The Price of Exchange Data*

Thierry Foucault† Vincent Maurin‡

This version: July 2026

Abstract

Exchanges increasingly generate revenues from selling transaction data, in addition to their traditional business of matching traders. Given exchanges' market power on their data, this raises questions about the quality and the pricing of these data, and exchanges' contribution to market transparency. We study the optimal design of an exchange that sells both access to its trading platform and transaction data generated by traders' activity. We show that the introduction of a market for transaction data reduces trading fees and increases volume and welfare. Caps on data fees can increase data quality but lower welfare.

JEL classification: G10, G14

Keywords: Exchange Design, Sale of Market Data, Market Transparency

*We thank Simon Gervais, Sophie Moinas and seminar participants at HEC, the 2026 Summer Meeting of the Finance Theory Group at INSEAD and the Online FTG Working Group on Market Microstructure for helpful comments.

†HEC. E-mail: foucault@hec.fr.

‡HEC Paris. E-mail: maurin@hec.fr

1 Introduction

> *"The current system is heavily tilted toward the exchanges, who can sell their prop data feeds at increasingly high prices without any meaningful competition from the public feeds or otherwise. The investing public pays the price."* Commissioner Caroline Crenshaw, SEC (2020).

Data have become a fundamental input into economic activity, leading to the emergence of markets for information, in which data are sold by brokers or digital platforms such as Google, Meta, and Amazon (Bergemann and Bonatti, 2019). In line with this evolution, exchanges, which provide platforms for matching traders, increasingly derive revenue from the sale of data generated by trading activity. This trend has become controversial as some market participants claim that exchanges charge excessive fees for their data.[1] As a result, whether exchanges should be allowed to sell data and which data should remain proprietary have become key questions on the regulatory agenda in both the United States and Europe.[2]

The market for exchange data and the market for matching services—exchanges' core business—are intrinsically linked. The value of the data depends on trading activity on the platform, while access to data affects trading activity. Assessing policy intervention therefore requires a joint analysis of exchange data and trading services. Yet, the existing theoretical literature typically studies these two markets in isolation.

To bridge this gap, we develop a model in which an exchange jointly sells trading services and trading data. We use this model to analyze the effects of data sales on trading costs, trading volume, transparency, and welfare—outcomes that are central to both policy debates and empirical analyses of exchange data markets.

[1] See, for instance, "Wall Street goes to war over market data and access", The Trade, Jan. 3, 2019 or Securities Industry and Financial Markets Association (2020) and Chapter 4 in Duffie et al. (2022).

[2] See U.S. Securities and Exchange Commission (2020) and European Commission (2021).

Our analysis rests on the premise that exchange data convey information about asset fundamentals, which can be used to speculate or to reduce exposure to adverse selection.[3] This observation generates two insights about exchange data economics. First, the quality of data depends on the informativeness of the order flow, which is jointly shaped by exchanges' data policies and trading fees. Second, the usage value of data also depends on trading fees: who participates in trading determines the usefulness of acquiring data. Therefore, accounting for fee adjustments is key to our analysis of the market for exchange data.

We capture these insights in a model where an exchange operates two related markets for correlated assets.[4] Each market features three types of risk-neutral, profit-maximizing traders (specialized by market): (i) speculators who can acquire private information about the asset payoff in their market, (ii) liquidity traders with private gains from trading, and (iii) a market maker who posts bid and ask quotes at which they execute other traders' unit sell and buy orders. Trading in the first ("upstream") market generates order-flow information that is valuable to participants in the second ("downstream") market. The exchange can thus sell a proprietary data signal, choosing its informativeness—called quality—up to the informational content of the order flow. The exchange sets trading fees for each market, data quality, and a data fee (its "design") to maximize profits from trading fees and data sales.

We first solve for the equilibrium in both markets, taking the exchange design as given. In the upstream market, the market maker charges a spread that reflects both its pricing power and its exposure to adverse selection from upstream speculators. Importantly, in the

[3]For instance, Bartlett et al. (2025) show that data on odd-lot orders contain incremental information about future price changes relative to so-called "SIP" data about trades. SIP data are available to the general public, while odd-lot data are sold by exchanges to sophisticated investors.

[4]One asset might be an index ETF (e.g., the iShares MSCI World ETF traded on NYSE Arca), and the other asset a futures contract on that index (e.g., the ICE MSCI World Futures traded on the Intercontinental Exchange, which owns the NYSE). Alternatively, one asset might be an option written on the other asset. For instance, Nasdaq operates three option exchanges in the US and sells proprietary options data products.

downstream market, we show that only the market maker buys exchange data, to reduce its exposure to adverse selection from downstream speculators.[5] In equilibrium, the market maker's quotes reveal the exchange data to other participants. Thus, exchange data dissemination endogenously narrows the information gap between speculators and other traders.

This observation leads to our first main result. Relative to a benchmark without data dissemination, exchange data curbs speculation in the downstream market. This reduces adverse selection, increases trading volume, and increases (utilitarian) welfare. Furthermore, the exchange lowers its trading fees because trading services and exchange data are complements.[6] Lower trading fees increase profits from data sales for two reasons. Lower upstream fees stimulate upstream speculation, which increases the informativeness of the order flow (higher data quality).[7] Lower downstream fees help maintain downstream speculation, preserving the market maker's exposure to adverse selection and its willingness to pay for data. Despite the increase in speculation from lower fees, total welfare increases as data dissemination improves public information revelation about fundamentals.

Our second main result relates to the imposition of a cap on the exchange's data fees. We consider an extreme cap by comparing the laissez-faire outcome with another benchmark in which the exchange can disseminate data, but cannot monetize it. Without monetization, the exchange typically chooses a higher (not lower) data quality than under laissez-faire. This counterintuitive result arises for two reasons. First, without monetization, the exchange

[5]This is consistent with the fact that major market making firms (Citadel Securities, Jane Street, Virtu, Optiver, etc.) buy proprietary data from most exchanges. Moreover, in practice, not all trading firms buy market data. For instance, Hendershott et al. (2025) find that 20% of trading firms (brokers, investment banks, prop trading firms or market making firms) in their sample do not buy any data from the NYSE. This is also consistent with our model: speculators and liquidity traders choose not to buy the data in equilibrium.

[6]This point is different from that in Glosten (2020) who argues that the datafeed from different exchanges are complementary goods, not that trading services and exchange data are complements.

[7]This is not obvious, as lower trading fees could increase participation from both speculators and liquidity traders. However, the second effect is muted in equilibrium when speculation is endogenous, as in our model.

benefits from disseminating high-quality data to curb adverse selection in the downstream market and increase total volume. Data monetization, on the other hand, incentivizes the exchange to limit data quality. Indeed, higher data quality reduces downstream speculation, which ultimately erodes the value of data.[8] These forces imply that a cap on data fees increases data quality for a large range of parameters. For these parameters, holding trading fees fixed, our model predicts that total trading costs fall and volume and welfare increase.

However, such results fail to account for the exchange's adjustment to other policies when faced with a cap on data fees. We show that the exchange would react to the cap by increasing trading fees in the downstream market. Without monetization, the exchange does not keep trading fees low, because it no longer captures the value of information sold to the downstream market maker. Under a large set of parameters, this reaction to the cap on data fees can outweigh the welfare benefits from greater market informativeness due to the increase in data quality. As we discuss below, these results suggest that empirical analyses of data policy changes should account for their likely impact on exchanges' pricing of access.

Finally, we consider two variants of the baseline model. First, we study the case in which the exchange's trading fees are constrained to be zero. This "competition for order flow" case captures the following idea: Even if exchanges face competition from other platforms for asset trading[9], they would retain market power over their data because proprietary data feeds, even for the same asset, are imperfect substitutes (Spatt, 2020). Competition in trading fees weakens incentives to sell high-quality data, as the exchange no longer benefits from higher

[8]Exchanges have argued that monetizing their data increases their incentives to improve market data products. See, for instance, Nasdaq's comments on Proposed Rule on Market Data Infrastructure. To the extent that data sold by an exchange is used on its platform, our analysis challenges this assertion: The exchange provides higher-quality data when it cannot monetize them in our model.

[9]A given stock in the U.S. typically trades on multiple competing trading platforms. However, in European equity markets, switching to alternative trading venues is costly for some investors (e.g., due to differences in clearing arrangements). As another example, derivatives markets offering proprietary contracts, such as SPX or VIX options at Cboe, or Brent crude oil futures at ICE are not readily substitutable across venues.

volume on its platform. Nevertheless, we find that the dominant effect of competition for order flow is to increase order flow informativeness, thereby pushing the technological frontier for data quality. However, welfare can be lower with zero trading fees, precisely because they result in excessive investment in private information acquisition in the upstream market.[10]

Second, we study the case in which the two markets are operated by distinct exchanges. In this case, exchanges do not internalize complementarities between trading and data revenues, resulting in lower data quality, higher trading fees, and lower welfare than under an integrated exchange. These findings suggest that exchanges enjoy economies of scope by better exploiting complementarities between the sale of data and the sale of trading services.

Our model offers a framework for interpreting recent empirical evidence. Some studies analyze changes in exchanges' data quality. Hendershott et al. (2025) find that the introduction of a more granular NYSE limit order book data product increased trading activity by both subscribers and non-subscribers. In our model, an increase in data quality indeed predicts higher trading volume and lower informational asymmetries. Conversely, Hendershott and Jones (2005) study an episode in which a trading platform stopped disseminating data for three exchange-traded funds (ETFs), as in the no-data benchmark of our model. They find that trading volume declined, particularly among liquidity traders, while trading costs increased, which is again consistent with our predictions.

Brogaard et al. (2024) find that charging users for data (data monetization) reduces trading volume and price discovery while increasing trading costs. Our model rationalizes these results when treating trading fees as fixed, which is likely appropriate in the short run[11] However, our model suggests that trading fees would ultimately adjust such that

[10]Trading fees act as a (welfare-enhancing) tax on speculation. With zero trading fees, the jump in speculation can outweigh the benefits of higher data quality under competition for order flow. These results do not contradict our findings about the benefits of exchange data; there the exchange chooses trading fees.

[11]Brogaard et al. (2024) find no major changes in trading fees around the events considered in their study.

trading volume would fall without monetization, reversing the fixed-trading-fee conclusion.

The rest of the paper is organized as follows. In the next section, we position our contribution relative to related theoretical papers. Section 3 presents the model. Section 4 solves for the equilibrium for a fixed exchange design. Section 5 studies the interaction between the exchange's data sales and the sale of matching services. Section 6 analyzes optimal exchange policies and a cap on data fees. Section 7 considers two extensions with competition for order flow and market fragmentation, respectively. Finally, Section 8 concludes.

2 Contribution to the Literature

Only a few theoretical papers study data sales by exchanges. Cespa and Foucault (2014) and Easley et al. (2016) study sales of past transaction prices by exchanges in a noisy rational framework. In these studies, exchanges do not charge trading fees, which precludes an analysis of the interactions between the market for data and the market for trading services, which are key for our implications.[12] In addition, liquidity traders in these models are noise traders, which precludes an analysis of exchange pricing policies on welfare.

Models of pure data sellers (e.g., Admati and Pfleiderer, 1986; Cespa, 2008; Huang et al., 2022; García and Sangiorgi, 2011; Cong et al., 2026) cannot account for either the interaction between exchanges' business lines or the endogenous determination of the quality of the data sold by the exchange. In our model, the pure data seller case arises under competition for order flow, and we recover a key insight from this literature: the data seller optimally limits the informativeness of its data.

[12] In Easley et al. (2016), the exchange derives revenues only from data sales. In Cespa and Foucault (2014), the exchange charges both a fee for participation in the market and a fee for data. The participation fee, however, is a fixed lump-sum fee paid by traders before receiving any information, in contrast to trading fees considered in our paper (and in reality).

Our paper also relates to the broader literature on the industrial organization of securities exchanges (e.g., Gehrig, 1993; Degryse et al., 2009; Foucault and Parlour, 2004; Colliard and Foucault, 2012; Chao et al., 2018). A strand of this literature studies the sale of trading speed through co-location services (e.g., Pagnotta and Philippon, 2018; Cespa and Vives, 2021; Budish et al., 2024).[13] These papers emphasize that competition for order flow constrains trading fees, leading exchanges to exercise market power primarily through the sale of trading speed. Our analysis complements this literature by studying the sale of proprietary data instead of speed. Unlike co-location, proprietary data mitigate adverse selection arising from private information rather than from differential speeds of reaction to public information. In addition, we analyze the joint pricing of trading and data services, with perfect competition for order flow as a special case of our framework.

Last, our paper is related to the vast theoretical literature on transparency in securities markets (e.g., Madhavan, 1995; Pagano and Roell, 1996; Bolton et al., 2016; Asriyan et al., 2017; Vairo and Dworczak, 2023; Cespa and Vives, 2026). Unlike this literature, which treats transparency as an exogenous disclosure choice, we study profit-maximizing exchanges' endogenous contribution to transparency through the sale of trading data.

3 The Model

We present the model setup in Section 3.1 and discuss our modeling choices in Section 3.2.

[13] Exchanges sell co-location services, which enable traders to access the platform more quickly. Co-location is distinct from proprietary data feeds, and market participants may purchase one service without the other. See Mackintosh and Normyle (2024), Figure 11.

3.1 Setup

The economy has four dates $t \in \{0, 1, 2, 3\}$. An exchange runs two markets that open sequentially: the "upstream" market at $t = 1$ and the "downstream" market at $t = 2$. Each market features three types of specialized risk-neutral traders: a market maker, liquidity traders and speculators. The assets traded in these markets are correlated, and, for simplicity, we assume that they have the same payoff $v \in \{-\sigma, +\sigma\}$ at $t = 3$.

Traders

Liquidity traders In each market $t \in \{1, 2\}$, a continuum of short-lived liquidity traders either wish to buy or sell the asset. Let $\tilde{m}_t^{L+}$ and $\tilde{m}_t^{L-}$ be the mass of liquidity buyers and liquidity sellers at date t, respectively. Both masses are uniformly and independently distributed on $[0, \bar{m}]$. A liquidity buyer (seller) derives a private benefit ε from holding (shorting) the asset, over and above the financial profit from the trade.[14]

Speculators Speculators also belong to one of two generations, exogenously assigned to trade either in the upstream market at $t = 1$ or the downstream market at $t = 2$. Speculators enjoy no private benefit for the asset, but they can acquire private information about its payoff. At $t = 0$, speculators of either market $t \in \{1, 2\}$ can pay a cost c to learn the asset payoff v. The mass of market-t speculators who acquire information is denoted μ_t.

All traders incur an idiosyncratic cost $k \sim \mathcal{U}[0, \bar{k}]$ from trading, which is independent across time. This cost may represent the return from some alternative private use of capital. For our analysis, it plays the technical role of ensuring that demand and supply are smooth.

Market Makers In each market $t \in \{1, 2\}$, a short-lived market maker (MM) sets a bid price, b_t, at which it executes all sell orders and an ask price, a_t, at which it executes all

[14] The private benefit can be interpreted as a reduced-form proxy for idiosyncratic hedging needs.

buy orders. After observing these quotes, market-t traders can submit a unit order to buy or to sell the asset. Let $\tilde{D}_t^I$ ($\tilde{S}_t^I$) be the resulting aggregate demand (supply) from informed speculators and $\tilde{D}_t^L$ ($\tilde{S}_t^L$) be the aggregate demand (supply) from liquidity traders. In each market, the short-lived market maker is a monopolist (see discussion below).

The Exchange As the operator of both markets, the exchange is the only long-lived agent. It generates revenues from matching traders and from the sale of data. For matching services, the exchange charges a per-unit trading fee f_t at date t, paid by traders.[15]

In addition, the exchange sells data about the upstream market order flow to the downstream market participants. We model data as a random posterior belief, p_2, that the asset value is high. To motivate this approach, observe that the aggregate demand for the asset in market $t-1$, $\tilde{D}_{t-1} = \tilde{D}_{t-1}^I + \tilde{D}_{t-1}^L$, which aggregates speculators' positions, is informationally equivalent to the posterior $\bar{p}_t \equiv \Pr(v = \sigma \mid \tilde{D}_{t-1})$. Obviously, $\bar{p}_1 = \frac{1}{2}$ since there is no market at date 0. We conjecture and verify (see Corollary 1) that $\bar{p}_2$ is distributed as follows:

$$\bar{p}_2|v = \begin{cases} \frac{1}{2} & (\text{prob. } 1-\bar{q}_2) \\ \frac{1}{2}\left(\frac{v}{\sigma}+1\right) & (\text{prob. } \bar{q}_2) \end{cases} \tag{1}$$

Equation (1) states that the aggregate demand in the upstream market either fully reveals the asset value (with probability $\bar{q}_2$) or is completely uninformative (with probability $1-\bar{q}_2$). We refer to the probability $\bar{q}_2$ as the informativeness of the (upstream market's) order flow.[16]

To capture exchanges' ability to choose the granularity of their data in practice, we

[15] The analysis is unchanged if trading fees are paid by the MM instead. Trading fees can be viewed as taxes and, as is well known, statutory incidence is irrelevant from an economic perspective.

[16] As shown below, this all-or-nothing structure results from the uniform distribution of aggregate liquidity-trader demand. See Dugast and Foucault (2018) or Foucault and Fresard (2014) for related analyses.

assume that the exchange can sell data p_2 that contains less information than the order flow signal $\bar{p}_2$. Exchange data are a garbled signal of the upstream market's order flow, with

$$p_2 = \begin{cases} \frac{1}{2} & (\text{prob. } 1 - q_2) \\ \frac{1}{2}\left(\frac{v}{\sigma} + 1\right) & (\text{prob. } q_2) \end{cases}, \quad q_2 \leq \bar{q}_2. \tag{2}$$

Similar to the raw aggregate demand, the exchange data either reveals the asset value or reveals nothing.[17] The constraint $q_2 \leq \bar{q}_2$ reflects the fact that the exchange can garble the order flow signal but data quality cannot exceed the informativeness of trades in the upstream market. The case $q_2 = 0$ captures the situation with no exchange data.[18]

To focus on the interesting cases for our analysis, we make the following assumptions.

Assumption 1. $\varepsilon < \sigma$

Assumption 2. $K \equiv \sqrt{2\bar{k}c} \geq \frac{\sigma - \varepsilon + \sqrt{(\sigma-\varepsilon)(2\sigma-\varepsilon)}}{2}$

Assumption 3. $\bar{k} \geq 2\sigma$

Assumption 1 avoids the situation in which speculation never arises. Assumption 2 is a technical assumption imposed to guarantee that the MM's optimal quoting strategy is in pure strategies.[19] Similarly, Assumption 3 simplifies the analysis of the trading equilibrium.

[17] Specification (2) disciplines how the exchange can add noise to the signal by ruling out "false-positives". The exchange can omit information, that is, report $p_2 = \frac{1}{2}$ when $\bar{p}_2 \in \{0, 1\}$, but it cannot fabricate information, for instance, report $p_2 = 1$ when $\bar{p}_2 = 0$. This specification is both convenient theoretically and appealing empirically. In practice, exchanges do sell data feeds with different levels of granularity, but they could be liable for misreporting information.

[18] Instead of selling data about demand D_1^* only, the exchange could sell data about demand and supply (D_1^*, S_1^*). The joint realization of demand and supply has similar qualitative properties as D_1^*: It either reveals the asset value or nothing but it is more informative. Hence, we assume that the quality of exchange data is bounded by the quality of the signal contained in D_1^* rather than that conveyed by (D_1^*, S_1^*). This makes the analysis more tractable without changing the economics of the exchange problem.

[19] Without this assumption, under some exchange design choices, the MM would randomize between quoting and not quoting in equilibrium. While this feature would not affect the qualitative properties of the model, it significantly complicates the analysis of the exchange problem.

Timing. Figure 1 depicts the timing of actions. At date 0, there are two stages. First, the exchange chooses its design, that is, its trading fees $\boldsymbol{f} = (f_1, f_2)$, its data quality q_2 and its data fee, Φ. Then, speculators in each market $t \in \{1, 2\}$ decide whether to acquire private information, and the downstream market participants privately decide whether to buy the exchange data or not. Next, the two trading periods take place, as described previously. After upstream trading, the exchange releases data p_2, derived from the upstream order flow $\bar{p}_2$, to downstream market participants who bought its data.

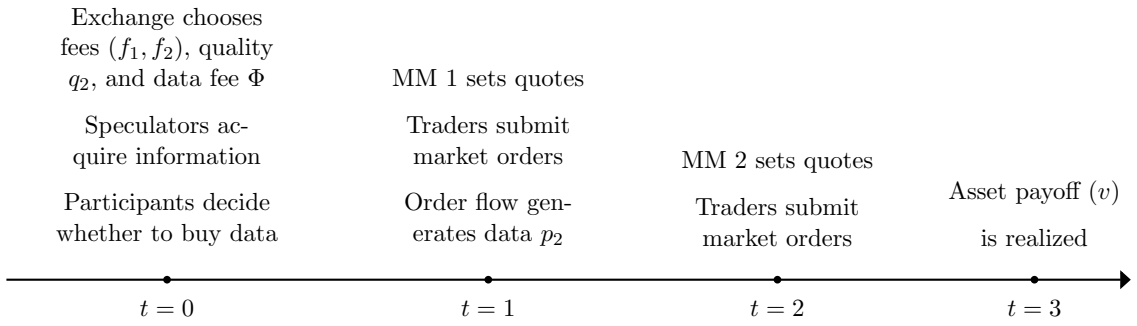


Figure 1. Timing

3.2 Discussion of Modeling Choices

In this section, we discuss our main modeling choices and their robustness.

Market power of exchanges. We assume that the exchange has market power over both matching services and transaction data. For robustness, Section 7.1 proxies for competition in matching services, by setting the exchange's trading fees to zero. Pricing power over data, however, is central to our model. Such pricing power would arise if different venues capture different order flows, so that their transaction data are not perfect substitutes. In practice, exchanges charge significant fees for proprietary data feeds.[20]

[20]Some market participants such as brokers or other exchanges have a de facto need to observe data from

Market power of market makers. We model market makers as local monopolists to simplify the characterization of optimal quotes and of their decision to purchase data from the exchange. What matters for our analysis is only that market makers retain some market power, consistent with the profitability of market making in practice.[21]

Short-lived participants. Traders and market makers are short-lived in the model. This assumption rules out across-market strategic considerations by agents other than the exchange and lets us cleanly decompose the key economic forces.

Pre-trade vs. post-trade transparency. MM quotes are observable at no cost in our model, so the market is fully transparent before trade. The data market therefore concerns transaction data, or post-trade transparency. In practice, exchanges may sell both pre-trade and post-trade information. What matters is that the exchange controls access to payoff-relevant information about the downstream asset generated by upstream trading activity.

Trading Game. Finally, the extensive form of the trading game follows Admati and Pfleiderer (1989), Zhu (2014), and Biais et al. (2015): the market maker posts quotes at which it simultaneously processes all orders from a continuum of traders. Alternatively, there could be one trade per period with a random trader type, so that quotes would be contingent on the order flow, as in Glosten and Milgrom (1985). We use the continuum specification because it yields a simple distribution for the signal generated by order flow, given by equation (1). Apart from this simplification, we expect the same economic forces to be present with sequential trading.

several venues, which gives each venue market power over the information it generates.

[21] If market makers compete à la Bertrand for order flow, there is no equilibrium in pure strategies in which both MMs buy data from the exchange at a positive price. Hence, since the exchange chooses the data fee, the equilibrium must feature randomization in MMs' decision to acquire information.

4 Equilibrium

In this section, we derive the continuation equilibrium induced by a fixed exchange design. The goal is to express the key market outcomes—quotes, order-flow informativeness, the value of data and trading volume—as functions of the exchange's policy variables. We proceed in two steps. First, in Section 4.1, we solve for traders' optimal order submissions and equilibrium quotes in each market . Then, in Section 4.2, we solve for speculators' optimal decision to acquire information at date 0 and market participants' decision to buy exchange data. These decisions to acquire information (private signal or exchange data) are not observed but are correctly anticipated by all market participants in equilibrium.

In the equilibrium we analyze, MM2 is the only buyer of exchange data in the downstream market. This result follows from two steps. First, in any equilibrium, either MM2 buys the data or no agent buys it. MM2 is large relative to atomistic speculators and liquidity traders in the downstream market. Hence, any data fee Φ such that the latter buy data is a negligible cost to MM2. Second, the result hinges on a mild equilibrium selection criterion for the downstream market. Since MM2 buys the data, that trading stage is a signaling game: MM2 moves first by setting quotes, which can signal the information it obtained from exchange data to uninformed traders.[22] We select the unique separating equilibrium of this game. Since MM2's quotes reveal its information, liquidity traders and speculators do not buy exchange data at date 0 as they observe it for free in equilibrium. Appendix B characterizes this equilibrium and shows that it corresponds to the equilibrium in which exchange data is public information.[23] Hence, we can solve for the downstream market equilibrium as if the

[22]The case in which market makers are informed has received little attention in the market microstructure literature. One exception is Calcagno and Lovo (2006).

[23]These nonstandard results arise because the MM makes a two-sided market with a bid and an ask. As we show in Appendix B, this exposes deviating MMs to predatory behavior from speculators, which helps to guarantee existence. Uniqueness rests on the result that a MM with any type $p_2 \in \{0, 1/2, 1\}$ can attain the

exchange data signal $p_2 \in \{0, 1/2, 1\}$ were observable to all its participants.

4.1 Trading Equilibrium

In this section, we solve for the equilibrium quotes of the MM at date t, taking as given the exchange design and the speculator mass, μ_t, in that market. We then derive the order flow informativeness, the value of data and the equilibrium trading volume.

To derive the optimal MM quotes, we characterize traders' order submission choices. Consider first liquidity traders. In each market, in equilibrium, these traders have the same information as the market maker (including in the downstream market, for the reasons given above). Thus, they assign the same probability p_t as MMt to the value of the asset being high ($v = \sigma$) where $p_1 = \frac{1}{2}$ and $p_2 \in \{0, \frac{1}{2}, 1\}$. We denote by $v(p_t)$ the expected payoff of the asset given the market maker and liquidity traders' posterior in market $t \in \{1, 2\}$:

$$v(p_t) \equiv 2\sigma p_t - \sigma, \tag{3}$$

A liquidity buyer in market t buys the asset if and only if his valuation exceeds the total trading cost, which comprises his private trading cost, k, the ask price, a_t, and the exchange fee, f_t. That is, he buys if and only if $v(p_t) + \varepsilon \geq a_t + f_t + k$. Symmetrically, a liquidity seller sells the asset if and only if $b_t - f_t - k \geq v(p_t) - \varepsilon$. We deduce that the aggregate demand and supply from liquidity traders at date t are:

$$\tilde{D}_t^L(a_t, p_t) = \widetilde{m}_t^{L+} \Pr[k \leq v(p_t) + \varepsilon - a_t - f_t] = \widetilde{m}_t^{L+} G\big(v(p_t) + \varepsilon - a_t - f_t\big), \tag{4}$$

$$\tilde{S}_t^L(b_t, p_t) = \widetilde{m}_t^{L-} G\big(\varepsilon - v(p_t) + b_t - f_t\big), \tag{5}$$

payoff of the game with public exchange data by choosing the optimal quotes from that game, independently of how liquidity traders map (possibly out-of-equilibrium) quotes to beliefs.

where G is the CDF of the private trading cost $k \sim \mathcal{U}[0, \bar{k}]$.

Now, consider speculators. In market t, speculators, who are perfectly informed, buy the asset when $v = \sigma$ only if $\sigma > a_t + f_t + k$ and sell the asset when $v = -\sigma$ only if $-\sigma < b_t - f_t - k$. Hence, at date t, the demand and supply from the mass μ_t of speculators are given by:

$$D_t^I(a_t, v) = \begin{cases} 0 & \text{if} \quad v = -\sigma \\ \mu_t G(\sigma - a_t - f_t) & \text{if} \quad v = \sigma \end{cases}, \qquad S_t^I(b_t, v) = \begin{cases} \mu_t G(\sigma + b_t - f_t) & \text{if} \quad v = -\sigma \\ 0 & \text{if} \quad v = \sigma \end{cases}. \tag{6}$$

The MM for market $t \in \{1, 2\}$ sets its ask, a_t, and its bid, b_t, to maximize its profit:

$$\max_{(a_t, b_t)} \Pi_t^{MM}(a_t, b_t; p_t) \equiv \mathbb{E}\Big[\tilde{D}_t(a_t - \tilde{v}) + \tilde{S}_t(\tilde{v} - b_t) \,|\, p_t \,\Big], \tag{7}$$

anticipating the effect of the quotes on demand $\tilde{D}_t = \tilde{D}_t^L + \tilde{D}_t^I$ and supply $\tilde{S}_t = \tilde{S}_t^L + \tilde{S}_t^I$.

Proposition 1. ***MM quotes.*** *The MM's optimal quotes are characterized by the half-spread*

$$s_t^*(p_t) \equiv v(p_t) - b_t^*(p_t) = a_t^*(p_t) - v(p_t) = \begin{cases} \frac{1}{2}(\varepsilon - f_t) & \textit{if} \quad p_t \in \{0, 1\} \\ \frac{1}{2}(\varepsilon - f_t) + \frac{1}{2}\frac{\mu_t}{\bar{m} + \mu_t}(2\sigma - \varepsilon) & \textit{if} \quad p_t = \frac{1}{2} \end{cases}, \tag{8}$$

provided that the trading fee satisfies $f_t < \varepsilon$ and the MM's profit at $p = \frac{1}{2}$ is positive.

For a posterior p_t, $v(p_t)$ would be the MM quote if it had no market power and did not face adverse selection. Its actual quotes are characterized by a symmetric half-spread $s_t^*(p_t)$, relative to $v(p_t)$. This spread has two components. When $p_t = 1$ or $p_t = 0$ (which can happen in the downstream market), speculators' demand or supply is zero because they have no information advantage. The MM's spread is then just determined by market power over liquidity traders, and it increases with liquidity traders' willingness to pay ε. The market

maker effectively rebates half of the exchange trading fee f_t to traders to limit the negative impact of an increase in f_t on liquidity traders' demand and supply. When $p_t = \frac{1}{2}$ (which is always the case in the upstream market), the MM's spread features an additional component, $\frac{1}{2}\frac{\mu_t}{\bar{m}+\mu_t}(2\sigma - \varepsilon)$, due to adverse selection, which increases with the mass of speculators μ_t.[24]

The equilibrium quotes determine the composition of the upstream order flow. We use Proposition 1 to confirm our conjecture that the posterior probability that the asset value is high, $\bar{p}_2(\tilde{D}_1^*)$ given the aggregate upstream demand, D_1^*, is distributed as in equation (1).

Corollary 1. ***Order Flow Signal.*** *Given exchange fee f_1 and a mass μ_1 of speculators in the upstream market, the equilibrium order flow signal $\bar{p}_2$ is given by (1), that is, it either perfectly reveals the asset value or is completely uninformative. The probability that it reveals the asset value (order flow informativeness) is*

$$\bar{q}_2(\mu_1, f_1) = \min\left\{\frac{\mu_1}{\bar{m}}\frac{\sigma - s_1^*(1/2) - f_1}{\varepsilon - s_1^*(1/2) - f_1}, 1\right\}. \tag{9}$$

With uniformly distributed liquidity demand, the order flow is fully revealing in the tail realizations that cannot be generated by liquidity traders alone. Otherwise, it does not reveal anything, which is why $\bar{p}_2$ is given by (1). The size of these tails determines the order-flow informativeness $\bar{q}_2$, which is proportional to informed demand relative to liquidity demand. In particular, informativeness increases with the mass of upstream speculators, μ_1. This order flow informativeness, $\bar{q}_2$ is an upper bound for the quality q_2 of exchange data.

Next, consider the value of exchange data for the downstream market maker. This value,

[24] The characterization of equilibrium quotes when $p_t = \frac{1}{2}$ is valid only if the conditions stated in Proposition 1 hold. Condition $f_t < \varepsilon$ is obviously satisfied in equilibrium because the exchange would never choose $f_t > \varepsilon$, since there would be no trading. If the second condition (positive MM profit) fails, the market breaks down due to adverse selection. While this outcome could arise if μ_t is large, we show in the proof of Lemma 1 below that it is ruled out in equilibrium under our maintained assumptions when μ_t is endogenized.

denoted $VD(\mu_2, f_2, q_2)$, corresponds to the incremental expected profit for MM2 from buying exchange data, gross of the data fee. The following result follows from Proposition 1.

Corollary 2. *Given data quality q_2, downstream trading fee f_2, and speculator mass μ_2 in the downstream market, the value of exchange data for MM2 is*

$$VD(\mu_2, f_2, q_2) = \frac{\mu_2 q_2}{4\bar{k}} \left[\frac{\bar{m}}{\bar{m} + \mu_2} (2\sigma - \varepsilon)^2 - f_2^2 \right]. \tag{10}$$

Intuitively, the value of the exchange data increases with its quality, q_2, because more precise data allow MM2 to better mitigate adverse selection vis-à-vis speculators in the downstream market. The value of data increases with the mass of speculators, μ_2, as this increases MM2's exposure to adverse selection.[25] Indeed, as shown in the proof of Corollary 2, the scaled data value, VD/q_2, measures the increase in MM profit from being perfectly informed, thus facing no adverse selection, relative to being uninformed. As the uninformed profit itself decreases with the mass of speculators, the value of data increases with μ_2.

Corollary 3. ***Equilibrium Trading Volume.*** *At date $t \in \{1, 2\}$, in equilibrium, the expected trading volume and informed volume share are respectively*

$$EV_t(\mu_t, f_t, q_t) \equiv \mathbb{E}(\tilde{D}_t + \tilde{S}_t) = \frac{\bar{m}}{2\bar{k}} (\varepsilon - f_t) - \frac{\mu_t}{2\bar{k}} (1 - q_t) f_t. \tag{11}$$

$$IVS_t(\mu_t, f_t, q_t) = \frac{\mathbb{E}(\tilde{D}_t^I + \tilde{S}_t^I)}{EV_t(\mu_t, f_t, q_t)} \tag{12}$$

[25] Formally, taking the partial derivative of VD with respect to μ_2, we obtain

$$\frac{\partial VD}{\partial \mu_2} = \frac{q_2}{4\bar{k}} \left[\frac{\bar{m}^2}{(\bar{m} + \mu_2)^2} (2\sigma - \varepsilon)^2 - f_2^2 \right],$$

which is positive when $s_2^*(1/2) < \sigma - f_2$. This condition is necessary for equilibrium speculation, and we will verify that it holds under our maintained assumptions.

The expected trading volume in market $t \in \{1, 2\}$ decreases with the exchange's trading fee f_t, which is akin to a trading tax. Expected trading volume also decreases with the mass of speculators in the market, μ_t. As μ_t increases, the MM charges a larger spread (Proposition 1), which more than offsets the positive effect of speculation on trading volume. This explains why expected trading volume in the downstream market increases with data quality, q_2: speculators trade only when quotes do not reveal the asset value (probability $1-q_2$). The informed volume share, defined in (12), is the fraction of order flow generated by speculators. It is a model counterpart of the PIN measure in empirical market microstructure (Easley et al., 1996).

4.2 Speculative Entry and the Data-Quality Frontier

Now, we endogenize the mass of speculators in each market, which was taken as given in Section 4.1. This step completes the characterization of the equilibrium and allows us to express all relevant equilibrium objects as functions of the exchange's policy variables.

To determine the equilibrium mass of speculators, consider their decision to acquire information about the asset value at date 0, at cost c. Speculators in market t profit from their private information only when exchange data does not reveal v, which happens with probability $(1-q_t)$. In this case, speculators trade if their information advantage, σ, exceeds their total trading cost, given by the sum of the MM spread, $s^*(\frac{1}{2})$, given by (8), the exchange fee, f_t, and their private trading cost, $k \sim \mathcal{U}[0, \bar{k}]$. The net profit from information acquisition for date-t speculators is thus

$$(1-q_t)\int_0^{\sigma - s^*(1/2) - f_t} \left[\sigma - s_t^*(1/2) - f_t - k\right] \frac{\mathrm{d}k}{\bar{k}} - c \tag{13}$$

In equilibrium, two cases are possible. Either speculation is strictly unprofitable and $\mu_t = 0$, or the mass of speculators adjusts such that the net profit from private information is 0. We recall that $K \equiv \sqrt{2\bar{k}c}$.

Lemma 1. *Given trading fee f_t and data quality q_t, the mass of market-t speculators is*

$$\mu_t^*(f_t, q_t) = \max\left\{0, \bar{m}\left(\frac{2\sigma - \varepsilon}{2K(1-q_t)^{-1/2} + f_t} - 1\right)\right\} \tag{14}$$

The equilibrium mass of speculators increases with the volatility of the asset, σ, the amount of liquidity trading, $\bar{m}$, and decreases with liquidity traders' willingness to pay, ε, and the exchange trading fee, f_t. Equation (14) further highlights a key force of our model: in the downstream market, speculation decreases with the quality of exchange data, q_2. As the quality of data purchased by MM2 increases, its information gap with perfectly informed speculators narrows. This reduces speculators' willingness to pay for a private signal, and ultimately the mass of speculators in equilibrium, μ_2^*, via the entry condition.

Lemma 1 closes the preliminary steps of the analysis. Indeed, the equilibrium objects derived in Section 4.1 can now be expressed as functions of the exchange policies. In particular, we characterize the data-quality frontier for the exchange by expressing the informativeness of the upstream order flow as a function of the exchange's upstream trading fee alone.

Corollary 4. ***Order Flow Informativeness.*** *Given exchange fee f_1, the equilibrium informativeness of the order flow in the upstream market is*

$$\bar{q}_2(f_1) \equiv \bar{q}_2(\mu_1^*(f_1, 0), f_1) = \begin{cases} \min\left\{\frac{K}{K-\sigma+\varepsilon}\left(\frac{2\sigma-\varepsilon}{2K+f_1} - 1\right), 1\right\} & \text{if} \quad f_1 < 2(\sigma - K) - \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The informativeness of the upstream order flow increases with σ, which measures profit

opportunities for speculators, and it decreases with ε—the willingness to pay of liquidity traders. Corollary 4 further shows that informativeness decreases with the exchange fee, f_1. This nonobvious result obtains because the mass of speculators in the upstream market, μ_1, is endogenous. Holding μ_1 fixed, both speculative and liquidity volume decrease with f_1, with an ambiguous impact on order flow informativeness. With endogenous speculation, liquidity volume does not change with the trading fee. To see this, consider the liquidity trader who is indifferent between trading and abstaining, who has private trading cost

$$\hat{k}_1^L \equiv \varepsilon - s^*(1/2) - f_1 \overset{\mu_1^*>0}{=} \sigma - K, \tag{16}$$

where the equality follows from the zero-profit condition for speculators when $\mu_1^* > 0$. Economically, when the exchange increases its trading fee f_1, adverse selection decreases, which induces the MM to lower its spread $s^*(1/2)$. In equilibrium, these two conflicting effects on total trading costs exactly offset each other. Hence, a higher trading fee reduces speculative volume but leaves liquidity volume unchanged, which reduces order flow informativeness.

To conclude, the trading fee in the upstream market controls the technological frontier for data quality. Lowering this fee increases the informed share of the order flow, which relaxes the constraint $q_2 \leq \bar{q}_2(f_1)$ on data quality.

5 Selling Access and Data

In this section, we study how data sales affect the exchange's business model and market quality. Section 5.1 shows that data quality affects both trading-fee revenue and data revenue. Section 5.2 instead characterizes the welfare impact of the data market. Sections 5.1 and 5.2 are useful in their own right and they also provide the building blocks for the analysis

of optimal data policies and regulation in Section 6.

5.1 Exchange Profit Sources

The exchange derives revenues from trading fees, denoted Π^E_{trad}, and from selling data at price Φ. The expected revenues from trading fees are given by

$$\Pi^E_{trad}(f_1, f_2, q_2) = EV(\mu_1^*, f_1, 0)\, f_1 + EV(\mu_2^*, f_2, q_2)\, f_2 \tag{17}$$

with $EV_t(\mu_t^*, f_t, q_t)$ the expected trading volume in market t, given by (11). It depends on the data quality, q_t (with $q_1 = 0$), the trading fee, f_t and the equilibrium mass of speculators μ_t^*. Next, under laissez-faire, the monopolistic exchange sets the data fee, Φ, equal to the value of the data for MM2, that is, $\Phi^*(f_1, f_2, q_2) = VD(\mu_2^*, f_2, q_2)$, given by (10).

Overall, the exchange chooses policies to maximize profits from trading and data services:

$$(f_1^*, f_2^*, q_2^*) = \underset{(f_1, f_2, q_2)}{\arg\max} \;\; \Pi^E(f_1, f_2, q_2) \equiv \Pi^E_{trad}(f_1, f_2, q_2) + VD(\mu_2^*, f_2, q_2) \tag{18}$$

$$\text{s.t.} \qquad q_2 \in [0, \bar{q}_2(f_1)], \tag{19}$$

where the equilibrium mass of speculators, μ_t^*, is expressed as a function of the exchange policies in (14). The order flow informativeness, $\bar{q}_2(f_1)$, is given by (15). We recall that constraint (19) states that data quality cannot exceed the order flow informativeness.

Before solving this problem in Section 6, we study how each policy instrument affects the exchange's two profit sources.

Lemma 2. *The profit from trading fees, Π^E_{trad}, is single-peaked in trading fees (f_1, f_2) and strictly increasing with data quality q_2. The optimal data fee, Φ^*, is single-peaked in data*

quality q_2 and strictly (weakly) decreasing with trading fee f_2 (f_1).

The exchange's profits from trading fees, Π^E_{trad}, are hump-shaped in these fees because expected trading volume in market t, EV_t, is decreasing with the trading fee f_t (Corollary 3). This profit component from trading fees also depends on the exchange's data policy, because Π^E_{trad} increases with data quality q_2. Higher data quality sold to MM2 discourages speculation in the downstream market (Lemma 1), which increases trading volume (Corollary 3) and thus revenues from trading fees. Hence, disseminating high-quality data supports the exchange's traditional business of matching traders, even if it cannot monetize data. This effect will be at the center of the analysis of a cap on data fees in Section 6.2.

When the exchange monetizes data, its profit has a new component, Φ^*, equal to the MM's value for data in the downstream market. The value of data does not increase monotonically with quality because it reflects MM2's exposure to adverse selection. Higher data quality reduces this exposure and can erode the value of data because it reduces the mass of speculators in the downstream market (Lemma 1). Equation (14) shows that above a certain quality, data has no value, as speculators stop acquiring private information. Hence, monetization does not necessarily induce transparency: if the order flow is too informative (high $\bar{q}_2$), the exchange could prefer to degrade the data quality, that is, set $q_2 \leq \bar{q}_2$.[26]

Finally, Lemma 2 shows that data revenues are decreasing with the trading fees (f_1, f_2). The downstream fee, f_2, reduces MM2's value for data via two channels. First, it directly lowers MM2's profit and its incentives to acquire information. Next, it reduces speculation in the downstream market and thus MM2's exposure to adverse selection. This reduces MM2's willingness to buy data to reduce its information gap vis-à-vis speculators (see Corollary 2).

[26]This observation relates to a result by Admati and Pfleiderer (1986) who show that a data seller optimally adds noise to a trading signal. In particular, we share the observation that the value of data increases with speculators' information advantage, and that excessive data quality erodes this information advantage. In our model, however, a third party, MM2, buys the data, not speculators themselves.

Finally, the upstream trading fee f_1 weakly lowers data revenues via its negative impact on the order flow informativeness, $\bar{q}_2(f_1)$—the upper bound for data quality (Corollary 4).

5.2 Market Efficiency

The exchange maximizes profit, but regulation is concerned with market efficiency. We consider two metrics for efficiency: information revelation and overall social welfare.

Data Quality as Information Revelation. Regulators often view the production of information about fundamentals as a key role of financial markets. Although information revelation matters in the model only through welfare, it is useful to track separately because public information about fundamentals can affect allocative decisions outside the model. Our model-based measure of information revelation about the asset value is data quality, q_2, which measures the probability that quotes publicly reveal the asset value before date 3.[27]

Utilitarian Welfare. Next, consider (utilitarian) welfare, which captures all aspects of market efficiency. Because agents are risk neutral, transfers cancel under a utilitarian objective. Welfare is realized surplus from liquidity trading, net of traders' private trading costs and speculators' information-acquisition costs. To express welfare, it is useful to introduce the marginal traders in the liquidity and speculator groups. Extending equation (16), these

[27] An alternative definition could be total information revealed by the order flow. However, we emphasize that all information produced need not be publicly and freely observable, at least in real time. Instead, data quality q_2 corresponds to the informativeness of public information about the asset value.

marginal traders have private trading cost

$$\hat{k}_t^{\theta}(p_t) \equiv \begin{cases} \varepsilon - s_t^*(p_t) - f_t & \text{if} \quad \theta = L \\ \sigma - s_t^*(p_t) - f_t & \text{if} \quad \theta = I \end{cases}, \qquad p_t \in \{0, 1/2, 1\}, \quad t \in \{1, 2\} \tag{20}$$

where p_t is the realization of the data sold by the exchange. Social welfare in market t is

$$W_t = \frac{\bar{m}}{\bar{k}} \mathbb{E}_{p_t} \left[\int_0^{\hat{k}_t^L(p_t)} (\varepsilon - k) dk \right] - (1 - q_t) \frac{\mu_t}{\bar{k}} \int_0^{\hat{k}_t^I(p_t)} k dk - \mu_t c, \tag{21}$$

with the expectation taken over realization of the MM data, p_t, which is degenerate in the upstream market ($p_1 = 1/2$) and distributed according to (2) in the downstream market. The first term of (21) is liquidity-trader surplus net of trading costs. The second term measures trading costs incurred by speculators: speculative trading creates no social benefit and occurs only when MM data do not reveal the asset value. The final term measures aggregate costs from private information acquisition by speculators.

As when we studied the impact of the trading fees and data quality on the exchange's profit in Section 5.1, we characterize the effect of these policies on social welfare.

Lemma 3. *Social welfare W_t, given by (21), increases with data quality q_t. The date-t trading fee, f_t has an ambiguous impact on W_t. In the upstream market, W_1 strictly increases with f_1, provided that there is speculation in equilibrium, that is, when $\mu_1^*(f_1, 0) > 0$.*

In equilibrium, data purchased by the MM is revealed via its quotes. Thus, data quality levels the playing field between uninformed liquidity traders and speculators. Welfare increases via two related channels. First, for a given mass of speculators, μ_t, the MM is more likely to be informed as data quality increases. The informed MM spread is lower than when it is uninformed ($s^*(0) < s^*(1/2)$), which crowds in liquidity traders, and speculators do

not trade in this case, which avoids wasting resources to pay private trading costs. Second, higher data quality reduces speculation, which reduces resources spent on socially wasteful private information.

Next, Lemma 3 shows that trading fees have an ambiguous effect on welfare. On the one hand, exchange trading fees increase total participation costs for liquidity traders, which reduce welfare. On the other hand, higher trading fees deter speculators from acquiring socially wasteful private information. Welfare increases with trading fee f_1 because only the latter effect is present in the upstream market. This result follows again from the observation that the marginal upstream liquidity trader, defined in (20), does not change with f_1. As explained in the discussion of Corollary 4, the decline in MM spread due to reduced speculators' entry exactly offsets the impact of the exchange fee. In the downstream market, however, both channels are active, which explains why the result is ambiguous. If data quality q_2 is so high that it eliminates downstream speculation ($\mu_2^* = 0$), the total trading cost is $s_2^* + f_2 = \frac{1}{2}(\varepsilon + f_2)$ and we obtain the classic effect that higher trading fees reduce participation by liquidity traders.

6 Optimal Data Policy and Regulation

This section studies the problem of an exchange that jointly chooses its trading fees and its data policy. The solution to Problem (18)-(19) can be characterized analytically, albeit not in closed form, save for specific cases. Instead, we study exchanges' data policies by comparing the laissez-faire outcome with two benchmarks. In Section 6.1, the *no-data benchmark* shuts down data dissemination completely. In Section 6.2, the *no-monetization benchmark* allows dissemination but prevents the exchange from monetizing the data.

6.1 Data Sales Lower Trading Fees

In this section, we study the overall impact of the exchange's data business by comparing the laissez-faire outcome to the no-data benchmark. This benchmark allows us to capture technological or regulatory impediments to the dissemination of data. The no-data benchmark in our model corresponds to data quality $q_2 = 0$. In what follows, we use subscript $|_{q_2=0}$ to index equilibrium variables in the no-data benchmark.

Proposition 2. *Relative to the no-data benchmark ($q_2 = 0$), exchange data lowers exchange trading fees f_t^* and total (expected) trading costs, $\mathbb{E}[s_t^*(p_t)] + f_t^*$ in both markets $t \in \{1, 2\}$. Yet, the upstream (downstream) MM spread s_1^* ($\mathbb{E}[s_2^*(p_2)]$) increases (may increase).*

The key result from Proposition 2 is that the exchange lowers its trading fees when it sells data, relative to the no-data benchmark. As shown in the proof, this also implies that total trading costs—exchange fees plus MM spreads—decrease. It is worth pointing out that MM spreads move in the opposite direction. In the upstream market, any reduction in the trading fee f_1 is offset one for one by an increase in the MM spread.[28] Hence, empirical measures based on spreads alone can misidentify the effect of data sales on trading costs.

Now, we explain why the exchange lowers trading fees when it sells data. First, lower upstream trading fees increase the quality of data by making the order flow more informative. By definition, the no-data trading fee $f_1^*|_{q_2=0}$ maximizes upstream profits from trading fees. With data sales, the exchange would only alter this fee to relax the technological constraint $q_2 \leq \bar{q}(f_1)$ if it wants to set data quality q_2^* above the baseline level $\bar{q}(f_1^*|_{q_2=0})$. Instead, if the target data quality satisfies $q_2^* < \bar{q}(f_1^*|_{q_2=0})$, the exchange sets $f_1^* = f_1^*|_{q_2=0}$ and garbles the

[28] This observation implies that the total upstream trading cost $s_1^*(1/2) + f_1$ does not change, relative to the no-data benchmark. Hence, the statement in Proposition 2 about total trading costs is strong only for date 2.

order flow, which avoids reducing upstream profits. The result then follows from Corollary 4, which shows that pushing the data frontier requires decreasing the upstream fee, f_1.

The exchange also sets the downstream trading fee f_2^* below the benchmark level $f_2^*|_{q_2=0}$, this time because the value of data to MM2 decreases with f_2 (Lemma 2). For a given mass of downstream speculators μ_2, a lower fee has a direct positive effect on MM2's willingness to pay. In addition, a lower fee also stimulates speculation, which increases MM2's exposure to adverse selection and ultimately its willingness to buy exchange data. These two forces induce the exchange to lower its trading fee f_2, relative to the benchmark level $f_2^*|_{q_2=0}$.[29]

These results suggest that exchange data increases trading volume, which we verify next.

Corollary 5. *Trading volume increases in both markets relative to the no-data benchmark. The informed volume share in the upstream (downstream) market increases (may decrease).*

When the exchange sells data, trading volume weakly increases in the upstream market, as the exchange may lower its trading fee to stimulate speculative trading (Proposition 2). As this change does not affect participation by liquidity traders (see Corollary 4), the informed share of the order flow also increases. In the downstream market, total volume increases relative to the no-data benchmark because data increases trading volume (Corollary 3) and the exchange lowers its trading fee (Proposition 2). Finally, the effect of the data market on the informed volume share of the downstream market is ambiguous. Data sales crowd out downstream speculators, as we explained above. However, when it sells data, the exchange lowers its downstream fee to stimulate speculation and increase the value of its data for MM2 (Proposition 2). The net effect on speculative activity is ambiguous, as confirmed by the

[29]For completeness, there is a conflicting effect that incentivizes the exchange to increase f_2 when it sells data. Indeed, when $q_2 > 0$, Corollary 3 shows that expected volume is less sensitive to the trading fee because speculators enter less and trade less; formally, $\frac{\partial^2 EV}{\partial f_2 \partial q_2} > 0$. However, we show in the proof of Proposition 2 that the negative effect of trading fees on revenues from data sales dominates.

numerical analysis reported in Figure 2.

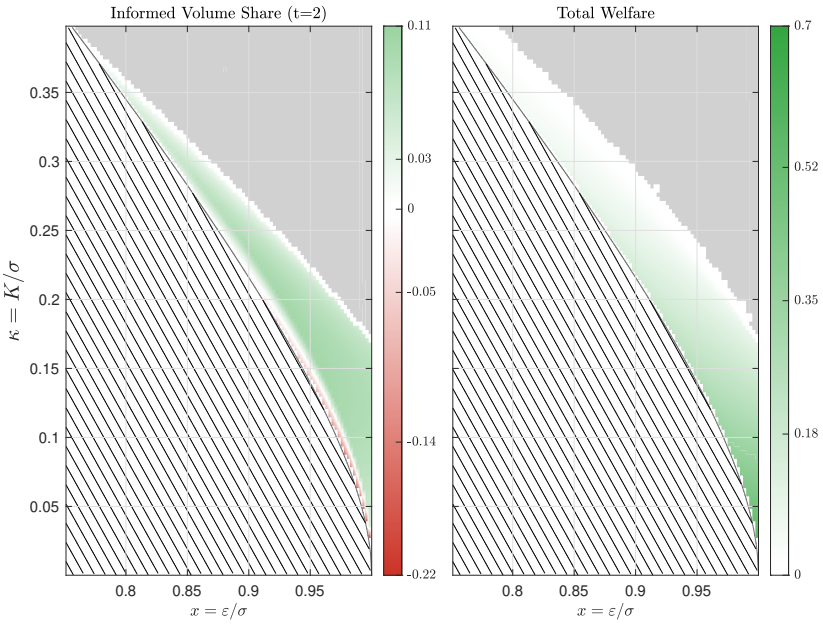


Figure 2. Laissez-faire vs. no-data benchmark
In all panels, the x-axis (y-axis) is the normalized liquidity traders' private benefit ε/σ (private trading cost parameter $\kappa = K/\sigma$). Each panel plots the difference between the laissez-faire variable and the no-data benchmark variable. Green (red) values indicate that the laissez-faire variable is higher (lower) than the benchmark variable. The total welfare panel reports $\omega_1 + \omega_2$, where $\omega_t = \frac{2\bar{k}}{\bar{m}\sigma^2} W_t$ is market-t normalized welfare. For parameters in the gray area, there is no speculation in both cases. Parameters in the striped area violate Assumption 2.

Finally, we discuss the impact of exchange data on our two metrics for market efficiency. Since data quality is the measure of information revelation, the latter improves by construction when exchange data becomes available, as data quality increases from 0 to $q_2^* > 0$. On the other hand, the welfare impact of data sales is a priori more nuanced. First, Lemma 3 implies that welfare in the upstream market decreases, because the drop in trading fee f_1 only attracts more speculators. Next, the welfare impact of exchange data sales in the downstream market is ambiguous. Data quality increases from 0 to q_2^*, which increases welfare for a given trading fee (Lemma 3). However, the exchange lowers its downstream fee to stimu-

late speculation, which has an ambiguous effect on welfare (again, Lemma 3). The numerical analysis in Figure 2 reveals that total welfare aggregated across the two markets increases relative to the no-data benchmark. This implies that the positive effect of data sales on the downstream market's welfare outweighs the negative impact on upstream welfare.

6.2 Data Monetization and Data Quality

Regulators view the creation of low-cost data feeds as key to market efficiency. Yet, exchanges often contend that such interventions weaken their ability to provide high-quality data. To analyze this tension, we compare the laissez-faire outcome to a *no-monetization* benchmark. In this benchmark, the exchange can disseminate data, but it cannot charge data buyers, that is, $\Phi = 0$. This constraint can proxy for an extreme regulatory cap on data fees.

First, contrary to the argument often advanced by exchanges, our model predicts that an exchange would still disseminate high-quality data even without monetization.

Proposition 3. *When the exchange can disseminate data but cannot monetize them, it sets the maximum data quality* $q_2^*|_{\Phi=0} = \bar{q}(f_1^*|_{\Phi=0})$ *and sets the upstream trading fee below the no-data level to increase data quality, that is,* $f_1^*|_{\Phi=0} < f_1^*|_{q_2=0}$.

Proposition 3 follows from the result in Lemma 2 that high-quality data increases the exchange's profit from trading fees. Remember that higher data quality crowds out speculation, which increases trading volume (Corollary 3). Hence, without data monetization, the exchange's objective strictly increases with data quality. Thus, it sets data quality equal to the informativeness of the order flow: $q_2^*|_{\Phi=0} = \bar{q}(f_1^*|_{\Phi=0})$. Furthermore, the exchange strictly lowers its upstream fee below the no-data benchmark to relax constraint (19) on data quality, that is, $f_1^*|_{\Phi=0} < f_1^*|_{q_2=0}$, which reduces its upstream profit from trading fees.

Hence, even absent monetization, the exchange disseminates high-quality data, not only as a free byproduct of operations, but also to raise its profits from downstream trading fees.

Proposition 3 should be contrasted with Lemma 2, which shows that the exchange may instead have an incentive to degrade data quality when it can monetize it. Indeed, data of excessive quality loses its value to MM2 because it shuts down downstream speculation and removes any exposure to adverse selection. This latter observation and Proposition 3 suggest that a cap on data fees may increase rather than decrease data quality. The numerical analysis reported in Figure 3 confirms that the ranking between q_2^* and $q_2^*|_{\Phi=0}$ is ambiguous. The highest data quality obtains without monetization for a large range of parameters.

Even if data-fee regulation increases data quality and thus information revelation, a complete assessment requires considering the adjustment of trading fees by the exchange. It follows from the discussion above that when $q_2^* < q_2^*|_{\Phi=0}$, the exchange's upstream trading fee is lower without monetization than under laissez-faire, as the exchange seeks to attract speculators to maximize the quality of its data. However, the effect of the data-fee cap on the downstream trading fee is ambiguous. In particular, absent data monetization, the exchange no longer has incentives to lower this fee to increase MM2's value for data. As a result, when the exchange cannot monetize the data, it may increase its downstream trading fee relative to laissez-faire, with a potentially adverse impact on welfare (Lemma 3).

To determine the overall effect of a cap on data fees, we solve the model and the no-monetization benchmark numerically. Figure 3 reports the results of this analysis by showing a heat map of the difference between variables' values in the laissez-faire equilibrium and their counterparts without monetization. The top panel represents the exchange's policies: data quality and trading fees in both markets. The bottom panel depicts statistics of interest: the total volume, the informed volume share, and social welfare.

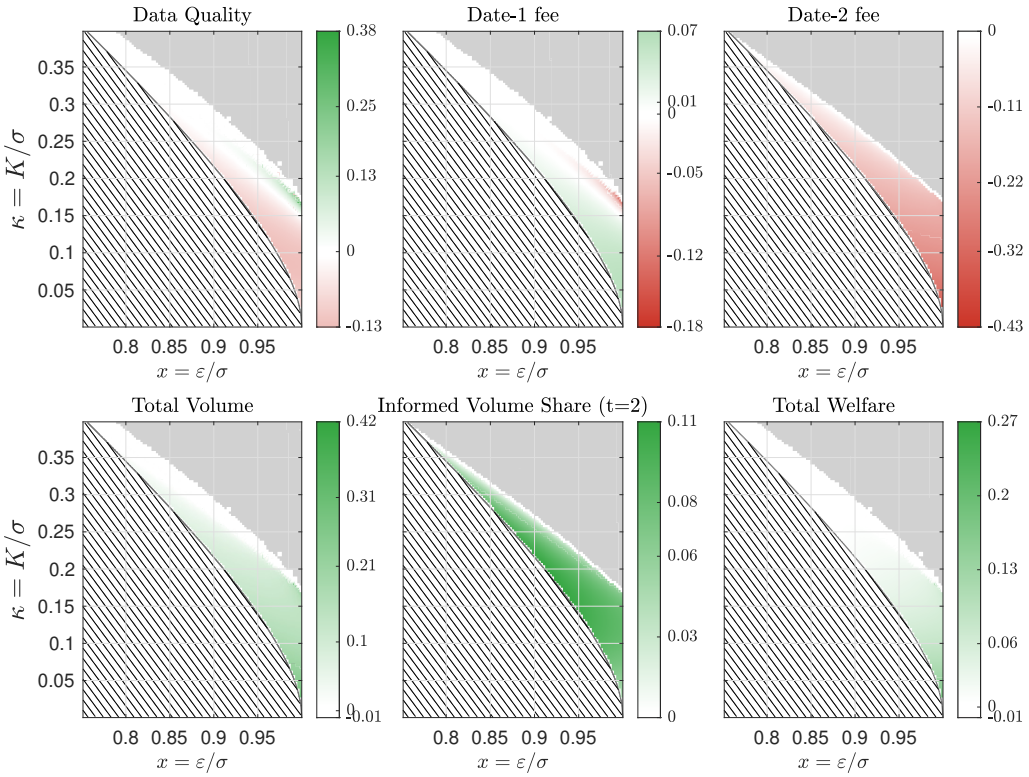


Figure 3. Laissez-faire vs. no-monetization benchmark
The x-axis (y-axis) is the normalized liquidity traders' private benefit ε/σ (private trading cost parameter $\kappa = K/\sigma$). Each panel plots the difference between the laissez-faire variable and the no-monetization benchmark variable. Green (red) values indicate that the laissez-faire variable is higher (lower) than the benchmark variable. Welfare (volume) is normalized by the scaling factor $\bar{m}\sigma^2/(2\bar{k})$ ($\bar{m}\sigma/(2\bar{k})$). For parameters in the gray area, there is no speculation in both cases. Parameters in the striped area violate Assumption 2.

The top panel of Figure 3 confirms that, while the ranking is ambiguous, data quality is lower with monetization for a significant range of parameters. The comparison for upstream trading fees mirrors the pattern for data quality, as upstream fees are reduced to achieve higher data quality. Interestingly, the numerical analysis confirms the intuition above that the exchange tends to increase its downstream trading fee when it cannot charge for data services. These observations explain the patterns observed in the bottom panel of Figure 3.

The bottom panel of Figure 3 reveals that total volume and welfare are higher with data

monetization for most parameter values. This result also holds in the parameter region where data quality is lower with monetization. This might seem counterintuitive, as volume and welfare increase with data quality, by Corollary 5 and Lemma 3, respectively. However, data monetization incentivizes the exchange to keep the downstream trading fee low. Numerical results show that the no-monetization constraint raises that fee enough to offset the welfare and volume gains from higher data quality. Again, this result speaks to the importance of accounting for the joint determination of data fees and trading fees by exchanges.

7 Extensions

Finally, we consider two variants of our baseline model. In Section 7.1, we characterize the data policy of an exchange that enjoys market power over the sale of data but competes for trading services. In Section 7.2, we compare the baseline model with common ownership of both markets to the situation in which each market is run by a distinct exchange.

7.1 Competitive Trading Services

In our model, the exchange has market power over both trading services and data services. The model allows us to proxy for competition in trading services by setting the exchange trading fees to zero in both markets.[30] In this benchmark, the exchange derives profits only from the sale of data. Formally, it solves Problem (18)-(19) facing the constraint that trading fees are equal to zero, in both the upstream and downstream markets.[31]

[30]Setting $\boldsymbol{f} = (f_1, f_2) = (0, 0)$ is only a proxy for competition in trading services, as our framework does not incorporate a venue choice for traders. In particular, profits from data services could theoretically allow competing exchanges to pay market participants for trading, that is, to set negative trading fees.

[31]The converse analysis of an exchange with market power over trading services but competitive data pricing is partly covered by our analysis in Section 6.2 in which we impose a zero data fee.

We compare the outcome under monopolistic trading fees to the competitive trading outcome, with a focus on optimal data quality. We denote $q_2^*|_{\boldsymbol{f}=0}$ the optimal data quality in the competitive trading benchmark.

Proposition 4. *Under competitive trading services, the optimal data quality is*

$$q_2^*|_{\boldsymbol{f}=0} = \min\{\bar{q}_2(0), q_{2c}\}, \quad \textit{where} \quad \frac{1-q_{2c}/2}{(1-q_{2c})^{3/2}} = \max\left\{1, \frac{2\sigma-\varepsilon}{K}\right\}. \tag{22}$$

Data quality increases with asset volatility σ and decreases with liquidity traders' private benefit ε and private information cost c. When $K \leq \sigma - \frac{3}{4}\varepsilon$, the upstream order flow is fully revealing ($\bar{q}_2(0) = 1$) but the exchange data is optimally noisy.

Proposition 4 characterizes the optimal data quality chosen by an exchange that provides competitive trading services. Unlike in the general case, a tight characterization of $q_2^*|_{\boldsymbol{f}=0}$ exists because the exchange problem becomes one-dimensional. The comparative statics reflect the negative impact of exchange data on speculative trading in this model. As explained before, data sales reduce speculative trading, which ultimately lowers the value of the data to MM2. As a result, the exchange optimally reduces data quality as speculation becomes costlier (e.g. lower σ or higher c). The same intuition explains why the exchange data never fully reveal the asset value if the upstream order flow itself is fully informative. Doing so would shut down downstream speculation, so data would have zero value to MM2. While incentives to limit the quality of data also exist with noncompetitive trading, these are stronger here, because the upstream order flow is very informative. Indeed, Corollary 1 shows that the informativeness of the order flow is decreasing with the upstream trading fee; it is thus maximized with competitive trading services because $f_1 = 0$ in that benchmark.

The discussion above suggests that exchanges produce better data in a competitive envi-

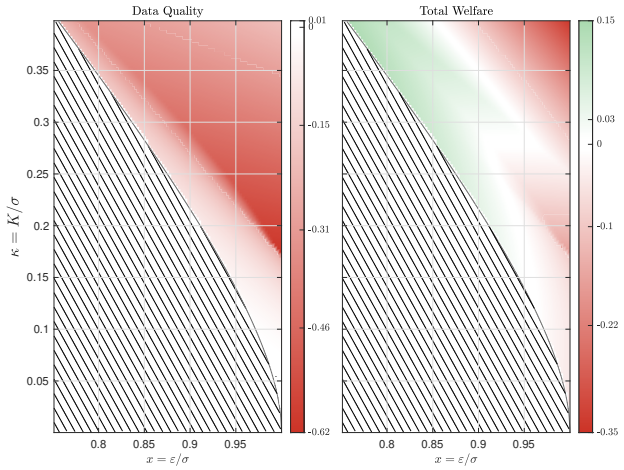


Figure 4. Market power in trading services and efficiency
The x-axis (y-axis) is the normalized liquidity traders' private benefit ε/σ (private trading cost parameter $\kappa = K/\sigma$). Each panel represents, as a heat map, the difference between the variable in the main model (with market power in trading services) and the competitive trading services benchmark. Welfare is normalized by the scaling factor $\bar{m}\sigma^2/(2\bar{k})$. Parameters in the striped area violate Assumption 2.

ronment. Indeed, the order-flow informativeness, which is an upper bound on data quality, decreases with the exchange's upstream trading fee, f_1. Competition in trading services thus expands the data frontier. However, the impact on the ultimate data quality sold by the exchange is not straightforward. Indeed, the exchange can degrade the data quality, and optimally does so when the order flow is very informative (Proposition 4). Furthermore, Lemma 2 shows that, for a fixed downstream trading fee, the exchange's target data quality is higher when it enjoys market power in trading services, because data quality increases trading-fee profits. We use a numerical analysis to determine which force dominates. The results reported in Figure 4 show that, for most parameter values, the exchange sells higher-quality data when it competes on trading services. That is, the positive effect of competitive trading fees on the technological frontier dominates. Still, the result can flip for K and σ low enough; then, data quality is highest with market power in trading services.

The second panel of Figure 4 reports the effect of market power in trading services on

welfare. It shows that social welfare can increase with market power in trading services, although data quality tends to be lower. This result arises because noncompetitive trading fees can increase welfare by acting as a tax on speculation (Lemma 3). In particular, while a competitive trading fee in the upstream market increases the order flow informativeness and thus data quality, it may foster too much speculation. If that effect is strong, welfare may decrease as the market for trading services becomes more competitive.

7.2 Separate Exchange Ownership

So far, the model featured a single exchange running both markets. However, data sold by an exchange can be used on other trading platforms. To isolate the role of exchange ownership, we compare the common-ownership baseline with the case of separate ownership. Under separate ownership, an upstream (downstream) exchange controls the upstream (downstream) market. We solve for a Nash equilibrium in exchanges' policy choices.

Exchange ownership has implications for data policy. Under separate ownership, the upstream exchange no longer internalizes the effect of data sales on the downstream exchange's profit.[32] As we explained, common ownership induces the exchange to increase data quality, because it raises downstream profits. The next result follows from this observation.

Lemma 4. *For a given downstream trading fee, f_2, data quality $q_2^*(f_2)$ is lower under separate exchange ownership than under common ownership.*

As explained above, under common ownership, the exchange has strong incentives to sell high-quality data because it internalizes the positive effect of data on its downstream profit

[32]Observe that without a data market, exchange ownership plays no role in our model, because data is the only element that connects trading across markets. Even with a data market, exchange ownership is also irrelevant if the market for trading services is competitive as in Section 7.1. Hence, the analysis is nontrivial only if both the upstream and the downstream exchanges enjoy some market power over trading services.

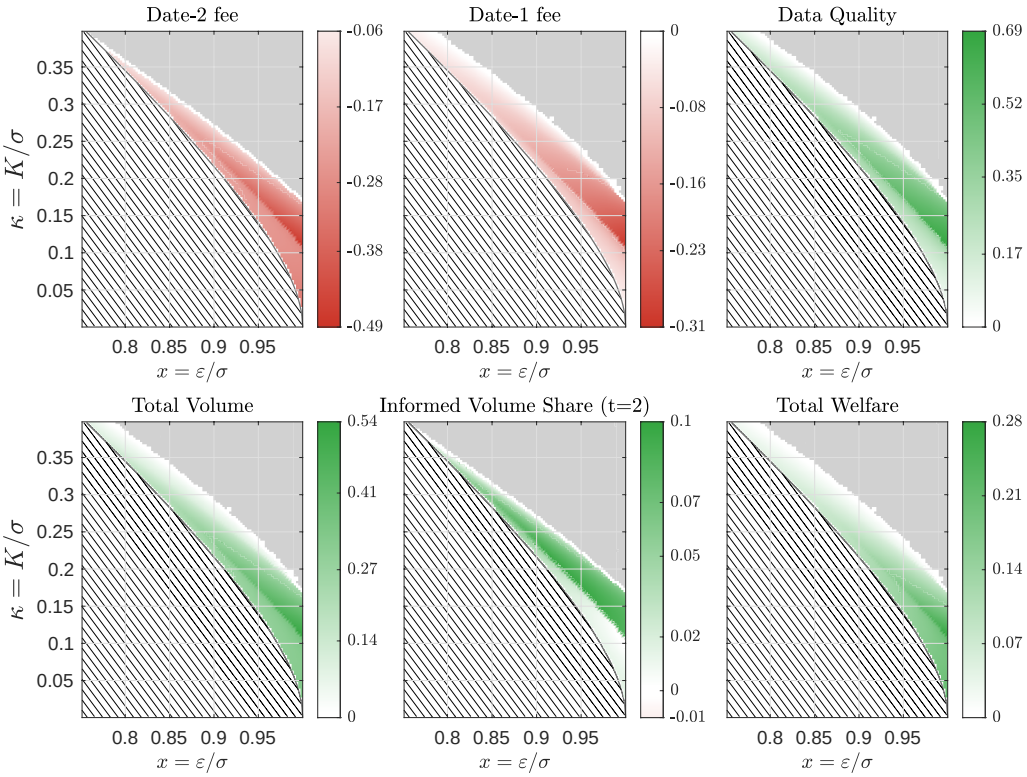


Figure 5. Common exchange ownership vs. separate ownership
The x-axis (y-axis) is the normalized liquidity traders' private benefit ε/σ (private trading cost parameter $\kappa = K/\sigma$). Each plot represents as a heat map the difference between the variable value under common ownership and its value under separate ownership. Welfare (volume) is normalized by the scaling factor $\bar{m}\sigma^2/(2\bar{k})$ ($\bar{m}\sigma/(2\bar{k})$). Parameters in the striped area violate Assumption 2.

from trading fees. With separate ownership, these incentives are no longer present for the upstream exchange. This alone may not imply that equilibrium data quality is higher under common ownership. Indeed, Lemma 4 keeps the downstream trading fee constant, but this fee depends on the ownership regime. Hence, the effect of exchange fragmentation on data quality is more subtle. Still, the numerical analysis reported in Figure 5 shows that the intuition from Lemma 4 applies: data quality is highest under common exchange ownership.

To complete the comparison between ownership regimes, Figure 5 reports additional statistics of interest. In particular, it shows that the downstream trading fee is typically

lower under common ownership.[33] Since the integrated exchange sells higher-quality data and sets lower trading fees, total volume is higher in this case. Finally, we find that social welfare is higher under common ownership than under separate ownership.

8 Conclusion

This paper studies the joint pricing of data services and trading services by exchanges. Our analysis shows that these two businesses are intimately linked: trading fees affect the informativeness of order flow, and data dissemination affects revenues from trading services. The analysis delivers two main insights. First, the ability to sell data induces the exchange to lower trading fees, which increases trading volume and welfare. More surprisingly, we find that a cap on data fees may raise data quality, but the exchange can respond by increasing trading fees, with adverse effects on volume and welfare. These results can help reassess regulation of exchanges' data policies and empirical findings related to data policy changes by exchanges. Our analysis leaves several directions for future research. In particular, unlike in our model, exchanges sell differentiated data products to heterogeneous data users. In addition, future research could study data market outcomes in a setting where exchanges compete for the production and provision of data.

[33] For a fixed data quality q_2, we can show that the integrated exchange chooses a lower downstream trading fee than the downstream exchange under separate ownership, because the former internalizes the negative effect of f_2 on the value of data.

References

Admati, Anat R and Paul Pfleiderer, "A monopolistic market for information," *Journal of Economic Theory*, 1986, *39* (2), 400–438.

Admati, Anat R. and Paul Pfleiderer, "Divide and Conquer: A Theory of Intraday and Day-of-the-Week Mean Effects," *The Review of Financial Studies*, 04 1989, *2* (2), 189–223.

Asriyan, Vladimir, William Fuchs, and Brett Green, "Information Spillovers in Asset Markets with Correlated Values," *American Economic Review*, July 2017, *107* (7), 2007–40.

Bartlett, Robert P, Justin McCrary, and Maureen O'Hara, "The Market Inside the Market: Odd-Lot Quotes," *The Review of Financial Studies*, 03 2025, *38* (3), 661–711.

Bergemann, Dirk and Alessandro Bonatti, "Markets for Information: An Introduction," *Annual Review of Economics*, 2019, *11* (Volume 11, 2019), 85–107.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, "Equilibrium fast trading," *Journal of Financial Economics*, 2015, *116* (2), 292–313.

Bolton, Patrick, Tano Santos, and Jose A. Scheinkman, "Cream-Skimming in Financial Markets," *The Journal of Finance*, 2016, *71* (2), 709–736.

Brogaard, Jonathan, James Brugler, and Dominik Rösch, "Competition and Exchange Data Fees," SSRN Working Paper 2024. Version dated October 10, 2024; accessed June 23, 2026.

Budish, Eric, Robin S. Lee, and John J. Shim, "A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?," *Journal of Political Economy*, 2024, *132* (4), 1209–1246.

Calcagno, Riccardo and Stefano Lovo, "Bid-Ask Price Competition with Asymmetric Information between Market-Makers," *The Review of Economic Studies*, 2006, *73* (2), 329–355.

Cespa, Giovanni, "Information Sales and Insider Trading with Long-Lived Information," *The Journal of Finance*, 2008, *63* (2), 639–672.

_ **and Thierry Foucault**, "Sale of Price Information by Exchanges: Does It Promote Price Discovery?," *Management Science*, 2014, *60* (1), 148–165.

_ **and Xavier Vives**, "Exchange Competition, Entry, and Welfare," *The Review of Financial Studies*, 09 2021, *35* (5), 2570–2624.

_ **and** _ , "Market Opacity and Fragility: Why Liquidity Evaporates When It Is Most Needed," *American Economic Review*, July 2026, *116* (7), 2454–2503.

Chao, Yong, Chen Yao, and Mao Ye, "Why Discrete Price Fragments U.S. Stock Exchanges and Disperses Their Fee Structures," *The Review of Financial Studies*, 07 2018, *32* (3), 1068–1101.

Colliard, Jean-Edouard and Thierry Foucault, "Trading Fees and Efficiency in Limit Order Markets," *The Review of Financial Studies*, 2012, *25* (11), 3389–3421.

Cong, Lin William, Siguang Li, and Mingzhe Zheng, "Open Data and Financial Market Quality," 2026. Working paper. First draft: June 2025; this draft: January 2026.

Degryse, Hans, Mark Van Achter, and Gunther Wuyts, "Dynamic order submission strategies with competition between a dealer market and a crossing network," *Journal of Financial Economics*, 2009, *91* (3), 319–338.

Duffie, Darrell, Thierry Foucault, Laura Veldkamp, and Xavier Vives, *Barcelona 4: Technology and Finance*, Paris and London: CEPR Press, 2022.

Dugast, Jérôme and Thierry Foucault, "Data abundance and asset price informativeness," *Journal of Financial Economics*, 2018, *130* (2), 367–391.

Easley, David, Maureen O'Hara, and Liyan Yang, "Differential Access to Price Information in Financial Markets," *Journal of Financial and Quantitative Analysis*, 2016, *51* (4), 1071–1110.

_ **, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman**, "Liquidity, Information, and Infrequently Traded Stocks," *The Journal of Finance*, 1996, *51* (4), 1405–1436.

European Commission, "Capital Markets Union: Commission Proposes New Measures to Boost Europe's Capital Markets," https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6251 November 2021. Press release IP/21/6251, November 25, 2021; accessed June 23, 2026.

Foucault, Thierry and Christine A. Parlour, "Competition for Listings," *The RAND Journal of Economics*, 2004, *35* (2), 329–355.

_ **and Laurent Fresard**, "Learning from peers' stock prices and corporate investment," *Journal of Financial Economics*, 2014, *111* (3), 554–577.

García, Diego and Francesco Sangiorgi, "Information Sales and Strategic Trading," *The Review of Financial Studies*, 05 2011, *24* (9), 3069–3104.

Gehrig, Thomas, "Intermediation in Search Markets," *Journal of Economics & Management Strategy*, 1993, *2* (1), 97–120.

Glosten, Lawrence R., "Economics of the Stock Exchange Business: Proprietary Market Data," 2020. Available at SSRN: https://ssrn.com/abstract=3533525.

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

Hendershott, Terrence and Charles M. Jones, "Island Goes Dark: Transparency, Fragmentation, and Regulation," *The Review of Financial Studies*, 01 2005, *18* (3), 743–793.

_ , Marc Rysman, and Rainer Schwabe, "Stock exchanges as platforms for data and trading," *Journal of Financial Markets*, 2025, *75*, 100986.

Huang, Shiyang, Yan Xiong, and Liyan Yang, "Skill Acquisition and Data Sales," *Management Science*, 2022, *68* (8), 6116–6144.

Mackintosh, Phil and Michael Normyle, "How Exchanges Compete: An Economic Analysis of Platform Competition," Technical Report, Nasdaq 2024.

Madhavan, Ananth, "Consolidation, Fragmentation, and the Disclosure of Trading Information," *The Review of Financial Studies*, 1995, *8* (3), 579–603.

Pagano, Marco and Ailsa Roell, "Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading," *The Journal of Finance*, 1996, *51* (2), 579–611.

Pagnotta, Emiliano S. and Thomas Philippon, "Competing on Speed," *Econometrica*, 2018, *86* (3), 1067–1115.

Securities Industry and Financial Markets Association, "SIFMA Letter on Market Data Infrastructure Proposed Rule," Comment letter submitted to the U.S. Securities and Exchange Commission May 2020. File No. S7-03-20; published May 26, 2020; accessed June 30, 2026.

Spatt, Chester S., "Is Equity Market Exchange Structure Anti-Competitive?," Technical Report, Carnegie Mellon University 2020.

U.S. Securities and Exchange Commission, "SEC Proposes to Modernize Key Market Infrastructure Responsible for Collecting, Consolidating, and Disseminating Securities Market Data," https://www.sec.gov/news/press-release/2020-34 February 2020. Press Release No. 2020-34, accessed June 23, 2026.

Vairo, Maren and Piotr Dworczak, "Optimal Transparency in OTC Markets," June 2023. Working Paper.

Zhu, Haoxiang, "Do Dark Pools Harm Price Discovery?," *The Review of Financial Studies*, 2014, *27* (3), 747–789.

Appendix

A Proofs

A.1 Proof of Proposition 1

The MM profit function is separable in the bid and ask price, that is,

$$\Pi_t^{MM}(a_t, b_t; p_t) = \underbrace{\mathbb{E}\big[(\tilde{D}_t^I + \tilde{D}_t^L)(a_t - \tilde{v}) \,|\, p_t\,\big]}_{\equiv \Pi_{ask,t}^{MM}(a_t; p_t)} + \underbrace{\mathbb{E}\big[(\tilde{S}_t^I + \tilde{S}_t^L)(\tilde{v} - b_t) \,|\, p_t\,\big]}_{\equiv \Pi_{bid,t}^{MM}(b_t; p_t)} \tag{A.1}$$

By symmetry,, for any $p_t \in \{0, 1/2, 1\}$, we have $\Pi_{bid}^{MM}(b_t; p_t) = \pi_{ask}^{MM}(-b_t, 1-p_t)$. This implies that the optimal bid price is given by $b^*(p) = -a^*(1-p)$. We thus solve for the optimal ask price $a^*(p)$ below and obtaint the optimal bid from that relationship.

First, consider the cases $p_t \in \{0, 1\}$. Since speculators have no information advantage in this case, no speculator trades when $b^*(p) - f < v(p) < a^*(p) + f$, which holds because the MM is a monopolist. When the MM sells only to liquidity traders, the optimal ask solves

$$\max_{a_t} \frac{\bar{m}}{2} G(\varepsilon - a_t + v(p_t) - f_t)(a_t - v(p_t)) = \frac{\bar{m}}{2\bar{k}}(\varepsilon - a_t - f_t)(a_t - \sigma) \tag{A.2}$$

We obtain $a^*(p_t) = v(p_t) + \frac{1}{2}(\varepsilon - f)$ and thus $b^*(p_t) = -a^*(1-p_t) = v(p_t) - \frac{1}{2}(\varepsilon - f)$ for $p_t \in \{0, 1\}$. As claimed above, speculators do not trade, because the bid (ask) is strictly lower (higher) than the asset value, $v(p)$.

Next, consider the case $p_t = \frac{1}{2}$ in which the MM is uninformed about the asset value. Under Assumption 1 , two equilibrium outcomes are possible. Either the MM does not post quotes and enjoys zero profit or it sets $a_t \in (0, \varepsilon - f)$ and enjoys positive profit. In this

second case, given $v(1/2) = 0$ the MM profit from the ask component is

$$\begin{aligned}\Pi^{MM}_{ask,t}(a_t; 1/2) &= \frac{\bar{m}}{2} G(\varepsilon - a_t - f_t) a_t + \frac{1}{2}\mu_t G(\sigma - a_t - f_t)(a_t - \sigma) \\ &= \frac{\bar{m}}{2\bar{k}}(\varepsilon - a_t - f_t) a_t + \frac{\mu_t}{2\bar{k}}(\sigma - a_t - f_t)(a_t - \sigma) \end{aligned} \tag{A.3}$$

This expression is concave in a_t so the first-order condition is necessary and sufficient:

$$0 = \bar{m}(\varepsilon - f_t - 2a_t) + \mu_t(2\sigma - f_t - 2a_t)$$

We obtain $a^*_t(1/2)$ as given by equation (8), and $b^*_t(1/2) = -a^*_t(1/2)$ from the argument above.

We will verify the conjectures that $s^*_t(1/2) < \varepsilon - f_t$ and the MM finds it optimal to quote $(b^*_t(1/2), a^*_t(1/2))$ instead of not trading at date $t \in \{1, 2\}$ in the proof of Lemma 1 in which we derive the equilibrium mass of speculators. For now, we observe that the first condition writes

$$\varepsilon - f_t \geq \frac{\mu_t}{\mu_t + \bar{m}}(2\sigma - \varepsilon), \tag{A.4}$$

and that the profit of the uninformed MM when quoting $(b^*_t(1/2), a^*_t(1/2))$ is given by

$$\Pi^{MM}_t(b^*_t(1/2), a^*_t(1/2); 1/2) = \bar{k}^{-1}\left[(\bar{m} + \mu_t)\big(s^*_t(1/2)\big)^2 - \mu_t\sigma(\sigma - f_t)\right] \tag{A.5}$$

where $s^*_t(1/2)$ is given by equation (8).

A.2 Proof of Corollary 1

In equilibrium, speculative demand at $t = 1$ when is given by

$$D_1^{I*}(v) = \begin{cases} \frac{\mu_1}{\bar{k}}(\sigma - s_1^*(1/2) - f_1) & \text{if} \quad v = \sigma \\ 0 & \text{if} \quad v = -\sigma \end{cases} \tag{A.6}$$

The equilibrium demand from liquidity traders is given by

$$\tilde{D}_1^{L*} = \frac{\tilde{m}_1^{L+}}{\bar{k}}(\varepsilon - s_1^*(1/2) - f_1), \quad \tilde{m}_1^{L+} \sim \mathcal{U}[0, \bar{m}] \tag{A.7}$$

From (A.6) and (A.7), the conditional density of total equilibrium demand, D_1^*, when $v = -\sigma$ is given by:

$$h(D_1^* \,|\, v = -\sigma\,) = \begin{cases} \left(\frac{\bar{m}}{\bar{k}}(\varepsilon - s_1^*(1/2) - f_1)\right)^{-1} & \text{if} \quad D_1^* \in \left[0, \frac{\bar{m}}{\bar{k}}(\varepsilon - s_1^*(1/2) - f_1)\right] \\ 0 & \text{otherwise} \end{cases}, \tag{A.8}$$

while the conditional density when the asset value is high is

$$h(D_1^* \,|\, v = \sigma\,) = h\big(D_1^* - D_1^{I*}(\sigma) \,|\, v = -\sigma\,\big) \tag{A.9}$$

By Bayes law, the posterior that the asset value is high satisfies

$$\bar{p}_2 \equiv \Pr[v = \sigma \,|\, D_1^*] = \frac{h(D_1^* \,|\, v = \sigma\,)\Pr[v = \sigma]}{h(D_1^* \,|\, v = \sigma\,)\Pr[v = \sigma] + h(D_1^* \,|\, v = -\sigma\,)\Pr[v = -\sigma]} \tag{A.10}$$

$$= \frac{h(D_1^* \,|\, v = \sigma\,)}{h(D_1^* \,|\, v = \sigma\,) + h(D_1^* \,|\, v = -\sigma\,)} \tag{A.11}$$

From (A.8) and (A.9), we obtain

$$\bar{p}_2(D_1^*) = \begin{cases} 0 & \text{if} \quad D_1^* \in [0, D_1^{I*}(\sigma)] \\ \frac{1}{2} & \text{if} \quad D_1^* \in \left[D_1^{I*}(\sigma), \frac{\bar{m}}{k}(\varepsilon - s_1^*(1/2) - f_1)\right] \\ 1 & \text{if} \quad D_1^* \in \left[\frac{\bar{m}}{k}(\varepsilon - s_1^*(1/2) - f_1), \frac{\bar{m}}{k}(\varepsilon - s_1^*(1/2) - f_1) + D_1^{I*}(\sigma)\right] \end{cases} \tag{A.12}$$

This proves the conjecture that the order flow either perfectly reveals the asset value (when $\bar{p}_2 \in \{0, 1\}$) or provides no information (when $\bar{p}_2 = 1/2$).

To complete the proof, we characterize the conditional distribution of $\bar{p}_2$ in equilibrium. Conditional on v, $\bar{p}_2$ takes either value $1/2$ or value $\frac{1}{2} + \frac{v}{2\sigma}$. Furthermore, by symmetry

$$\Pr[\bar{p}_2 = 1 \,|\, v = \sigma] = \Pr[\bar{p}_2 = 0 \,|\, v = \sigma] \equiv \bar{q}_2 \tag{A.13}$$

Hence, the order flow informativeness, $\bar{q}_2$, is given by

$$\bar{q}_2 = \Pr[\bar{p}_2 = 1 \,|\, v = \sigma] = \Pr\left[D_1^* \geq \frac{\bar{m}}{k}(\varepsilon - s_1^*(1/2) - f_1) \;\middle|\; v = \sigma\right] \tag{A.14}$$

$$= \min\left\{\frac{D_1^{I*}(\sigma)}{\frac{\bar{m}}{k}(\varepsilon - s_1^*(1/2) - f_1)}, 1\right\} = \min\left\{\frac{\mu_1}{\bar{m}}\frac{\sigma - s_1^*(1/2) - f_1}{\varepsilon - s_1^*(1/2) - f_1}, 1\right\} \tag{A.15}$$

This concludes the proof of Corollary 1.

A.3 Proof of Corollary 2

Due to the symmetry of the market maker' problem on the ask and bid sides, the equilibrium profit of MM2 with data $p_2 \in \{0, 1/2, 1\}$ is

$$\Pi_2^{MM}(p_2) \equiv \Pi_2^{MM}\big(a_2^*(p_2), b_2^*(p_2); p_2\big) = 2\,\Pi_{\text{ask},2}^{MM}\big(a_2^*(p_2), p_2\big), \tag{A.16}$$

where $\Pi_2^{MM}(a, b; p_2)$ is given by (7) and $\Pi_{ask,2}^{MM}$, the ask component, is defined in (A.1).

First, consider the cases $p = 0$ and $p = 1$. We showed that speculators do not trade. From equation (A.2), the ask component of the MM profit is

$$\Pi_{ask,2}^{MM}\left(a_2^*(1), 1\right) = \frac{\bar{m}}{2\bar{k}}(\varepsilon - a_2^*(1) - f_2)a_2^*(1) = \frac{\bar{m}}{8\varepsilon}(\varepsilon - f_2)^2 \tag{A.17}$$

Next, consider the case in which $p_2 = \frac{1}{2}$. From equation (A.3), we obtain

$$\begin{aligned}
\Pi_{\text{ask},2}^{MM}\big(a_2^*\big(\tfrac{1}{2}\big)\,;\tfrac{1}{2}\big) &= \frac{\bar{m}}{2\bar{k}}\left(\varepsilon - s_2^*\big(\tfrac{1}{2}\big) - f_2\right) s_2^*\big(\tfrac{1}{2}\big) + \frac{\mu_2}{2\bar{k}}\left(\sigma - s_2^*\big(\tfrac{1}{2}\big) - f_2\right)\left(s_2^*\big(\tfrac{1}{2}\big) - \sigma\right) \\
&= \frac{1}{2\bar{k}}\left[(\bar{m} + \mu_2)\left(s_2^*\big(\tfrac{1}{2}\big)\right)^2 - \mu_2\sigma(\sigma - f_2)\right] \qquad \text{(A.18)} \\
&= \frac{\bar{m} + \mu_2}{8\bar{k}}\left[\varepsilon - f_2 + \frac{\mu_2}{\mu_2 + \bar{m}}(2\sigma - \varepsilon)\right]^2 - \frac{\mu_2}{2\bar{k}}\sigma(\sigma - f_2) \qquad \text{(A.19)}
\end{aligned}$$

where we obtain (A.19) from (A.18) by substituting $s_2^*\big(\frac{1}{2}\big)$ thanks to equation (8).

The value of data to MM2 is its incremental profit relative to the case when it is uninformed. When MM2 purchases the data, its gross profit is

$$\Pi_2^{MM}(\mu_2, f_2, q_2) \equiv 2\left\{(1 - q_2)\Pi_{ask,2}^{MM}(a^*(1/2), 1/2) + q_2\Pi_{ask,2}^{MM}(a^*(1), 1)\right\}$$

Its gross profit when it does not buy data is given by $\Pi_2^{MM}(\mu_2, f_2, 0)$ as having no data is

equivalent to having data with zero quality. Hence, the value of exchange data to MM2 is:

$$\begin{aligned} VD(\mu_2, f_2, q_2) &= \Pi^{MM}(\mu_2, f_2, q_2) - \Pi^{MM}(\mu_2, f_2, 0), \\ &= q_2 \left\{ \frac{\bar{m}}{4\bar{k}}(\varepsilon - f_2)^2 - \frac{\bar{m}+\mu_2}{4\bar{k}} \left[\varepsilon - f_2 + \frac{\mu_2}{\mu_2+\bar{m}}(2\sigma - \varepsilon) \right]^2 + \frac{\mu_2}{\bar{k}}\sigma(\sigma - f_2) \right\}, \\ &= \frac{\mu_2 q_2}{4\bar{k}} \left[-(\varepsilon - f_2)^2 - 2(\varepsilon - f_2)(2\sigma - \varepsilon) - \frac{\mu_2}{\mu_2+\bar{m}}(2\sigma - \varepsilon)^2 + 4\sigma(\sigma - f_2) \right], \\ &= \frac{\mu_2 q_2}{4\bar{k}} \left[\frac{\bar{m}}{\bar{m}+\mu_2}(2\sigma - \varepsilon)^2 - f_2^2 \right]. \end{aligned} \tag{A.20}$$

The second equality follows from (A.17) and (A.19). The final expression is equivalent to expression (10). This concludes the proof.

A.4 Proof of Corollary 3

Consider first the expected trading volume for a given data realization $p_t \in \{0, \frac{1}{2}, 1\}$ at date t. The expected volume from liquidity traders is

$$EV_t^L(p_t) \equiv \mathbb{E}(\tilde{D}^L + \tilde{S}^L \mid p_t) = \frac{\bar{m}}{\bar{k}} \left(\varepsilon - s_t^*(p_t) - f_t \right) \tag{A.21}$$

$$= \frac{\bar{m}}{2\bar{k}} \left(\varepsilon - f_t \right) - \frac{\bar{m}}{2\bar{k}} \frac{\mu_t}{\mu_t + \bar{m}} (2\sigma - \varepsilon) \mathbb{1}_{p_t = \frac{1}{2}} \tag{A.22}$$

Informed speculators submit buy or sell orders (depending on the value of v) only if $p_t = \frac{1}{2}$. Thus, the informed volume as a function of the data realization $p_t \in \{0, 1/2, 1\}$ is

$$V_t^I(1) = V_t^I(0) = 0 \tag{A.23}$$

$$V_t^I\left(\frac{1}{2}\right) = \frac{\mu}{\bar{k}}\left(\sigma - s^*\left(\frac{1}{2}\right) - f_t\right) = \frac{\mu}{\bar{k}}\left(\sigma - \frac{1}{2}(\varepsilon - f_t) - \frac{1}{2}\frac{\mu}{\mu + \bar{m}}(2\sigma - \varepsilon) - f_t\right), \tag{A.24}$$

$$= \frac{\mu}{2\bar{k}}\left(\frac{\bar{m}}{\mu + \bar{m}}(2\sigma - \varepsilon) - f_t\right). \tag{A.25}$$

Summing $V_t^L(p_t)$ and $V_t^I(p_t)$ and taking expectations over p_t, we obtain equation (11), using $p_1 = 1$ and equation (2) for the distribution of p_2.

A.5 Proof of Lemma 1

In Step 1, we prove the result in Lemma 1. In Step 2, we verify our claim that the necessary conditions stated in Proposition 1 hold when the mass of speculators is endogenized. For this proof, we define $K_t \equiv K(1-q)^{-1/2}$.

Step 1: Endogenous Speculator Mass

As mentioned in the main text, two outcomes are possible. First, suppose that there is no speculator entry, that is, $\mu_t = 0$. Proposition 1 implies that $s^*\left(\frac{1}{2}\right)\Big|_{\mu=0} = \frac{1}{2}(\varepsilon - f)$. Given such spread, speculators indeed find it optimal not to enter if and only if the net profit from private information, given by (13), is negative. This condition writes

$$0 \geq (1 - q_t)\frac{1}{2\bar{k}}\left(\sigma - s_t^*\left(\tfrac{1}{2}\right)\Big|_{\mu_t=0} - f_t\right)^2 - c \tag{A.26}$$

$$2K_t \geq \sigma - \frac{1}{2}(\varepsilon + f_t), \tag{A.27}$$

which is consistent with formula (14) in Lemma 1.

Next, if condition (A.27) does not hold, there must be positive entry by speculators in equilibrium. In this case, μ_t is pinned down by setting the net profit from information acquisition in (13) to zero. We obtain

$$K_t = \sigma - \frac{1}{2}(\varepsilon + f_t) - \frac{1}{2}\frac{\mu}{\mu + \bar{m}}(2\sigma - \varepsilon) \tag{A.28}$$

$$\Leftrightarrow \quad 2K_t = \frac{\bar{m}}{\mu + \bar{m}}(2\sigma - \varepsilon) - f_t \tag{A.29}$$

Equation (14) follows directly from this last equality.

Step 2: Necessary conditions for Proposition 1

Now, we show that the characterization of the equilibrium in Proposition 1 is valid when μ_t is given by (14). For simplicity, we write $\mu_t^*(f_t, K_t)$ for $\mu_t^*(f_t, q_t)$. The quote $s_t^*(1/2)$ given by (8) is optimal only if $s_t^*(1/2) < \varepsilon - f_t$ and MMt achieves a positive profit, $\Pi_t^{MM}(s_t^*(1/2); 1/2)$. In that proof, we maintain $\varepsilon > f_t$. This condition holds because f_t is chosen by the profit-maximizing exchange and there is no trading if it does not hold.

Case 1: $\mu_t^*(f_t, K_t) = 0$

Setting $\mu_t = 0$ in equation (A.5) and substituting $s_t^*(1/2)|_{\mu_t=0} = \frac{1}{2}(\varepsilon - f_t)$, we obtain

$$\Pi_t^{MM,*} = \frac{\bar{m}}{4\bar{k}}(\varepsilon - f_t)^2$$

which is strictly positive when $\varepsilon > f_t$.

Case 2: $\mu_t^*(f_t, K_t) > 0$

By the speculators' zero-profit condition, we have $s_t^*(1/2) = \sigma - f_t - K_t$, which implies

$$s_t^*(1/2) \leq \varepsilon - f_t \quad \Leftrightarrow \quad K_t > \sigma - \varepsilon \tag{A.30}$$

Observe that $K_t \geq K$ for any $q_t \geq 0$, so the inequality above is implied by Assumption 2.

From equation (A.5), the MM profit at quote $s^*\left(\frac{1}{2}\right)\big|_{\mu_t=\mu_t^*(q_t,f_t)}$ is given by

$$\Pi^{MM,*} = (\bar{k})^{-1}\left[(\bar{m}+\mu_t^*(q_t,f_t))\Big(s^*(1/2)\Big)^2 - \mu_t^*(q_t,f_t)\sigma(\sigma-f_t)\right] \tag{A.31}$$

$$= (\bar{k})^{-1}\left[(\bar{m}+\mu_t^*(q_t,f_t))\Big(\sigma-f_t-K_t\Big)^2 - \mu_t^*(q_t,f_t)\sigma(\sigma-f_t)\right] \tag{A.32}$$

$$= \frac{\bar{m}}{\bar{k}}\left[\frac{2\sigma-\varepsilon}{2K_t+f_t}\Big(\sigma-f_t-K_t\Big)^2 - \left(\frac{2\sigma-\varepsilon}{2K_t+f_t)}-1\right)\sigma(\sigma-f_t)\right] \tag{A.33}$$

$$= \frac{\bar{m}}{\bar{k}}\left[\sigma(\sigma-f_t) - \frac{2\sigma-\varepsilon}{2K_t+f_t}\left\{\sigma(\sigma-f_t) - \Big(\sigma-f_t-K_t\Big)^2\right\}\right] \tag{A.34}$$

After simplification, the condition $\Pi^{MM,*} \geq 0$ is equivalent to $Q(f_t, K_t) \geq 0$ where

$$Q(f,K) = f^2(\sigma-\varepsilon) - f(\sigma-\varepsilon)(\sigma-2K) + K\Big[K(2\sigma-\varepsilon) - 2\sigma(\sigma-\varepsilon)\Big] \tag{A.35}$$

Next, we verify that this inequality holds for all $f_t \in [0, 2(\sigma-K_t)-\varepsilon]$ where the upper bound on f_t is implied by $\mu_t^* \geq 0$. $Q(f,K)$ is a quadratic polynomial in f with $Q_{ff} > 0$. Hence, its minimum is attained at $\underline{f}_t = \sigma/2 - K_t$. Two cases are possible.

First if $K_t \leq \sigma/2$, $\underline{f}_t \in [0, 2(\sigma-K_t)-\varepsilon]$. Hence, the mimimum of $Q(., K_t)$ in the interval $[0, 2(\sigma-K_t)-\varepsilon]$ is attained at $\underline{f}_t$. Using the fact that $Q(f,K)$ is quadratic in f, we obtain

$$\begin{aligned} Q(\underline{f}_t, K_t) &= -(\sigma-\varepsilon)\left(\frac{\sigma}{2}-K_t\right)^2 + K_t\Big[K_t(2\sigma-\varepsilon) - 2\sigma(\sigma-\varepsilon)\Big] \\ &= \sigma K_t^2 - \sigma(\sigma-\varepsilon)K_t - \frac{\sigma^2}{4}(\sigma-\varepsilon) \end{aligned}$$

This expression is quadratic in K_t with $Q_{KK} > 0$ and $Q(\underline{f}_t, K_t) < 0$. Hence, when $K_t \geq 0$, it is positive if and only if K_t is above the positive root, which is the lower bound on K imposed by Assumption 2. The conclusion follows from $K_t > K$.

Finally, suppose that $K_t \geq \sigma/2$, which implies that $\underline{f}_t < 0$. Hence, the minimum of $Q(., K_t)$ in the interval $[0, 2(\sigma - K_t) - \varepsilon]$ is attained at $f_t = 0$. We have

$$Q(0, K_t) \geq 0 \quad \Leftrightarrow \quad K_t \geq \frac{2\sigma(\sigma - \varepsilon)}{2\sigma - \varepsilon} \tag{A.36}$$

Since, again, $K_t \geq K$, the desired result follows from $\sigma/2 \geq \frac{2\sigma(\sigma-\varepsilon)}{2\sigma-\varepsilon}$.

A.6 Proof of Corollary 4

From equation (9), we obtain

$$\bar{q}_2(f_1) = \bar{q}_2\big(\mu_1^*(f_1, 0), f_1\big)\big) = \min\left\{\frac{\mu_1^*(f_1, 0)}{\bar{m}} \frac{\sigma - s_1^*(1/2) - f_1}{\varepsilon - s_1^*(1/2) - f_1}, 1\right\} \tag{A.37}$$

If $f_1 \geq 2(\sigma - K) - \varepsilon$ and thus $\mu_1^*(f_1, 0) = 0$, then $\bar{q}_2(f_1) = 0$. If instead $f_1 \leq 2(\sigma - K) - \varepsilon$ and thus $\mu_1^*(f_1, 0) > 0$, use the zero-profit condition of speculators to obtain:

$$\bar{q}_2(f_1) = \min\left\{\frac{\mu_1^*(f_1, 0)}{\bar{m}} \frac{K}{\varepsilon - \sigma + K}, 1\right\} \tag{A.38}$$

Substituting for $\mu_1^*(f_1, 0)$ thanks to equation (14), we then obtain equation (15).

A.7 Proof of Lemma 2

Part 1: Profit component from trading fees

Consider first the profits from trading fees at date t

$$\Pi^E_{trad,t}(f_t, q_t) = EV(\mu^*_t(f_t, q_t), f_t, q_t) f_t \tag{A.39}$$

$$= \left\{ \frac{\bar{m}}{2\bar{k}} (\varepsilon - f_t) - \frac{\mu^*_t(f_t, q_t)}{2\bar{k}} (1 - q_t) f_t \right\} f_t \tag{A.40}$$

Case 1: $\mu^*_t(f_t, q_t) = 0$

Then $\Pi^E_{trad,t}(f_t, q_t)$ is independent of q_t and a quadratic and concave function in f_t. This proves that $\Pi^E_{trad,t}$ is single-peaked in f_t.

Case 1: $\mu^*_t(f_t, q_t) > 0$

In this case, we have

$$\Pi^E_{trad,t}(f_t, q_t) = \frac{\bar{m}}{2\bar{k}} \left\{ \varepsilon - f_t - \left(\frac{2\sigma - \varepsilon}{2K(1 - q_t)^{-1/2} + f_t} - 1 \right) (1 - q_t) f_t \right\} f_t$$

Thus,

$$\frac{\partial \Pi^E_{trad,t}}{\partial f_t} \propto \varepsilon - 2q_t f_t - (1 - q_t)(2\sigma - \varepsilon) \frac{\left[4K(1 - q_t)^{-1/2} + f_t\right] f_t}{\left(2K(1 - q_t)^{-1/2} + f_t\right)^2} \tag{A.41}$$

$$= \varepsilon - 2q_t f_t - (1 - q_t)(2\sigma - \varepsilon) + \frac{4K^2(2\sigma - \varepsilon)}{\left(2K(1 - q_t)^{-1/2} + f_t\right)^2} \tag{A.42}$$

It follows that $\frac{\partial^2 \Pi^E_{trad,t}}{\partial f_t^2} < 0$, which means that $\Pi^E_{trad,t}$ is concave in f_t. To prove that it is single-peaked, observe that

$$\lim_{f_t \to \infty} \frac{\partial \Pi^E_{\text{trad},t}}{\partial f_t} > 0, \qquad \lim_{f_t \to 0^+} \frac{\partial \Pi^E_{\text{trad},t}}{\partial f_t} < 0 \tag{A.43}$$

Next, to see why $\Pi^E_{trad,t}(f_t, q_t)$ increases with q_t observe that

$$q_t \to (1-q_t)\mu^*_t(f_t, q_t) = (1-q_t)\left(\frac{2\sigma-\varepsilon}{2K(1-q_t)^{-1/2}+f_t} - 1\right)$$

is decreasing with q_t.

Part 2: Profit component from data sales

Next, consider the effect of the exchange policies on profits from selling data. Corollary 2 shows that data has no value when $\mu^*_2 = 0$. We are thus left to consider the case in which $\mu^*_2 > 0$ at date 2. From equation (10), we have

$$VD(f_2, q_2) = \frac{\mu^*_2(f_2, q_2)q_2}{4\bar{k}}\left[\frac{\bar{m}}{\bar{m}+\mu^*_2(f_2,q_2)}(2\sigma-\varepsilon)^2 - f_2^2\right] \tag{A.44}$$

First, to see that $VD(f_2, q_2)$ decreases with f_2, observe that

$$VD_{f_2} = \frac{\partial VD}{\partial \mu_2}\frac{\partial \mu^*_2(f_2,q_2)}{\partial f_2} + \frac{\partial VD}{\partial f_2} \tag{A.45}$$

Lemma 1 implies that $\frac{\partial \mu^*_2(f_2,q_2)}{\partial f_2} < 0$ from and $\frac{\partial VD}{\partial f_2} < 0$ follows immediately from equation (A.44). Hence, to prove that $VD_{f_2} < 0$, it is sufficient to show that $\frac{\partial VD}{\partial \mu_2} > 0$. We have

$$\frac{\partial VD}{\partial \mu_2} = \frac{q_2}{4\bar{k}}\left[\frac{\bar{m}^2}{(\bar{m}+\mu_2)^2}(2\sigma-\varepsilon)^2 - f_2^2\right]$$

In any equilibrium of the trading game with speculation, we have

$$0 \leq \sigma - s_2^*(1/2) - f_2 = \sigma - \frac{1}{2}(\varepsilon - f_2) - \frac{1}{2}\frac{\mu_2}{\mu_2 + \bar{m}}(2\sigma - \varepsilon) - f_2$$
$$= \frac{1}{2}\left[\frac{\bar{m}}{\mu_2 + \bar{m}}(2\sigma - \varepsilon) - f_2\right]$$

This implies that $\frac{\partial VD}{\partial \mu_2} \geq 0$ and thus that $VD_{f_2} < 0$.

Finally, consider the impact of data quality on profits from data sales. From Lemma 1, for every $f_2 \in [0, 2(\sigma - K) - \varepsilon]$, there exists $\hat{q}(f_2)$ such that $\mu_2^*(f_2, q_2) = 0$ for all $q_2 \geq \hat{q}(f_2)$. This implies that

$$\forall\, q_2 \geq \hat{q}(f_2), \quad VD(f_2, q_2) = VD(f_2, 0) = 0 \tag{A.46}$$

We are left to consider the interval $[0, \hat{q}(f_2)]$. Equation (A.46) implies that VD is single-peaked on $[0, \hat{q}(f_2)]$ with a strict interior maximizer if it is concave. To prove concavity, we use the following parameterization:

$$x = \frac{\varepsilon}{\sigma}, \qquad A \equiv 2 - x, \qquad \phi = \frac{f_2}{\sigma}, \qquad \kappa = \frac{K}{\sigma}, \qquad z = \frac{\mu_2}{\bar{m}}, \tag{A.47}$$

Normalizing the value of data, $VD(f_2, q_2)$ in equation (A.44), we obtain

$$vd(\phi, q) \equiv \frac{2\bar{k}}{\bar{m}\sigma^2} VD(f_2, q_2) = \frac{z(\phi, q)q}{2}\left[\frac{A^2}{1 + z(\phi, q)} - \phi^2\right], \tag{A.48}$$

with

$$z(\phi, q) = \frac{A}{\phi + 2\kappa(1 - q)^{-1/2}} - 1 \tag{A.49}$$

Consider the following change of variables $y = (1-q)^{1/2}$, such that

$$z(\phi, y) = \frac{Ay}{\phi y + 2\kappa} - 1 = \frac{(A-\phi)y - 2\kappa}{\phi y + 2\kappa}.$$

Hence, we can write the normalized value of data as

$$vd(\phi, q(y)) = F(\phi, y) \equiv \frac{1-y^2}{2} \frac{[(A-\phi)y - 2\kappa]\,[2A\kappa + \phi(A-\phi)y]}{y(\phi y + 2\kappa)}.$$

Using the chain rule, we have

$$vd_{qq}(\phi, q(y)) = \frac{1}{4y^3}\left[yF_{yy}(\phi, y) - F_y(\phi, y)\right]$$

Cumbersome but direct calculations give

$$vd_{qq}(\phi, q(y)) = -\frac{A\kappa^2\left[2\kappa^2 y^2 + 6\kappa^2 + 3\kappa\phi y^3 + 9\kappa\phi y + 4\phi^2 y^2\right]}{y^5(2\kappa + \phi y)^3} < 0$$

This shows that VD is concave in data quality for any $q_2 \in [0, \hat{q}_2(f_2))$, which corresponds to the speculation region with $\mu_2^*(f_2, q_2) > 0$.

A.8 Proof of Lemma 3

Part 1: Impact of f_t on W_t

When $\mu_t^* = 0$

$$W_t = \frac{\bar{m}}{\bar{k}} \int_0^{\frac{1}{2}(\varepsilon - f_t)} (\varepsilon - k)dk$$

which is strictly decreasing in f_t.

When instead $\mu_t^* > 0$, the free entry condition for speculators implies

$$c = (1-q_t)\frac{\mu_t}{\bar{k}}\int_0^{\hat{k}_t^I(1/2)]}\left[\hat{k}_t^I(1/2) - k\right]dk = (1-q_t)\frac{\mu_t}{\bar{k}}\int_0^{\hat{k}_t^I(1/2)} k dk$$

Hence, we can rewrite equation (21) as follows

$$W_t = \frac{\bar{m}}{\bar{k}}\left[\int_0^{\hat{k}_t^L(1/2)}(\varepsilon - k)dk + q_t\int_{\hat{k}_t^L(1/2)}^{\hat{k}_t^L(0)}(\varepsilon - k)dk\right] - 2\mu_t c \tag{A.50}$$

Equation (20) shows that $\hat{k}_t^L(1/2)$ is independent from the trading fee f_t. Hence, we have

$$\frac{\partial W_t}{\partial f_t} = \frac{\bar{m}}{\bar{k}}q_t\frac{\partial \hat{k}_t^L(0)}{\partial f_t}\left[\varepsilon - \hat{k}_t^L(0)\right] - 2\frac{\partial \mu_t}{\partial f_t}c = -\frac{\bar{m}}{4\bar{k}}(\varepsilon - f_t)q_t - 2\frac{\partial \mu_t}{\partial f_t}c$$

where the second equality obtains by definition of $\hat{k}_t^L(0) = \varepsilon - s_t^*(0) - f_t = \frac{1}{2}(\varepsilon - f_t)$. Hence, since $\frac{\partial \mu_t}{\partial f_t} < 0$ by Lemma 1, the sign of the derivative $\frac{\partial W_t}{\partial f_t}$ is ambiguous. At date 1, however, $q_1 = 0$, which implies that the derivative is positive, as claimed in Lemma 3.

Part 2: Impact of q_t on W_t

Observe that $\hat{k}_t^L(0)$ is independent of q_t. Differentiating the right-hand side of equation (A.50), we obtain

$$\frac{\partial W_t}{\partial q_t} = \frac{\bar{m}}{\bar{k}}\left[\int_{\hat{k}_t^L(1/2)}^{\hat{k}_t^L(0)}(\varepsilon - k)dk + (1-q_t)\frac{\partial \hat{k}_t^L(1/2)}{\partial q_t}\left(\varepsilon - \hat{k}_t^L(1/2)\right)\right] - 2\frac{\partial \mu_t}{\partial q_t}c \tag{A.51}$$

We know that $\frac{\partial \mu_t}{\partial q_t} < 0$ from equation (14). Besides, from speculators' free entry condition,

$$\hat{k}_t^L(1/2) = \varepsilon - s_t^*(1/2) - f_t = \varepsilon - \sigma + K(1-q_t)^{-1/2}, \tag{A.52}$$

which implies that $\frac{\partial \hat{k}_t^L(1/2)}{\partial q_t} > 0$. Since the first term of (A.51) is strictly positive as well, it follows that $\frac{\partial W_t}{\partial q_t} > 0$, which concludes the proof.

A.9 Proof of Proposition 2

Let $(f_1^*(q), f_2^*(q))$ be the optimal fees for data quality $q \in [0,1]$. The exchange's objective function is concave in (f_1, f_2), so the first-order conditions are necessary and sufficient.

Part 1: Optimal trading fees $(f_1^*(0), f_2^*(0))$ without data sales

We determine the optimal fees when $q_2 = 0$. In this case, the exchange problem is identical across dates, so we abstract from the time subscript. We have

$$f_1^*(0) = f_2^*(0) = f^*(0) \equiv \arg\max_f EV(\mu^*(f,0), f, 0)f = \frac{\bar{m}}{2\bar{k}}(\varepsilon - f)f - \frac{\mu^*(f,0)}{2\bar{k}}f^2 \quad \text{(A.53)}$$

Lemma A.1 (Optimal trading fees without data sales)**.** *If $q_2 = 0$, the optimal trading fee is*

$$f^*(0) = \begin{cases} K\left(\sqrt{\frac{2(2\sigma-\varepsilon)}{\sigma-\varepsilon}} - 2\right), & \text{if } K < K^I, \\ 2\sigma - \varepsilon - 2K, & \text{if } K^I \leq K \leq K^N, \\ \frac{\varepsilon}{2}, & \text{if } K > K^N, \end{cases}$$

with

$$K^I \equiv \sqrt{\frac{(2\sigma-\varepsilon)(\sigma-\varepsilon)}{2}}, \qquad K^N \equiv \sigma - \frac{3}{4}\varepsilon.$$

Speculation is strictly positive at the optimum, that is, $\mu^*(f^*(0), 0) > 0$ if and only if $K < K^I$.

Proof. Given that the objective is concave by part, there are three candidate solutions.

Case i) $\mu^* = 0$ and $f^*(0) = \varepsilon/2$

This case is valid if $\mu^*(\varepsilon/2, 0) = 0$ which is equivalent to $K > \sigma - \frac{3}{4}\varepsilon$

Case ii) $\mu^*(f^*(0), 0) > 0$

Substitute for $\mu^*(f, 0)$ in equation (A.53) thanks to equation (14) to obtain

$$EV(\mu^*(f,0), f, 0)f = \frac{\bar{m}}{2\bar{k}}\left[(\varepsilon - f)f - \frac{2\sigma - \varepsilon}{2K + f}f^2 + f^2\right] \tag{A.54}$$

$$= \frac{\bar{m}}{2\bar{k}}\left[2(\varepsilon - \sigma)f + \frac{2K(2\sigma - \varepsilon)f}{2K + f}\right] \tag{A.55}$$

The first-order condition writes

$$0 = -2(\sigma - \varepsilon) + \frac{4K^2(2\sigma - \varepsilon)}{[2K + f^*(0)]^2} \quad \Leftrightarrow \quad f^*(0) = K\left[\sqrt{\frac{2(2\sigma - \varepsilon)}{\sigma - \varepsilon}} - 1\right] \tag{A.56}$$

This conjecture is valid if $\mu^*(f^*(0), 0) > 0$, which is equivalent to

$$f^*(0) \leq 2\sigma - \varepsilon - 2K \quad \Leftrightarrow \quad 4K\sqrt{\frac{2\sigma - \varepsilon}{\sigma - \varepsilon}} \leq 2\sigma - \varepsilon \tag{A.57}$$

This holds if and only if $K \leq \sqrt{\frac{(2\sigma-\varepsilon)(\sigma-\varepsilon)}{2}}$, which is strictly lower than $\sigma - \frac{3}{4}\varepsilon$

Case iii) $K \in \left[\sqrt{\frac{(2\sigma-\varepsilon)(\sigma-\varepsilon)}{2}}, \sigma - \frac{3}{4}\varepsilon\right]$ In this case, the optimal fee is determined by the condition $\mu^*(0, f^*(0)) = 0$. Using equation (14), we get

$$f^*(0) = 2\sigma - \varepsilon - 2K \tag{A.58}$$

□

Part 2: Data Sales Lower Exchange Trading Fees

Upstream Trading Fee The first-order condition with respect to f_1 is

$$0 = \frac{\partial \Pi^E_{trad}(f_1, f_2, q_2)}{\partial f_1} + \lambda \frac{d\bar{q}_2}{df_1} = \underbrace{\frac{\partial}{\partial f_1}\Big[f_1 EV(\mu_1(f_1, 0), f_1, 0)\Big]}_{\equiv H(f_1)} + \lambda \frac{d\bar{q}_2}{df_1}, \tag{A.59}$$

where $\lambda \geq 0$ is the Lagrange multiplier on the constraint $q_2 \leq \bar{q}_2(f_1)$. By definition, $H(f_1^*(0)) = 0$. Since $\frac{d\bar{q}_2}{df_1} < 0$ by Corollary 4, the right-hand side of (A.59) is weakly negative at $f_1^*(0)$. By concavity of the objective function, this implies that $f_1^*(q) \leq f_1^*(0)$ for all $q \geq 0$.

Downstream Trading Fee

First, suppose that $\mu_2^*(f_2^*(0), 0) = 0$, that is, there is no speculation at date 2 without data sales. Them since μ^* decreases with its first argument, it must be that for any $f_2 \geq f_2^*(0)$, we have $\mu_2^*(f_2^*(0), q) = 0$. This implies that for any $f_2 > f_2^*(0)$, the exchange can strictly benefit by lowering f_2 until $f_2 = f_2^*(0)$. Hence, the downstream fee with data sales must be lower than $f_2^*(0)$.

Now, consider the more interesting case in which $\mu_2^*(f_2^*(0), 0) > 0$. For the rest of this proof, let $q \equiv q_2$ and $K_q \equiv K(1-q)^{-1/2}$ and $A \equiv 2\sigma - \varepsilon$. The profit component of (19)

relevant for the choice of f_2 is

$$\Pi_2^E(f_2,q) = EV(\mu_2^*(f_2,q),f_2,q)\, f_2 + VD(\mu_2^*(f_2,q),f_2,q) \tag{A.60}$$

$$= \frac{\bar{m}f_2}{2\bar{k}}\left[\varepsilon - f_2 - (1-q)f_2\frac{A-2K_q-f_2}{2K_q+f_2}\right] \tag{A.61}$$

$$+ \frac{q\bar{m}}{4\bar{k}}\left[A(A-2K_q-f_2) - f_2^2\frac{A-2K_q-f_2}{2K_q+f_2}\right] \tag{A.62}$$

$$= \frac{f_2\bar{m}}{2\bar{k}}\left[\varepsilon - qf_2 - A(1-q) + (1-q)\frac{2AK_q}{2K_q+f_2}\right] \tag{A.63}$$

$$+ \frac{q\bar{m}}{4\bar{k}}\left[A(A-2K_q-f_2) + f_2^2 - \frac{A}{2K_q+f_2}f_2^2\right] \tag{A.64}$$

The first-order derivative of the exchange profit with respect to f_2 is given by

$$\frac{\partial\Pi_2^E}{\partial f_2} = \frac{\bar{m}}{2\bar{k}}\left[\varepsilon - A(1-q) - 2qf_2 + \frac{4(1-q)AK_q^2}{(2K_q+f_2)^2}\right]$$

$$+ \frac{q_2\bar{m}}{2\bar{k}(2K_q+f_2)^2}\Big[f_2(f_2+2K_q)^2 - 2AK^2(q)f_2\Big] \tag{A.65}$$

$$= -\frac{\bar{m}}{2\bar{k}(2K_q+f_2)^2}\Psi(f_2,q), \quad \text{where} \tag{A.66}$$

$$\Psi(f_2,q) \equiv (\sigma-\varepsilon)f_2\Big[2f_2+8K_q\Big] - 4\varepsilon K^2(q) + q\Big(f_2(f_2+2K_q)^2 + 2(2\sigma-\varepsilon)K^2(q)\Big). \tag{A.67}$$

By definition, $\Psi(f_2^*(q),q) = 0$. This implies that $f_2^*(q) \le f_2^*(0)$ if and only if $\Psi(f_2^*(0),q) \ge 0$ since $\Pi_2^E(f_2,q)$ is concave with respect to f_2. We use the following result

Lemma A.2. *Let* $r = (1-q)^{-1/2}$ *and* $\xi = f_2^*(0)/K$ *and* $y = \frac{K}{\sigma-\varepsilon}$, *then*

$$\Psi(f_2^*(0),q) = \frac{(\sigma-\varepsilon)K^2(0)(r-1)}{r^2}\Big(r^3c_3 + r^2c_2 + rc_1 + c_0\Big), \quad \textit{with} \tag{A.68}$$

$$c_3 = -\xi^2 + 4 - 4\xi + 4, \tag{A.69}$$

$$c_2 = 4\xi^2 y - \xi^2 + 4 + 4\xi + 4, \tag{A.70}$$

$$c_1 = \xi^2 y(\xi + 4), \tag{A.71}$$

$$c_0 = \xi^3 y. \tag{A.72}$$

Proof. Substituting in equation (A.67), we obtain

$$\begin{aligned}
\Psi(f_2^*(0), q) &= (\sigma - \varepsilon)K\xi\Big[2K\xi + 8Kr\Big] - 4\varepsilon K^2 r^2 + \frac{r^2 - 1}{r^2}\Big(K^3\xi(\xi + 2r)^2 + 2(2\sigma - \varepsilon)K^2 r^2\Big) \\
&= \frac{K^2}{r^2}\left\{r^2(\sigma - \varepsilon)\xi\Big[2\xi + 8r\Big] - 4\varepsilon r^4 + (r^2 - 1)\Big(K\xi(\xi + 2r)^2 + 2(2\sigma - \varepsilon)r^2\Big)\right\} \\
&= \frac{K^2}{r^2}\left\{-K\xi^3 - 4K\xi^2 r + \Big[2(\sigma - \varepsilon)\xi^2 + K\xi^3 - 4K\xi - 2(2\sigma - \varepsilon)\Big]r^2\right\} \\
&= +\frac{K^2}{r^2}\left\{\Big[8\xi(\sigma - \varepsilon) + 4\xi^2 K\Big]r^3 + \Big[-4\varepsilon + 2(2\sigma - \varepsilon) + 4\Big]r^4\right\}
\end{aligned}$$

Using the first-order condition for $f_2^*(0)$, given by $\Psi(f_2^*(0), 0) = 0$, which can be written

$$(\sigma - \varepsilon)(\xi^2 + 4\xi) = 2\varepsilon \tag{A.73}$$

and identifying power terms, we obtain equation (A.68). □

Lemma A.2 implies that the desired result follows if $c_i \geq 0$ for all $i \in \{0, 1, 2, 3\}$. This result holds trivially for c_0 and c_1. To prove the result for c_2 and c_3, observe that Assumption 2 implies that

$$y = \frac{K_0}{\sigma - \varepsilon} \geq \frac{1}{2}\left(1 + \sqrt{\frac{2\sigma - \varepsilon}{\sigma - \varepsilon}}\right) = \frac{1}{2}\left(1 + \frac{\xi + 2}{\sqrt{2}}\right), \tag{A.74}$$

We then obtain

$$c_3 \geq -\xi^2 + \xi(2 + 2\sqrt{2}) + \xi^2\sqrt{2} - 4\xi + 4 \geq 0$$
$$c_2 \geq 2\xi^2(1 + \sqrt{2}) - \xi^2 + 4 + 4\xi + 4 \geq 0$$

Hence, $\frac{\partial \Pi_2^E}{\partial f_2}(f_2^*(0), q) \leq 0$ which implies that $f_2^*(0) > f_2^*(q)$ for all $q \geq 0$, by concavity of Π_2^E. This concludes the proof for Part 1.

Part 2: Total Trading Costs

Now, we prove that total trading costs also decrease when the exchange sells data. At date 1, when $\mu_1^* > 0$, the zero profit condition for speculators implies

$$s_1^*(1/2) + f_1^* = \sigma - K \tag{A.75}$$

This proves that total trading costs, $s_1^*(1/2) + f_1^*$ are unchanged at $t = 1$ and that the upstream spread, $s_1^*(1/2)$, increases strictly relative to the no-data benchmark when $f_1^*(q^*) > f_1^*(0)$.

Finally, consider total expected trading costs decease at date 2:

$$\mathbb{E}[s_2^*(p_2)] + f_2^* = s_2^*(1/2) + f_2 - q_2^*\big[s_2^*(1/2) - s^*(0)\big] \tag{A.76}$$
$$< \sigma - K(1 - q_2^*)^{-1/2} \tag{A.77}$$
$$< \sigma - K \leq (\mathbb{E}[s_2^*(p_2)] + f_2^*)|_{q_2=0} \tag{A.78}$$

The inequality in (A.77) follows from Proposition 1 and the free-entry condition for speculators. The inequality in (A.78) follows from $q_2^* \geq 0$ and the final inequality follows again

from the free-entry condition for speculators when $q_2 = 0$. This concludes the proof.

A.10 Proof of Corollary 5

Consider first expected volume at date 1. Combining equations (14) and (11), we obtain

$$EV_1^*(f_1) = \frac{\bar{m}}{\bar{k}}\left[\varepsilon - \sigma + K\frac{\mu^*(f_1, 0) + \bar{m}}{\bar{m}}\right]$$

Given that $\mu^*(f_1, 0)$ is decreasing with f_1 and $f_1^* < f_1^*|_{q_2=0}$, it follows that expected volume at date 1, $EV_1(f_1^*)$ is weakly higher when the exchange can sell data, relative to $q_2 = 0$.

Next, consider date 2. From equation (11), we obtain

$$EV(f_2, q_2) = \frac{\bar{m}}{2\bar{k}}\left[\varepsilon - f_2 - (1 - q_2)\frac{\mu_2^*(f_2, q_2)}{\bar{m}}f_2\right] \tag{A.79}$$

$$= \frac{\bar{m}}{2\bar{k}}\left[\varepsilon - f_2 - (1 - q_2)f_2\left(\frac{2\sigma - \varepsilon}{2K(1 - q_2)^{-1/2} + f_2} - 1\right)\right]. \tag{A.80}$$

The first-order derivatives are characterized by

$$EV_{q_2}(f_2, q_2) \propto -\frac{1}{2\bar{k}}\frac{d\big[(1 - q_2)\mu_2^*(f_2, q_2)\big]}{dq_2}f_2 > 0 \tag{A.81}$$

$$EV_{f_2}(f_2, q_2) \propto -q_2 - 2K(q_2)(1 - q_2)\frac{2\sigma - \varepsilon}{(2K(q_2) + f_2)^2} < 0 \tag{A.82}$$

Relative to the no-data benchmark, $q_2^* > 0$ and $f_2^* < f_2^*|_{q_2=0}$. Equations (A.81) and (A.82) thus imply that $EV_2^* > EV_2^*|_{q_2=0}$, which proves the result for date 2 and concludes the proof.

A.11 Proof of Proposition 4

First, consider the case without speculation at date 1. If $K \geq \sigma - \varepsilon/2$, we obtain $\mu_1^*(0,0) = 0$ which implies that $\bar{q}_2(0) = 0$ by Corollary 4 and thus $q_2^*|_{\boldsymbol{f}=0} = 0$.

Next, consider the case $K < \sigma - \varepsilon/2$. When $f_1 = f_2 = 0$, the exchange solves $\max_{q_2 \in [0, \bar{q}_2(0)]} VD(\mu_2^*(0,q_2), 0, q_2)$. By Lemma 1,

$$\frac{\mu_2^*(0,q_2)}{\bar{m}} = \max\left\{0, \frac{(2\sigma-\varepsilon)(1-q_2)^{1/2}}{2K} - 1\right\}.$$

Thus, speculation is positive if and only if

$$q_2 < q_2^{nospec} \equiv 1 - \left(\frac{2K}{2\sigma-\varepsilon}\right)^2. \tag{A.83}$$

Given that the value of data is zero when $\mu_2^* = 0$ and that this is the only source of revenue for the exchange, it must be that $q_2^*|_{\boldsymbol{f}=0} = 0 < q_2^{nospec}$. Using the expression for the value of data from Corollary 2 and substituting for $\mu_2^*(0,q_2)$ using Lemma 1, we obtain

$$\begin{aligned} VD(\mu_2^*(0,q_2),0,q_2) &= \frac{\bar{m}}{4\bar{k}} q_2 (2\sigma-\varepsilon)^2 \frac{\mu_2^*(0,q_2)/\bar{m}}{1+\mu_2^*(0,q_2)/\bar{m}} \\ &= \frac{\bar{m}}{4\bar{k}} q_2 (2\sigma-\varepsilon)^2 \left(1 - \frac{2K}{(2\sigma-\varepsilon)\sqrt{1-q_2}}\right). \end{aligned}$$

We show in the proof of Lemma 2 that this expression is concave in q_2, and that the optimum is characterized by the first-order condition:

$$0 = \frac{\partial VD}{\partial q_2} = \frac{\bar{m}(2\sigma-\varepsilon)^2}{4\bar{k}} \left[1 - \frac{2K}{2\sigma-\varepsilon} \frac{1-q_2/2}{(1-q_2)^{3/2}}\right], \tag{A.84}$$

which leads to the implicit definition of q_2^c in equation (22) for the case $K \leq \sigma - \varepsilon/2$ considered

here. Imposing the feasibility constraint $q_2 \leq \bar{q}_2(0)$ therefore gives the stated result.

Next, we establish the comparative statics. Let $h(q) \equiv \frac{1-q/2}{(1-q)^{3/2}}$, which is strictly increasing with q. Hence, when $q_2^*|_{\boldsymbol{f}=0} < \bar{q}_2(0)$, the strict monotonicity follows directly from the implicit function theorem applied to $q_2^*|_{\boldsymbol{f}=0}$ in equation (22). When instead $q_2^*|_{\boldsymbol{f}=0} = \bar{q}_2(0)$, it must be that $\bar{q}_2(0) < 1$ as $q_2^*|_{\boldsymbol{f}=0} < q_2^{nospec} < 1$ from the argument above. Then, in this case, equation (15) in Corollary 4 implies that

$$\bar{q}_2(0) = \frac{2\sigma - \varepsilon - 2K}{2(K - \sigma + \varepsilon)}.$$

This expression is indeed increasing with σ and it is decreasing with ε and K.

Finally, the result that $\bar{q}_2(0) = 1$ when $K \leq \sigma - \frac{3}{4}\varepsilon$ follows from equation (15). In this case, the upstream order flow is fully revealing and, by the argument above, $q_2^*|_{\boldsymbol{f}=0} < 1 = \bar{q}_2(0)$, which means that the exchange data is strictly less precise than the order flow. This concludes the proof.

B Separating Equilibrium of the Signaling Game

In this section, we prove the claim that the unique separating equilibrium of the signaling game is the equilibrium of the game in which the exchange data $p \in \{0, 1/2, 1\}$ is public information. Proposition 1 describes the equilibrium of this latter game, which we henceforth refer to as the public data game.

B.1 Description of the Signaling Game

In the signaling game, the exchange data $p \in \{0, 1/2, 1\}$ is the MM's private information, which we call the MM type. Equation (2) implies that the probability distribution over this type is given by $(q/2, 1-q, q/2)$ for some $q \in [0, 1]$. Liquidity traders form beliefs $\beta : \mathbb{R}^2 \mapsto [0, 1]$ about this type for each pair of quotes $(a, b) \in \mathbb{R}^2$, including out-of-equilibrium quotes. Speculators, on the other hand, know the asset value perfectly, so the MM's quotes possess no informational content for them. Given a belief system $\boldsymbol{\beta}$ of liquidity traders, the MM with type p who quotes (a, b) earns payoff

$$\begin{aligned}\Pi^{MM}(a, b; p, \boldsymbol{\beta}) = &\frac{\bar{m}}{2} G\left(1 - \frac{a - v(\beta(a,b)) + f}{\varepsilon}\right)(a - v(p)) \\ &+ \frac{\bar{m}}{2} G\left(1 - \frac{-b + v(\beta(a,b)) + f}{\varepsilon}\right)(v(p) - b) \\ &+ p\mu G\left(\frac{\sigma - a - f}{\bar{k}}\right)(a - \sigma) + (1-p)\mu G\left(\frac{\sigma + b - f}{\bar{k}}\right)(-\sigma - b), \quad \text{(B.1)}\end{aligned}$$

where G is the CDF of a uniform random variable on $[0, 1]$. The last line of (B.19) corresponds to the MM payoff component with speculators, denoted $\Pi_S^{MM}(a, b; p)$ for future reference. It is identical to the expression in the public data game, and independent of $\boldsymbol{\beta}$ because the speculators are perfectly informed about the asset value. Let $\Pi_L^{MM}(a, b; p, \boldsymbol{\beta})$ be the MM

payoff component with liquidity traders, such that[34]

$$\Pi^{MM}(a,b;p,\boldsymbol{\beta}) = \Pi_L^{MM}(a,b;p,\boldsymbol{\beta}) + \Pi_S^{MM}(a,b;p) \tag{B.2}$$

In what follows, we use superscript $*$ $(**)$ to refer to equilibrium variables of the public-data (signaling) game. A pure strategy equilibrium of the signaling game is given by a set of quotes $\mathcal{Q}^{**} = \{(b^{**}(p), a^{**}(p))\}_{p\in\{0,1/2,1\}}$ and a belief mapping $\beta^{**} : \mathbb{R}^2 \mapsto [0,1]$ such that (i) quotes $(b^{**}(p), a^{**}(p))$ maximize $\Pi^{MM}(a,b;p,\beta^{**})$ for each $p \in \{0,1/2,1\}$, and (ii) belief mapping $\beta^{**} : \mathbb{R}^2 \mapsto [0,1]$ satisfies Bayes rule for any quotes $(a,b) \in \mathcal{Q}^{**}$.

B.2 Existence and Uniqueness of a Separating Equilibrium

Below, we state the main result that proves the statement in the main text.

Proposition B.1. *The only separating equilibrium of the signaling game is the equilibrium of the public data game, characterized in Proposition 1*

We prove Proposition B.1 using a series of intermediate claims. For expositional clarity, we first state these claims and their implications and then provide the proofs in Section **??**.

Claim B.1. *For each $p \in \{0,1/2,1\}$, the MM payoff in the public data game, $\Pi^{MM,*}(p)$ is a lower bound for the MM payoff in the signaling game, denoted $\Pi^{MM,**}(p)$.*

Claim B.1 implies that the only separating equilibrium is the equilibrium of the public data game. Indeed, in a separating equilibrium, the MM's type is revealed in equilibrium.

[34]The expression of the MM payoff in the public data game in (**??**) can be obtained from (B.19), substituting $\beta(a,b)$ with p, since, in the public data game, liquidity traders know the MM data $p \in \{0,1/2,1\}$.

This implies, that an upper bound for the MM's payoff in a separating equilibrium is

$$\Pi^{MM,**}(p) \leq \max_{a,b} \Pi^{MM}(a,b,p,p) = \Pi^{MM,*}(p), \tag{B.3}$$

The rightmost equality obtains by definition of optimality in the public data game. Combining this observation with Claim B.1, which states that $\Pi^{MM,*}(p)$ is a also a lower bound for $\Pi^{MM,**}(p)$, it follows that in a separating equilibrium $\Pi^{MM,**}(p) = \Pi^{MM,*}(p)$. Unicity of the separating equilibrium then follows from the fact that $(a^*(p), b^*(p))$ is the unique pair of optimal quotes for the MM with type p in the public data game.

Thus, the separating equilibrium is unique if it exists. The next result establishes existence by characterizing a belief mapping that supports the public-data game quotes $\{(a^*(p), b^*(p))\}$ as a separating equilibrium of the signaling game.

Claim B.2. *The set of quotes Q^* constitute a separating equilibrium of the signaling game together with the clipped-midpoint belief mapping*

$$\beta^*(a,b) \equiv \begin{cases} 0 & \text{if} \quad \frac{1}{2} + \frac{m(a,b)}{2\sigma} < 0, \\ \frac{1}{2} + \frac{m(a,b)}{2\sigma} & \text{if} \quad \frac{1}{2} + \frac{m(a,b)}{2\sigma} \in [0,1], \\ 1 & \text{if} \quad \frac{1}{2} + \frac{m(a,b)}{2\sigma} > 1 \end{cases} \tag{B.4}$$

where $m(a,b) = \frac{a+b}{2}$ is the midpoint of the (a,b).

B.3 Proofs of Intermediate Results

B.3.1 Proof of Claim B.1

To prove this result, we compute the MM profit at the equilibrium quotes $(b^*(p), a^*(p))$ of the public data game and show that it is weakly higher than $\Pi^{MM,*}(p)$. For ease of notation, we write $\beta^*(p) = \beta(b^*(p), a^*(p))$.

First, observe that the MM profit component with speculators, $\Pi_S^{MM}(a, b; p)$ depends only on the quotes. Hence, by quoting $(b^*(p), a^*(p))$ in the signaling game the MM with type p achieves the same profit with speculators as in the public data game, that is,

$$\Pi_S^{MM}(a^*(p), b^*(p); p) = \Pi_S^{MM,*}$$

To show that the MM achieves a weakly larger profit with liquidity traders in the signaling game than in the public data game with quotes $(b^*(p), a^*(p))$, we prove the following result:

$$\forall\, p \in \{0, 1/2, 1\} \quad \forall\, \beta \in [0, 1], \qquad \Pi_L^{MM}(b^*(p), a^*(p); p, \beta) \geq \Pi_L^{MM}(b^*(p), a^*(p); p, p)$$

Observing that the public data game quotes $(b^*(p), a^*(p))$ are symmetric around the midpoint $v(p)$, and denoting $s^*(p) = a^*(p) - v(p) = v(p) - b^*(p)$, we obtain

$$\begin{aligned}\Pi_L^{MM}(a^*(p), b^*(p); p, \beta) = &\frac{\bar{m}}{2} s^*(p) G\left(1 - \frac{s^*(p) + v(p) - v(\beta^*(p)) + f}{\varepsilon}\right) \\ &+ \frac{\bar{m}}{2} s^*(p) G\left(1 - \frac{s^*(p) - v(p) + v(\beta^*(p)) + f}{\varepsilon}\right) \qquad \text{(B.5)}\end{aligned}$$

We show that this expression is minimized for $\beta = p$ to prove our claim. Let

$$u(p) \equiv \varepsilon - f - s^*(p), \quad x(p) \equiv v(p) - v(\beta^*(p)) \tag{B.6}$$

Thus, condition (B.5) can be written:

$$\frac{\bar{m}}{2} s^*(p) \left\{ G\left(\frac{u(p) + x(p)}{\varepsilon} \right) + G\left(\frac{u(p) - x(p)}{\varepsilon} \right) \right\} \geq \bar{m} s^*(p) \left(\frac{u(p)}{\varepsilon} \right) \tag{B.7}$$

Let $H_p(x) = G\left(\frac{u(p)+x}{\varepsilon}\right)$. By symmetry, condition (B.7) holds if

$$\forall\, x \geq 0, \quad H_p(x) + H_p(-x) \geq 2H_p(0) \tag{B.8}$$

Proposition 1 implies that $u(p) \in \left[0, \frac{\varepsilon}{2}\right]$, wich means that there are three cases to consider.

1. if $x \in [0, u(p)]$, H_p is linear, which implies that (B.8) holds as an equality.

2. If $x \in [u(p), \varepsilon - u(p)]$, we have $H_p(-x) = 0$ and

$$H_p(x) = \frac{u(p) + x}{\varepsilon} \geq \frac{2u(p)}{\varepsilon} = 2H_p(0) \tag{B.9}$$

3. if $x \geq \varepsilon - u(p)$, condition (B.8) follows from $H_p(-x) = 0$, $H_p(x) = 1$, and $H_p(0) \leq \frac{1}{2}$

Hence, condition (B.8) holds, which implies that the sufficient condition (B.7) holds as well. This concludes the proof of Claim B.1.

B.3.2 Proof of Claim B.2

First, the belief mapping given by (B.4) is consistent with Bayes rule on the conjectured equilibrium path. Indeed since, $m(a^*(p), b^*(p)) = v(p)$, we have

$$\beta^*(a^*(p), b^*(p)) = p, \quad \forall p \in \{0, 1/2, 1\} \tag{B.10}$$

by definition of that belief.

The key step of the proof is to verify that no type $p \in \{0, 1/2, 1\}$ wants to deviate from quotes $(a^*(p), b^*(p))$ when facing the clipped-midpoint belief mapping in (B.4). Given the symmetry of the problem, the proof for type $p = 0$ is identical to the proof for type $p = 1$, so we only present the latter. For future reference, define three subsets of quotes:

$$\mathcal{Q}_0 = \left\{(a, b) \in \mathbb{R}^2 \,|\beta^*(a, b) = 0\right\} \tag{B.11}$$

$$\mathcal{Q}_{int} = \left\{(a, b) \in \mathbb{R}^2 \,|\beta^*(a, b) \in (0, 1)\right\} \tag{B.12}$$

$$\mathcal{Q}_1 = \left\{(a, b) \in \mathbb{R}^2 \,|\beta^*(a, b) = 1\right\} \tag{B.13}$$

No Deviation for MM with type $p = 1$

First, by optimality of $(a^*(1), b^*(1))$ in the public-data game, these quotes are optimal when $\beta(a, b) = 1$ for all $(a, b) \in \mathbb{R}^2$. This implies that they dominate all quotes in $\mathcal{Q}_1$, and thus, that no deviation in $\mathcal{Q}_1$ is strictly profitable for the MM with type $p = 1$.

Next, consider quotes $(a, b) \in \mathcal{Q}_0$. We show that the constraint $m(a, b) \leq -\sigma$ binds, which leaves only the case $(a, b) \in \mathcal{Q}_{int}$ to analyze. To prove this result, consider the payoff

of the MM with type $p = 1$ when $\beta(a, b) = 0$:

$$\Pi^{MM}(a, b; 1, \beta = 0) = \frac{\bar{m}}{2} G\left(1 - \frac{a + \sigma + f}{\varepsilon}\right)(a - \sigma) + \frac{\bar{m}}{2} G\left(1 - \frac{-b - \sigma + f}{\varepsilon}\right)(\sigma - b) + \mu G\left(\frac{\sigma - a - f}{\bar{k}}\right)(a - \sigma) \tag{B.14}$$

Let $(\tilde{b}, \tilde{a})$ be the quotes that maximize (B.14), abstracting from the constraint $m(a, b) \leq -\sigma$. The bid $\tilde{b}$ that maximizes the expression above is given by

$$\tilde{b} = \min\{-\sigma + f, -\frac{1}{2}(\varepsilon - f)\} \tag{B.15}$$

Due to the presence of speculators, the ask price $\tilde{a}$ that maximizes (B.14) satisfies

$$\tilde{a} \geq \frac{1}{2}(\varepsilon - f) \tag{B.16}$$

Hence, the unconstrained maximizer $(\tilde{a}, \tilde{b})$ of (B.14) satisfies $m(\tilde{a}, \tilde{b}) > 0$. This implies that the best possible deviation in $\mathcal{Q}_0$ satisfies constraint $m(a, b) \leq -\sigma$ as an equality, and thus, that it also belongs to $\mathcal{Q}_{int}$. The next step rules out deviation in $\mathcal{Q}_{int}$.

Finally, consider quotes $(a, b) \in \mathcal{Q}_{int}$. To prove the desired result, we show that the MM obtains a weakly lower payoff with liquidity traders than in the conjectured equilibrium. For all $(a, b) \in \mathcal{Q}_{int}$, the payoff component with liquidity traders is given by

$$\Pi_L^{MM}(a, b; 1, \beta^*) = \frac{\bar{m}}{2} G\left(1 - \frac{a - m(a, b) + f}{\varepsilon}\right)(a - \sigma) + \frac{\bar{m}}{2} G\left(1 - \frac{-b + m(a, b) + f}{\varepsilon}\right)(\sigma - b), \tag{B.17}$$

where we used that $v(\beta^*(a, b)) = m(a, b)$ for all $(a, b) \in \mathcal{Q}_{int}$. Letting $s(a, b) = a - m(a, b) =$

$m(a,b) - b$ be the half-spread relative to the midpoint for quotes (a,b), we have

$$\Pi_L^{MM}(a,b;1,\beta^*) = \bar{m}G\left(1 - \frac{s(a,b)+f}{\varepsilon}\right)s(a,b)$$

That payoff is maximized for $s = \frac{1}{2}(\varepsilon - f)$, which is equal to $s^*(1)$ in the public data game. This proves that the MM cannot increase its payoff with liquidity traders by deviating to some $(a,b) \in \mathcal{Q}_{int}$. On the other hand, such a deviation causes a weakly greater loss from trading with speculators, relative to the conjectured equilibrium quotes $(a^*(1), b^*(1))$. Indeed, the payoff component with speculators, $\Pi_S^{MM}(a,b;1)$ is weakly negative, and the upper bound of 0 is attained for the conjectured equilibrium quotes. Overall, this shows that

$$\forall\,(a,b) \in \mathcal{Q}_{int}, \quad \Pi^{MM}(a,b;1,\boldsymbol{\beta}^*) \leq \Pi^{MM}(a^*(1),b^*(1);1,\boldsymbol{\beta}^*),$$

which concludes the proof for the MM with type $p = 1$ and by symmetry for the MM with type $p = 0$.

No Deviation for MM with type $p = \frac{1}{2}$

Consider first a deviation to $(a,b) \in \mathcal{Q}_1$. By symmetry, the same proof applies to deviation to $(a,b) \in \mathcal{Q}_0$. We use a similar argument as above to establish that the best possible deviation in $\mathcal{Q}_1$ also belongs to $\mathcal{Q}_{int}$. For this, we show that the best deviation in $\mathcal{Q}_1$ satisfies $m(a,b) \geq \sigma$ as an equality. Consider the payoff of MM with type $p = 1/2$ when

the liquidity traders assign belief 1 to all quotes:

$$\Pi^{MM}\left(a,b;\frac{1}{2},\beta=1\right)=\frac{\bar{m}}{2}G\left(1-\frac{a-\sigma+f}{\varepsilon}\right)a-\frac{\bar{m}}{2}G\left(1-\frac{-b+\sigma+f}{\varepsilon}\right)b \tag{B.18}$$

$$-\frac{1}{2}\mu G\left(\frac{\sigma+b-f}{\bar{k}}\right)(\sigma+b)+\frac{1}{2}\mu G\left(\frac{\sigma-a-f}{\bar{k}}\right)(a-\sigma) \tag{B.19}$$

Observe that the optimal ask $\tilde{a}$ satisfies $\tilde{a}\leq\sigma$. Indeed, the ask price that maximizes the profit component with the liquidity traders is $\frac{\sigma+\varepsilon-f}{2}<\sigma$. This implies that the MM will never ask more than σ since for any $a\geq\sigma$, no speculator buys the asset and its profit with liquidity traders is decreasing with a. The unconstrained best bid satisfies

$$\tilde{b}<\frac{\sigma-\varepsilon+f}{2}$$

Together these observations imply that, when liquidity traders assign belief $\beta=1$ to any deviation, the best deviation satisfies $m(\tilde{a},\tilde{b})\leq\sigma$. Hence, the constraint $m(\tilde{a},\tilde{b})\geq\sigma$ binds for the best deviation in $\mathcal{Q}_1$, which implies that such deviation also belongs to $\mathcal{Q}_{int}$. The argument below rules out such deviations.

Finally, consider a deviation to $(a,b)\in\mathcal{Q}_{int}$. Under the clipped-midpoint belief β^*, we can reparametrize the MM profit as a function of the midpoint $m(a,b)=\frac{a+b}{2}$ and the spread $s=a-m(a,b)=b-m(a,b)$. The payoff component with liquidity traders is:

$$\Pi_L^{MM}\left(a,b,\frac{1}{2},\beta^*\right)=\frac{\bar{m}}{2}G\left(1-\frac{a-m(a,b)+f}{\varepsilon}\right)a-\frac{\bar{m}}{2}G\left(1-\frac{-b+m(a,b)+f}{\varepsilon}\right)b \tag{B.20}$$

$$=\bar{m}G\left(1-\frac{s(a,b)+f}{\varepsilon}\right)s(a,b) \tag{B.21}$$

The payoff component with speculators as a function of $(m(a,b), s(a,b))$ is

$$\Pi_S^{MM}\left(a, b, \frac{1}{2}, \beta^*\right) = \frac{1}{2}\mu G\left(\frac{\sigma - a - f}{\bar{k}}\right)(a - \sigma) - \frac{1}{2}\mu G\left(\frac{\sigma + b - f}{\bar{k}}\right)(\sigma + b)$$

$$= \frac{1}{2}\mu\left[\underbrace{G\left(\frac{\sigma - m - s - f}{\bar{k}}\right)(m + s - \sigma)}_{\equiv F(m,s)} + \underbrace{G\left(\frac{\sigma + m - s - f}{\bar{k}}\right)(s - \sigma - m)}_{\equiv F(-m,s)}\right]$$

Thus for any quotes $(m, s) \in \mathcal{Q}_{int}$,

$$\Pi^{MM}\left(m, s; \frac{1}{2}, \beta^{**}\right) = \bar{m}G\left(1 - \frac{s + f}{\varepsilon}\right)s + \frac{1}{2}\mu\big(F(-m,s) + F(m,s)\big) \tag{B.22}$$

We are left to show that this expression is maximized for the equilibrium quotes of the public-date game, that is, for

$$(m^*(1/2), s^*(1/2)) = \left(0, \frac{1}{2}(\varepsilon - f) + \frac{1}{2}\frac{\mu}{\bar{m} + \mu}(2\sigma - \epsilon)\right).$$

By definition of optimality in the public data game,

$$s^*(1/2) = \arg\max_s \Pi^{MM}\left(0, s; \frac{1}{2}, \beta^{**}\right) \tag{B.23}$$

Hence, to prove the desired result, the following claim is sufficient.

Claim B.3. *For any $s \geq 0$, $m \geq 0$ $F(m,s) + F(-m,s) \leq 2F(0,s)$*

Proof. By symmetry it is enough to consider $m \geq 0$. Let us define the following two thresh-

olds

$$\hat{m}_a(s) = \sigma - s - f \tag{B.24}$$

$$\hat{m}_b(s) = \bar{k} - (\sigma - s - f) = \bar{k} - \hat{m}_a(s) \tag{B.25}$$

Assumption 3 implies that for any $s \geq 0$ and $f \geq 0$, $\hat{m}_b(s) \geq \hat{m}_a(s)$. We show that $m \to F(m,s) + F(-m,s)$ is decreasing on the three relevant intervals.

When $m \in [0, \hat{m}_a(s)]$, $F(m;s)$ is a quadratic function of m, which implies

$$F(m,s) + F(-m,s) = 2F(0;s) + \frac{\partial^2 F}{\partial m^2} m^2 = 2F(0;s) - \frac{2}{\bar{k}} m^2 \tag{B.26}$$

This proves that $m \to F(m,s) + F(-m,s)$ is decreasing for all $m \in [0, \hat{m}_a(s)]$.

Next, for $m \in [\hat{m}_a(s), \hat{m}_b(s)]$, we have

$$F(m,s) + F(-m,s) = \frac{\sigma + m - s - f}{\bar{k}}(s - \sigma - m) \tag{B.27}$$

The term $s - \sigma - m$ is negative and decreasing with m while the term $\sigma + m - s - f$ is positive and increasing with m. □

From Claim B.3, we conclude that the conjectured equilibrium quotes $(m^*(1/2), s^*(1/2)) = \left(0, \frac{1}{2}(\varepsilon - f) + \frac{1}{2}\frac{\mu}{\bar{m}+\mu}(2\sigma - \epsilon)\right)$ are optimal among quotes in $\mathcal{Q}_{int}$. This proves that no strictly profitable deviation exists for the MM with type $p = 1/2$ under the clipped-midpoint belief mapping given by (B.4), which concludes the proof of Claim B.2.