.1b}$$

with Λ_1, Λ_{21}, and Λ_2 to be determined in equilibrium. In the second period a new mass of liquidity traders with risk-tolerance coefficient $\gamma^L > 0$ enter the market. Because of CARA and normality, the objective function of a second period liquidity trader is given by

$$E[-\exp\{-\pi_2^L/\gamma^L\}|\Omega_2^L] = -\exp\left\{-\frac{1}{\gamma^L}\left(E[\pi_2^L|\Omega_2^L] - \frac{1}{2\gamma^L}\mathrm{Var}[\pi_2^L|\Omega_2^L]\right)\right\}, \tag{A.2}$$

where $\Omega_2^L = \{u_1, u_2\}$, and $\pi_2^L \equiv (v - p_2)x_2^L + u_2 v$. Maximizing (A.2) with respect to x_2^L, yields:

$$X_2^L(u_1, u_2) = \gamma^L \frac{E[v - p_2|\Omega_2^L]}{\mathrm{Var}[v - p_2|\Omega_2^L]} - \frac{\mathrm{Cov}[v - p_2, v|\Omega_2^L]}{\mathrm{Var}[v - p_2|\Omega_2^L]} u_2. \tag{A.3}$$

Using (A.1a):

$$E[v - p_2|\Omega_2^L] = \Lambda_2 u_2 - \Lambda_{21} u_1 \tag{A.4a}$$

$$\mathrm{Var}[v - p_2|\Omega_2^L] = \mathrm{Cov}[v - p_2, v|\Omega_2^L] = \frac{1}{\tau_v}. \tag{A.4b}$$

Substituting (A.4a) and (A.4b) in (A.3) yields

$$X_2^L(u_1, u_2) = a_2 u_2 + b u_1, \tag{A.5}$$

where

$$a_2 = \gamma^L \tau_v \Lambda_2 - 1 \tag{A.6a}$$

$$b = -\gamma^L \tau_v \Lambda_{21}. \tag{A.6b}$$

Consider the sequence of market clearing equations

$$\mu x_1^{FD} + (1-\mu) x_1^{SD} + x_1^L = 0 \tag{A.7a}$$

$$\mu (x_2^{FD} - x_1^{FD}) + x_2^L = 0. \tag{A.7b}$$

Condition (A.7b) highlights the fact that since first period liquidity traders and SD only participate at the first trading round, their positions do not change across dates. Rearrange (A.7a) as follows:

$$(1-\mu) x_1^{SD} + x_1^L = -\mu x_1^{FD}.$$

Substitute the latter in (A.7b):

$$\mu x_2^{FD} + x_2^L + (1-\mu) x_1^{SD} + x_1^L = 0. \tag{A.8}$$

To pin down p_2, we need the second period strategy of FD and the first period strategies of SD and liquidity traders. Starting from the former, because of CARA and normality, the expected utility of a FD is given by:

$$E\Big[-\exp\Big\{ -\frac{1}{\gamma}\Big((p_2 - p_1) x_1^{FD} + (v - p_2) x_2^{FD} \Big) \Big\} | p_1, p_2 \Big] = \tag{A.9}$$
$$= \exp\Big\{ -\frac{1}{\gamma}(p_2 - p_1) x_1^{FD} \Big\} \Big(-\exp\Big\{ -\frac{1}{\gamma}\Big(E[v - p_2 | p_1, p_2] x_2^{FD} - \frac{(x_2^{FD})^2}{2\gamma} \mathrm{Var}[v - p_2 | p_1, p_2] \Big) \Big\} \Big),$$

For given x_1^{FD} the above is a concave function of x_2^{FD}. Maximizing with respect to x_2^{FD} yields:

$$X_2^{FD}(p_1, p_2) = -\gamma \tau_v p_2. \tag{A.10}$$

Similarly, due to CARA and normality, in the first period a traditional market maker maximizes

$$E\left[-\exp\left\{-\frac{1}{\gamma}(v-p_1)x_1^{SD}\right\}|p_1\right]=-\exp\left\{-\frac{1}{\gamma}\left(E[v-p_1|p_1]x_1^{SD}-\frac{(x_1^{SD})^2}{2\gamma}\mathrm{Var}[v-p_1|p_1]\right)\right\}. \tag{A.11}$$

Hence, his strategy is given by

$$X_1^{SD}(p_1)=-\gamma\tau_v p_1. \tag{A.12}$$

Finally, consider a first period liquidity trader. CARA and normality imply

$$E[-\exp\{-\pi_1^L/\gamma^L\}]=-\exp\left\{-\frac{1}{\gamma}\left(E[\pi_1^L|u_1]-\frac{1}{2\gamma^L}\mathrm{Var}[\pi_1^L|u_1]\right)\right\}, \tag{A.13}$$

where $\pi_1^L \equiv (v-p_1)x_1^L+u_1v$. Maximizing (A.13) with respect to x_1^L, and solving for the optimal strategy, yields

$$X_1^L(u_1)=\gamma^L\frac{E[v-p_1|u_1]}{\mathrm{Var}[v-p_1|u_1]}-\frac{\mathrm{Cov}[v-p_1,v|u_1]}{\mathrm{Var}[v-p_1|u_1]}u_1. \tag{A.14}$$

Using (A.1b):

$$E[v-p_1|u_1]=\Lambda_1 u_1 \tag{A.15a}$$

$$\mathrm{Cov}[v-p_1,v|u_1]=\frac{1}{\tau_v}. \tag{A.15b}$$

Substituting the above in (A.14) yields

$$X_1^L(u_1)=a_1u_1, \tag{A.16}$$

where

$$a_1=\gamma^L\tau_v\Lambda_1-1. \tag{A.17}$$

Substituting (A.5), (A.10), (A.12), and (A.16) in (A.8) and solving for p_2 yields

$$p_2=-\underbrace{\frac{1-\gamma^L\tau_v\Lambda_2}{\mu\gamma\tau_v}}_{\Lambda_2}u_2+\underbrace{\frac{((1-\mu)\gamma+\gamma^L)\tau_v\Lambda_1-1-\gamma^L\tau_v\Lambda_{21}}{\mu\gamma\tau_v}}_{\Lambda_{21}}u_1. \tag{A.18}$$

Identifying the price coefficients:

$$\Lambda_2 = \frac{1}{(\mu\gamma + \gamma^L)\tau_v} \tag{A.19a}$$

$$\Lambda_{21} = \Lambda_2 \left(((1-\mu)\gamma + \gamma^L)\tau_v \Lambda_1 - 1\right). \tag{A.19b}$$

Substituting the above expressions in (A.18), and using (A.12) yields:

$$p_2 = -\Lambda_2 u_2 + \Lambda_2 \left((1-\mu)x_1^{SD} + x_1^L\right).$$

Consider now the first period. We start by characterizing the strategy of a FD. Substituting (A.10) in (A.9), rearranging, and applying Lemma 1 yields the following expression for the first period objective function of a FD:

$$E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})|u_1] = -\left(1 + \frac{\mathrm{Var}[p_2|u_1]}{\mathrm{Var}[v]}\right)^{-1/2} \times \tag{A.20}$$
$$\exp\left\{-\frac{1}{\gamma}\left(\frac{\gamma\tau_v}{2}\nu^2 + (\nu - p_1)x_1^{FD} - \frac{(x_1^{FD} + \gamma\tau_v\nu)^2}{2\gamma}\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)^{-1}\right)\right\},$$

where, due to (A.1a) and (A.1b)

$$\nu \equiv E[p_2|u_1] = \Lambda_{21}u_1 \tag{A.21a}$$

$$\mathrm{Var}[p_2|u_1] = \frac{\Lambda_2^2}{\tau_u}. \tag{A.21b}$$

Maximizing (A.20) with respect to x_1^{FD} and solving for the first period strategy yields

$$X_1^{FD}(p_1) = \gamma\frac{E[p_2|u_1]}{\mathrm{Var}[p_2|u_1]} - \gamma\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)p_1 \tag{A.22}$$
$$= \gamma\frac{\Lambda_{21}\tau_u}{\Lambda_2^2}u_1 - \gamma\frac{\tau_u + \Lambda_2^2\tau_v}{\Lambda_2^2}p_1.$$

Substituting (A.12), (A.16), and (A.22) in (A.7a) and solving for the price yields $p_1 = -\Lambda_1 u_1$, where

$$\Lambda_1 = \left(\left(1 + \frac{\mu\gamma^L\tau_u}{\Lambda_2 + \mu\gamma\tau_u}\right)\gamma + \gamma^L\right)^{-1}\frac{1}{\tau_v}. \tag{A.23}$$

The remaining equilibrium coefficients are as follows:

$$a_1 = \gamma^L \Lambda_1 \tau_v - 1 \tag{A.24}$$

$$a_2 = -\frac{\mu\gamma}{\mu\gamma + \gamma^L} \tag{A.25}$$

$$b = -\gamma^L \tau_v \Lambda_{21} \tag{A.26}$$

$$\Lambda_{21} = -\frac{\mu\gamma(\Lambda_2^2 \tau_v + \tau_u)}{\mu\gamma\tau_u + \Lambda_2}\Lambda_1 \tag{A.27}$$

$$\text{Var}[p_2|u_1] = \frac{\Lambda_2^2}{\tau_u}, \tag{A.28}$$

where Λ_2 is given by (A.19a). An explicit expression for Λ_1 can be obtained substituting (A.19a) into (A.23):

$$\Lambda_1 = \frac{1 + (\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v}{(\gamma + \gamma^L + (\gamma + 2\gamma^L)(\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v)\tau_v}. \tag{A.29}$$

Finally, substituting (A.19a) and (A.29) in (A.27) yields

$$\Lambda_{21} + \Lambda_1 = \frac{\gamma^L}{\tau_v(\gamma\mu + \gamma^L)(\gamma\mu\tau_u\tau_v(\gamma + 2\gamma^L)(\gamma\mu + \gamma^L) + \gamma + \gamma^L)} > 0. \tag{A.30}$$

□

Proof of Corollary 1

The comparative static effect of μ and γ on Λ_2 follows immediately from (A.19a). For Λ_1, differentiating (A.29) with respect to μ and γ yields:

$$\frac{\partial \Lambda_1}{\partial \mu} = -\frac{(2\mu\gamma + \gamma^L)\gamma\gamma^L\tau_u}{(\gamma + \gamma^L + (\gamma + 2\gamma^L)(\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v)^2} < 0$$

$$\frac{\partial \Lambda_1}{\partial \gamma} = -\frac{1 + (\gamma^2\mu\tau_u\tau_v(\gamma\mu + \gamma^L)^2 + 2\gamma^2\mu + 2\gamma\gamma^L\mu + 2\gamma\gamma^L + (\gamma^L)^2)\mu\tau_u\tau_v}{\tau_v(\gamma\mu\tau_u\tau_v(\gamma + 2\gamma^L)(\gamma\mu + \gamma^L) + \gamma + \gamma^L)^2} < 0,$$

which proves our result.

□

Proof of Corollary 2

The first part of the corollary follows from (10). Also, since Λ_t is decreasing in μ, because of (6d), $|a_t|$ is increasing in μ. Finally, substituting (A.27) in (A.26) and rearranging yields

$$b = \frac{\mu\gamma\gamma^L(1 + (\mu\gamma + \gamma^L)^2\tau_u\tau_v)}{(\mu\gamma + \gamma^L)(\gamma + \gamma^L + (\gamma + 2\gamma^L)\mu\gamma\tau_u\tau_v)},$$

which is increasing in μ. □

Proof of Corollary 3

Computing the covariance between first and second period returns and using (A.23), and (A.27) yields

$$\begin{aligned}\text{Cov}[p_2 - p_1, p_1] &= -\Lambda_1 (\Lambda_1 + \Lambda_{21}) \tau_u^{-1} \\ &= -\frac{\gamma^L\Lambda_1\Lambda_2}{(\gamma + \gamma^L + (\gamma + 2\gamma^L)(\mu\gamma + \gamma^L)\mu\gamma\tau_u\tau_v)\tau_u},\end{aligned}$$

which, in view of the fact that Λ_t^* is decreasing in μ, proves the result. □

Proof of Proposition 2

We start by obtaining an expression for the unconditional expected utility of SDs and FDs. Because of CARA and normality, a dealer's conditional expected utility evaluated at the optimal strategy is given by

$$\begin{aligned}E[U((v - p_1)x_1^{SD})|p_1] &= -\exp\left\{-\frac{(E[v|p_1] - p_1)^2}{2\text{Var}[v]}\right\} \\ &= -\exp\left\{-\frac{\tau_v\Lambda_1^2}{2}u_1^2\right\}. \end{aligned} \tag{A.31}$$

Thus, traditional dealers derive utility from the expected, long term capital gain obtained supplying liquidity to first period hedgers.

$$\begin{aligned}EU^{SD} \equiv E\left[U\left((v - p_1)x_1^{SD}\right)\right] &= -\left(1 + \frac{\text{Var}[p_1]}{\text{Var}[v]}\right)^{-1/2} \\ &= -\left(\frac{\tau_{u_1}}{\tau_{u_1} + \tau_v\Lambda_1^2}\right)^{1/2}, \end{aligned} \tag{A.32}$$

and

$$CE^{SD} = \frac{\gamma}{2} \ln \left(1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right). \tag{A.33}$$

Differentiating CE^{SD} with respect to μ yields:

$$\begin{aligned} \frac{\partial CE^{SD}}{\partial \mu} &= \frac{\gamma \tau_v}{2} \left(1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right)^{-1} \frac{\partial \text{Var}[p_1]}{\partial \mu} \\ &= \frac{\gamma \tau_v}{2\tau_{u_1}} \left(1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right)^{-1} 2\Lambda_1 \frac{\partial \Lambda_1}{\partial \mu} < 0, \end{aligned} \tag{A.34}$$

since Λ_1 is decreasing in μ.

Turning to FDs. Replacing (A.22) in (A.20) and rearranging yields

$$E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})|u_1] = - \left(1 + \frac{\text{Var}[p_2|u_1]}{\text{Var}[v]} \right)^{-1/2} \times \exp \left\{ -\frac{g(u_1)}{\gamma} \right\}, \tag{A.35}$$

where

$$g(u_1) = \frac{\gamma}{2} \left(\frac{(E[p_2|p_1] - p_1)^2}{\text{Var}[p_2|p_1]} + \frac{(E[v|p_1] - p_1)^2}{\text{Var}[v]} \right).$$

The argument at the exponential of (A.35) is a quadratic form of the first period endowment shock. We can therefore apply Lemma 1 and obtain

$$\begin{aligned} EU^{FD} &\equiv E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})] = \\ &= - \left(1 + \frac{\text{Var}[p_2|p_1]}{\text{Var}[v]} \right)^{-1/2} \left(1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} + \frac{\text{Var}[E[p_2|p_1] - p_1]}{\text{Var}[p_2|p_1]} \right)^{-1/2}, \end{aligned} \tag{A.36}$$

where, because of (A.21a),

$$\text{Var}\left[E[p_2 - p_1|p_1]\right] = \left(\Lambda_{21} + \Lambda_1\right)^2 \tau_u^{-1}, \tag{A.37}$$

so that:

$$\frac{\text{Var}[E[p_2 - p_1|u_1]]}{\text{Var}[p_2|u_1]} = \left(\frac{\Lambda_{21} + \Lambda_1}{\Lambda_2} \right)^2.$$

Therefore, we obtain

$$CE^{FD} = \frac{\gamma}{2}\left\{\ln\left(1+\frac{(\Lambda_2)^2\tau_v}{\tau_u}\right)+\ln\left(1+\frac{(\Lambda_1)^2\tau_v}{\tau_u}+\left(\frac{\Lambda_{21}+\Lambda_1}{\Lambda_2}\right)^2\right)\right\}. \tag{A.38}$$

Computing,

$$\frac{\Lambda_{21}+\Lambda_1}{\Lambda_2} = \frac{\gamma^L}{\gamma+\gamma^L+(\gamma+2\gamma^L)(\mu\gamma+\gamma^L)\mu\gamma\tau_u\tau_v}. \tag{A.39}$$

Thus, the arguments of the logarithms in (A.38) are decreasing in μ, which proves that CE^{FD} is decreasing in μ.

Finally, note that taking the limits for $\mu \to 0$ and $\mu \to 1$ in (A.33) and (A.38) yields

$$\lim_{\mu\to 0} CE^{SD} = \frac{\gamma}{2}\ln\left(1+\frac{1}{(\gamma+\gamma^L)^2\tau_u\tau_v}\right)$$
$$\lim_{\mu\to 1} CE^{FD} = \frac{\gamma}{2}\left\{\ln\left(1+\frac{1}{(\gamma+\gamma^L)^2\tau_u\tau_v}\right)+\ln\left(1+\frac{(\Lambda_1)^2\tau_v}{\tau_u}+\left(\frac{\Lambda_{21}+\Lambda_1}{\Lambda_2}\right)^2\right)\right\},$$

which proves the last part of the corollary. □

Proof of Proposition 3

Consider now first period liquidity traders. Evaluating the objective function at optimum and rearranging yields

$$-\exp\left\{-\frac{1}{\gamma^L}\left(E[\pi_1^L|u_1]-\frac{1}{2\gamma^L}\mathrm{Var}[\pi_1^L|u_1]\right)\right\} = -\exp\left\{-\frac{u_1^2}{\gamma^L}\left(\frac{a_1^2-1}{2\gamma^L\tau_v}\right)\right\},$$

where $u_1 \sim N(0, \tau_{u_1}^{-1})$. The argument at the exponential is a quadratic form of a normal random variable. Therefore, applying again Lemma 1 yields

$$E\left[-\exp\left\{\frac{\pi_1^L}{\gamma^L}\right\}\right] = -\left(\frac{(\gamma^L)^2\tau_u\tau_v}{(\gamma^L)^2\tau_u\tau_v-1+a_1^2}\right)^{1/2}, \tag{A.40}$$

so that

$$CE_1^L = \frac{\gamma^L}{2}\ln\left(1+\frac{a_1^2-1}{(\gamma^L)^2\tau_u\tau_v}\right). \tag{A.41}$$

Note that a higher a_1^2 increases traders' expected utility, and thus increases their payoff.

Next, for second period liquidity traders, substituting the optimal strategy (A.3) in the objective function (A.2) yields

$$
\begin{aligned}
E\Bigg[- \exp\left\{-\frac{\pi_2^L}{\gamma^L}\right\}|\Omega_2^L\Bigg] &= -\exp\left\{-\frac{1}{\gamma^L}\left(\frac{(x_2^L)^2 - u_2^2}{2\gamma^L\tau_v}\right)\right\} \\
&= -\exp\left\{-\frac{1}{\gamma^L}\begin{pmatrix} x_2^L & u_2 \end{pmatrix}\left(\frac{1}{2\gamma_2^L\tau_v}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right)\begin{pmatrix} x_2^L \\ u_2 \end{pmatrix}\right\}.
\end{aligned} \tag{A.42}
$$

The argument of the exponential is a quadratic form of the normally distributed random vector

$$
\begin{pmatrix} x_2^L & u_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 & 0 \end{pmatrix}, \Sigma\right),
$$

where

$$
\Sigma \equiv \begin{pmatrix} \mathrm{Var}[x_2^L] & a_2\mathrm{Var}[u_2] \\ a_2\mathrm{Var}[u_2] & \mathrm{Var}[u_2] \end{pmatrix}. \tag{A.43}
$$

Therefore, we can again apply Lemma 1 to (A.42), obtaining

$$
E\left[E\left[-\exp\left\{-\frac{\pi_2^L}{\gamma^L}\right\}|\Omega_2^L\right]\right] = -|I + (2/\gamma^L)\Sigma A|^{-1/2}, \tag{A.44}
$$

where

$$
A \equiv \frac{1}{2\gamma^L\tau_v}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{A.45}
$$

$$
\mathrm{Var}[x_2^L] = \frac{a_2^2 + b^2}{\tau_u}. \tag{A.46}
$$

Substituting (A.43), (A.45), and (A.46) in (A.44) and computing the certainty equivalent, yields:

$$
CE_2^L = \frac{\gamma^L}{2}\ln\left(1 + \frac{a_2^2 - 1}{(\gamma^L)^2\tau_u\tau_v} + \frac{b^2((\gamma^L)^2\tau_u\tau_v - 1)}{(\gamma^L)^4\tau_u^2\tau_v^2}\right). \tag{A.47}
$$

For $\mu = 0$, $b = 0$ and, in view of Corollary 2, $CE_1^L > CE_2^L$. The same condition holds when evaluating (A.41) and (A.47) at $\mu = 1$. As CE_t^L is increasing in μ, we have that for all $\mu \in (0, 1]$, $CE_1^L(\mu) > CE_2^L(\mu)$. □

Proof of Corollary 4

We need to prove that:

$$
\frac{\partial CE_1^L(\mu)}{\partial\mu} > -\frac{\partial CE^{SD}(\mu)}{\partial\mu}.
$$

Computing:

$$\frac{\partial CE_1^L(\mu)}{\partial \mu} = \frac{\gamma^L a_1 a_1'}{(\gamma^L)^2 \tau_u \tau_v - 1 + a_1^2}, \tag{A.48}$$

where a_1' the partial derivative of a_1 with respect to μ, and

$$\frac{\partial CE^{SD}(\mu)}{\partial \mu} = \frac{\gamma(1+a_1)a_1'}{(\gamma^L)^2 \tau_u \tau_v + (1+a_1)^2}. \tag{A.49}$$

First, note that the denominator in (A.49) is higher than the one in (A.48), and they are both positive. Next, comparing the numerators in the above expressions yields:

$$\gamma^L a_1 a_1' > -\gamma(1+a_1)a_1' \iff \underbrace{(\gamma^L a_1 + \gamma(1+a_1))}_{<0} \underbrace{a_1'}_{<0} > 0,$$

as can be checked by substituting (A.24) in the above. Thus, the LHS of the inequality to be proved has a higher (positive) nominator and a lower (positive) denominator compared to the (positive) nominator and denominator of the RHS, and the inequality follows. □

Proof of Corollary 5

The first part of the result follows immediately from (17), and Corollary 4. Next, because of Propositions 2 and 3, $GW(1) > \lim_{\mu \to 0} GW(\mu)$, which rules out the possibility that gross welfare is maximized at $\mu \approx 0$. □

Proof of Corollary 6

Note that because of (A.39), we can write

$$\frac{\Lambda_{21} + \Lambda_1}{\Lambda_2} = \frac{\Lambda_1 \gamma^L \tau_v}{1 + \mu\gamma(\mu\gamma + \gamma^L)\tau_u \tau_v}.$$

Therefore, substituting the expressions for dealers' payoffs in (18), we have:

$$\begin{aligned} \phi(\mu) &= CE^{FD} - CE^D \\ &= \frac{\gamma}{2}\left\{ \ln\left(1 + \frac{\Lambda_2^2 \tau_v}{\tau_u}\right) + \ln\left(1 + \frac{\Lambda_1^2 \tau_v}{\tau_u} K\right) - \ln\left(1 + \frac{\Lambda_1^2 \tau_v}{\tau_u}\right) \right\} > 0. \end{aligned} \tag{A.50}$$

where $K = 1 + (\gamma^L/(1+\mu\gamma(\mu\gamma+\gamma^L)\tau_u\tau_v))^2\tau_u\tau_v > 1$, and decreasing in μ. The first term inside curly braces in the above expression is decreasing in μ since Λ_2 is decreasing in μ. The difference between the second and third terms can be written as follows:

$$\ln\left(1+\frac{\Lambda_1^2\tau_v}{\tau_u}K\right) - \ln\left(1+\frac{\Lambda_1^2\tau_v}{\tau_u}\right) = \ln\left(\frac{\tau_u+\Lambda_1^2\tau_v K}{\tau_u+\Lambda_1^2\tau_v}\right).$$

Differentiating the above logarithm and rearranging yields:

$$\frac{\tau_v\Lambda_1}{(\tau_u+\Lambda_1^2\tau_v K)(\tau_u+\Lambda_1^2\tau_v)}\left(2(K-1)\tau_u\frac{\partial\Lambda_1}{\partial\mu}+(\tau_u+\Lambda_1^2\tau_u)\Lambda_1\frac{\partial K}{\partial\mu}\right)<0,$$

since $K > 1$, and both Λ_1 and K are decreasing in μ. □

Proof of Corollary 7

As shown in the text following Corollary 7, the derivative of $\phi(\mu;\gamma)$ with respect to γ can be written as the sum of two components, the first one being positive. Consider now the effect of the change in γ on the dealers' expected utilities' ratio. Computing this derivative yields:

$$\begin{aligned}\frac{\partial(EU^{FD}/EU^{SD})}{\partial\gamma} \propto \Big(& -\tau_u\tau_v(\gamma\mu+\gamma^L)^3\big(\gamma\mu\tau_u^2\tau_v^2(\gamma\mu+\gamma^L)\big(\gamma^2\big(\mu^2+2\big)+\gamma\gamma^L(\mu+6)+4(\gamma^L)^2\big) \\ & +\gamma^2\mu^2\tau_u^3\tau_v^3(\gamma+2\gamma^L)^2(\gamma\mu+\gamma^L)^2+\tau_u\tau_v\big(2\gamma^2\mu^2+2\gamma\gamma^L\mu+(\gamma+\gamma^L)^2\big)+1\big)^2\Big)^{-1}\times \\ \Big(2\big(&\gamma^4\mu^5\tau_u^6\tau_v^6(\gamma+2\gamma^L)^4(\gamma\mu+\gamma^L)^4+\mu\tau_u\tau_v\big(\gamma^2\big(4\mu^2+2\big) \\ & +4\gamma\gamma^L(\mu+1)+3(\gamma^L)^2\big)+\tau_u^2\tau_v^2\big(\gamma^4\big(6\mu^5+8\mu^3+\mu\big)+4\gamma^3\gamma^L\mu(\mu(\mu+1)(3\mu+2)+ \\ & +2\gamma^2(\gamma^L)^2\mu(\mu(11\mu+10)+4)+\gamma(\gamma^L)^3(\mu(17\mu+7)+1)+(\gamma^L)^4(3\mu+1)\big)+ \\ & \gamma\mu^2\tau_u^5\tau_v^5(\gamma+2\gamma^L)(\gamma\mu+\gamma^L)^3\big(2\gamma^5\mu^2\big(\mu^2+2\big)+2\gamma^4\gamma^L\mu^2\big(2\mu^2+\mu+10\big)+ \\ & \gamma^3(\gamma^L)^2\mu^2(4\mu+35)+5\gamma^2(\gamma^L)^3\mu(4\mu+1)+2\gamma(\gamma^L)^4(3\mu+1)+2(\gamma^L)^5\big)+ \\ & \tau_u^3\tau_v^3(\gamma\mu+\gamma^L)\big(4\gamma^5\big(\mu^6+3\mu^4+\mu^2\big)+4\gamma^4\gamma^L\mu^2(\mu(\mu(2\mu+9)+3)+5)+ \\ & \gamma^3(\gamma^L)^2\mu^2(\mu(35\mu+36)+43)+\gamma^2(\gamma^L)^3\mu(\mu(35\mu+44)+5)+ \\ & \gamma(\gamma^L)^4(\mu(21\mu+8)+1)+(\gamma^L)^5(3\mu+1)\big)+\mu\tau_u^4\tau_v^4(\gamma\mu+\gamma^L)^2\big(\gamma^6\mu^2\big(\mu^4+8\mu^2+6\big)+ \\ & 2\gamma^5\gamma^L\mu^2(\mu(\mu(\mu+14)+4)+18)+\gamma^4(\gamma^L)^2\mu^2(\mu(27\mu+28)+86)+ \\ & \gamma^3(\gamma^L)^3\mu(\mu(29\mu+95)+9)+\gamma^2(\gamma^L)^4(\mu(44\mu+21)+3)+ \\ & \gamma(\gamma^L)^5(11\mu+6)+2(\gamma^L)^6\big)+\mu\big)\Big)<0, \qquad \text{(A.51)}\end{aligned}$$

proving our claim. $\square$

Proof of Proposition 5

In the First Best case, for given μ, the objective function (28) is decreasing in N. Thus, to economise on fixed costs, the planner allows a monopolistic exchange to provide trading services, $N^{FB} = 1$. To see that $\mu^{FB} > \mu^{M}$ (equal in the case where $\mu^{M} = 1$), evaluate the derivative of $\mathcal{P}(\mu, 1)$ at μ^{M} to obtain:

$$\left.\frac{\partial \mathcal{P}(\mu,1)}{\partial \mu}\right|_{\mu=\mu^{M}} = \underbrace{\frac{\partial \pi}{\partial \mu}}_{=0} + \underbrace{\frac{\partial \psi(\mu)}{\partial \mu}}_{>0} > 0 \implies \mu^{FB} > \mu^{M},$$

where $\partial\psi(\mu)/\partial\mu$ is positive by Proposition 3 and Corollary 4.

To see why $N^{BEH} = 1$ and $\mu^{FB} > \mu^{BEH}$ if at μ^{FB} the monopoly profit is negative, look at the proofs of Lemma 2 and Proposition 7 below.

We now show that $\mu^{BEH} \geq \mu^{M}$. Given that at the monopoly optimal μ^{M} the monopoly profit is positive, $(\phi(\mu^{M}) - c)\mu^{M} > f$. Suppose by contradiction that $\mu^{BEH} < \mu^{M}$. The pair (μ^{BEH}, N^{BEH}) satisfies the two BEH constraints: $(\phi(\mu^{BEH}) - c)\mu^{BEH}/N^{BEH} \geq f \geq (\phi(\mu^{BEH}) - c)\mu^{BEH}/(N^{BEH} + 1)$. Then, if the right constraint does not bind, by continuity, single-peakedness of monopoly profit at μ^{M} and $\mu^{BEH} < \mu^{M}$ there exists $\epsilon > 0$ small enough such that $\mu^{BEH} + \epsilon \leq \mu^{M}$ and $f > (\phi(\mu^{BEH} + \epsilon) - c)(\mu^{BEH} + \epsilon)/(N^{BEH} + 1)$. $\mu^{BEH} + \epsilon \leq \mu^{M}$ guarantees that $(\phi(\mu^{BEH} + \epsilon) - c)(\mu^{BEH} + \epsilon)/N^{BEH} > f$. Thus, the BEH constaints are satisfied at $(\mu^{BEH} + \epsilon, N^{BEH})$ and given that $\mu^{BEH} + \epsilon \leq \mu^{M} \leq \mu^{FB}$ and single-peakedness of $\mathcal{P}(\mu)$ at μ^{FB}, $\mathcal{P}(\mu^{BEH} + \epsilon, N^{BEH}) > \mathcal{P}(\mu^{BEH}, N^{BEH})$, a contradiction to (μ^{BEH}, N^{BEH}) being the BEH solution. If the right constraint binds, $(\phi(\mu^{BEH}) - c)\mu^{BEH}/N^{BEH} > f = (\phi(\mu^{BEH}) - c)\mu^{BEH}/(N^{BEH} + 1)$ and the planner's function takes the value $\mathcal{P}(\mu^{BEH}, N^{BEH}) = \psi(\mu^{BEH}) + f$. We can increase N^{BEH} by one and μ^{BEH} to $\mu^{BEH'} > \mu^{BEH}$ such that $(\phi(\mu^{BEH'}) - c)\mu^{BEH'}/(N^{BEH} + 1) \geq f \geq (\phi(\mu^{BEH'}) - c)\mu^{BEH'}/(N^{BEH} + 2)$ and the planner's function will take the value $\mathcal{P}(\mu^{BEH'}, N^{BEH} + 1) \geq \psi(\mu^{BEH'}) + f > \psi(\mu^{BEH}) + f = \mathcal{P}(\mu^{BEH}, N^{BEH})$, given that $\psi(\mu)' > 0$ and $\mu^{BEH'} > \mu^{BEH}$, a contradiction.

Last, for the case where at μ^{FB} the monopoly profit is non-positive, $N^{BEH} = 1$. The case $\mu^{BEH} < \mu^{M}$ has already been rejected. Suppose by contradiction that $\mu^{BEH} = \mu^{M} <$

μ^{FB}, and evaluate the derivative of $\mathcal{P}(\mu, 1)$ with respect to μ at μ^{BEH} to obtain:

$$\begin{aligned}\left.\frac{\partial \mathcal{P}(\mu,1)}{\partial \mu}\right|_{\mu=\mu^{BEH}} &= \left.\left[\frac{\partial((\phi(\mu)-c)\mu - f)}{\partial \mu} + \frac{\partial \psi(\mu)}{\partial \mu}\right]\right|_{\mu=\mu^{BEH}} \\ &= \left.\left[\underbrace{\frac{\partial((\phi(\mu)-c)\mu)}{\partial \mu}}_{=0} + \underbrace{\frac{\partial \psi(\mu)}{\partial \mu}}_{>0}\right]\right|_{\mu=\mu^{BEH}} > 0,\end{aligned}$$

where the first term is zero given that $\mu^{BEH} = \mu^M$. Given that at $\mu^{BEH} = \mu^M$ profits are positive and $N^{BEH} = 1$, there exists $\epsilon > 0$ such that (i) by continuity at $(\mu^{BEH} + \epsilon, 1)$ the firm makes positive profits so the left BEH constraint is satisfied, (ii) by single-peakedness of monopoly profit and since $\mu^{BEH} = \mu^M$, $(\mu^{BEH} + \epsilon, 1)$ also satisfies the right BEH constraint, and (iii) $\mu^{BEH} < \mu^{BEH} + \epsilon \leq \mu^{FB}$, so that by single-peakedness of $\mathcal{P}(\mu, 1)$, $\mathcal{P}(\mu^{BEH} + \epsilon, 1) > \mathcal{P}(\mu^{BEH}, 1)$, a contradiction to $(\mu^{BEH}, 1)$ being the BEH solution. □

Proof of Proposition 6

Let $\mu^C(N)$ denote the total co-location capacity at a symmetric Cournot equilibrium for a given number of exchanges N. The objective function of a planner that controls entry can be written as follows:

$$\mathcal{P}(\mu^C(N), N) = N\pi_i(\mu^C(N)) + \psi(\mu^C(N)), \tag{A.52}$$

where $\psi(\mu^C(N))$ denotes the welfare of other market participants at the Cournot solution:

$$\psi(\mu^C(N)) = CE^{SD}(\mu^C(N)) + CE_1^L(\mu^C(N)) + CE_2^L(\mu^C(N)).$$

Consider now the derivative of the planner's objective function with respect to N, and evaluate it at N^{CFE}:

$$\begin{aligned}\left.\frac{\partial \mathcal{P}(\mu^C(N), N)}{\partial N}\right|_{N=N^{CFE}} &= \left.\underbrace{\pi_i(\mu^C(N), N)}_{=0}\right|_{N=N^{CFE}} \\ &\quad + N^{CFE} \left.\underbrace{\frac{\partial \pi_i(\mu^C(N), N)}{\partial N}}_{<0}\right|_{N=N^{CFE}} + \psi'(\mu^C(N)) \left.\underbrace{\frac{\partial \mu^C(N)}{\partial N}}_{>0}\right|_{N=N^{CFE}}.\end{aligned} \tag{A.53}$$

The first term on the right hand side of (A.53) is null at N^{CFE} (modulo the integer constraint). At a stable, symmetric Cournot equilibrium, an increase in N has a negative impact on the profit of each exchange, and a positive impact on the aggregate technological capacity (see, e.g., Vives (1999)). Therefore, the second and third terms are respectively negative and positive. Given our definitions, N^{CFE} is the largest N such that platforms break even. N^{STR}, instead, reflects the planner's choice of N in Cournot equilibria that keep exchanges from making negative profits and maximizes welfare. Hence, it can only be that

$$N^{CFE} \geq N^{STR} \text{ and } \mu^{CFE} \geq \mu^{STR},$$

since a planner can decide to restrict entry. At a $USTR$, the planner can make side payments to an unprofitable exchange. This has two implications: first, the planner can push entry beyond the level at which platforms break even, so that

$$N^{USTR} \geq N^{STR} \text{ and } \mu^{USTR} \geq \mu^{STR}.$$

Additionally, depending on which of the two terms in (A.53) prevails, we have

$$\left.\frac{\partial \mathcal{P}(\mu^C(N), N)}{\partial N}\right|_{N=N^{CFE}} \gtreqless 0 \implies N^{CFE} \lesseqgtr N^{USTR}.$$

Finally, $\mu^C(N) \geq \mu^M$, for $N \geq 1$ because at a stable CFE the total capacity is an increasing function of the number of platforms. A similar argument holds at both the STR and USTR, since in this case the planner picks N subject to μ being a Cournot equilibrium

We have that $\mathcal{P}^{USTR} \geq \mathcal{P}^{STR}$, because STR imposes an additional constraint on the planner's objective function compared to STR. Finally, $\mathcal{P}^{STR} \geq \mathcal{P}^{CFE}$, because CFE does not account for other traders' welfare, and the planner may choose to favour these market participants when at the margin this creates a larger increase in $GW(\mu)$. □

Lemma 2. $\pi^M(\mu^{FB}) \leq 0 \implies \pi^{BEH}(\mu^{BEH}) = 0$ *and the converse is also true generically.*

Proof. First we prove the direction $\implies$. Since $\pi^M(\mu^{FB}) \leq 0$, then, given that the monopoly profit is single-peaked, the BEH constraints can only be satisfied for $\mu \leq \mu^{FB}$. Note that for a given (aggregate) μ, the profit (given that it is non-negative) of an exchange is non-increasing in N, so for a given μ, $N = 1$ maximizes profit. Then, given that $\mathcal{P}(\mu)$ is single-peaked at μ^{FB}, it is optimal for μ^{BEH} to be set as large as possible with $N^{BEH} = 1$, so that $\pi^{BEH}(\mu^{BEH}) = 0$.

Next we prove the opposite direction ($\Longleftarrow$) generically by proving the contrapositive. Suppose that at μ^{FB} the monopoly profit is positive, that is $(\phi(\mu^{FB}) - c)\mu^{FB} > f$, then:

1. If $(\phi(\mu^{FB}) - c)\mu^{FB}/2 \leq f$, then $\mu^{BEH} = \mu^{FB}$, $N^{BEH} = N^{FB} = 1$ and thus $\pi^{BEH}(\mu^{BEH}) > 0$.

2. If $(\phi(\mu^{FB}) - c)\mu^{FB}/2 > f$, then given that from Proposition 5 we know that $\mu^{BEH} \geq \mu^{M}$, and monopoly profit is single peaked at μ^{M} (thus, we work in the decreasing part of monopoly profit), we only need to examine whether it is optimal to choose $N^{BEH} > 1$ and/or $\mu^{BEH} > \mu^{FB}$ in order to satisfy the right BEH constraint.

   (a) Assume that for $N > 2$, we do not have that $(\phi(\mu^{FB}) - c)\mu^{FB}/N = f$. We prove that it cannot be $N^{BEH} > 1$ with $\mu^{BEH} \leq \mu^{FB}$. Suppose by contradiction that the latter holds. Then with $\mu^{BEH} = \mu^{FB}$, the left BEH constraint cannot bind and $\pi^{BEH}(\mu^{BEH}) > 0$. If $\mu^{BEH} < \mu^{FB}$, then the left BEH constraint must bind (and the right not): $(\phi(\mu^{BEH}) - c)\mu^{BEH}/N^{BEH} = f > (\phi(\mu^{BEH}) - c)\mu^{BEH}/(N^{BEH} + 1)$. (To see this, observe that if the left did not bind, we could increase μ^{BEH} to bring it closer to μ^{FB} with both constraints still satisfied.) But then consider a new candidate BEH solution resulting from reducing N^{BEH} by one and increasing μ^{BEH} to $\mu^{BEH'} > \mu^{BEH}$. From the previous left BEH constraint we know that the new right BEH constraint will not bind. Thus, it has to either be that $\mu^{BEH'} = \mu^{FB}$, in which case (μ^{BEH}, N^{BEH}) is rejected as a solution and we have a contradiction, or that the new left BEH constraint will bind—to see the latter, it suffices to observe that if neither constraint binds and $\mu^{BEH'} \neq \mu^{FB}$, there is $\epsilon > 0$ small enough such that either $\mu^{BEH'} + \epsilon$ or $\mu^{BEH'} - \epsilon$ increases the planner's function. In the case that the new left constraint binds, we have that $(\phi(\mu^{BEH'}) - c)\mu^{BEH'}/(N^{BEH} - 1) = f > (\phi(\mu^{BEH'}) - c)\mu^{BEH'}/(N^{BEH})$, so $\mu^{BEH'} < \mu^{FB}$ (consider a similar argument of reducing $\mu^{BEH'}$ by ϵ to exclude $\mu^{BEH'} > \mu^{FB}$). This case also induces $P(\mu^{BEH'}, N^{BEH} - 1) > P(\mu^{BEH}, N^{BEH})$, as $\mu^{BEH} < \mu^{BEH'} < \mu^{FB}$. We conclude that it cannot be that $N^{BEH} > 1$ with some $\mu^{BEH} < \mu^{FB}$.

   (b) Now consider the case $N^{BEH} \geq 1$ with $\mu^{BEH} > \mu^{FB}$. Then the right BEH constraint must bind (and the left not): $(\phi(\mu^{BEH}) - c)\mu^{BEH}/N^{BEH} > f = (\phi(\mu^{BEH}) - c)\mu^{BEH}/(N^{BEH} + 1)$. To see this, observe that if the right did not bind, we could reduce μ^{BEH} to bring it closer to μ^{FB} with both constraints still satisfied. Thus, $\pi^{BEH}(\mu^{BEH}) = (\phi(\mu^{BEH}) - c)\mu^{BEH}/N^{BEH} - f > 0$.

$\square$

Proof of Proposition 7

Saying that profits are zero at the BEH solution means that the left BEH constraint binds and, given that monopoly profit is negative at μ^{FB} (so it cannot be $\mu^{FB} \leq \mu^{BEH}$), it will be $N^{BEH} = 1$ and $\mu^{FB} > \mu^{BEH}$. From $\pi^M(\mu^{FB}) < 0$ it follows that:

$$(\phi(\mu^{FB}) - c)\mu^{FB} < f \implies \phi(1) - c < f \implies \frac{\phi(1) - c}{N} < f \; \forall N \in \mathbb{N}$$

so it cannot be $\mu^{CFE} = 1$, thus $\mu^{CFE} < 1$. We now prove that $\mu^{BEH} > \mu^{CFE}$. Suppose, by contradiction, that $\mu^{CFE} \geq \mu^{BEH}$. From $\pi^M(\mu^{FB}) < 0$ we know that at the Behavioral Second Best $N^{BEH} = 1$ and the exchange breaks even, so given that $\psi'(\mu) > 0$ we have:

$$(\phi(\mu^{CFE}) - c)\mu^{CFE} \leq f. \tag{A.54}$$

We first deal with the case where only one firm enters at CFE, and then the one where more enter. At a CFE with $N = 1$ exchanges, we have $\mu^{CFE} = \mu^M$ and thus profit is positive:

$$(\phi(\mu^{CFE}) - c)\frac{\mu^{CFE}}{N} = (\phi(\mu^{CFE}) - c)\mu^{CFE} = (\phi(\mu^M) - c)\mu^M > f. \tag{A.55}$$

Putting together (A.54) and (A.55) leads to a contradiction.

At a CFE with $N > 1$ exchanges, we have:

$$(\phi(\mu^{CFE}) - c)\frac{\mu^{CFE}}{N} \geq f. \tag{A.56}$$

Putting together (A.54) and (A.56) yields

$$f \leq (\phi(\mu^{CFE}) - c)\frac{\mu^{CFE}}{N} < (\phi(\mu^{CFE}) - c)\mu^{CFE} \leq f,$$

a contradiction. Thus, we must have $\mu^{BEH} > \mu^{CFE}$.

From Proposition 6 we have $\mu^{CFE} \geq \mu^{STR}$. Together with what we have proved above, it must be that $\mu^{BEH} > \mu^{CFE} \geq \mu^{STR}$. Now, since $N^{BEH} = 1$, it follows that $\mathcal{P}^{FB} > \mathcal{P}^{BEH} > \mathcal{P}^M$ since $\mu^{FB} > \mu^{BEH} > \mu^M$ and $\mathcal{P}(\mu, 1)$ is single-peaked at μ^{FB}.

Also, we have that $\mathcal{P}(\mu, 1) = GW(\mu) - c\mu - f$ is single-peaked in μ at μ^{FB}, which means that $GW(\mu) - c\mu$ is so. Thus, since $\mu^{BEH} > \mu^{CFE} \geq \mu^{STR}$ we have:

$$\begin{aligned}\mathcal{P}^{BEH} = GW(\mu^{BEH}) - c\mu^{BEH} - f &> GW(\mu^{STR}) - c\mu^{STR} - f \\ &\geq GW(\mu^{STR}) - c\mu^{STR} - fN^{STR} = \mathcal{P}^{STR}\end{aligned}$$

and so $\mathcal{P}^{BEH} > \mathcal{P}^{STR} \geq \mathcal{P}^{CFE}$, where the weak inequality follows from the fact that the CFE solution is always available in solving the STR problem. For the same reason, $\mathcal{P}^{STR} \geq \mathcal{P}^{M}$. Thus, overall we have:

$$\mathcal{P}^{FB} > \mathcal{P}^{BEH} > \mathcal{P}^{STR} \geq \max\left\{\mathcal{P}^{CFE}, \mathcal{P}^{M}\right\} \geq \min\left\{\mathcal{P}^{CFE}, \mathcal{P}^{M}\right\}.$$

Last, evaluate the derivative of welfare with respect to μ at the $USTR$ solution:

$$\left.\frac{\partial \mathcal{P}(\mu, N)}{\partial \mu}\right|_{(\mu,N)=(\mu^{USTR},N^{USTR})} = \left.[\phi'(\mu)\mu + \phi(\mu) - c]\right|_{\mu=\mu^{USTR}}$$

The FOC of a firm i in $USTR$ reads:

$$\left.\left[\frac{\phi'(\mu)\mu}{N} + \phi(\mu) - c\right]\right|_{(\mu,N)=(\mu^{USTR},N^{USTR})} = 0.$$

Combining this with the above we have:

$$\left.\frac{\partial \mathcal{P}(\mu, N)}{\partial \mu}\right|_{(\mu,N)=(\mu^{USTR},N^{USTR})} = \left.\phi'(\mu)\mu\frac{N-1}{N}\right|_{(\mu,N)=(\mu^{USTR},N^{USTR})} < 0$$

so the $USTR$ solution does not maximize welfare given that the FB solution is interior (and thus, for the $USTR$ solution to maximize welfare the derivative above should have been zero), so it must be $\mathcal{P}^{FB} > \mathcal{P}^{USTR}$. □

B A model with SD at the second round

In this appendix we consider a variation of the model presented in Section 3, in which we assume that SD enter the market at the second round of the liquidity determination stage of the game (the proofs of the results involve minor variations from the ones in Appendix A, and are available upon request). This captures the intuition that through technological services FD are quicker in accommodating liquidity traders' demand shocks than SD. In this case, the market clearing conditions in periods 1 and 2 are given respectively by $x_1^L + \mu x_1^{FD} = 0$ and $x_2^L + \mu(x_2^{FD} - x_1^{FD}) + (1-\mu)x_2^{SD} = 0$ (see Figure 6 for the modified timeline). We restrict attention to linear equilibria where

$$p_1 = -\tilde{\Lambda}_1 u_1 \tag{B.1a}$$

$$p_2 = -\tilde{\Lambda}_2 u_2 + \tilde{\Lambda}_{21} u_1, \tag{B.1b}$$

where we use $\sim$ to denote variables related to the model with SD entering at the second round.

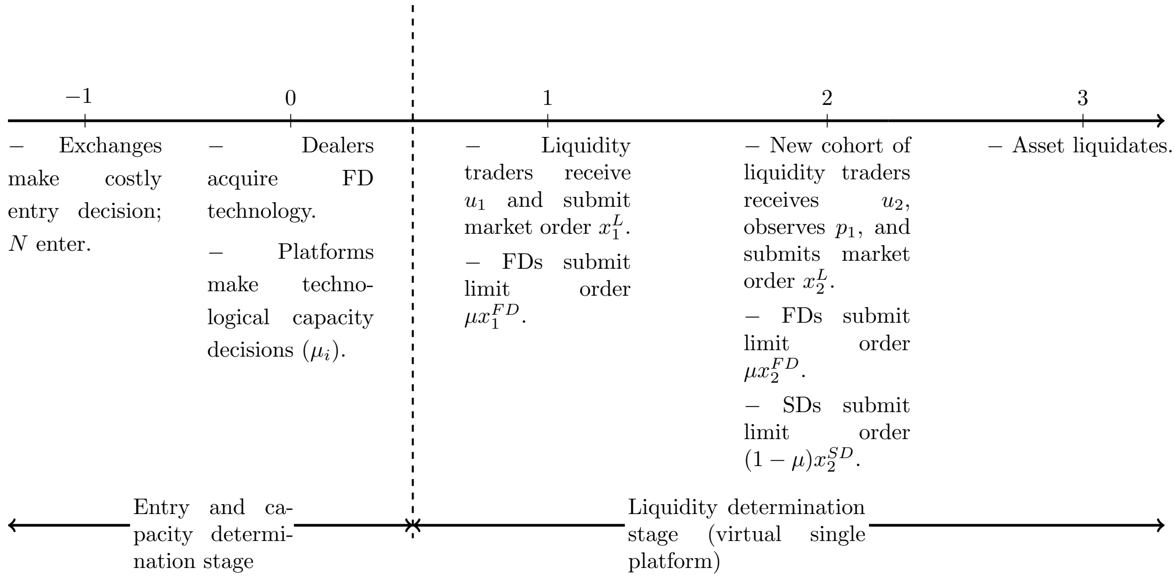


Figure 6: Timeline in the model where SD enter at the second round.

We obtain the following result:

Proposition 8. *For $\mu \in (0, 1]$, there exists a unique equilibrium in linear strategies in the stock market where SD enter at the second round. Compared to the baseline case, $\tilde{\Lambda}_1 > \Lambda_1$, $\Lambda_1 < \tilde{\Lambda}_2 < \Lambda_2$, and $|\tilde{\Lambda}_{21}| > |\Lambda_{21}|$.*

Thus, SD entry at the second round reduces (increases) the competitive pressure faced by FD at the first (second) round, explaining the decrease (increase) in first (second) period liquidity. Comparing dealers' payoffs across the two models, we find

Proposition 9. *$\widetilde{CE}^{FD} > \widetilde{CE}^{SD}$, and SD have a higher payoff when entering in the second round, whereas the result for FD is ambiguous: $\widetilde{CE}^{SD} > CE^{SD}$, and $\widetilde{CE}^{FD} \gtreqless CE^{FD}$.*

As in the baseline model, more access to the liquidity supply market has value for dealers. In the baseline model, in the first round FD supply liquidity anticipating the possibility to rebalance their position at the second round. This heightens the competitive pressure they exert on SD compared to the model studied in this section, explaining why $\widetilde{CE}^{SD} > CE^{SD}$. Conversely, the payoff comparison for FD is less clear cut. Indeed, compared to the baseline model, liquidity is lower (higher) at the first (second) round. We define the demand for technological services as $\tilde{\phi}(\mu) = \widetilde{CE}^{FD} - \widetilde{CE}^{SD}$.

Proposition 10. *In the model where SD enter at the second round, $\tilde{\phi}(\mu)$ is decreasing in μ.*

Furthermore, numerical simulations show that even in this case, the demand for technological services can be log-convex, implying that strategic complementarities in platform capacity decisions can arise.

C A permanent shock to dealers' risk tolerance

In this appendix we use our model to study the effect of a permanent shock to liquidity providers' risk tolerance. As is usual in a setup where dealers are risk averse, a lower γ reduces market liquidity (Corollary 1). However, liquidity also depends positively on the proportion of FDs (Corollary 1), which is pinned down by the equilibrium arising at the technological capacity determination stage of the game. This implies that if a lower γ leads to a positive shift in the demand for technological services (Corollary 7) it can also have an indirect, *positive* effect on liquidity, via its effect on platforms' capacity decisions.

To fix ideas, consider the unregulated monopoly case. At the optimum, the monopolist supplies

$$\mu^M = \frac{\phi - c}{-\phi'}.$$

Differentiating the above expression with respect to γ (and denoting ϕ' by $\partial\phi/\partial\mu$) yields

$$\frac{\partial \mu^M}{\partial \gamma} = \frac{1}{(\partial\phi/\partial\mu)^2}\left(- \frac{\partial\phi}{\partial\gamma}\frac{\partial\phi}{\partial\mu} + (\phi - c)\frac{\partial^2\phi}{\partial\mu\partial\gamma}\right) \begin{cases} > 0 & \text{amplification} \\ < 0 & \text{attenuation} \end{cases}$$

If a permanent shock to γ shifts ϕ upwards ($\partial\phi/\partial\gamma > 0$) and makes it flatter ($\partial^2\phi/\partial\gamma\partial\mu < 0$) it leads the monopolist to increase its supply of technological services, inducing an amplification of the initial shock. Conversely, a sufficient condition for attenuation is $\partial\phi/\partial\gamma < 0$ and $\partial^2\phi/\partial\gamma\partial\mu > 0$.

We run simulations to gauge the effect of platform capacity decisions on liquidity, in the presence of a permanent reduction in dealers' risk aversion. For γ decreasing from γ^{orig} to $\hat{\gamma} < \gamma^{orig}$, we define the percentage of the direct positive effect on Λ_t mitigated by the indirect-platform competition effect as follows:

$$\%\text{mit} \equiv 1 - \frac{\text{Total Effect}}{\text{Direct Effect}} = 1 - \frac{\Lambda_t(\hat{\gamma}, \mu(\hat{\gamma})) - \Lambda_t(\gamma^{orig}, \mu(\gamma^{orig}))}{\Lambda_t(\hat{\gamma}, \mu(\gamma^{orig})) - \Lambda_t(\gamma^{orig}, \mu(\gamma^{orig}))},$$

where the Direct Effect at the denominator in the above expression captures the change in liquidity that obtains when only γ changes, and μ is kept at the value it had prior to the the shock to risk-tolerance. Accordingly, if $\%\text{mit} < 0$ (> 0), the direct effect is enhanced (mitigated) by the indirect effect, and if $\%\text{mit} > 1$, the indirect effect *overturns* the direct effect. Figure 7 displays the result of a simulation in which the mitigation effect can be strong enough to overturn the direct negative liquidity impact of a reduction of dealers' risk tolerance for Λ_2.

Initial parametrization	Alternative parameter values		
	c	γ	γ_L
$c = 0.002$, $\gamma = 0.5$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v = 25$	0.001	$\{0.45, 0.35, 0.3, 0.25\}$	0.15
$c = 0.002$, $\gamma = 0.5$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v = 3$	0.003	$\{0.45, 0.35, 0.3, 0.25\}$	0.15
$c = 2$, $\gamma = 25$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$	2.5	$\{22.5, 17.5, 15, 12.5\}$	18

Table 1: Parametrizations used in the simulations.

Numerical Result 5. *For second period liquidity, with standard risk aversion, a 10% decrease in* γ *leads to*

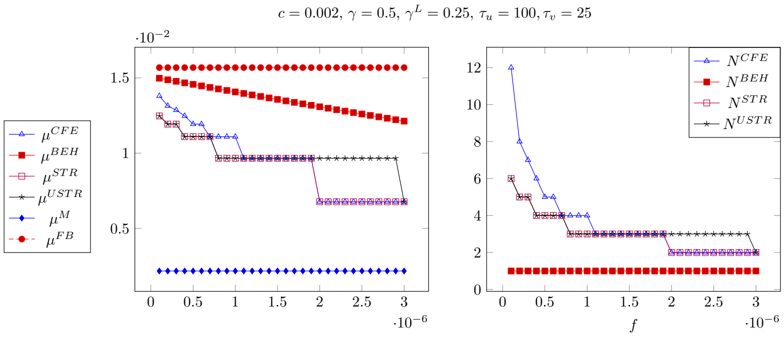
1. *Mitigation in the case of the unregulated monopolist.*
2. *Both mitigation and amplification are possible at the CFE when payoff volatility is low.*

A shock to γ does not lead to a parallel shift in ϕ (see Figure 7, Panel (a)). Thus, its ultimate effect on second period liquidity (amplification vs. attenuation) depends on the value of μ^M or μ^{CFE} pre-shock. For example, for a range of values close to the origin, to which μ^M belongs, a 10% shock shifts ϕ up, and flattens the inverse demand curve. These two effects are responsible for the observed attenuation at the monopoly solution which occurs at all f, since μ^M is independent of f (see Figure 7, Panel (b)).

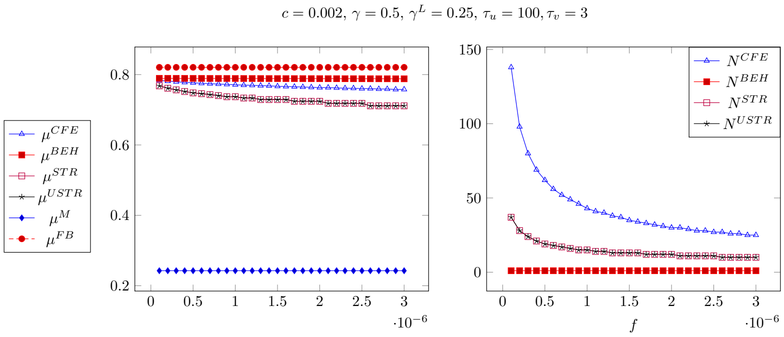
For larger values of μ, to which μ^{CFE} belongs, the effect of the shock on liquidity is more complicated because the pre-shock value of μ depends on (i) f and (ii) N. For the values of μ^{CFE} that correspond to $f \in \{1 \times 10^{-7}, 2 \times 10^{-7}\}$, the shock shifts ϕ mildly down ($\partial\phi/\partial\gamma > 0$) and makes it steeper ($\partial^2\phi/\partial\gamma\partial\mu < 0$). Each platform faces a smaller mark-up and a steeper demand curve, and cuts down on μ_i. This induces a profit increase that prompts entry. The paradoxical result is that we observe entry with a *reduction* in aggregate μ, and thus amplification (see Figure 7, Panels (c) and (d)). As f increases, both attenuation and amplification can obtain, because the pre-shock value of μ shrinks, but stays in the region where the shock to γ has a complex effect on ϕ. When parameter values are such that the industry supply increases and entry occurs (as in the cases $f \in \{2 \times 10^{-6}, 2.1 \times 10^{-6}, 2.2 \times 10^{-6}\}$, in Figure 7, Panel (c)) attenuation is so strong that a reduction in γ leads to an increase in second period liquidity (see Figure 7, Panel (d)).

Result	Order of action	Parametrization
$N^{CFE} < N^{USTR}$	OO	$c = 0.002$, $\gamma \in \{0.45, 0.5\}$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v = 25$ $c = 2$, $\gamma \in \{25, 22.5, 17.5, 15, 12.5\}$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$ $c = 2$, $\gamma = 25$, $\gamma_L = 18$, $\tau_u = 0.1$, $\tau_v = 0.1$
	RO	$c = 2$, $\gamma \in \{25, 22.5, 17.5, 15, 12.5\}$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$ $c = 2$, $\gamma = 25$, $\gamma_L = 18$, $\tau_u = 0.1$, $\tau_v = 0.1$ $c = 2.5$, $\gamma = 25$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$
$N^{CFE} < N^{USTR} - 1$	OO RO	$c = 2$, $\gamma \in \{25, 22.5, 17.5, 15, 12.5\}$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$ $c = 2$, $\gamma \in \{17.5, 15, 12.5\}$, $\gamma_L = 12$, $\tau_u = 0.1$, $\tau_v = 0.1$
$\gamma \downarrow \rightarrow \Lambda_2 \downarrow$ for CFE	OO	$c = 0.002$, $\gamma = 0.5 \rightarrow \hat{\gamma} = 0.45$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v = 25$
$\pi^M\left(\mu^{FB}\right) > 0$, $\mu^{BEH} > \mu^{CFE} > \mu^{FB}$	RO	$c = 0.002$, $\gamma = 0.5$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v \in \{3, 25\}$ and all shifts around presented in Table 1
$\pi^M(\mu^{FB}) \leq 0$	OO	$\{(\gamma, \gamma_L, \tau_u, \tau_v, c, f) : \gamma, \gamma_L \in \{1, 2, \ldots, 25\}, \tau_u, \tau_v \in \{0.1, 0.2, \ldots, 10\}, c \in \{0.01, 0.02, \ldots, 0.5\}, f \in \{0.001, 0.002, \ldots, 0.1\}\}$
SC in Cournot equilibrium	RO	$N = 2$, $c = 0.002$, $\gamma = 0.5$, $\gamma_L = 0.25$, $\tau_u = 100$, $\tau_v \in \{25, 3\}$
$\gamma_L \downarrow \rightarrow \mu, N \uparrow$	OO RO	The three parametrizations in Table 1 with the corresponding shocks to γ_L
$c \downarrow \rightarrow \mu, N \uparrow$	OO	The three parametrizations in Table 1 with the corresponding shocks to γ_L
$c \downarrow \rightarrow \mu \uparrow, N \downarrow$	RO	The three parametrizations in Table 1 with the corresponding shocks to γ_L
$N^{BEH} = 1$	OO RO	The three parametrizations in Table 1 with the corresponding shocks to γ_L
Single-peakedness of $\mathcal{P}(\mu, 1)$ and $\pi(\mu)$	OO RO	The three parametrizations in Table 1 with the corresponding shocks to γ_L

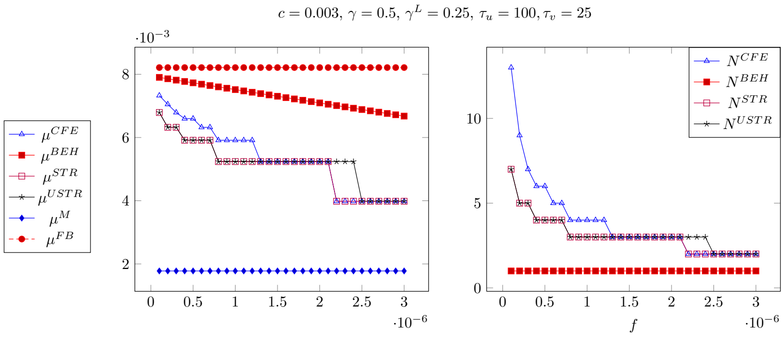
Table 2: Cases where various phenomena are observed. Notation: *(i)* $\gamma \downarrow \rightarrow \Lambda_2 \downarrow$ means that a reduction in dealers' risk tolerance leads to an increase in liquidity in period 2 in CFE; *(ii)* OO and RO refer to the original and reverse order of action models, respectively; *(iii)* $\gamma_L \downarrow \rightarrow \mu, N \uparrow$ means that a decrease in γ_L leads to an increase in μ, N. The rest of the results on the effects of shocks to γ_L and c are read in a similar way; *(iv)* SC means strategic complementarity.



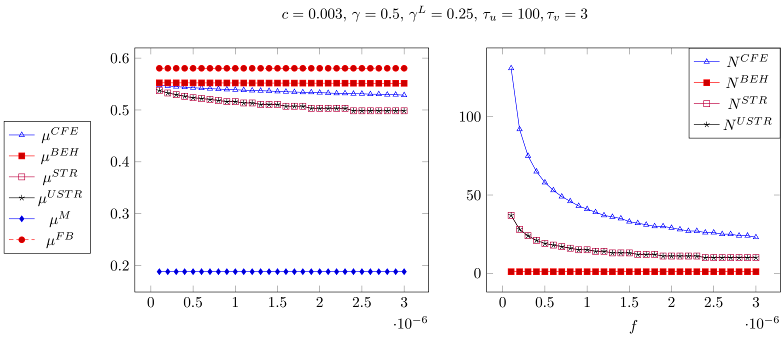
(a)



(b)



(c)



(d)

Figure 5: Panels (a) and (c) illustrate two cases in which insufficient entry occurs. In Panel (b) and (d), entry is always excessive.

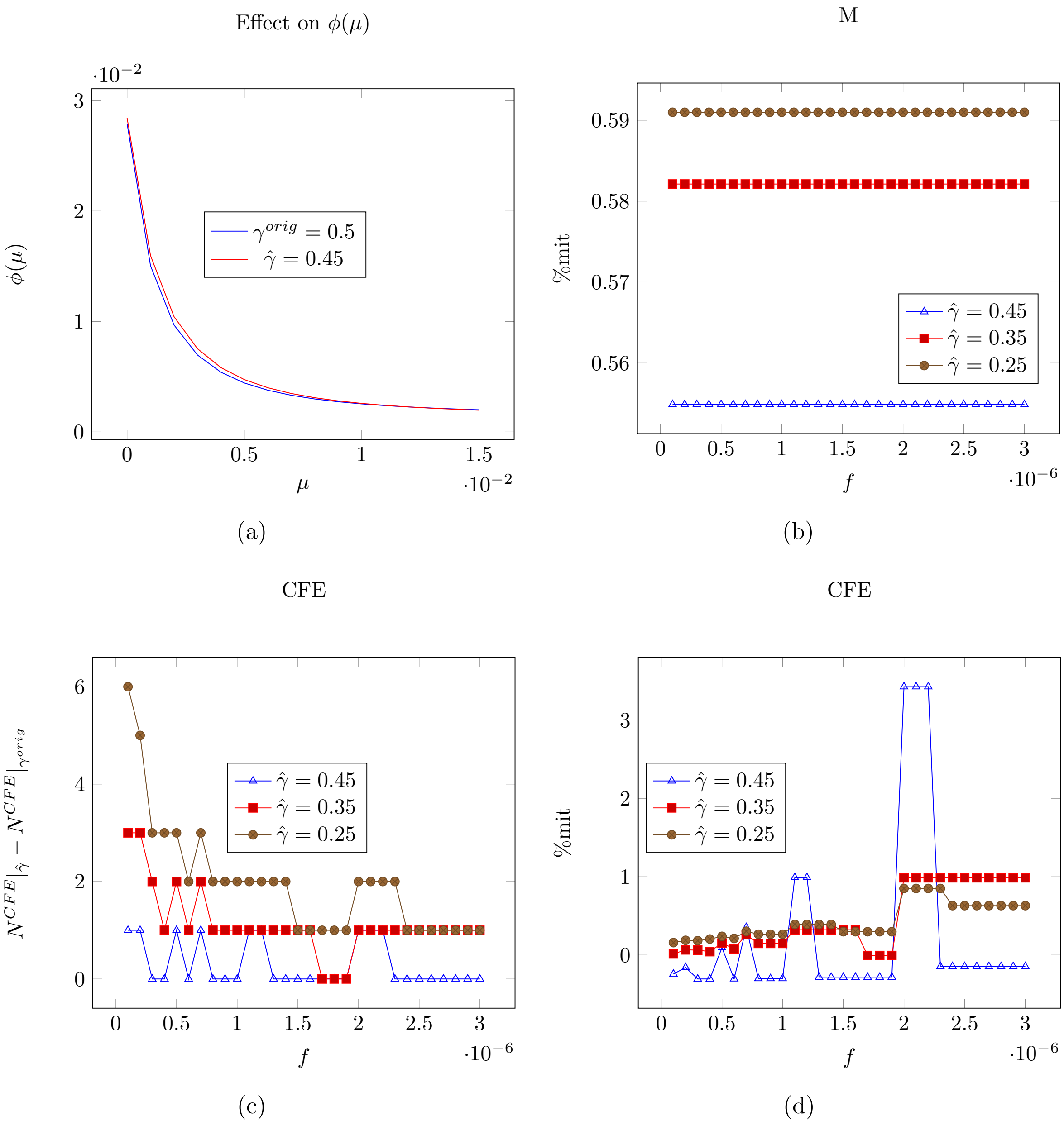


Figure 7: Liquidity impact of a reduction in dealers' risk tolerance. In panel (a) we plot the effect of the shock on the demand for technological services; in panel (b) we plot the mitigation effect in the M case. In panels (c) and (d) we plot the impact on entry and the mitigation effect in the CFE case. Parameter values: $c = 0.002$, $\gamma^{orig} = 0.5$, $\gamma^{L} = 0.25$, $\tau_u = 100$, $\tau_v = 25$.