Exchange Competition, Entry, and Welfare∗

Giovanni Cespa† and Xavier Vives‡

March 2018

Abstract

We assess the consequences for market quality and welfare of different entry regimes and exchange pricing policies in a context of limited market participation. To this end we integrate a two-period market microstructure model with an exchange competition model with entry in which exchanges supply technological services, and have market power. We find that dealers’ demand for technological services is log-convex for a wide range of model parameters and that this may induce strategic complementarity in platform competition. Free entry of platforms delivers a superior outcome in terms of liquidity and welfare compared to the case of an unregulated monopoly. Controlling entry or, even better, platform fees may theoretically further increase liquidity and welfare. However, regulation requires fine tuning since the market may deliver excessive or insufficient entry, and both structural and fee regulation have high informational requirements and are subject to rent-seeking efforts by market participants.

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, Thierry Foucault, Felix Suntheim, and seminar participants at ICEF (Moscow), Universidad Carlos III (Madrid), and Imperial College (London) for useful comments and suggestions.
†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).

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 of market fragmentation (see Figure). 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]). 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

---

1 See Foucault et al. [2013], Chapter 1.
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]).
steered their business models towards the provision of technological services e.g., proprietary data, and co-location space—see Figure 7 in the Appendix).³

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

Such a paradigm shift has raised a number of concerns. Indeed, market participants allege that exchanges exercise market power in the provision of technological services.⁴ 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.

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

³Increasing 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, from 2014 to 2016 ICE (the mother company of the NYSE)’s revenues from data services almost tripled from M$691 to Bn$1.978 (ICE, 10-k filing, 2017). Finally, 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“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.”)

⁵Responding 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.) See also Okuliar (2014) on whether US competition authorities should intervene more in financial exchange consolidation.
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. 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. We also assume that trading fees are set at the competitive level by the exchanges.

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. 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 compared to

---

6 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 to the trading platform that posts the best price and the exchanges have to provide accessible electronic data about their price quotations. The aim is to enforce price priority in all markets. In Europe there is no order protection rule similar to RegNMS. Foucault and Menkveld (2008) show empirically the existence of trade-throughs in Amsterdam and London markets. Hendershott and Jones (2005) find that in the US price protection rules improve market quality.

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

8 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.
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, inducing a positive “liquidity externality” on their welfare. On the other hand, it imposes a negative externality on 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, implying that on balance the positive effect of the liquidity externality prevails. In turn, this contributes to make total 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 profitably enter the market, she achieves the Structural Second Best solution.

Insulated from competition, a monopolistic exchange seeks to restrict the supply of technological services to increase the fees it extracts from FDs. Thus, the market at a free entry Cournot equilibrium delivers a superior outcome in terms of liquidity and welfare. However, 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.
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. 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 inferior compared to the alternative of regulating the technological service fee charged by a monopolistic exchange. This is because in this case the planner minimizes the setup cost borne by the industry and forces the monopolistic exchange to charge the lowest possible technological service that is compatible with a break-even condition.

Overall, our analysis suggests that fee regulation achieves the outcome that is closest to the First Best, since it minimizes entry costs and forces the next to highest provision of technological services. However, fee regulation is subject to rent-seeking efforts by market participants which suggests that entry regulation appears as a realistic alternative instrument. 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.

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. 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.

---

10 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.

11 The evidence presented in footnote 5 suggests that regulators’ ability to weigh on the technological fee-setting process is far from perfect.

12 Pagnotta and Philippon (2018) also study the market integration impact of RegNMS.
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 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 $v \sim N(0, \tau^{-1})$, and a risk-less asset with unit return are exchanged during two trading rounds. 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 trading services, owing to a best price rule, trading can seamlessly occur either in the incumbent venue, or in one of $N-1$ potential entrants at either 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 $(0, \mu)$; these traders are active at both dates, and in both types of venues. Second, competitive, risk-averse “Standard Dealers” (denoted by SD) in the interval $[\mu, 1]$, who instead are active only in the first period. 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. When venues compete in the provision of trading services, we assume
that a best price rule ensures that the price at which orders are executed is the same across all venues. We now illustrate the preferences and orders of the different players.

2.1 Trading venues

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, they 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. Suppose that each venue \( i = 1, 2, \ldots, N \) produces a technological service capacity \( \mu_i \), 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, we assume that trading fees are set to the competitive level.

2.2 Liquidity providers

A FD has CARA preferences, with risk-tolerance \( \gamma \), and submits price-contingent orders \( x_{t}^{FD} \), to maximize the expected utility of his final wealth: \( W^{FD} = (v - p_t)x_{t}^{FD} + (p_2 - p_1)x_{t}^{FD} \), where \( p_t \) denotes the equilibrium price at date \( t \in \{1, 2\} \). A SD also has CARA preferences with risk-tolerance \( \gamma \), 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 \( W^{SD} = (v - p_1)x_{1}^{SD} \). 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.

2.3 Liquidity demanders

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. Recent research documents the existence of a sizeable proportion of market participants who

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

We assume \( u_1 \sim N(0, \tau_u^{-1}) \), and \( \text{Cov}[u_1, v] = 0 \). First period traders have identical CARA preferences with common risk-tolerance coefficient \( \gamma^L \). Formally, a trader posts a market order \( x_1^L \) to maximize the expected utility of his profit \( \pi_1^L = u_1v + (v - p_1)x_1^L : \)

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

In period 2, a new unit mass of traders enters the market. A second period trader observes \( p_1 \), receives a random endowment of the risky asset \( u_2 \sim N(0, \tau_u^{-1}) \), and takes a position \( x_2^L \) in the asset. We assume \( \text{Cov}[u_2, v] = \text{Cov}[u_2, u_1] = 0 \). Second period traders have identical CARA preferences with risk-tolerance \( \gamma^L \), and submit a market order to maximize the expected utility of their profit \( \pi_2^L = u_2v + (v - p_2)x_2^L : \)

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

To ensure that the payoff functions of the liquidity demanders are well defined (see Section 4.1), we impose

\[
(\gamma^L)^2\tau_u \tau_v > 1,
\]

an assumption that is common in the literature (see, e.g. Vayanos and Wang (2012)). Finally, we assume that trading fees are nil, an assumption that is consistent with the systematic decline in the cost of trading brought about by the automation of the trading process (see, e.g. Menkveld (2016)).

### 2.4 Market clearing and prices

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

\[
p_1 = -\Lambda_1 u_1 \quad \text{(5a)}
\]

\[
p_2 = -\Lambda_2 u_2 + \Lambda_21 u_1 \quad \text{(5b)}
\]

where the price impact of endowment shocks \( \Lambda_1, \Lambda_2, \) and \( \Lambda_21 \) are determined in equilibrium. Thus, 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.
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 à la [Grossman and Miller, 1988] where dealers only accommodate endowment shocks, but where all traders are expected utility maximizers. We prove existence and uniqueness of an equilibrium in linear strategies, and analyze the equilibrium properties.

**Proposition 1.** For $\mu \in (0,1]$, there exists a unique equilibrium in linear strategies in the stock market, where

\[ x_1^{SD} = -\gamma \tau_v p_1, \quad x_1^{FD} = \gamma \tau_u \Lambda_2^{-2}(\Lambda_2 + \Lambda_1)u_1 - \gamma \tau_v p_1, \quad x_2^{FD} = -\gamma \tau_v p_2, \quad x_1^L = a_1 u_1, \quad x_2^L = a_2 u_2 + b u_1, \]

\[
p_1 = -\Lambda_1 u_1 \tag{6a}
\]
\[
p_2 = -\Lambda_2 u_2 + \Lambda_{21} u_1, \tag{6b}
\]

\[
\Lambda_1 = \left( 1 - \left( 1 + a_1 + \mu \gamma \tau_u \frac{\Lambda_{21} + \Lambda_1}{\Lambda_2} \right) \right) \frac{1}{\gamma \tau_v} > 0 \tag{7a}
\]
\[
\Lambda_2 = -\frac{a_2}{\mu \gamma \tau_v} > 0 \tag{7b}
\]
\[
\Lambda_{21} = -(1 - ((1 - \mu)\gamma + \gamma^L)\tau_v \Lambda_1) \Lambda_2 < 0, \tag{7c}
\]

where

\[ \Lambda_{21} + \Lambda_1 > 0. \tag{8} \]
The coefficient $\Lambda_t$ in (6a) and (6b) denotes the period $t$ endowment shock’s negative price impact, and is our measure of illiquidity:

$$\Lambda_t = -\frac{\partial p_t}{\partial u_t}. \tag{9}$$

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

$$X^{L_1}_1(u_1) = \gamma L E\left[v - p_1|u_1\right] - u_1$$

$$X^{L_2}_2(u_1, u_2) = \gamma L E\left[v - p_2|u_1, u_2\right] - u_2.$$

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. Substituting the equilibrium prices (6a) and (6b) in the above expressions implies that the trading aggressiveness is given by $|a_t|:

$$a_t = \gamma L \tau_v \Lambda_t - 1 \in (-1, 0). \tag{10}$$

Additionally, second period traders put a positive weight $b$ on the first period endowment shock:

$$b = -\gamma L \tau_v \Lambda_{21} \in (0, 1). \tag{11}$$

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,

$$x^{FD}_1 = \gamma E\left[p_2 - p_1|u_1\right] - \gamma \tau_v p_1.$$

To interpret the above expression, suppose $u_1 > 0$. Then, liquidity traders sell the asset, depressing its price (see (6a)) 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 to second period traders, the effect of the first period price pressure has not completely disappeared (see (7c)), explaining the positive sign of the coefficient $b$ in (11).

Thus, 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. In view of this, $\Lambda_1$ in (7a) 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:

$$\Lambda_1 = \left(1 - \left(1 + a_1 + \frac{\mu \gamma \tau_v \Lambda_{21} + \Lambda_1}{\Lambda_2^2}\right)\right) \frac{1}{\gamma \tau_v}.$$

14This 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)).
In the second period, liquidity traders hedge a portion $a_2$ of their order, which is absorbed by a mass $\mu$ of FDs, thereby explaining the expression for $\Lambda_2$ in (7b).

Therefore, at both trading rounds, an increase in $\mu$ increases the risk bearing capacity of the market, leading to a lower illiquidity:

**Corollary 1.** An increase in the proportion of FDs reduces the illiquidity of both trading rounds: $\frac{\partial \Lambda_t}{\partial \mu} < 0$, for $t \in \{1, 2\}$.

According to (6b) and (7c), 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 (7c) in (6b), and rearranging yields:

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

(12)

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

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 also provide liquidity by posting contrarian market orders, this implies that

$$\Lambda_1 < \Lambda_2,$$

(13)

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 $\mu$ improves liquidity at both dates, but also increases the portion of the first period endowment shock absorbed by FD (see (12)). This, in turn, leads second period liquidity traders to step up their response to $u_1$.

In view of (7a) and (8), 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: $\frac{\partial |\text{Cov}\{p_2 - p_1, p_1\}|}{\partial \mu} < 0$.

Summarizing, an increase in $\mu$ 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. 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.
**Remark 1.** The variance of the first period price is given by $\text{Var}[p_1] = \Lambda^2 \tau^{-1}$. Therefore, a more illiquid market increases price volatility.

4 Traders’ welfare, capacity 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:

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

where $EU^j, j \in \{SD, FD\}$ and $EU^L_t, t \in \{1, 2\}$ 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 4.** The payoffs of a SD and a FD are given by

$$CE^{SD} = \frac{\gamma}{2} \ln \left( 1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} \right) \tag{14a}$$

$$CE^{FD} = \frac{\gamma}{2} \left( \ln \left( 1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} \right) + \frac{\text{Var}[E[p_2 - p_1|p_1]]}{\text{Var}[p_2 - p_1|p_1]} \right) + \ln \left( 1 + \frac{\text{Var}[E[v - p_2|p_1, p_2]]}{\text{Var}[v - p_2|p_1, p_2]} \right) \right). \tag{14b}$$

Furthermore:

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

According to (14a) and (14b), dealers’ payoffs reflect the precision with which these agents can anticipate the unit profits from their strategies. 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 in both venues, 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 \( \text{Var}[E[v - p_2|p_1, p_2]]/\text{Var}[v - p_2|p_1, p_2] \). As FD can trade twice, they enjoy a higher expected utility.

Substituting (10) and (11) in (14a) and (14b), and rearranging yields:

\[
 CE^{SD} = \frac{\gamma}{2} \ln \left( 1 + \frac{(1 + a_1)^2}{\gamma^L} \right) + \ln \left( 1 + \frac{(1 + a_2)^2}{\gamma^L} \right).
\]  

(15a)

An increase in \( \mu \) 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 5.** Suppose (4) holds, then the payoffs of first and second period traders are given by

\[
 CE^{L_1} = \frac{\gamma}{2} \ln \left( 1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} + \frac{2 \text{Cov}[p_1, u_1]}{\gamma^L} \right).
\]  

(16a)

\[
 CE^{L_2} = \frac{\gamma}{2} \ln \left( 1 + \frac{\text{Var}[E[v - p_2|p_1, p_2]]}{\text{Var}[v - p_2|p_1, p_2]} + \frac{2 \text{Cov}[p_2, u_2|p_1]}{\gamma^L} + \frac{\text{Var}[E[v - p_2|p_1]]}{\text{Var}[v]} - \left( \frac{\text{Cov}[p_2, u_1]}{\gamma^L} \right)^2 \right).
\]  

(16b)

Furthermore:

1. \( CE^{L_1} \) and \( CE^{L_2} \) are increasing in \( \mu \).

2. For all \( \mu \in (0, 1] \), \( CE^{L_1} > CE^{L_2} \).

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 (16a) and (16b)). Differently from SDs, these traders are however exposed to a random endowment shock. As a more illiquid market increases hedging costs, it negatively affects their payoff (\( \text{Cov}[p_1, u_1] = -\Lambda_1 \gamma u^{-1} \), and \( \text{Cov}[p_2, u_2|p_1] = -\Lambda_2 \gamma u^{-1} \)). Finally, (16b) 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 \( u_1 \), (\( \text{Cov}[p_2, u_1]^2 \)), and indicates that a trader hedges little of his endowment shock \( u_2 \), and keeps a large exposure to the asset risk, which reduces his payoff.
Substituting (10) and (11) in (16a) and (16b), and rearranging yields:

\[
CE_1^L = \frac{\gamma}{2} \ln \left( 1 + \frac{a_1^2 - 1}{(\gamma L)^2 \tau u \tau v} \right),
\]

(17)

\[
CE_2^L = \frac{\gamma}{2} \ln \left( 1 + \frac{a_2^2 - 1 + b^2((\gamma L)^2 \tau u \tau v - 1)}{(\gamma L)^4 \tau u^2 \tau v^2} \right).
\]

(18)

A higher \( \mu \) makes the market more liquid, leading traders to hedge and speculate more aggressively. This, in turn, makes their payoffs increasing in \( \mu \). Together with the second part of Proposition 4, this implies that an increase in the proportion of FDs induces a “liquidity externality,” which affects positively both liquidity traders’ cohorts, and negatively SDs. In the second period, liquidity traders can also speculate but face a less liquid market (see (13)). This dampens their payoff compared to their first period peers.

We conclude this section by showing that the positive externality exerted on traders’ payoffs by an increase in \( \mu \) swamps the negative externality it imposes on SDs:

**Corollary 6.** Suppose (4) holds. 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:

\[
\frac{\partial CE_1^L}{\partial \mu} \geq - \frac{\partial CE_{SD}}{\partial \mu},
\]

(19)

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

An increase in the proportion of FDs leads traders to hedge more aggressively (Corollary 2), benefiting first period traders (Proposition 5). At the same time, it heightens the competitive pressure faced by SDs, lowering their payoffs (Proposition 4). 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 liquidity externality prevails (see (19)).

Aggregating across market participants’ welfare yields the following total welfare function:

\[
TW(\mu) = \mu CE_{FD} + (1 - \mu)CE_{SD} + CE_1^L + CE_2^L
\]

(20)

\[
= \mu(CE_{FD} - CE_{SD}) + CE_{SD} + CE_1^L + CE_2^L
\]

Surplus earned by FDs Welfare of other market participants

**Corollary 7.** Suppose (4) holds:

1. The welfare of market participants other than FDs is increasing in \( \mu \).

2. Total welfare is higher at \( \mu = 1 \) than at \( \mu \approx 0 \).

The first part of the above result is a direct consequence of Corollary 6 as \( \mu \) increases, SDs’ losses due to heightened competition are more than compensated by traders’ gains due to higher liquidity. The second part, follows from Proposition 4 (part 3), and Proposition 5. Note that it rules out the possibility that the payoff decline experienced by FDs as \( \mu \) increases,
leads total 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.** Suppose (4) holds. Numerical simulations show that $TW(\mu)$ is monotone in $\mu$. Therefore, $\mu = 1$ is the unique maximum of total welfare $TW(\mu)$.

In view of Corollaries 1 and 3, total welfare is maximal when illiquidity and the mean reversion in returns are at their lowest levels.$^{15}$ Furthermore, because of monotonicity, the above market quality indicators, become “measurable” welfare indexes.

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 (14a) and (14b), this is given by:

$$
\phi(\mu) \equiv CE^{FD} - CE^{SD} = \gamma \left( \ln \left( 1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} + \frac{\text{Var}[E[p_2 - p_1|p_1]]}{\text{Var}[p_2 - p_1|p_1]} \right) - \ln \left( 1 + \frac{\text{Var}[E[v - p_1|p_1]]}{\text{Var}[v - p_1|p_1]} \right) \right)
$$

Competition

$$
+ \ln \left( 1 + \frac{\text{Var}[E[v - p_2|p_1,p_2]]}{\text{Var}[v - p_2|p_1,p_2]} \right).
$$

Liquidity supply

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, and according to the next result it is a decreasing function of $\mu$.$^{16}$

**Corollary 8.** The inverse demand for technological services $\phi(\mu)$ is decreasing in $\mu$.

A marginal increase in $\mu$ heightens the competition FDs face in the established and new venues, as well as the competition between FDs and SDs for the orders accruing to the established venue. Other things equal, 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 $\mu$.

**Numerical Result 2.** Assuming (4), when $\mu$, $\tau_u$, and $\tau_v$ are sufficiently large and $\gamma$ is large above $\gamma^L$, $\phi(\mu)$ is log-convex in $\mu$:

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

$^{15}$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).

$^{16}$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.
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 \( \mu \) becomes increasingly smaller as \( \mu \) increases.\(^{17}\) As we argue in the next section, this property of the demand function for technological services can have important implications for the nature of platform competition.

4.3 The supply of technological services and exchange equilibrium

Depending on market organization, 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

\[
\pi(\mu) = (\phi(\mu) - c)\mu,
\]

where \( c \) denotes the marginal cost of producing a capacity \( \mu \). We denote by \( \mu^M \) the optimal capacity decision of the monopolist exchange:

\[
\mu^M \in \arg\max_{\mu \in (0,1]} (\phi(\mu) - c)\mu.
\]

In the latter case, denoting by \( \mu_i \) 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 \( \mu_i \), an exchange \( i \)'s profit function is given by

\[
\pi(\mu_i, \mu_{-i}) = (\phi(\mu) - c)\mu_i - f.
\]

We define a symmetric Cournot equilibrium as follows:

**Definition 1.** A symmetric Cournot equilibrium in technological service capacities is a set of capacities \( \mu^C_i \in (0,1], i = 1, 2, \ldots, N \), such that (i) each \( \mu^C_i \) maximizes (25), for given capacity choice of other exchanges \( \mu^C_{-i} \):

\[
\mu^C_i \in \arg\max_{\mu_i} \pi(\mu_i, \mu^C_{-i}),
\]

(ii) \( \mu^C_1 = \mu^C_2 = \cdots = \mu^C_N \), and (iii) \( \sum_{i=1}^N \mu^C_i = \mu^C(N) \).

We have the following result:

**Proposition 9.** There exists at least one symmetric Cournot equilibrium in technological service capacities and no asymmetric ones. Numerical simulations show that the equilibrium is unique and stable.\(^{18}\)

**Proof.** See Amir (2018), Proposition 7, and Vives (1999), Section 4.1. \( \square \)

\(^{17}\)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 \( \mu, \tau_u, \) or \( \tau_v \) low, and for \( \gamma^L > \gamma \) (e.g., this happens when \( \tau_u = 1, \tau_v = 6, \mu = 0.2, \) and \( \gamma^L = 0.41, \gamma = 0.01 \)).

\(^{18}\)In our setup, a sufficient condition for stability (Section 4.3 in Vives (1999)) is that the elasticity of the
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 \((\mu_i, \mu_{-i})\). 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 \((\mu_i, \mu_{-i})\) since this is equivalent to \(\phi\) 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 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

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

increasing in its rivals’ choices (see Figure 3, panel (b)).

**Numerical Result 3.** When \(N = 2\), strategic complementarities in capacity decisions can arise for some range of exchanges’ best response (see (27)).

More in detail, assuming \(c > 0\), our numerical results show that two parameter configurations lead to strategic complementarities\(^{19}\)

1. \(\gamma \geq \gamma^L\), and low \(\tau_u\), and \(\tau_v\). In this case, liquidity traders have a high demand for liquidity (on account of a low risk tolerance and high payoff and endowment risk), and exchanges step up capacity to exploit FDs’ liquidity provision opportunities.

2. \(\gamma < \gamma^L\), and intermediate \(\tau_u\), and \(\tau_v\). In this case, liquidity traders have a low demand for liquidity, but in the second period can speculate more aggressively (due to the relatively

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

Numerical simulations where conducted using the following grid: \(c \in \{0.01, 0.02\}, N = 2, 3, \ldots, 30, \gamma \in \{0.01, 0.02, \ldots, 1\}, \tau_u, \tau_v \in \{1, 2, \ldots, 10\}\), 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).

\(^{19}\)Numerical simulations were conducted assuming \(c \in \{0.001, 0.005, 0.01, 0.08\}, \tau_v, \tau_u \in \{5, 10\}, \gamma, \gamma^L \in \{0.1, 0.2, \ldots, 1\}\).
higher risk tolerance, and low payoff and endowment risk), and exchanges step up capacity to exploit FD ability to speculate on short term returns.

The case where $\gamma \geq \gamma^L$ is empirically more reasonable, since liquidity providers are likely to be better capitalized than traders, and thus exhibit a higher risk tolerance.

Comparative statics analysis shows that for $N = 2$, other things equal, an increase in $\tau_u$, $\tau_v$, $\gamma^L$ or $c$ leads to a lower capacity provision (Figure 4, panels (a)-(c), and (e)). Conversely, a higher $\gamma$ leads to a higher $\mu^C_i$ (Figure 4, panel (d)). Thus, when market participants need less the market (either because they face lower payoff or endowment risk, or because they have a higher tolerance for risk), or when the marginal cost of installing capacity increases, platforms scale down their capacity provision. When, instead, liquidity providers have a higher risk bearing capacity, the opposite effect obtains.

Figure 3: Log-convexity of the demand function and strategic complementarities in platforms’ capacity decisions.

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:

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

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 (28) 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:

\[
\frac{\partial \mu^C(N)}{\partial N} \geq 0 \quad (29a)
\]

\[
\frac{\partial \pi_i(\mu)}{\partial N} \bigg|_{\mu = \mu^C(N)} \leq 0. \quad (29b)
\]

If the number of competing platforms is exogenously determined, condition (29a) implies that spurring competition in the intermediation industry has positive effects in terms of liquidity and total welfare (Proposition 1 and Numerical Result 1):

**Corollary 10.** At a stable Cournot equilibrium, an exogenous increase in the number of competing exchanges has a positive impact on liquidity and total welfare: \( \partial \Lambda_t/\partial N < 0, \partial TW/\partial N > 0 \).

Degryse et al. (2015) study 52 Dutch stocks in 2006-2009 (listed on Euronext Amsterdam and trading on Chi-X, Deutsche Börse, Turquoise, BATS, Nasdaq 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 (29a), and the resulting impact this has on liquidity and total welfare.\(^{20}\) To illustrate this effect, in Figure 5 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 up their individual capacity, with a positive effect on liquidity (panels (b) and (c)), and welfare (panel (d)). Note that the stronger is the effect on each platform technological capacity, the higher are the liquidity and welfare improvements.

### 5 Endogenous platform entry and welfare

In this section we endogenize platform entry, and study its welfare implications.\(^{21}\) Assuming that platforms’ technological capacities are identical (\( \mu = N \mu_c \)), a social planner who takes into

---

\(^{20}\)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)).

\(^{21}\)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.
account the costs incurred by the exchanges faces the following objective function:

\[
P(\mu, N) \equiv TW(\mu) - c\mu - fN = \pi(\mu)N + \phi(\mu). \tag{30}
\]

Expression (30) 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:

\[
\pi(\mu)N = (\phi(\mu) - c)\mu - fN = \phi(\mu)\mu - c\mu - fN.
\]

Therefore, FDs surplus only contributes indirectly to the planner’s function, via platforms’ total profit. The second component in (30) reflects the welfare of other market participants:

\[
\phi(\mu) = CE_{SD} + CE_{L1} + CE_{L2},
\]

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

We consider four possible outcomes:

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

\[
(\phi(\mu_C(N)) - c)\mu_C(N) \geq f > (\phi(\mu_C(N + 1)) - c)\mu_C(N + 1), \tag{31}
\]

which pins down \( N \). We denote by \( \mu_{CFE} \), and \( N_{CFE} \) the pair that solves the Cournot case, and note that, given Proposition 9 and (29b), a unique Cournot equilibrium with free entry obtains in our setup.

2. Structural Second Best (STR). In this case we posit that the planner has no control over the fee that competing exchanges charge to FDs, but 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:

\[
\max_{N \geq 1} P(\mu_C(N), N) \text{ s . t. } \mu_C(N) \text{ is a Cournot equilibrium with } \pi_i^C(N) \geq 0, \tag{32}
\]

and denote by \( \mu_{ST} \), and \( N_{ST} \) the pair that solves (32).

3. Behavioral Second Best (BEH). In this case, we assume that the planner can set the fee that exchanges charge to FDs, taking into account a break-even condition. Because of Corollary 8, 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).
\( \phi(\mu) \) is invertible in \( \mu \), implying that setting the fee is equivalent to choosing the aggregate technological capacity \( \mu \). Thus, we look for a solution to the following problem:

\[
\max_{\mu \in (0,1]} \mathcal{P}(\mu, N) \text{ s.t. } \left( \phi(\mu(N)) - c \right) \frac{\mu(N)}{N} \geq f > \left( \phi(\mu(N + 1)) - c \right) \frac{\mu(N + 1)}{N + 1},
\]

and denote by \( \mu^{BEH} \) and \( N^{BEH} \) the pair that solves (33).

4. First Best (FB). In this case, we assume that the planner can regulate the market choosing \( \mu \) and \( N \):

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

We denote by \( \mu^{FB} \) and \( N^{FB} \) the pair that solves (34).

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

Our first set of results relates to the case in which the planner can regulate the technological service fee:

**Proposition 11.** When the planner can regulate the technological service fee:

1. \( N^{FB} = N^{BEH} = 1 \).

2. \( \mu^{FB} \geq \mu^{BEH} \geq \mu^{M} \). Therefore, \( \Lambda_t(\mu^{M}) \geq \Lambda_t(\mu^{BEH}) \geq \Lambda_t(\mu^{FB}) \).

**Proof.** In the First Best case, for given \( \mu \), the objective function (30) is decreasing in \( N \). Thus, to economise on fixed costs, the planner allows a monopolistic exchange to provide trading services. Similarly, at a Behavioral Second Best, exchanges break even, so that the planner chooses \( \mu \) to maximize the welfare of other market participants:

\[
\mu^{BEH} \in \arg \max_{\mu \in (0,1]} \pi(\mu) + \varphi(\mu).
\]

Since \( \varphi(t(\mu)) > 0 \), and \( (\phi(\mu) - c)(\mu/N) \) is decreasing in \( N \), the second best outcome is achieved by increasing \( \mu \) to the point that only a monopolistic exchange breaks even.

Compare now the co-location capacity decisions in the two planner’s problems. At an interior First Best, since \( N^{FB} = 1 \), we have

\[
\pi'(\mu) + \varphi'(\mu) = 0.
\]

Since \( \varphi'(\mu) > 0 \), the above condition implies that the planner increases co-location capacity to the point that the marginal welfare gain this yields on other market participants is offset by the marginal welfare loss experienced by the exchange. At a Behavioral Second Best, instead, such increase is limited by the break even condition, and \( \mu^{FB} \geq \mu^{BEH} \). Finally, because of profit maximization, a monopolist always sets \( \mu^{M} \) lower than the value at which it breaks even. \( \square \)
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 (31)). We thus compare this outcome to the Structural Second Best. Evaluating the first order condition of the planner at $N = N_{CFE}$ yields:

$$\frac{\partial P(\mu^C(N), N)}{\partial N} \bigg|_{N = N_{CFE}} = \pi_i(\mu^C(N), N) \bigg|_{N = N_{CFE}} = 0$$

$$+ N_{CFE} \frac{\partial \pi_i(\mu^C(N), N)}{\partial N} \bigg|_{N = N_{CFE}}$$

Profitability depression $< 0$

$$+ \varphi'(\mu) \frac{\partial \mu^C(N)}{\partial N} \bigg|_{N = N_{CFE}}$$

Liquidity creation $> 0$

According to (35), if the Cournot equilibrium is stable, platform entry has two countervailing welfare effects. The first one is a “profitability depression” effect, and captures the profit decline associated with the demand reduction faced by each exchange 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.

Those effects differ from the standard ones in a Cournot equilibrium with free entry (Mankiw and Whinston (1986)). The same profitability depressing effect of increasing the number of firms exists; the counteracting effect in the traditional Cournot model being 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 if there is business stealing (that is, individual output decreasing with the number of firms) then the profitability depressing effect of entry always dominates. In our case the profitability depressing effect of entry is counteracted by the liquidity impact that increased platform competition has on the welfare of market participants.

23 This is because at a stable equilibrium (29a) and (29b) hold.
We thus have the following result:

**Proposition 12.** When the planner cannot regulate the technological service fee:

1. If the Cournot equilibrium is stable, \( N^{\text{CFE}} > N^{\text{STR}} \), \( \mu^{\text{CFE}} > \mu^{\text{STR}} \), and \( \Lambda_t(\mu^{\text{CFE}}) < \Lambda_t(\mu^{\text{STR}}) \), whenever the profitability depression effect is stronger than the liquidity creation effect. Otherwise, the opposite inequalities hold.

2. The technological capacity at either the Cournot equilibrium with free entry or at the Structural Second Best is higher than at the unregulated monopoly: \( \min\{\mu^{\text{STR}}, \mu^{\text{CFE}}\} \geq \mu^{M} \). Therefore, \( \Lambda_t(\mu^{M}) \geq \max\{\Lambda_t(\mu^{\text{STR}}), \Lambda_t(\mu^{\text{CFE}})\} \).

To verify the possibility of excessive or insufficient entry compared to the Structural Second Best, we run numerical simulations.\(^{24}\)

**Numerical Result 4.** According to numerical simulations, both excessive and insufficient entry compared to the Structural Second Best can occur. Furthermore, at all solutions \( N \) and \( \mu \) are decreasing in \( f \).

Figure 6 illustrates the output of two simulations in which entry can be insufficient or excessive, depending on parameter values. Insufficient entry is more likely to obtain when marginal and fixed costs are high, the dispersion in traders’ endowment shock and liquidation value volatility is low, and traders’ risk tolerance is high. As argued above, this outcome arises because the liquidity creation effect weighs relatively more than profitability depression. The simulations suggest that this occurs when second period traders speculate more aggressively, or when the cost of providing technological services is high. Conversely, excessive entry tends to obtain when the provision of technological services is relatively cheap, and traders’ need more the market (in view of a higher risk aversion, and liquidation value/endowment risk).

Finally, the next result compares outcomes depending on whether fee regulation is available or not:

**Proposition 13.** The technological service capacity with fee regulation is higher than with entry regulation or with Cournot free entry, and the monopolist outcome features the lowest technological capacity:

\[ \mu^{\text{BEH}} \geq \max\{\mu^{\text{STR}}, \mu^{\text{CFE}}\} \geq \mu^{M}. \]  

Therefore,

\[ \Lambda_t(\mu^{FB}) \leq \Lambda_t(\mu^{\text{BEH}}) \leq \min\{\Lambda_t(\mu^{\text{STR}}), \Lambda_t(\mu^{\text{CFE}})\} \leq \Lambda_t(\mu^{M}). \]  

The number of exchanges entering the market with Cournot free entry or with entry regulation is higher than with fee regulation:

\[ \min\{N^{\text{CFE}}, N^{\text{STR}}\} \geq 1. \]  

\(^{24}\)The parameter space adopted in these simulations is as follows: \( f \in \{0.00001, 0.00002, \ldots, 0.00031\} \), \( c \in \{0.01, 0.02\} \), \( \gamma = 1 \), \( \gamma^L \in \{0.5, 0.6, \ldots, 1\} \), \( \tau_v, \tau_u \in \{5, 6, \ldots, 10\} \).
Whenever $TW(\mu)$ is increasing in $\mu$,

$$P^{FB} \geq P^{BEH} \geq P^{STR} \geq P^{CFE} \geq P^{M}. \quad (39)$$

The above result suggests that fee regulation achieves the outcome that is closest to the First Best, since it minimizes entry costs and forces the next to highest provision of technological services. However, the evidence presented in the introduction suggests that regulators’ ability to weigh on the technological fee-setting process is far from perfect. Thus, entry regulation appears as a realistic alternative instrument. Indeed, spurring entry achieves two objectives. First, it works as a corrective against exchanges’ market power in the provision of technological services, thereby stemming the monopolist temptation to restrict the supply of technological services to heighten their fees; additionally, by creating competitive pressure, it achieves the objective of keeping exchanges’ trading fees in check.

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 supply liquidity in 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 proportion of 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. Thus, allowing competition among trading platforms has a beneficial impact on liquidity and welfare. Regulatory intervention through controlled entry, or even better, technological service fee determination has the power to further improve welfare and liquidity. However, fee regulation is subject to rent-seeking efforts by market participants which suggests that entry regulation, even though inferior to fee regulation, appears as a viable alternative.

Our results suggest that strategic interaction among exchanges is an important driver of market liquidity, adding to the usual factors (e.g., arbitrage capital, risk bearing capacity of the market) the literature has recently highlighted. 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.
References


A Appendix

The following is a standard results (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

$$p_2 = -\Lambda_2 u_2 + \Lambda_{21} u_1$$  (A.1a)
$$p_1 = -\Lambda_1 u_1,$$  (A.1b)

with $\Lambda_1$, $\Lambda_{21}$, and $\Lambda_2$ 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

$$E[-\exp\{-\pi^L_2/\gamma^L\}|\Omega^L_2] = -\exp\left\{ -\frac{1}{\gamma^L} \left( E[\pi^L_2|\Omega^L_2] - \frac{1}{2\gamma^L} \text{Var}[\pi^L_2|\Omega^L_2] \right) \right\},$$  (A.2)

where $\Omega^L_2 = \{u_1, u_2\}$, and $\pi^L_2 \equiv (v - p_2)x^L_2 + u_2v$. Maximizing (A.2) with respect to $x^L_2$, yields:

$$X^L_2(u_1, u_2) = \gamma^L \frac{E[v - p_2|\Omega^L_2]}{\text{Var}[v - p_2|\Omega^L_2]} - \frac{\text{Cov}[v - p_2, v|\Omega^L_2]}{\text{Var}[v - p_2|\Omega^L_2]} u_2.$$  (A.3)

Using (A.1a):

$$E[v - p_2|\Omega^L_2] = \Lambda_2 u_2 - \Lambda_{21} u_1$$  (A.4a)
$$\text{Var}[v - p_2|\Omega^L_2] = \text{Cov}[v - p_2, v|\Omega^L_2] = \frac{1}{\tau_v}.$$  (A.4b)

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

$$X^L_2(u_1, u_2) = a_2 u_2 + bu_1,$$  (A.5)

where

$$a_2 = \gamma^L \tau_v \Lambda_2 - 1$$  (A.6a)
$$b = -\gamma^L \tau_v \Lambda_{21}.$$  (A.6b)
Consider the sequence of market clearing equations
\begin{align}
\mu x_1^{FD} + (1 - \mu)x_1^{SD} + x_1^{L} &= 0 \quad (A.7a) \\
\mu(x_2^{FD} - x_1^{FD}) + x_2^{L} &= 0. \quad (A.7b)
\end{align}

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):
\[
\mu x_2^{FD} + x_2^{L} + (1 - \mu)x_1^{SD} + x_1^{L} = 0. \quad (A.8)
\]

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:
\[
E[-\exp\left\{\frac{1}{\gamma} (p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD}\right\} | p_1, p_2] = \exp\left\{ -\frac{1}{\gamma} \left( E[v - p_2 | p_1, p_2] x_2^{FD} - \frac{(x_2^{FD})^2}{2\gamma} \text{Var}[v - p_2 | p_1, p_2] \right) \right\}. \quad (A.9)
\]

For given \(x_1^{FD}\) the above is a concave function of \(x_2^{FD}\). Maximizing with respect to \(x_2^{FD}\) yields:
\[
X_2^{FD}(p_1, p_2) = -\gamma \tau p_2. \quad (A.10)
\]

Similarly, due to CARA and normality, in the first period a traditional market maker maximizes
\[
E[-\exp\left\{ -\frac{1}{\gamma} (v - p_1)x_1^{SD}\right\} | p_1] = \exp\left\{ -\frac{1}{\gamma} \left( E[v - p_1 | p_1] x_1^{SD} - \frac{(x_1^{SD})^2}{2\gamma} \text{Var}[v - p_1 | p_1] \right) \right\}. \quad (A.11)
\]

Hence, his strategy is given by
\[
X_1^{SD}(p_1) = -\gamma \tau v p_1. \quad (A.12)
\]

Finally, consider a first period liquidity trader. CARA and normality imply
\[
E[-\exp\left\{ -\pi_1^{L} / \gamma \right\}] = \exp\left\{ -\frac{1}{\gamma} \left( E[\pi_1^{L} | u_1] - \frac{1}{2\gamma} \text{Var}[\pi_1^{L} | u_1] \right) \right\}, \quad (A.13)
\]

where \(\pi_1^{L} \equiv (v - p_1)x_1^{L} + u_1 v\). Maximizing (A.13) with respect to \(x_1^{L}\), and solving for the optimal strategy, yields
\[
X_1^{L}(u_1) = \gamma \frac{E[v - p_1 | u_1]}{\text{Var}[v - p_1 | u_1]} - \frac{\text{Cov}[v - p_1, v | u_1]}{\text{Var}[v - p_1 | u_1]} u_1. \quad (A.14)
\]
Using (A.1b):

\[ E[v - p_1|u_1] = \Lambda_1 u_1 \]  
(A.15a)

\[ \text{Cov}[v - p_1, v|u_1] = \frac{1}{\tau_v}. \]  
(A.15b)

Substituting the above in (A.14) yields

\[ X_1^L(u_1) = a_1 u_1, \]  
(A.16)

where

\[ a_1 = \gamma^L \tau_v \Lambda_1 - 1. \]  
(A.17)

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

\[ p_2 = -\frac{1 - \gamma^L \tau_v \Lambda_2}{\mu \gamma \tau_v} u_2 + \frac{(1 - \mu) \gamma + \gamma^L \tau_v \Lambda_1 - 1 - \gamma^L \tau_v \Lambda_2}{\mu \gamma \tau_v} u_1. \]  
(A.18)

Identifying the price coefficients:

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

\[ \Lambda_{21} = \Lambda_2 \left( (1 - \mu) \gamma + \gamma^L \tau_v \Lambda_1 - 1 \right). \]  
(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:

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

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

\[ \nu = E[p_2|u_1] = \Lambda_{21} u_1 \]  
(A.21a)

\[ \text{Var}[p_2|u_1] = \frac{\Lambda_2^2}{\tau_v}. \]  
(A.21b)
Maximizing (A.20) with respect to $x_{FD}^1$ and solving for the first period strategy yields

\begin{align*}
X_{FD}^1(p_1) &= \gamma \frac{E[p_2|u_1]}{\Var[p_2|u_1]} - \gamma \left( \frac{1}{\Var[p_2|u_1]} + \frac{1}{\Var[v]} \right) p_1 \\
&= \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.
\end{align*}

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

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

The remaining equilibrium coefficients are as follows:

\begin{align*}
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} \\
\Var[p_2|u_1] &= \frac{\Lambda_2^2}{\tau_u} \tag{A.28}
\end{align*}

where $\Lambda_2$ is given by (A.19a).

**Proof of Corollary 2**

The first part of the corollary follows from (13). Also, since $\Lambda_t$ is decreasing in $\mu$, because of (10), $|a_t|$ is increasing in $\mu$. Finally, substituting (A.27) in (A.26) and rearranging yields

\begin{equation*}
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)},
\end{equation*}

which is increasing in $\mu$. \hfill \Box

**Proof of Corollary 3**

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

\begin{equation*}
\Cov[p_2 - p_1, p_1] = -\Lambda_1 (\Lambda_1 + \Lambda_{21}) \tau_u^{-1} \\
= -\gamma^L \Lambda_1 \Lambda_2 \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{equation*}

which, in view of the fact that $\Lambda_t^*$ is decreasing in $\mu$, proves the result. \hfill \Box
Proof of Proposition 4

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

\[
E\left[U((v - p_1)x_1^{SD})|p_1\right] = -\exp\left\{ -\frac{(E[v|p_1] - p_1)^2}{2\text{Var}[v]} \right\} \\
= -\exp\left\{ -\frac{\tau_v \Lambda_1^2}{2} u_1^2 \right\}.
\] (A.29)

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

\[
EU^{SD} \equiv E\left[U((v - p_1)x_1^{SD})\right] = -\left( 1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right)^{-1/2} \\
= -\left( \frac{\tau_u}{\tau_u + \tau_v \Lambda_1^2} \right)^{1/2},
\] (A.30)

and

\[
CE^{SD} = \frac{\gamma}{2} \ln \left( 1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right).
\] (A.31)

Differentiating \( CE^D \) with respect to \( \mu \) yields:

\[
\frac{\partial CE^{SD}}{\partial \mu} = \frac{\gamma \tau_v}{2} \left( 1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right)^{-1} \frac{\partial \text{Var}[p_1]}{\partial \mu} \\
= \frac{\gamma \tau_v}{2 \tau_u} \left( 1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} \right)^{-1} 2 \Lambda_1 \frac{\partial \Lambda_1}{\partial \mu} < 0,
\] (A.32)

since \( \Lambda_1 \) is decreasing in \( \mu \).

Turning to FDs. Replacing \( \text{A.22} \) in \( \text{A.20} \) and rearranging yields

\[
E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})|u_1] = -\left( 1 + \frac{\text{Var}[p_2|u_1]}{\text{Var}[v]} \right)^{-1/2} \times \exp\left\{ -\frac{g(u_1)}{\gamma} \right\},
\] (A.33)

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 \( \text{A.33} \) is a quadratic form of the first period endowment shock. We can therefore apply Lemma 1 and obtain

\[
EU^{FD} \equiv E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})] = \\
= -\left( 1 + \frac{\text{Var}[p_2|p_1]}{\text{Var}[v]} \right)^{-1/2} \left( 1 + \frac{\text{Var}[p_1]}{\text{Var}[v]} + \frac{\text{Var}[E[p_2|p_1] - p_1]}{\text{Var}[p_2|p_1]} \right)^{-1/2},
\] (A.34)
where, because of (A.21a),
\[
\text{Var} \left[ E[p_2 - p_1] \right] = (\Lambda_{21} + \Lambda_1)^2 \tau_u^{-1},
\]
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
\[
CE^{FD} = \frac{\gamma}{2} \left\{ \ln \left( 1 + \frac{(\Lambda_2)^2 \tau_v}{\tau_u} \right) + \ln \left( 1 + \frac{(\Lambda_1)^2 \tau_v}{\tau_u} + \left( \frac{\Lambda_{21} + \Lambda_1}{\Lambda_2} \right)^2 \right) \right\}. \tag{A.36}
\]
Computing,
\[
\frac{\Lambda_{21} + \Lambda_1}{\Lambda_2} = \frac{\gamma^L}{\gamma + \gamma^L + (\gamma + 2\gamma^L)(\mu \gamma + \gamma^L) \mu \gamma \tau_u \tau_v}. \tag{A.37}
\]
Thus, the arguments of the logarithms in (A.36) are decreasing in \( \mu \), which proves that \( CE^{FD} \) is decreasing in \( \mu \).

Finally, note that taking the limits for \( \mu \to 0 \) and \( \mu \to 1 \) in (A.31) and (A.36) 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. \( \square \)

**Proof of Proposition 5**

Consider now first period liquidity traders. Evaluating the objective function at optimum and rearranging yields
\[
-\exp \left\{ -\frac{1}{\gamma L} \left( E[\pi^L_1|u_1] - \frac{1}{2 \gamma L} \text{Var}[\pi^L_1|u_1] \right) \right\} = -\exp \left\{ -\frac{u_1^2}{\gamma L} \left( \frac{a_1^2}{2 \gamma^L \tau_v} - 1 \right) \right\},
\]
where \( u_1 \sim N(0, \tau_u^{-1}) \). The argument at the exponential is a quadratic form of a normal random variable. Therefore, applying again Lemma 1 yields
\[
E \left[ -\exp \left\{ \frac{\pi^L_1}{\gamma^L} \right\} \right] = -\left( \frac{(\gamma^L)^2 \tau_u \tau_v}{(\gamma^L)^2 \tau_u \tau_v - 1 + a_1^2} \right)^{1/2}, \tag{A.38}
\]
so that
\[
CE^{L}_1 = \frac{\gamma^L}{2} \ln \left( 1 + \frac{a_1^2 - 1}{(\gamma^L)^2 \tau_u \tau_v} \right). \tag{A.39}
\]
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

\[
E \left[ - \exp \left\{ - \frac{\pi L}{\gamma L} \right\} | \Omega_L^t \right] = - \exp \left\{ - \frac{1}{\gamma L} \left( \frac{(x_2^L)^2 - u_2^2}{2\gamma L \tau_v} \right) \right\} \\
= - \exp \left\{ - \frac{1}{\gamma L} \left( x_2^L u_2 \right) \left( \frac{1}{2\gamma L \tau_v} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) \left( \begin{pmatrix} x_2^L \\ u_2 \end{pmatrix} \right) \right\}.
\]

(A.40)

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

\[
\Sigma = \begin{pmatrix} \text{Var}[x_2^L] & a_2 \text{Var}[u_2] \\ a_2 \text{Var}[u_2] & \text{Var}[u_2] \end{pmatrix}.
\]

(A.41)

Therefore, we can again apply Lemma 1 to (A.40), obtaining

\[
E \left[ E \left[ - \exp \left\{ \frac{\pi L}{\gamma L} \right\} | \Omega_L^t \right] \right] = - | I + (2/\gamma L) \Sigma A |^{-1/2},
\]

where

\[
A = \frac{1}{2\gamma L \tau_v} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
\]

\[
\text{Var}[x_2^L] = \frac{a_2^2 + b^2}{\tau_u}.
\]

(A.43)

Substituting (A.41), (A.43), and (A.44) in (A.42) and computing the certainty equivalent, yields:

\[
CE^L_2 = \frac{\gamma L}{2} \ln \left( 1 + \frac{a_2^2 - 1}{(\gamma L)^2 \tau_u \tau_v} + \frac{b^2((\gamma L)^2 \tau_u \tau_v - 1)}{(\gamma L)^4 \tau_u \tau_v^2} \right).
\]

(A.45)

For \( \mu = 0, b = 0 \) and, in view of Corollary 2 \( CE^L_1 > CE^L_2 \). The same condition holds when evaluating (A.39) and (A.45) at \( \mu = 1 \). As \( CE^L_t \) is increasing in \( \mu \), we have that for all \( \mu \in (0, 1) \), \( CE^L_1(\mu) > CE^L_2(\mu) \). \( \Box \)

**Proof of Corollary 6**

We need to prove that:

\[
\frac{\partial CE^L_1(\mu)}{\partial \mu} \geq - \frac{\partial CE^{SD}(\mu)}{\partial \mu}.
\]

Computing:

\[
\frac{\partial CE^L_1(\mu)}{\partial \mu} = \frac{\gamma L a_1 a'_1}{(\gamma L)^2 \tau_u \tau_v - 1 + a_1^2}
\]

(A.46)
and
\[
\frac{\partial C^{ESD}(\mu)}{\partial \mu} = \frac{\gamma(1 + a_1)a_1'}{(\gamma^L)^2\tau_u\tau_v + (1 + a_1)^2}.
\]

(A.47)

First, note that the denominator in (A.47) is higher than the one in (A.46). Next, comparing the numerators in the above expressions yields:
\[
\gamma^L a_1 a_1' > -\gamma(1 + a_1) a_1' \iff (\gamma^L a_1 + \gamma(1 + a_1)) a_1' > 0,
\]
as can be checked by substituting (A.24) in the above.

\[\square\]

Proof of Corollary 7

The first part of the result follows immediately from (20), and Corollary 6. Next, because of Propositions 4 and 5, \(TW(1) > \lim_{\mu \to 0} TW(\mu)\), which rules out the possibility that total welfare is maximized at \(\mu \approx 0\).

\[\square\]

Proof of Corollary 8

Note that because of (A.37), 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 (21), we have:
\[
\phi(\mu) = C^{FD} - C^D
\]
\[
= \frac{\gamma}{2} \left\{ \ln \left( 1 + \frac{\Lambda_2^2 \tau_v}{\tau_u} \right) + \ln \left( 1 + \frac{\Lambda_1^2 \tau_v}{\tau_u} K \right) - \ln \left( 1 + \frac{\Lambda_1^2 \tau_v}{\tau_u} \right) \right\} > 0.
\]

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 \(\mu\). The first term inside curly braces in the above expression is decreasing in \(\mu\) since \(\Lambda_2\) is decreasing in \(\mu\). The difference between the second and third terms can be written as follows:
\[
\ln \left( 1 + \frac{\Lambda_2^2 \tau_v}{\tau_u} K \right) - \ln \left( 1 + \frac{\Lambda_1^2 \tau_v}{\tau_u} \right) = \ln \left( \frac{\tau_u + \Lambda_1^2 \tau_v}{\tau_u + \Lambda_1^2 \tau_v} K \right).
\]

Differentiating the above logarithm and rearranging yields:
\[
\frac{\tau_v \Lambda_1}{(\tau_u + \Lambda_1^2 \tau_v K)(\tau_u + \Lambda_1^2 \tau_v)} \left( (2(K - 1)\tau_u \partial \Lambda_1 / \partial \mu + (\tau_u + \Lambda_1^2 \tau_v)\Lambda_1 \partial K / \partial \mu) < 0, \right.
\]
since \(K > 1\), and both \(\Lambda_1\) and \(K\) are decreasing in \(\mu\).

\[\square\]
Proof of Proposition 11

At a Behavioral Second Best, the planner sets $\mu$ to the highest possible value that allows an exchange to break even. This occurs when the exchange’s profit equates the fixed cost:

$$(\phi(\mu) - c) \frac{\mu}{N} = f.$$  \hfill (A.49)

Let $\mu^{BEH}$ be the highest $\mu$ that is compatible with (A.49) when $N = 1$. As the average profit declines in $N$, increasing competition implies that the planner has to choose $\mu < \mu^{BEH}$ to satisfy (A.49). However, this implies a lower welfare since $\varphi'(\mu) > 0$.

Proof of Proposition 12

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 solves for the Structural Second Best can be written as follows:

$$P(\mu^{C}(N), N) = N \pi_i(\mu^{C}(N)) + \varphi(\mu^{C}(N)),$$  \hfill (A.50)

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

$$\varphi(\mu^{C}(N)) = CE^{SD}(\mu^{C}(N)) + CE^{T}(\mu^{C}(N)) + CE^{L}(\mu^{C}(N)).$$

Consider now the first order condition of the planner, and evaluate it at $N^{CFE}$:

$$\frac{\partial P(\mu^{C}(N), N)}{\partial N} \bigg|_{N=N^{CFE}} = \pi_i(\mu^{C}(N^{CFE}), N) \bigg|_{=0} + N^{CFE} \frac{\partial \pi_i(\mu^{C}(N), N)}{\partial N} \bigg|_{N=N^{CFE}} < 0 + \varphi'(\mu^{C}(N)) \frac{\partial \mu^{C}(N)}{\partial N} \bigg|_{N=N^{CFE}} > 0.$$

The first term on the right hand side of (A.51) is null at $N^{CFE}$. 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 instead respectively negative and positive. Depending on which of the two terms prevails, we thus have

$$\frac{\partial P(\mu^{C}(N), N)}{\partial N} \bigg|_{N=N^{CFE}} \geq 0 \implies N^{CFE} \leq N^{STR}.$$

To compare the technological fee at a Cournot equilibrium with that of the monopoly
outcome, consider the first order conditions of these two problems:

\[
\text{CFE: } \mu \phi'(\mu) + N(\phi(\mu) - c) = 0 \quad \implies \quad \mu^C(N) = -\frac{(\phi(\mu) - c)N}{\phi'(\mu)} \quad (A.52a)
\]

\[
\text{M: } \mu \phi'(\mu) + \phi(\mu) - c = 0 \quad \implies \quad \mu^M = -\frac{\phi(\mu) - c}{\phi'(\mu)}. \quad (A.52b)
\]

Comparing (A.52a) with (A.52b) yields \( \mu^C(N) \geq \mu^M \), for \( N \geq 1 \). A similar argument holds at a Structural Second Best, since in this case the planner picks \( N \) subject to \( \mu \) being a Cournot equilibrium.

\[\square\]

**Proof of Proposition 13**

Suppose \( \mu^{CFE} > \mu^{BEH} \). Given that at a Behavioral Second Best exchanges break even, as \( \phi'(\mu) > 0 \), this implies that

\[
(\phi(\mu^{CFE}) - c)\mu^{CFE} < f. \quad (A.53)
\]

However, at a Cournot equilibrium with free entry with \( N > 1 \) exchanges, we have

\[
(\phi(\mu^{CFE}) - c)\frac{\mu^{CFE}}{N} = f. \quad (A.54)
\]

Putting together (A.53) and (A.54) yields

\[
f = (\phi(\mu^{CFE}) - c)\frac{\mu^{CFE}}{N} < (\phi(\mu^{CFE}) - c)\mu^{CFE} < f,
\]

which is impossible. Thus, we must have \( \mu^{BEH} \geq \mu^{CFE} \). The comparison between \( \mu^{BEH} \) and \( \mu^{STR} \) runs along similar lines since at a Structural Second Best the planner’s problem incorporates a break even condition for exchanges’ profits.

Finally, the welfare ranking follows from the fact that \( N^{FB} = N^{BEH} = 1 \leq \min\{N^{STR}, N^{CFE}\} \), and \( \mu^{FB} \geq \mu^{BEH} \geq \max\{\mu^{STR}, \mu^{CFE}\} \). Thus, when \( TW(\mu) \) is increasing in \( \mu \), \( P(\mu^{FB}, N^{FB}) \geq P(\mu^{BEH}, N^{BEH}) \). Because of (31) and (32), we also have \( P(\mu^{STR}, N^{STR}) \geq P(\mu^{CFE}, N^{CFE}) \). Thus, \( P(\mu^{BEH}, N^{BEH}) \geq P(\mu^{STR}, N^{STR}) \geq P(\mu^{CFE}, N^{CFE}) \), completing the ranking. \[\square\]
Figure 7: From 2016 ICE operates as two business segments “Trading and Clearing” (transaction-based execution and clearing) and “Data and Listings” (subscription-based data services and listings). The figure shows the evolution of the revenues from technological services. See ICE-10k (2017).
Figure 4: Comparative statics in the $N = 2$ Cournot platform capacity game.
Figure 5: Effect of entry on each platform capacity decisions (panel (a)), liquidity (panels (b) and (c)), and total welfare (panel (d)).
Figure 6: The top part of panel (a) illustrates a case in which insufficient entry occurs, plotting the number of exchanges obtaining in the Cournot Free Entry case (CFE, in green), those obtaining with entry regulation (STR, in red), and with fee regulation (BEH, in blue), as a function of $f$. The bottom part illustrates the corresponding capacity choices, including the Monopolist and First Best cases (respectively in black and yellow). Panel (b) illustrates a case in which entry is instead excessive, compared to all welfare benchmarks.