Why Trade Over the Counter?

Tomy Lee Chaojun Wang*

June 4, 2024

Abstract

Over-the-counter trading thrives despite opposing regulatory pressure. In our model, traders who are less likely to be informed optimally choose the OTC market, where a dealer offers a trader-specific discount. Restricting the dealer's ability to price discriminate causes some traders to exit, but induces others with larger gains from trade to enter. Overall, welfare strictly improves if adverse selection risk is low, even as aggregate trade volume falls and average bid-ask spread widens. In a benchmark case, closing the OTC market would raise welfare specifically where *most* trades are over the counter. A simple procedure implements the optimal Pigouvian tax.

JEL-Classification: D47, D61, D82, G14, G18

Keywords: Over-the-counter, exchanges, adverse selection, price discrimination, venue choice

*Lee is at the Central European University. Wang is at the Wharton School, University of Pennsylvania. We thank Philip Bond (the editor), the associate editor, two referees, Jordi Mondria, Xianwen Shi, Darrell Duffie, Ettore Damiano, Colin Stewart, Ádám Szeidl, Ádám Zawadowski, Yu An, Daniel Carvalho, Hongseok Choi, Peter Cziraki, Piotr Dworczak, Jim Goldman, Vincent Glode, Botond Kőszegi, Sam Langfield, Katya Malinova, Vincent Maurin, Andreas Park, Cecilia Parlatore, Christine Parlour, Semih Üslü, Adrian Walton, Liyan Yang, and Yao Zeng for invaluable advice and guidance. Amy Dai, William Fan, Clarise Huang, Dylan Marchlinski, Sifan Tao, and Ceyda Ustun provided excellent research assistance. Jiaqi Zou provided excellent copy editing. The Internet Appendix is available in the online version of the article on *The Journal of Finance* website. We have read *The Journal of Finance* disclosure policy and have no conflict of interest to disclose. Emails: leeso@ceu.edu; wangchj@wharton.upenn.edu.

Over-the-counter (OTC) markets are opaque and difficult to access.[1] Yet, they continue to host the vast majority of trades in financial assets, despite regulators seeking to shift OTC trades onto exchanges for over a decade. The EU's MiFID II bans OTC dealers from matching or rerouting equities trades under most circumstances. The US Dodd-Frank Act requires standardized swap trading platforms to offer the option of trading the swaps on an exchange. Nonetheless, 20% of EU equities trades remain over the counter—the same share as before MiFID II (Exton and Healey, 2022)—as are nearly all trades of standardized swaps in the US (Riggs, Onur, Reiffen, and Zhu, 2020). Dealers argue against regulatory intervention, citing the high and persistent OTC market shares and evidence of lower trading costs compared to exchanges.[2] Do these high market shares and low trading costs indicate that having OTC markets improve welfare? Should policymakers restrict OTC trading, and under which conditions?

In this paper, we show that trading cost, trade volume, and market share can starkly mislead policy. Restricting OTC trading can widen average bid-ask spread and reduce aggregate trade volume, yet strictly improve welfare. Moreover, a restriction may, under certain conditions, improve welfare *exactly where* the OTC market share is *high*. In contrast to these measures, adverse selection risk provides robust guidance. We show that restrictions on OTC trading strictly raise traders' utilitarian welfare whenever adverse selection risk

[1] In our context, "over-the-counter (OTC) markets" consist of all financial markets in which trades are executed non-anonymously between a client and a dealer. This definition includes traditional voice markets in which clients contact dealers one by one, and request-for-quote (RFQ) markets in which clients contact multiple dealers at a time. "Exchanges" cover all other markets, including limit order books, batch auctions, dark pools, and all-to-all request-for-quote platforms.

[2] A typical example is the comment of Securities Industry and Financial Markets Association (2018) on a US regulatory proposal to limit the dealers' access to their counterparties' identities (called post-trade name give-up). SIFMA suggests regulatory intervention is unwarranted, because (i) the OTC trades have smaller bid-ask spreads (citing Riggs et al. (2020) and Collin-Dufresne, Junge, and Trolle (2020)) and (ii) the swap traders overwhelmingly trade over the counter despite being "free to choose" an exchange. de Roure, Moench, Pelizzon, and Schneider (2021) provides further evidence for (i).

is low, even if the OTC market were frictionless. Beyond trade restrictions, we devise an optimal Pigouvian tax that can be implemented using a simple sufficient statistic.

We develop a model of trade that adopts the choice between an exchange and an OTC market from Seppi (1990) and Desgranges and Foucault (2005). OTC dealers can price discriminate among their customers, whereas exchanges cannot. Hence, the dealers can offer a discount to the traders who are less likely to be informed and cream skim them from the exchange. We nest this intuition in the otherwise standard framework of Glosten and Milgrom (1985). Doing so adds one crucial feature for welfare analysis: endogenous participation of traders who differ in their gains from trade.

Section I describes our model. A continuum of traders trade an asset with an uncertain payoff. Uninformed traders have heterogeneous hedging benefits that incentivize them to trade. Informed traders receive signals about the asset payoff and seek profit. Whether a trader is informed is her private information. However, each trader is publicly labeled as either *Likely Informed (LI)* or *Likely Uninformed (LU)*, which imperfectly indicates her true type. All traders optimally choose to buy or sell on an exchange, with a dealer over the counter, or to exit. The venues differ solely in that only the dealer may condition his prices on each trader's label. In equilibrium, the (informed and uninformed) LI traders endogenously choose the exchange, the LU traders choose the OTC market, and the exchange spread is wider than the LU traders' OTC spread. An informed (LI or LU) trader always trades and an uninformed trader trades if her hedging benefit exceeds her best half bid-ask spread.

Section II examines how restricting the dealer's ability to price discriminate affects utilitarian welfare, aggregate trade volume, and average bid-ask spread. Specifically, we reduce the accuracy of the traders' labels, which further blends together the informed and the uninformed traders in the eyes of the OTC dealer. Restricting the dealer this way until the LI and the LU traders are equally likely to be informed is equivalent to closing the OTC market. We show that restricting the dealer strictly raises welfare if the ratio β of the mass of informed traders to the mass of uninformed traders is low and reduces welfare if β is high. Yet, the restriction can *always* decrease the aggregate volume and widen the average spread. In sum, (i) how a policy affects the aggregate volume and the average spread is a poor indicator for its effect on welfare, and (ii) restricting the OTC dealer improves welfare whenever adverse selection risk is low.

Result (i) is driven by "cheap substitution." The exchange spread always exceeds the OTC spread. Therefore, the marginal uninformed traders on the exchange have larger hedging benefits than the marginal uninformed traders in the OTC market. Restricting the OTC dealer pulls the two spreads towards each other, narrowing the exchange spread and widening the OTC spread. Thus, the marginal uninformed traders with the larger hedging benefits enter the exchange while those with the smaller hedging benefits exit the OTC market. In welfare terms, the entrants substitute for the comparably "cheap" exiters. Due to this cheap substitution, welfare can increase even if the exiters outnumber the entrants.

Figure 1 depicts a striking example of cheap substitution. We let $m(\text{entrants})$ and $\bar{b}(\text{entrants})$ denote the mass and the average hedging benefit of entrants upon restricting the OTC dealer, and define $m(\text{exiters})$ and $\bar{b}(\text{exiters})$ analogously. The restriction generates a gross welfare gain worth $\bar{b}(\text{entrants}) \times m(\text{entrants})$, and destroys $\bar{b}(\text{exiters}) \times m(\text{exiters})$. On net,

$$\text{welfare increases if and only if } \underbrace{\frac{\bar{b}(\text{entrants})}{\bar{b}(\text{exiters})}}_{\substack{\text{Cheap}\\ \text{Substitution}}} > \underbrace{\frac{m(\text{exiters})}{m(\text{entrants})}}_{\text{Volume Effect}}. \tag{1}$$

The ratio $\frac{\bar{b}(\text{entrants})}{\bar{b}(\text{exiters})}$ measures the beneficial impact of cheap substitution on welfare. In Figure 1, all informed traders are initially correctly labeled as Likely Informed, and hence trade on the exchange. We then restrict the dealer's ability to price discriminate by mislabeling a small mass of informed traders as Likely Uninformed. The resulting entrants have an average hedging benefit between the initial exchange spread S_{E} and the new wider spread S'_{E}, $\bar{b}(\text{entrants}) \in (S'_{\text{E}}, S_{\text{E}}) > 0$, whereas the average exiter hardly loses, $\bar{b}(\text{exiters}) \in (S_{\text{O}}, S'_{\text{O}}) \approx 0$. Thus, cheap substitution $\frac{\bar{b}(\text{entrants})}{\bar{b}(\text{exiters})} \approx \frac{S_{\text{E}}}{S_{\text{O}}}$ approaches infinity, and welfare strictly increases no matter how much the aggregate volume declines. Going beyond this example, cheap substitution $\frac{\bar{b}(\text{entrants})}{\bar{b}(\text{exiters})}$ is generally finite, and therefore the restriction can raise or lower welfare.

Result (ii), that restricting the OTC dealer raises welfare when adverse selection risk is low, follows because cheap substitution persists as the risk becomes small, while the effect on volume vanishes. Our results are remarkably robust. They hold under any commonly used

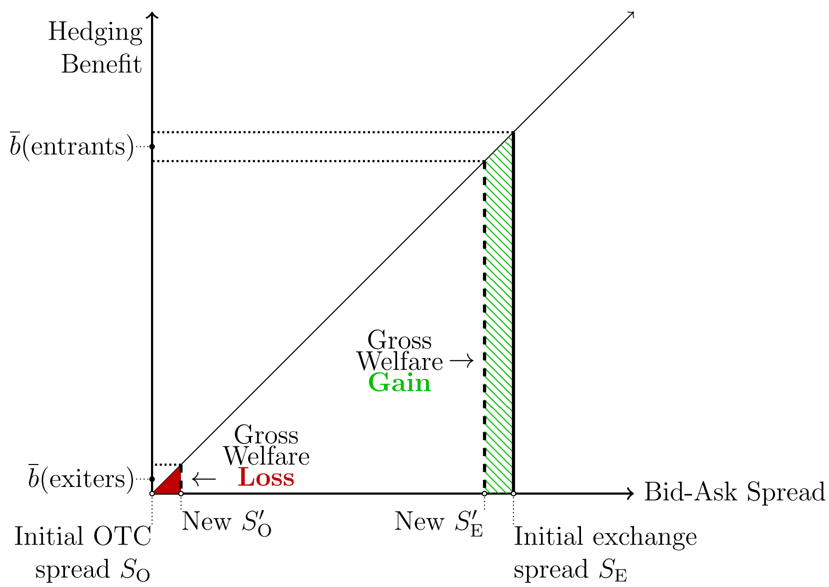


Figure 1: Restricting the OTC dealer can raise welfare

Initially, all informed traders are correctly labeled as Likely Informed. Then, we mislabel a small mass of them as Likely Uninformed. The average hedging benefit of the resulting entrants is $\bar{b}$(entrants) and of the exiters is $\bar{b}$(exiters).

distribution, for any change in the accuracy of traders' labels, and whether the informed traders have hedging benefits themselves (Internet Appendix IA.C).

Dealers frequently invoke the high market share of OTC markets to oppose regulatory intervention. Exchanges are indeed available for many OTC-dominated assets.[3] Perhaps, then, the high OTC market share is a competitive equilibrium outcome, which may well be socially optimal. Section II.C shows that a high OTC market share is *not* evidence against intervention. A higher ratio of informed to uninformed traders β mechanically reduces the share of Likely Uninformed traders, who choose the OTC market. Therefore, β and the OTC market share are negatively related under many distributions. Proposition 3 combines this negative relationship with our main result: When β drives the variation in the OTC market share, restricting or closing the OTC market *strictly improves* welfare if the OTC market share is sufficiently *high*. Thus, high OTC market share is not evidence against intervention; it is not even evidence for keeping the OTC market.

Venue choice generates the key externality underlying our results. The traders who opt to trade over the counter do not internalize that their choice worsens the spread on the exchange. Section III devises a Pigouvian tax that optimally penalizes this externality. A simple policy experiment suffices to implement the tax. First, one imposes a small tax T on OTC trades. Second, one computes the Weighted Spread Ratio $WSR \coloneqq |S_{\mathrm{E}}/S_{\mathrm{O}} \div (dV_{\mathrm{O}}/dV_{\mathrm{E}})|$. The ratio of spreads $S_{\mathrm{E}}/S_{\mathrm{O}}$ captures the cheap substitution effect, and the ratio of changes in trade volumes $|dV_{\mathrm{O}}/dV_{\mathrm{E}}|$ captures the volume effect. If the WSR exceeds 1, cheap substitution dominates, and raising the tax T would strictly improve welfare. If the WSR is below 1, the decline in aggregate volume dominates, and cutting the tax T would strictly improve welfare.

Our model generates two testable predictions. First, the OTC spread is narrower than

[3]Examples of exchanges for mostly OTC-traded assets include NYSE Bonds for corporate bonds, Saxo Bank SaxoTrader for EU government bonds, Tradeweb Dealerweb for repos, Refinitiv FXall for foreign exchange, and Bloomberg SEF CLOB and GFI Swaps Exchange for swaps. All these exchanges are open to any buy-side trading firm. In addition, several OTC trading platforms allow clients to anonymously request quotes (e.g., Open Trading on MarketAxess), which fall under "trading on exchange" in our definition.

the exchange spread (Proposition 0). Internet Appendix IA.E presents supporting evidence. Second, the exchange's market share and spread are positively correlated through the informed ratio β, because increasing the share of informed traders both raises the exchange market share and widens its spread. Internet Appendix IA.F documents a novel empirical pattern: The total market share of exchanges and their quoted spreads are positively correlated across US equities.

Whether price discrimination in OTC markets is socially beneficial is an increasingly relevant question. Beyond Dodd-Frank and MiFID II, several proposals to implement blockchain technology for recordkeeping of OTC trades are under consideration. Most such proposals would reveal traders' identities to selected dealers. Internet Appendix IA.A discusses these proposals through the lens of our model.

Section IV compares our mechanism to those in the literature. We make three contributions. (i) We introduce cheap substitution, a new mechanism that can overturn the effects of aggregate volume and average spread on welfare. (ii) We contribute the novel result that pooling improves welfare whenever adverse selection risk is low. (iii) We devise a simple sufficient statistic with which one can implement the optimal Pigouvian tax.

The remainder of the paper proceeds as follows. Section I describes our model. Section II states and explains our main results. Section III devises the optimal Pigouvian tax on OTC trades. Section IV compares our mechanism and results to existing work. Section V concludes with a discussion of frictions that are not captured by our model.

I. A Model of Venue Choice and Price Discrimination

Section I.A sets up a model in which each trader may trade over the counter, on an exchange, or exit. The OTC dealer can price discriminate across the traders' public labels, whereas the exchange dealer cannot. Section I.B interprets our assumptions. Section I.C derives the unique equilibrium.

A. *Trading Game*

A continuum of risk-neutral traders may trade an indivisible asset in a three-stage game. The asset is equally likely to pay $v = 1$ or -1 in the third stage. Each trader either exits without trading, or buys or sells 1 unit in one of two markets. In the OTC market, a dealer acts as the counterparty to the traders and absorbs net demand. Another dealer does so on the exchange. The OTC and the exchange dealers set prices such that their expected profit is zero, as we detail below.

A mass μ of the traders are informed and a mass 1 are uninformed. An informed trader has a private binary signal, which equals the true value v with probability $\alpha \in (1/2, 1)$ and $-v$ otherwise. Probability α is the accuracy of the informed traders' signals. Each uninformed trader is equally likely to be a buyer or a seller, and obtains a hedging benefit b_i upon trading in her desired direction. The hedging benefits are independently drawn from a distribution F, $b_i \overset{\text{iid}}{\sim} F$, with support $[0, 1]$. An uninformed trader's realized hedging benefit b_i and whether she is a buyer or a seller are her private information.

An informed trader is labeled $\ell_i = \text{LI}$ ("Likely Informed") with probability θ and LU ("Likely Uninformed") otherwise. An uninformed trader is LU with probability γ and LI otherwise. Hence, there are $\theta\mu$ informed LI traders and γ uninformed LU traders. We assume that the LI traders are more likely to be informed than the LU traders, $\theta > 1 - \gamma$, and that the labels are not perfectly revealing, $\theta < 1$ or $\gamma < 1$.

In Stage 1, the OTC dealer posts a bid to buy and an ask to sell one unit of the asset to every trader i. The OTC dealer's quote is the highest bid and the lowest ask that earn him zero expected profit conditional on the trader's label $\ell_i \in \{\text{LI}, \text{LU}\}$. Simultaneously, the exchange dealer posts the highest bid and the lowest ask that unconditionally earn him zero expected profit (as in Glosten and Milgrom (1985)). That is, the OTC market differs from the exchange in one way—the OTC dealer observes the label ℓ_i before setting the prices for trader i.

In Stage 2, every trader observes all prices, then makes two decisions: *whether* to buy, sell, or exit, and *where* to trade. Figure 2 summarizes the timing of our model. All distributions, parameters, and the structure of the game are common knowledge.

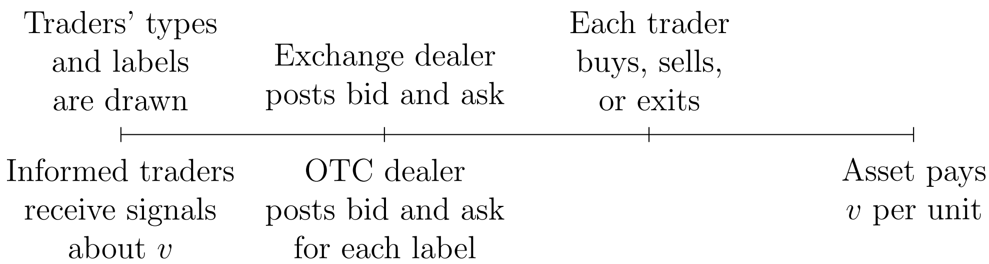


Figure 2: Timeline

We impose a regularity condition on the distribution F of hedging benefits.

Assumption 1. The hedging benefit distribution F has a pdf $f(s)$ that is analytic in some neighborhoods of 0 and $2\alpha - 1$.[4]

Assumption 1 precludes distributions whose pdfs oscillate between extreme values around the lower and upper bounds of the equilibrium bid-ask spread (which we derive in Section I.C). In practice, any commonly used distribution in economics satisfies Assumption 1.

We impose a tie-breaking rule to pin down a unique equilibrium.

Assumption 2. If a trader is indifferent between trading over the counter or on the exchange, she trades on the exchange.

Assumption 2 is purely expositional.[5] The rule is equivalent to imposing a small cost on OTC trades, which can represent the inconvenience of soliciting prices that is absent on exchanges.

[4]A function f is analytic at some x_0 if f is equal to its Taylor series at x_0 in a neighborhood of x_0,

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n \qquad \forall x \in (x_0 - \varepsilon, x_0 + \varepsilon) \text{ for some } \varepsilon > 0.$$

Analytic functions include the pdf of any beta distribution, all polynomials, the exponential function, all trigonometric functions, logarithms, and the power function, among many others.

[5]All equilibria that would exist without Assumption 2 are payoff equivalent. We can eliminate Assumption 2 and show that traders who are indifferent between trading on the exchange or in the OTC market must choose the exchange with some probability ρ. Our results only require that $\rho > 0$.

B. Interpretation

Prices. Our setup features competitive prices, as defined in Glosten and Milgrom (1985). That prices on exchanges are competitive is a good proxy of reality. However, search frictions and the dealers' market power limit price competition in OTC markets. We nonetheless assume competitive prices to show that, even when the OTC market is made artificially efficient, restricting or taxing it can still improve welfare. Introducing search frictions or market power would further raise the appeal of restricting the OTC dealer's ability to price discriminate.

Trading protocol. The OTC dealer spontaneously posts label-dependent prices in the model. In practice, an OTC trade occurs in two steps. A trader first requests quotes from her dealers, and then those dealers respond with trader-specific quotes. Internet Appendix IA.B presents an extension that incorporates request for quote and endogenizes the competition among dealers. Every trader requests quotes from all dealers in any extensive-form trembling-hand perfect equilibrium (Selten, 1975). The equilibrium allocation is identical to that of the base model described in the next section. We let the OTC dealer post label-dependent prices to simplify exposition.

Traders' labels. One can interpret a trader's label as a summary statistic of her reputation and observable characteristics. Such observables may include the trader's industry (e.g., hedge fund or insurer), marketing or public filings (e.g., active versus passive fund), name (e.g., "Two-Sigma" or "AIG"), and any other public fact that is informative about the trader's motive. We assume imperfectly informative labels, because reputation and observables are noisy signals about the true motive behind a trade. This assumption is consistent with the evidence in Cheng and Xiong (2014) from the commodities futures market. The US Commodity Futures Trading Commission labels traders as "hedgers" if they are commodity producers and their past trades are not consistently profitable. Cheng and Xiong (2014) finds that the hedgers' positions are far more volatile than their output, especially their short positions. These short positions are consistently profitable and uncorrelated with output, which suggests that the so-called hedgers sometimes trade for profit.

Mass of traders. We fix the mass of uninformed traders at 1 and vary the mass of informed traders μ in the welfare analysis. This normalization ensures that the maximum welfare that can be achieved is fixed and equal to the total hedging benefit of all uninformed traders.

C. Equilibrium

Definition 1. An equilibrium consists of the OTC dealer's bid-ask spread for the LI traders, his spread for the LU traders, the exchange dealer's spread, and the traders' venue choice and trading strategies. Each trader chooses the market that offers the lowest spread available to her, and her trading strategy maximizes her expected payoff given this spread. Given the traders' strategies, the OTC dealer sets each of his spreads to the lowest spread that earns zero profit in expectation conditional on the label, and the exchange dealer sets his spread to the lowest spread that earns him zero profit in expectation unconditionally.

An equilibrium bid is the negative of the corresponding ask, because the asset value v is symmetric around zero and the uninformed traders are equally likely to be a buyer or a seller. Thus, each pair of equilibrium bid and ask prices can be expressed as the half bid-ask spread s. For brevity, we hereon write "spread" to mean "half bid-ask spread."

The trader's equilibrium strategies are cutoff rules. An uninformed buyer or seller trades 1 unit in her corresponding direction at her smallest available spread if that spread is smaller than her hedging benefit, and exits otherwise. An informed trader buys or sells in the direction of her signal if her smallest spread is below $2\alpha - 1$, the expected value of this long or short position conditional on her signal.

We now pin down equilibrium spreads. If a dealer charges a spread s, an uninformed trader who chooses the dealer trades with probability $1 - F(s)$. Conditional on such a trade, the dealer's expected profit is s. On the other hand, the dealer's expected loss per trade with an informed trader is $(2\alpha - 1 - s)^+$. Altogether, the dealer earns zero profit in expectation if and only if

$$\underbrace{s \cdot (1 - F(s))}_{\substack{\text{Profit from} \\ \text{uninformed traders}}} = \underbrace{(2\alpha - 1 - s)^+ \cdot \beta}_{\text{Loss to informed traders}}, \tag{2}$$

where the *informed ratio* β is the mass of informed traders who choose the dealer per unit mass of uninformed traders who choose the same. The zero-profit condition (2) has a unique solution, which we denote $S(\beta)$.

Proposition 0. *(a) Without the OTC market, the spread on the exchange is the No-OTC spread $S_N = S(\mu)$. (b) With the OTC market, the exchange spread is $S_E = S\left(\frac{\theta}{1-\gamma}\mu\right)$, and the OTC spreads for LU and LI traders, respectively, are $S_O = S\left(\frac{1-\theta}{\gamma}\mu\right)$ and S_E. Every LI trader chooses the exchange and receives S_E, and every LU trader chooses the OTC market and receives S_O. (c) The exchange spread is strictly wider than the OTC spread and the No-OTC spread is strictly between the two, $S_E > S_N > S_O$.*

Part (a) is a standard result of Glosten and Milgrom (1985). Part (b) incorporates venue choice: The LI traders choose the exchange, and the LU traders choose the OTC market in equilibrium. Intuitively, traders of the same label choose the same market because, if they split, the prices they face in the two markets must be equal. Then, all traders with that label would choose the exchange by Assumption 2. Hence, the LU traders choose the OTC market whose spread is lower for them than the exchange spread, whereas the LI traders are indifferent between the two markets and choose the exchange due to Assumption 2. Our results remain unchanged if the LI traders choose the exchange with a strictly positive probability. Part (c) immediately follows from (2) and part (b). Internet Appendix IA.E cites recent evidence that trading costs are lower over the counter than on exchanges.

II. Welfare, Volume, and Spread

We analyze utilitarian welfare, aggregate trade volume, and average bid-ask spread upon restricting the OTC dealer's ability to price discriminate. To restrict the OTC dealer, we either lower the accuracy of the traders' labels, θ or γ, or close the OTC market altogether. The closure is equivalent to lowering θ or γ until $\theta = 1 - \gamma$, where the LU and the LI traders are equally likely to be informed.

Section II.A defines key terms. Section II.B pinpoints the general conditions under which the restriction raises or lowers welfare. Section II.C explains how adverse selection risk jointly determines market shares and welfare.

A. Definitions

Welfare W is the sum of all agents' *ex-ante* payoffs. It measures the total gains from trade in our model. Precisely, welfare W equals the sum of the hedging benefits that uninformed traders obtain through trade.

We let V_{O} denote the equilibrium volume of OTC trades, V_{E} the volume of trades on the exchange, and $V := V_{\mathrm{O}} + V_{\mathrm{E}}$ the aggregate trade volume.[6] *Average bid-ask spread* $\bar{S}$ is the volume-weighted average of bid-ask spreads in the OTC market and on the exchange,

$$\bar{S} := \frac{V_{\mathrm{O}}}{V} S_{\mathrm{O}} + \frac{V_{\mathrm{E}}}{V} S_{\mathrm{E}}.$$

We say "lower average spread $\bar{S}$" interchangeably with "higher aggregate volume V," because spread $\bar{S} \propto 1/V$ in equilibrium.[7]

Two statistics characterize the effects of restricting the OTC dealer on welfare and volume. *Marginal volume* Δ_V is the decrease $m(\text{exiters})$ in the mass of uninformed trades upon a marginal increase in the ratio of informed to uninformed traders β:

$$\Delta_V(\beta) := -\left(\int_{S(\beta)}^{1} f(s)\,\mathrm{d}s\right)' = \underbrace{S'(\beta) \cdot\ f(S(\beta))}_{\frac{\mathrm{d}m(\text{exiters})}{\mathrm{d}\beta}}. \tag{3}$$

Marginal welfare Δ_W is the decrease in welfare, $m(\text{exiters}) \times \bar{b}(\text{exiters})$, upon a marginal

[6] Explicitly, $V_{\mathrm{O}} = (1-\theta)\mu + \gamma \cdot (1 - S_{\mathrm{O}})$ and $V_{\mathrm{E}} = \theta\mu + (1-\gamma)\cdot(1 - S_{\mathrm{E}})$.

[7] Formally, the zero-profit condition (2) implies

$$\underbrace{V \cdot \bar{S}}_{\text{Revenue}} = \underbrace{(2\alpha - 1)\cdot \mu}_{\text{Gross loss}},$$

which yields the inverse relationship between $\bar{S}$ and V.

increase in β. It is equivalent to $\Delta_V(\beta)$ times the marginal exiters' hedging benefit $\bar{b}$(exiters):

$$\Delta_W(\beta) := -\left(\int_{S(\beta)}^{1} sf(s)\,\mathrm{d}s\right)' = \underbrace{S'(\beta)f(S(\beta))}_{\Delta_V(\beta)} \cdot \underbrace{S(\beta)}_{\bar{b}(\text{exiters})}. \tag{4}$$

We show in Appendix A.2 that both Δ_V and Δ_W are well-defined.

B. *Main Results*

The next proposition states the effects of restricting the OTC dealer on welfare, aggregate trade volume, and average bid-ask spread. Proofs are in Appendix A.

Proposition 1. *Given any α, we let each uninformed trader's hedging benefit $b_i \overset{iid}{\sim} F$ and its pdf f be analytic in some neighborhoods of 0 and $(2\alpha - 1)$ with support $[0, 1]$. Given any pairs $(\theta_l, \gamma_l) < (\theta_h, \gamma_h) < (1, 1)$ that satisfy $\theta_h + \gamma_l > 1$ and $\theta_l + \gamma_h > 1$, there exists cutoffs $\mu'_h > \mu'_l$, $\tilde{\mu}'_h > \tilde{\mu}'_l$, $\mu_h > \mu_l$, $\tilde{\mu}_h > \tilde{\mu}_l$, and $\bar{\mu} > \underline{\mu}$, all strictly positive, such that:*

- *(a) Marginally lowering the traders' label accuracy θ from θ_h to $\theta_h - d\theta$ strictly raises welfare W for all mass of informed traders $\mu < \mu'_l$ and strictly lowers W for all $\mu > \mu'_h$; marginally lowering γ from γ_h to $\gamma_h - d\gamma$ strictly raises W for all $\mu < \tilde{\mu}'_l$ and strictly lowers W for all $\mu > \tilde{\mu}'_h$.*

- *(b) Lowering the label accuracy θ from θ_h to θ_l strictly raises W for all $\mu < \mu_l$ and strictly lowers W for all $\mu > \mu_h$; lowering γ from γ_h to γ_l strictly raises W for all $\mu < \tilde{\mu}_l$ and strictly lowers W for all $\mu > \tilde{\mu}_h$.*

- *(c) Closing the OTC market strictly raises W for all $\mu < \underline{\mu}$ and strictly lowers W for all $\mu > \bar{\mu}$.*

Moreover:

- *(d) If the marginal welfare Δ_W (defined in (4)) is strictly quasiconcave, the cutoffs $\mu'_h = \mu'_l$, $\tilde{\mu}'_h = \tilde{\mu}'_l$, $\mu_h = \mu_l$, $\tilde{\mu}_h = \tilde{\mu}_l$, and $\bar{\mu} = \underline{\mu}$.*[8]

(e) *If the marginal volume Δ_V (defined in (3)) is strictly decreasing, lowering θ or γ (marginally or otherwise) or closing the OTC market strictly reduces the aggregate trade volume V and strictly widens the average bid-ask spread $\bar{S}$ for all $\mu > 0$.*

Proposition 1 has two messages. First, restricting the OTC dealer improves welfare if the mass of informed traders μ is small and harms welfare if μ is large. Second, in stark contrast, these interventions can *always* worsen the aggregate trade volume and the average bid-ask spread.

Precisely, Proposition 1 parts (a)–(c) say that, under any commonly used distribution, lowering label accuracy or closing the OTC market raises welfare if μ is sufficiently small and reduces welfare if μ is sufficiently large. Part (d) sharpens this result to a single cutoff on μ for distributions whose marginal welfares Δ_W are quasiconcave. Part (e) says that the restrictions reduce the aggregate volume and widen the average spread—no matter how they affect welfare—whenever the marginal volume Δ_V is decreasing. Thus, the aggregate measures of volume and spread are poor indicators of welfare. Proposition 2 below states the conditions on model primitives that are necessary and sufficient for parts (d) and (e), and lists broad classes of distributions that satisfy each condition.

Figure 3 illustrates Proposition 1 under the uniform distribution, $F = \mathbb{U}[0,1]$. The initial parameters are $\theta = \gamma = 0.9$ and $\alpha = 0.98$. We plot the changes in welfare W, the aggregate volume V, and the average spread $\bar{S}$ upon lowering the label accuracy θ to 0.6 as the proportion of informed traders $\mu/(1+\mu)$ varies. The changes are positive above the red line. Figure 3 confirms that restricting the OTC dealer raises welfare where μ is low, while it always reduces the aggregate volume and widens the average spread in this case. Adding search frictions or relaxing competitive prices in the OTC market would only expand the range of parameters under which the restriction raises welfare.

Intuition

Marginally lowering the label accuracy θ by $\mathrm{d}\theta$ raises welfare when the mass μ of informed traders is low, because cheap substitution persists yet the volume effect vanishes. (The

[8] A strictly quasiconcave Δ_W is strictly increasing up to a cutoff, then strictly decreasing thereafter. If Δ_W is weakly quasiconcave, Proposition 1 remains the same except that "strictly" becomes "weakly."

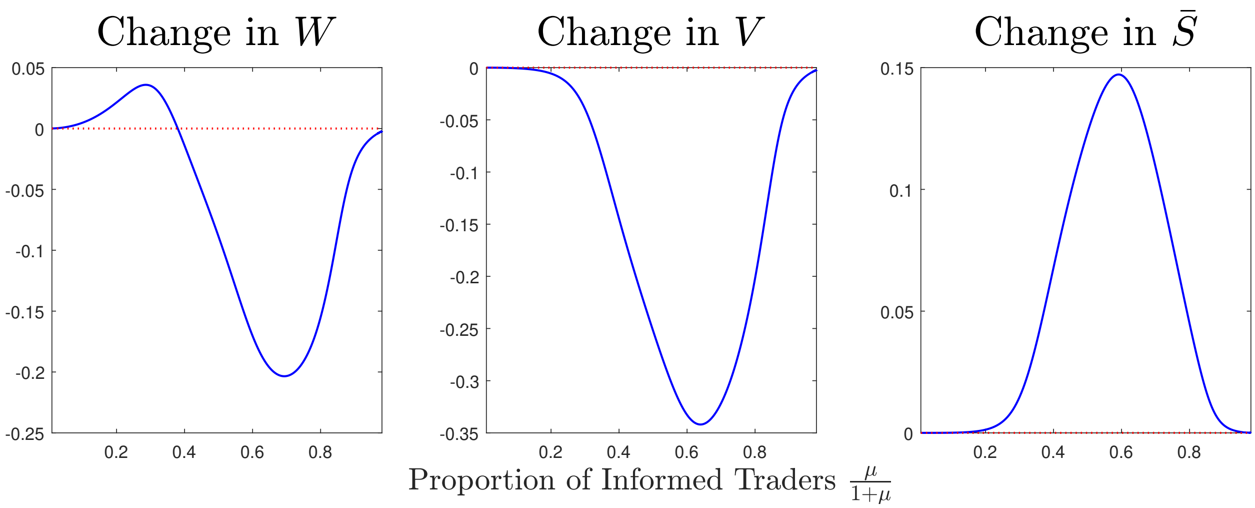


Figure 3: Effects of Restricting the OTC Dealer's Ability to Price Discriminate

intuition is analogous for marginally lowering γ, non-marginally lowering θ or γ, and closing the OTC market.) The key to the intuition is that cheap substitution depends on the relative *levels* of the spreads, while the volume effect depends on the relative *changes* in the spreads of the two markets.

To see that cheap substitution persists when μ is small, we note that cheap substitution equals $\bar{b}(\text{entrants})/\bar{b}(\text{exiters}) = S_{\text{E}}/S_{\text{O}}$. Here, the two spreads S_{E} and S_{O} are each proportional to the informed ratio β in their respective markets.[9] These β are μ scaled by a constant $\frac{\theta}{1-\gamma}$ on the exchange and by $\frac{1-\theta}{\gamma}$ in the OTC market. The constant is strictly larger on the exchange, because the uninformed traders are more likely to be labeled LU than the informed traders, $\gamma > 1-\theta$. Therefore, cheap substitution $S_{\text{E}}/S_{\text{O}}$ is bounded away from 1 as μ becomes small.

The volume effect $m(\text{exiters})/m(\text{entrants})$ vanishes as μ becomes small, because each dealer's adverse selection cost (gross loss to the informed traders) equals his gross revenue $V_{\text{M}}S_{\text{M}}$ in his market M. For small μ, V_{M} is approximately equal to the mass of uninformed traders who choose market M whether we lower θ or not. Lowering θ moves a

[9] Precisely, $S_{\text{M}} \sim S'(0)\beta_{\text{M}}$, for $\text{M} \in \{\text{E}, \text{O}\}$ as $\mu \to 0$, where $\beta_{\text{E}} = \frac{\theta}{1-\gamma}\mu$, and $\beta_{\text{O}} = \frac{1-\theta}{\gamma}\mu$.

fraction of informed traders from the exchange to the OTC market,[10] which transfers the corresponding adverse selection cost from the exchange to the OTC dealer. The resulting change in the dealer's gross loss $V_{\mathrm{M}}\, S_{\mathrm{M}}$ is approximately the same in the two markets and equal to $V_{\mathrm{M}}\, \mathrm{d}S_{\mathrm{M}}$. Therefore, the masses of exiters and entrants are about equal at small μ: $m(\text{exiters}) = \gamma|\mathrm{d}S_{\mathrm{O}}| \approx V_{\mathrm{O}}|\mathrm{d}S_{\mathrm{O}}| \approx V_{\mathrm{E}}|\mathrm{d}S_{\mathrm{E}}| \approx (1-\gamma)\mathrm{d}S_{\mathrm{E}} = m(\text{entrants})$.[11] The volume effect $m(\text{exiters})/m(\text{entrants})$ is thus close to 1.

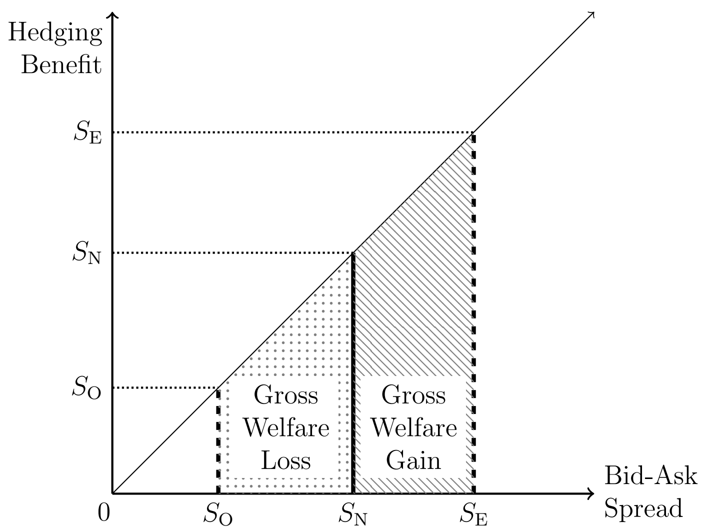


Figure 4: Effect of closing the OTC market on welfare

We mark the initial bid-ask spreads S_{O} and S_{E}, and the No-OTC spread S_{N} that prevails after the closure.

Equivalence between label accuracy and closure

Closing the OTC market converges all traders' bid-ask spreads to the No-OTC spread S_{N}. Hence, the closure is equivalent to lowering the traders' label accuracy until the LI and the

[10] As θ decreases, the number of informed traders in the exchange $\theta\mu$ decreases and the number of informed traders in the OTC market $(1-\theta)\mu$ increases.

[11] This sentence uses the case of the uniform distribution. For analytical distributions whose $\lim_{x\to 0} f'(x) > 0$, the approximation is $m(\text{exiters}) = \lim_{x\to 0} f(x)\gamma|\mathrm{d}S_{\mathrm{O}}| \approx \lim_{x\to 0} f(x)V_{\mathrm{O}}|\mathrm{d}S_{\mathrm{O}}| \approx \lim_{x\to 0} f(x)V_{\mathrm{E}}|\mathrm{d}S_{\mathrm{E}}| \approx \lim_{x\to 0} f(x)(1-\gamma)\mathrm{d}S_{\mathrm{E}} = m(\text{entrants})$. If the distribution has $\lim_{x\to 0} f'(x) = 0$ or ∞, the intuition is more involved. Our proof in Appendix A.2 allows all such cases.

LU traders are equally likely to be informed, $\theta = 1-\gamma$. Figure 4 shows $\bar{b}(\text{entrants}) \in (S_{\mathrm{N}}, S_{\mathrm{E}})$ and $\bar{b}(\text{exiters}) \in (S_{\mathrm{O}}, S_{\mathrm{N}})$ such that every entrant's hedging benefit is greater than any exiter's benefit.

Economic relevance

Most commonly used distributions have strictly quasiconcave marginal welfares Δ_W. Proposition 2 states the necessary and sufficient conditions on the pdf f for its Δ_W to be strictly quasiconcave and its marginal volume Δ_V to be decreasing.

Proposition 2. *(i) Marginal welfare Δ_W is strictly quasiconcave if and only if the function*

$$\frac{(2\alpha-1)(1-F(x))}{xf(x)(2\alpha-1-x)^2} - \frac{1}{2\alpha-1-x} \quad \textit{is strictly quasiconvex in } x \in (0, 2\alpha-1). \tag{5}$$

Any beta distribution Beta(a,b) for all $a, b > 0$ satisfies condition (5).

(ii) Marginal volume Δ_V is decreasing if and only if the function

$$\frac{(2\alpha-1)(1-F(x))}{f(x)(2\alpha-1-x)^2} - \frac{x}{2\alpha-1-x} \quad \textit{is increasing in } x \in (0, 2\alpha-1). \tag{6}$$

Any beta distribution Beta(a,b) for all $a \leq 1, b \leq 1$ satisfies condition (6).

Taylor expansions around $2\alpha-1$ verify that the uniform distribution $\mathbb{U}[0,1]$ and any beta distribution—for all parameter values—satisfy condition (5) and thereby have quasiconcave marginal welfare Δ_W.[12] To see what distributions are excluded by (5), we note that (5) holds at the extreme ends under *any* pdf f analytic in some neighborhoods of 0 and $2\alpha - 1$: The expression in (5) approaches infinity at the limit as $x \to 0$ and as $x \to 2\alpha-1$. Condition (5) only excludes pdfs that oscillate over moderate values in the support $(0, 2\alpha-1)$.

We further verify that (6) is satisfied for any uniform distribution and for beta distributions Beta(a,b) with parameters $a \leq 1, b \leq 1$. That a wide range of distributions satisfy

[12]The beta distribution Beta(a,b) (pdf $f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$) is a general class of bounded distributions that embeds the uniform distribution when $a = b = 1$. We can numerically verify that common unbounded distributions—normal, chi-squared, and gamma—satisfy condition (5) when truncated to $[0,1]$.

(6) shows how the aggregate volume and the average bid-ask spread are poor indicators of policies' impact on welfare.

C. Market Shares

Our results so far link welfare to adverse selection risk. Because regulatory debates often cite market shares, we analyze whether and how the OTC market share relates to welfare. Proposition 3 formalizes one message: Restricting the OTC dealer can strictly raise welfare specifically where the OTC market share is high.

Proposition 3. *Given any α, we let F and its pdf f have support $[0,1]$ and be such that*

$$x \cdot \left(\frac{1}{2\alpha - 1 - x} - \frac{f(x)}{1 - F(x)} \right) \quad \text{is strictly increasing in} \quad x \in (0, 2\alpha - 1),^{13} \tag{7}$$

and each uninformed trader's hedging benefit $b_i \overset{iid}{\sim} F$. Given any pairs $(\theta_l, \gamma_l) < (\theta_h, \gamma_h) < (1,1)$ that satisfy $\theta_h + \gamma_l > 1$ and $\theta_l + \gamma_h > 1$, there exist cutoffs $M, \widehat{M}, M^* > 0$ such that:

(a) OTC market share $\frac{V_O}{V}$ is strictly decreasing in μ.

(b) Lowering the label accuracy θ from θ_h to θ_l strictly raises welfare W for all $\frac{V_O}{V} > M$ and strictly reduces W for all $\frac{V_O}{V} < M$; lowering γ from γ_h to γ_l strictly raises welfare W for all $\frac{V_O}{V} > \widehat{M}$ and strictly reduces W for all $\frac{V_O}{V} < \widehat{M}$.

(c) Closing the OTC market strictly raises W for all $\frac{V_O}{V} > M^*$ and strictly reduces W for all $\frac{V_O}{V} < M^*$.

Intuitively, a larger μ mechanically raises the share of LI traders, who trade on the exchange. Under condition (7), this mechanical effect causes the OTC market share $\frac{V_O}{V}$ to strictly decrease in μ (part (a)). Parts (b) and (c) are corollaries of part (a) and Proposition 1:

[13]The uniform distribution, $F = \mathbb{U}[0,1]$, trivially meets (7). We numerically verify that, given any α, the beta distribution satisfies (7) for wide ranges of its parameters $a, b > 0$.

Holding all parameters constant except for μ, restricting the OTC dealer would raise welfare *if and only if* its market share $\frac{V_{\mathrm{O}}}{V}$ is sufficiently high.

Raising μ, all else equal, simultaneously reduces $\frac{V_{\mathrm{O}}}{V}$ and widens the exchange spread S_{E}. Stated empirically, adverse selection risk induces a positive correlation between the exchange spread S_{E} and its market share $\frac{V_{\mathrm{E}}}{V}$. We document a positive correlation between quoted spreads on exchanges and their market share across US-listed equities (Internet Appendix IA.F.3).

III. Optimal Pigouvian Tax

Section II provides policy guidance where one knows that adverse selection risk is low (restrict the OTC dealer) or high (keep as is). We turn to an optimal Pigouvian tax that can be implemented without knowing the severity of adverse selection risk. Instead, a simple sufficient statistic derived from a local experiment can determine whether the current tax is too low or too high. Computing this statistic, the Weighted Spread Ratio (WSR), requires only trade volumes and bid-ask spreads.

A. Characterizing the Optimal Pigouvian Tax

Gross tax revenue T is transferred from the OTC dealer to the exchange dealer.[14] The lump-sum tax T is common knowledge from the beginning of the game. We fix all parameters $\{\mu, \theta, \gamma, \alpha\}$ except the tax T.

All implementations of a Pigouvian tax levied on the dealer that raise the same gross revenue T are equivalent to each other, as they all lead to the same zero-profit conditions.[15]

[14]Taxing the exchange dealer is ineffective, as every LI trader would choose the OTC market and no trade would occur on the exchange.

[15]Examples include flat per-trade transaction tax, lump-sum tax, and non-linear levy on OTC volume. The same is true with multiple dealers in each market as long as the tax T is split between the dealers in proportion to their trade volumes.

Precisely, the OTC dealer's condition becomes

$$S_{\mathrm{O}}(T) \cdot \underbrace{\left[1 - F\left(S_{\mathrm{O}}(T)\right) + \beta_{\mathrm{O}}\right]\gamma}_{V_{\mathrm{O}}(T)} = (2\alpha - 1)\beta_{\mathrm{O}}\gamma + T, \tag{8}$$

and the exchange dealer's condition becomes

$$S_{\mathrm{E}}(T) \cdot \underbrace{\left[1 - F(S_{\mathrm{E}}(T)) + \beta_{\mathrm{E}}\right] \cdot (1 - \gamma)}_{V_{\mathrm{E}}(T)} = (2\alpha - 1)\beta_{\mathrm{E}} \cdot (1 - \gamma) - T. \tag{9}$$

Increasing tax T raises OTC spread $S_{\mathrm{O}}(T)$ and lowers exchange spread $S_{\mathrm{E}}(T)$.

We define the *optimal tax* as follows.

$$\text{The } \textit{optimal Pigouvian tax } T^* \text{ is the lump-sum tax } T \text{ that maximizes welfare } W. \tag{10}$$

Levying this optimal tax T^* attains a strictly higher welfare than closing the OTC market.

Proposition 4. *Given any triple (α, θ, γ): (a) Imposing the optimal tax T^* attains a strictly higher welfare W than closing the OTC market. (b) If $F = \mathbb{U}[0, 1]$, the optimal Pigouvian tax T^* (defined in (10)) is unique.*

The optimal Pigouvian tax *always* outperforms closing the OTC market. This is because cheap substitution vanishes under the extreme tax that effectively closes the OTC market. We define this *Zero-OTC* tax $\overline{T}$ as follows:

Zero-OTC tax $\overline{T}$ is the smallest tax such that the OTC trade volume is zero.

At the Zero-OTC tax $\overline{T}$, each trader is indifferent between trading over the counter or on the exchange. In particular,

$$S_{\mathrm{O}}(\overline{T}) = S_{\mathrm{E}}(\overline{T}).$$

Thereby, the Zero-OTC tax $\overline{T}$ equalizes the hedging benefits of the marginal uninformed traders in both markets. A marginal tax cut from $\overline{T}$ would create entrants and exiters with the same hedging benefits, eliminating cheap substitution. Given that a tax cut from $\overline{T}$ strictly increases the aggregate trade volume with the uniform distribution, the optimal tax here is strictly below the Zero-OTC tax, $T^* < \overline{T}$.

B. *Implementing the Optimal Pigouvian Tax*

We use the trade-off between the cheap substitution versus the volume effect (defined in (1)) to develop a simple precedure to implement the optimal Pigouvian tax T^*. Marginally increasing the tax T raises welfare W if and only if

$$\underbrace{\left|\frac{S_{\mathrm{E}}}{S_{\mathrm{O}}}\right|}_{\substack{\text{Cheap}\\ \text{Substitution}}} > \underbrace{\left|\frac{dV_{\mathrm{O}}}{dV_{\mathrm{E}}}\right|}_{\substack{\text{Volume}\\ \text{Effect}}}.$$

The optimal tax T^*, where $T^* > 0$, equalizes cheap substitution and the volume effect:[16]

$$\frac{S_{\mathrm{E}}(T^*)}{S_{\mathrm{O}}(T^*)} = \left|\frac{\mathrm{d}V_{\mathrm{O}}(T^*)}{\mathrm{d}V_{\mathrm{E}}(T^*)}\right|.$$

We define the ratio of cheap substitution to the volume effect as the *weighted spread ratio*:

$$\mathrm{WSR}(T) \coloneqq \left|\frac{S_{\mathrm{E}}(T)dV_{\mathrm{E}}(T)}{S_{\mathrm{O}}(T)dV_{\mathrm{O}}(T)}\right|. \tag{11}$$

[16] We differentiate (8) and (9) to compute $\mathrm{d}V_{\mathrm{O}}$ and $\mathrm{d}V_{\mathrm{E}}$. Doing so, we obtain

$$|\mathrm{d}V_{\mathrm{O}}(t)| = |\gamma S'_{\mathrm{O}}(t)|\mathrm{d}t = \frac{\mathrm{d}t}{1 - 2S_{\mathrm{O}}(t) + \beta_{\mathrm{O}}}$$
$$|\mathrm{d}V_{\mathrm{E}}(t)| = |(1-\gamma)S'_{\mathrm{E}}(t)|\mathrm{d}t = \frac{\mathrm{d}t}{1 - 2S_{\mathrm{E}}(t) + \beta_{\mathrm{E}}}.$$

Proposition 5. *We let the current Pigouvian tax be $T \geq 0$. If the tax marginally increases by $dT > 0$ and the consequent weighted spread ratio is* $\mathrm{WSR}(T)$ *(defined in (11)):*

(a) Welfare W increases ($dW/dT > 0$) if and only if $\mathrm{WSR}(T) > 1$*, and W decreases ($dW/dT < 0$) if and only if* $\mathrm{WSR}(T) < 1$.

(b) If $F = \mathbb{U}[0,1]$, the current tax is strictly lower than the optimal tax $T < T^*$ if and only if $\mathrm{WSR}(T) > 1$*, and $T > T^*$ if and only if* $\mathrm{WSR}(T) < 1$.

(c) If $F = \mathbb{U}[0,1]$, $\mathrm{WSR}(T)$ *is strictly decreasing in T.*

Proposition 5 spells out a simple "WSR Rule" for tax setting: Raise the tax T when $\mathrm{WSR}(T) > 1$, cut T when $\mathrm{WSR}(T) < 1$, and keep the current tax when $\mathrm{WSR}(T) = 1$. Shifting the tax T following the WSR Rule would strictly raise welfare. When $\mathrm{WSR}(T) = 1$, welfare is at a local maximum. The WSR Rule holds under any distribution F with support $[0,1]$. Moreover, under the uniform distribution $F = \mathbb{U}[0,1]$, iteratively following the WSR Rule obtains the optimal tax T^*.

Estimating the Weighted Spread Ratio

The volume term in the WSR, $|dV_{\mathrm{O}}/dV_{\mathrm{E}}|$, is easily computed using trade volumes just before and just after the change in the tax T. Already in the US and the EU, most trades must be reported in nearly real time and specify whether it was executed on an exchange or over the counter. The spread term $S_{\mathrm{E}}/S_{\mathrm{O}}$ is empirically the ratio of the average bid-ask spread on exchanges to the average spread over the counter just before the tax change. To estimate it, the exchange spread S_{E} can be the best quoted ask minus the best quoted bid, divided by the midpoint price. If no exchange has a bid or an ask, the effective spread from recent exchange trades can substitute for the quoted spread. In the OTC market, one can approximate the spread S_{O} with the effective spreads of similar assets (e.g., corporate bonds at the same firm with similar maturity) or adopt benchmarks widely used among traders (e.g., MarketAxess' Composite+ or Bloomberg's BVAL).

Applicability of the optimal Pigouvian tax

Our Pigouvian tax results apply to OTC-traded assets that are also traded via pre-trade anonymous methods. Such methods include limit order books, batch auctions, dark pools, and all-to-all requests for quote. Most important asset classes are actively traded via these methods, including equities, their futures and options, treasuries, corporate bonds, and repurchase agreements. Footnote 3 lists examples.

IV. Related Mechanisms

We find that price discrimination can harm social welfare via cheap substitution. That price discrimination can be inefficient is well-known. One might wonder if cheap substitution repackages a known mechanism. Below, we discuss whether prior mechanisms can generate our two main results: *(i)* Welfare can decline while the aggregate volume increases, and *(ii)* closing the OTC market raises welfare where adverse selection risk is low.

OTC versus exchange

We are most closely related to the literature on venue choice between OTC and centralized markets. One strand in this literature abstracts away from adverse selection and focuses on the presence of search frictions (Pagano, 1989; Rust and Hall, 2003; Vogel, 2019) or limited trading capacity (Dugast, Üslü, and Weill, 2022) in OTC markets. Others, like this paper, feature cream skimming driven by price discrimination (Seppi, 1990; Desgranges and Foucault, 2005). Seppi (1990) explains why trade sizes are larger over the counter than on exchanges. Desgranges and Foucault (2005) shows how endogenous dealer-client relationships can concentrate adverse selection risk on exchanges. They do not examine social welfare or aggregate volume. Because the combination of adverse selection risk and heterogeneous gains from trade is absent, the existing papers in this literature cannot generate either of our main results.

The original Akerlof (1970) framework

In Akerlof (1970), assets have varying common values, and an uninformed trader's private value from owning an asset is proportional to that asset's common value. Therefore, private values are heterogeneous across assets. Akerlof (1970) cannot obtain our main results for two reasons: (I) The private and the common values are perfectly correlated. (II) Each uninformed trader is unaware of her own private value; otherwise, she would learn the common value and face no adverse selection risk. Internet Appendix IA.D adds imperfect labels to Akerlof (1970) and shows how the combination of these two features *reverses* cheap substitution—upon pooling, every entrant has a *smaller* private value than any exiter. As the section shows, the minimum extension to the Akerlof (1970) framework necessary to generate our results is heterogeneity in private values that are (I') decoupled from the common value and (II') known to the uninformed traders.

Cream skimming

Bolton, Santos, and Scheinkman (2016) also features cream skimming by dealers. Nonetheless, Bolton et al. (2016) cannot generate our results, because cream skimming in their framework affects welfare through a tradeoff orthogonal to adverse selection risk. On the one hand, two exogenous frictions in the OTC market lower welfare: Their dealers (a) hold market power and (b) incur a deadweight cost to separate high-quality and low-quality assets. On the other hand, the endogenous opportunity to sell to an informed dealer incentivizes effort in origination, which raises welfare. Therefore, cream skimming necessarily raises welfare in absence of the two exogenous frictions, (a) and (b). We strip away effort in origination, and separate the private values of traders from the common value of the asset. Under this setup, we show that cream skimming strictly lowers welfare whenever adverse selection risk is low, even with a competitive dealer and zero deadweight costs.

Non-anonymity in financial markets

Cheap substitution requires different uninformed traders to be affected in opposite directions. In our case, price discrimination lowers the trading cost of uninformed LU traders and drives up that of uninformed LI traders. In Röell (1990), revealing the trade orders of certain uninformed traders benefits those traders and leaves others worse off. A large literature shows analogous effects (Admati and Pfleiderer, 1991; Forster and George, 1992; Fishman and Longstaff, 1992; Foucault, Moinas, and Theissen, 2007; Rindi, 2008). All such models include noise or liquidity traders who trade an exogenous quantity at any price. Consequently, these models cannot produce cheap substitution or our main results. Indeed, the literature focuses on measures of liquidity or price discovery, rather than social welfare.[17]

Binding minimum wage

Policies we examine lead the uninformed traders with smaller gains from trade to exit and those with larger gains to enter. Likewise, a higher minimum wage forces out the workers with the least surplus from employment, and can strictly raise social welfare under redistributive preferences (Allen, 1987; Guesnerie and Roberts, 1987; Boadway and Cuff, 2001; Lee and Saez, 2012). However, raising the minimum wage does not increase employment without additional features such as endogenous search or effort on the job (Clemens, 2021; Manning, 2021). Without such features, the minimum wage can only reduce aggregate employment and utilitarian welfare.

Third-degree price discrimination

That the dealer engages in price discrimination links our paper to the literature on third-degree price discrimination (e.g., Pigou, 1920; Aguirre, Cowan, and Vickers, 2010; Bergemann, Brooks, and Morris, 2015). Pigou (1920) establishes that allowing a monopolist producer to price discriminate can reduce the total surplus. He identifies a "misallocation effect" in which output is inefficiently distributed whenever different consumers are charged different prices (Aguirre et al., 2010). In their framework, adverse selection is absent and, instead, the distribution of private values determines the effect of price discrimination on welfare. Consequently, there can be *no* guidance over whether price discrimination would raise or lower welfare (Bergemann et al., 2015).[18] We show that adverse selection gives rise to a robust guidance: With minimal assumptions on the distribution of private values, price discrimination lowers welfare whenever adverse selection risk is low.

Broadly related literature

[17] The only exception is Admati and Pfleiderer (1991), whose result on welfare (their Proposition 1 (d)) is the opposite of ours.

[18] Specifically, price discrimination can generate any outcome under which the producer is at least as well off as with a uniform price, consumers receive non-negative payoffs, and the allocation is feasible (Bergemann et al., 2015).

We belong to the enduring literature that compares OTC and centralized markets. Benveniste, Marcus, and Wilhelm (1992), Pagano and Roell (1996), Biais, Foucault, and Salanié (1998), Malinova and Park (2013), and Glode and Opp (2019) compare the case of only having an exchange against only having the OTC market, and study outcomes unrelated to restricting price discrimination. More distantly related is the literature on how traders choose or split orders across multiple venues that do not feature price discrimination (Hendershott and Mendelson, 2000; Zhu, 2014; Pagnotta and Philippon, 2018; Lee, 2019; Chao, Yao, and Ye, 2019; Babus and Parlatore, 2024; Baldauf and Mollner, 2021).

V. Conclusion

We show that limiting price discrimination in the OTC market can improve utilitarian welfare, under the conservative setup of competitive prices. In practice, search frictions and the dealers' market power hamper price competition in OTC markets. As OTC trading moves onto electronic platforms, such frictions are dissipating (Hendershott and Madhavan, 2015; O'Hara and Zhou, 2021; Hau, Hoffmann, Langfield, and Timmer, 2021). Price discrimination by the dealers remains a fundamental feature of OTC trading.

Our model abstracts away from some important sources of inefficiency on exchanges. We do not consider how welfare might be affected by price impact (Vives, 2011) or sniping (Budish, Cramton, and Shim, 2015). Several papers already propose improvements to the design of exchanges that address such frictions.[19] Our focus on resolving the inefficiency of OTC trading complements this literature.

Previous work show that price discovery in secondary markets affects corporate investment decisions (for example, Goldstein and Guembel, 2008). We leave for future research the analysis of price discovery in the presence of an exchange and an OTC market for two reasons. First, in our model, price discovery *within* each market is uninteresting—price dis-

[19] For example, Malamud and Rostek (2017) and Chen and Duffie (2021) show that optimal market fragmentation can address price impact, and Budish et al. (2015) proposes frequent batch auctions to resolve sniping by fast traders.

covery is monotonically increasing in the ratio of informed to uninformed traders β. Second, analyzing the *aggregate* price discovery requires a stance on exactly how the quotes and the transaction prices are aggregated across the two markets. Any effect on aggregate price discovery would be driven by, for instance, the content and the timing of disclosures.

Appendix

A. Proofs

1. Proofs for Section I.C

Proof of Proposition 0. Part (a): It suffices to show that there exists at least one solution to the zero-profit condition (2) so that $S(\beta)$ is well-defined. We use Figure A.1, which plots the exchange dealer's payoff $s \cdot (1-s) - (2\alpha - 1 - s)^+\beta$ over the spread s. (With a general distribution F as in Section II.B, the payoff is $s \cdot [1 - F(s)] - (2\alpha - 1 - s)^+\beta$ and the proof is identical otherwise.) The payoff curve is continuous. Her payoff is negative at $s = 0$, as she is adversely selected yet has no revenue. It is positive at $s = 2\alpha - 1$, as she breaks-even on the trades against the informed and profits on the uninformed. The Intermediate Value Theorem implies that there exists at least one solution to the zero-profit condition.

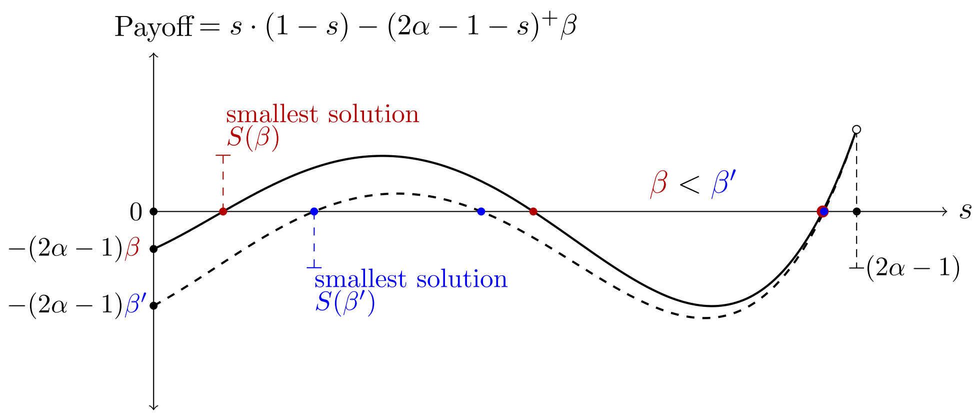


Figure A.1: Finding the equilibrium spread

Part (b): We proceed in three steps. First, we show that the spread function $S(\beta)$ is increasing in the informed ratio $\beta \in [0, \infty)$. We see this easily in Figure A.1: increasing β (to β') shifts the entire payoff curve downwards and the crossing point $S(\beta)$ to the right. Intuitively, as the informed traders impose losses on the exchange dealer, he requires a wider spread to break-even when there are more informed traders.

Second, we solve for the spreads in the OTC market. All traders with the same label share the same OTC spread, because they are indistinguishable to the OTC dealer. For an LU trader, the equilibrium OTC spread is $S(\beta_{\mathrm{LU}})$, where β_{LU} is the informed ratio of the LU traders. As LU traders consist of $(1-\theta)\mu$ informed and γ uninformed traders, their OTC spread is $S\left(\frac{1-\theta}{\gamma}\,\mu\right)$. Similarly, the LI traders' OTC spread is $S\left(\frac{\theta}{1-\gamma}\,\mu\right)$.

Third, we turn to the exchange spread S_{E}. If the exchange dealer sets $S_{\mathrm{E}} \in (S(\beta_{\mathrm{LU}}), S(\beta_{\mathrm{LI}})]$, all LU traders choose the OTC market, whereas all LI traders choose the exchange. Then the informed ratio on the exchange $\beta_{\mathrm{E}} = \beta_{\mathrm{LI}}$. The exchange dealer thus earns zero profit if and only if she sets $S_{\mathrm{E}} = S(\beta_{\mathrm{LI}})$ in this case. If the exchange dealer sets $S_{\mathrm{E}} \leq S(\beta_{\mathrm{LU}})$, then every trader chooses the exchange, implying $\beta_{\mathrm{E}} = \mu > \beta_{\mathrm{LU}}$, and thus the exchange dealer earns a non-zero profit. Therefore, in equilibrium, *(i)* the exchange spread is $S_{\mathrm{E}} = S(\beta_{\mathrm{LI}}) = S\left(\frac{\theta}{1-\gamma}\,\mu\right)$, the lowest spread that earns the exchange dealer a zero profit, and *(ii)* all LU traders choose the OTC market, whereas all LI traders choose the exchange.

Part (c): Since $\beta_{\mathrm{O}} < \beta_{\mathrm{E}}$ and the spread function $S(\beta)$ is strictly increasing in the informed ratio β, then $S_{\mathrm{O}} < S_{\mathrm{E}}$. $\square$

2. *Proofs for Section II*

The proof of Proposition 0 shows that the spread function $S(\beta)$ is increasing. As $S(\beta)$ is also left-continuous in β, then $S(\beta)$ is left-differentiable. We let $S'(\beta)$ be the left derivative. Marginal volume Δ_V and marginal welfare Δ_W as in (3) and (4) are thus well-defined.

Proof of Proposition 1. We first show that Δ_W begins at $\Delta_W(0) = 0$, from which Δ_W strictly increases before eventually strictly decreasing to zero. For this, we establish three properties: (i) $\lim_{s\downarrow 0} sf(s) = 0$, *(ii)* $\lim_{s\downarrow 0} (sf)'(s) \in \mathbb{R}^+ \cup \{\infty\}$, and *(iii)* $\lim_{s\downarrow 0} [\ln(sf)]'(s) = \infty$. (i) Since $sf(s)$ is analytic in a neighborhood $(0, \varepsilon)$ of 0, $\lim_{s\downarrow 0} sf(s)$ exists (the limit can but need not be infinite). Otherwise, $sf(s)$ would cross some strictly positive constant $c > 0$ infinitely often as $s \downarrow 0$, giving rise to an accumulation point of roots for the analytic function $sf(s) - c$. Then the Identity Theorem would imply that $sf(s) - c \equiv 0$ in the neighborhood $(0, \varepsilon)$ of 0, implying that $f(s) \sim c/s$ as $s \downarrow 0$. This contradicts the integrability of the pdf

f in $(0,\varepsilon)$. Further, it must be that $\lim_{s\downarrow 0} sf(s) = 0$ since f is integrable. (ii) Since $(sf)'$ is analytic in a neighborhood of 0, $\lim_{s\downarrow 0}(sf)'(s)$ exists. The limit must be either non-negative or infinity, since $\lim_{s\downarrow 0} sf(s) = 0$. (iii) Since $[\ln(sf)]'$ is analytic in a neighborhood of 0, $\lim_{s\downarrow 0}[\ln(sf)]'(s)$ exists. Since $\lim_{s\downarrow 0} sf(s) = 0$, it must be that $\lim_{s\downarrow 0}[\ln(sf)]'(s) = \infty$. As β approaches 0, $S(\beta)$ approaches 0. Then,

$$S'(\beta) = \frac{1}{\beta'(S(\beta))} = \frac{2\alpha - 1 - S(\beta)}{\frac{(2\alpha-1)[1-F(S(\beta))]}{2\alpha-1-S(\beta)} - S(\beta)f(S(\beta))} \xrightarrow{\beta\downarrow 0} 2\alpha - 1.$$

and thus property (i) implies that $\Delta_W(\beta) \to 0$ as $\beta \to 0$. One can verify that $\ln(\Delta_W)$ is differentiable in a neighborhood of $\beta = 0$ and properties (i)–(iii) imply that

$$(\ln \Delta_W)'(\beta) = [\ln(S')]'(\beta) + S'(\beta)\,[\ln(sf)]'(S(\beta)) \xrightarrow{\beta\downarrow 0} \infty.$$

Thus, $\ln(\Delta_W)$ is strictly increasing in a neighborhood of $\beta = 0$. That is, there exists some $\beta_l > 0$ below which Δ_W is strictly increasing. By a similar argument, there exists some $\beta_h > \beta_l$ above which $\Delta_W(\beta)$ is strictly decreasing to 0. Altogether, Δ_W begins at $\Delta_W(0) = 0$, from which Δ_W strictly increases before eventually strictly decreasing towards the lower limit of zero.

Closing the OTC market is equivalent to lowering label accuracy θ from the current level $\theta_h > 1-\gamma$ to the uninformative level $\theta_l \coloneqq 1-\gamma$. Hence, it suffices to establish Proposition 1 for when the label accuracy θ or γ falls.

Parts (a)–(c): As the label accuracy θ falls from θ_h to θ_l, the change in aggregate trade volume V is

$$\underbrace{(1-\gamma)\int_{S\left(\frac{\theta_l}{1-\gamma}\mu\right)}^{S\left(\frac{\theta_h}{1-\gamma}\mu\right)} f(s)\,\mathrm{d}s}_{\substack{\text{Entry by uninformed}\\ \text{LI-traders}}} - \underbrace{\gamma\int_{S\left(\frac{1-\theta_h}{\gamma}\mu\right)}^{S\left(\frac{1-\theta_l}{\gamma}\mu\right)} f(s)\,\mathrm{d}s}_{\substack{\text{Exit by uninformed}\\ \text{LU traders}}},$$

which is equal to

$$(1-\gamma)\int_{\frac{\theta_l}{1-\gamma}\mu}^{\frac{\theta_h}{1-\gamma}\mu} \Delta_V(\beta)\,\mathrm{d}\beta - \gamma\int_{\frac{1-\theta_h}{\gamma}\mu}^{\frac{1-\theta_l}{\gamma}\mu} \Delta_V(\beta)\,\mathrm{d}\beta. \tag{A.1}$$

Similarly, the change in welfare is

$$\underbrace{(1-\gamma)\int_{\frac{\theta_l}{1-\gamma}\mu}^{\frac{\theta_h}{1-\gamma}\mu} \Delta_W(\beta)\,\mathrm{d}\beta}_{\text{Gross welfare gain}} - \underbrace{\gamma\int_{\frac{1-\theta_h}{\gamma}\mu}^{\frac{1-\theta_l}{\gamma}\mu} \Delta_W(\beta)\,\mathrm{d}\beta}_{\text{Gross welfare loss}}. \tag{A.2}$$

The proofs are intuitive with the aid of graphs. Figure A.2 plots a generic Δ_W. We

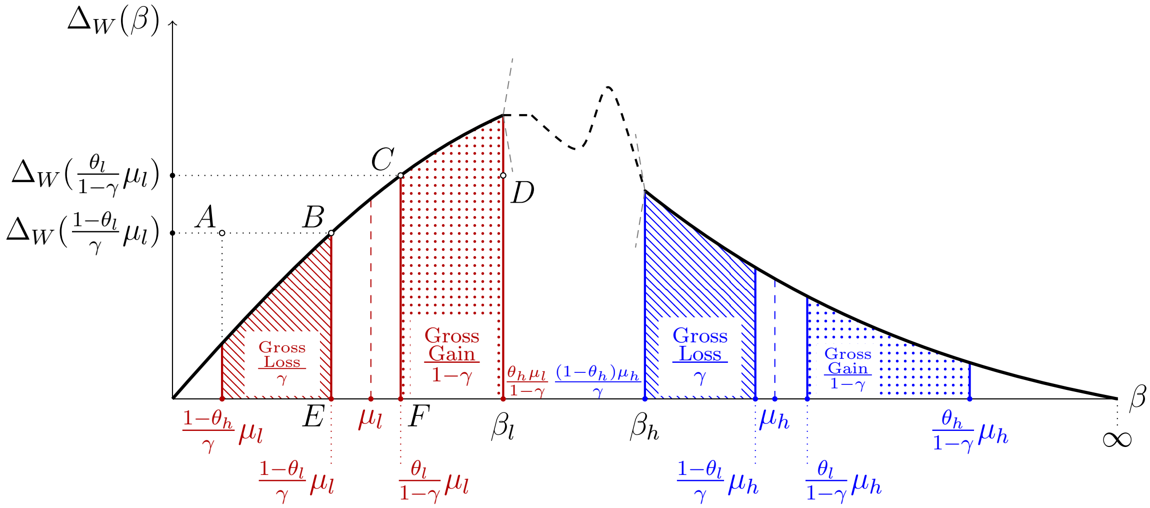


Figure A.2: Generic Δ_W

define μ_l such that $\frac{\theta_h}{1-\gamma}\mu_l = \beta_l$ (marked in Figure A.2). If $\mu = \mu_l$, the second integral in (A.2) (marked "Gross Loss/γ" in red) has a strict upper bound

$$\left(\frac{1-\theta_l}{\gamma}\mu_l - \frac{1-\theta_h}{\gamma}\mu_l\right)\cdot\Delta_W\left(\frac{1-\theta_l}{\gamma}\mu_l\right) = \frac{\theta_h-\theta_l}{\gamma}\mu_l\cdot||\overline{BE}||,$$

which corresponds to the area ABE. The first integral in (A.2) (marked "Gross Gain/$(1-\gamma)$" in red) has a strict lower bound

$$\frac{\theta_h-\theta_l}{1-\gamma}\mu_l\cdot||\overline{CF}||,$$

marked by the area DCF. Since the segment $\overline{CF}$ is longer than $\overline{BE}$ (because Δ_W is strictly increasing in $\beta\in[0,\beta_l]$), the Gross Gain in welfare is strictly larger than the Gross Loss. The same argument applies to any $\mu < \mu_l$, so that welfare rises if the mass of informed traders μ is small. Likewise, we choose μ_h such that $\frac{1-\theta_h}{\gamma}\,\mu_h = \beta_h$ and follow analogous steps to show that the Gross Loss in welfare (in blue) is larger than the Gross Gain if μ is large $\mu \geq \mu_h$. The above continues to hold in the limit where $\theta_l \uparrow \theta_h$, which corresponds to marginally lowering the label accuracy θ (part (a)).

As the label accuracy γ falls from γ_h to γ_l, the change in aggregate trade volume V is

$$\underbrace{(1-\gamma_h)\int_{\frac{\theta}{1-\gamma_l}\mu}^{\frac{\theta}{1-\gamma_h}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\substack{\text{Entry by uninformed}\\ \text{LI-traders}}} - \underbrace{\gamma_h \int_{\frac{1-\theta}{\gamma_h}\mu}^{\frac{1-\theta}{\gamma_l}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\substack{\text{Exit by uninformed}\\ \text{LU traders}}} - \underbrace{(\gamma_h-\gamma_l)\int_{\frac{1-\theta}{\gamma_l}\mu}^{\frac{\theta}{1-\gamma_l}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\text{Exit by relabeled traders}}\,. \tag{A.3}$$

Similarly, the change in welfare is

$$(1-\gamma_h)\int_{\frac{\theta}{1-\gamma_l}\mu}^{\frac{\theta}{1-\gamma_h}\mu} \Delta_W(\beta)\,\mathrm{d}\beta - \gamma_h \int_{\frac{1-\theta}{\gamma_h}\mu}^{\frac{1-\theta}{\gamma_l}\mu} \Delta_W(\beta)\,\mathrm{d}\beta - (\gamma_h-\gamma_l)\int_{\frac{1-\theta}{\gamma_l}\mu}^{\frac{\theta}{1-\gamma_l}\mu} \Delta_W(\beta)\,\mathrm{d}\beta. \tag{A.4}$$

We define $\tilde{\mu}_l$ such that $\frac{\theta}{1-\gamma_h}\,\tilde{\mu}_l = \beta_l$ (marked in Figure A.3). If $\mu = \tilde{\mu}_l$, the second integral

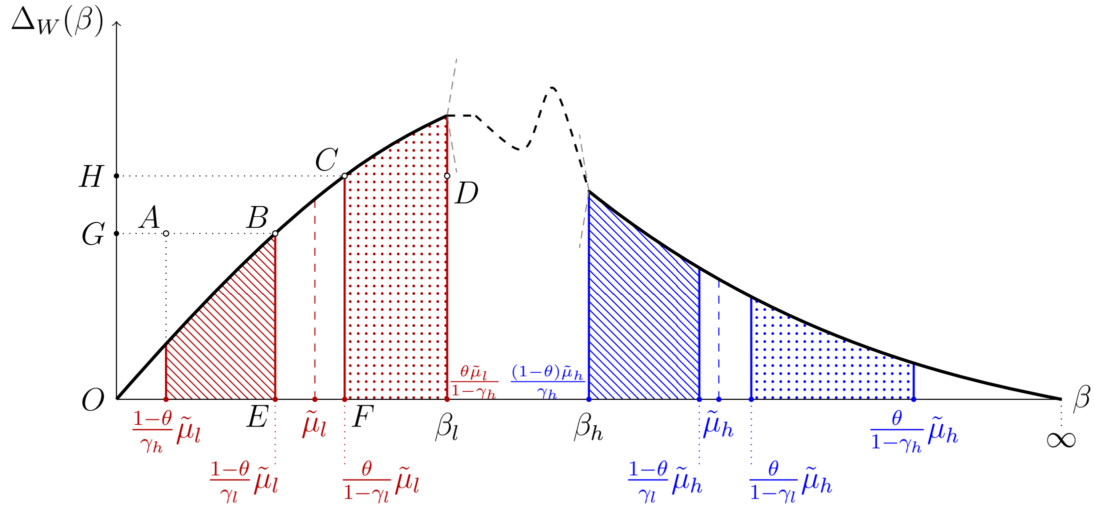


Figure A.3: Generic Δ_W

in (A.4) (marked in shaded red) has a strict upper bound

$$\left(\frac{1-\theta}{\gamma_l}\,\tilde{\mu}_l - \frac{1-\theta}{\gamma_h}\,\tilde{\mu}_l\right)\cdot\Delta_W\left(\frac{1-\theta}{\gamma_l}\,\tilde{\mu}_l\right) = (\gamma_h-\gamma_l)\,\frac{1-\theta}{\gamma_h\gamma_l}\,\tilde{\mu}_l\cdot||\overline{BE}||,$$

which corresponds to the area ABE. Thus, the second term has a strict upper bound

$$(\gamma_h-\gamma_l)\cdot\frac{1-\theta}{\gamma_l}\,\tilde{\mu}_l\cdot||\overline{BE}|| = (\gamma_h-\gamma_l)\cdot||\overline{GB}||\cdot||\overline{BE}|| = (\gamma_h-\gamma_l)\cdot||\overline{GBEO}||,$$

Similarly, the first term in (A.4) has a strict lower bound

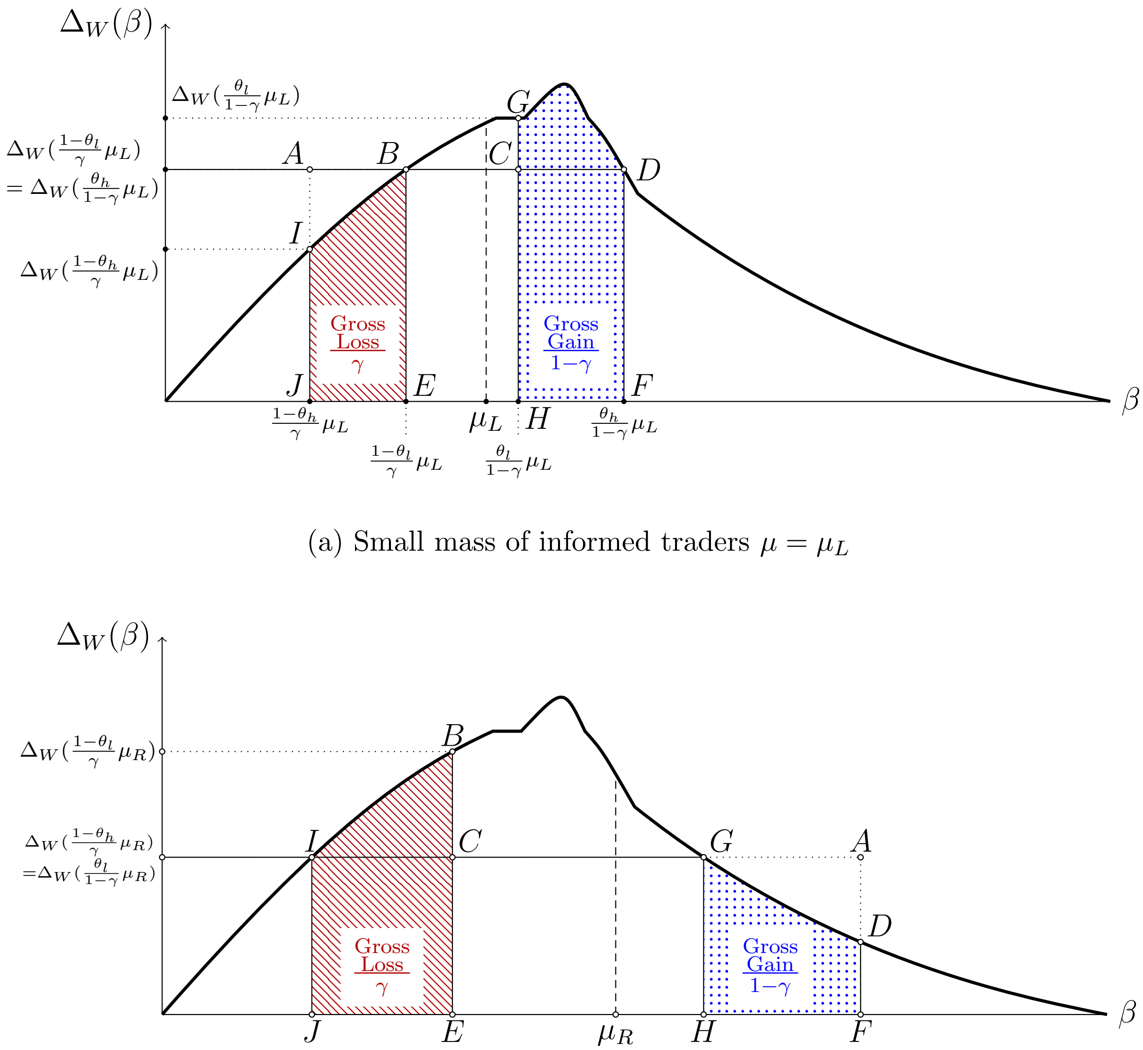
$$(\gamma_h-\gamma_l)\cdot||HCFO||,$$

The third term in (A.4) equals $(\gamma_h-\gamma_l)\cdot||BCFE||$. Then (A.4) is strictly more than

$$(\gamma_h-\gamma_l)\,(||HCFO|| - ||GBEO|| - ||BCFE||) > 0.$$

The same argument applies to any $\mu < \mu_l$, so that welfare rises if the mass of informed traders μ is small. Likewise, we choose μ_h such that $\frac{1-\theta}{\gamma_h}\,\tilde{\mu}_h = \beta_h$ and follow analogous steps to show that the welfare declines if μ is large $\mu \geq \mu_h$. The above continues to hold in the limit where $\gamma_l \uparrow \gamma_h$, which corresponds to marginally lowering the label accuracy γ (part (a)).

Part (d): Figure A.4 plots a quasiconcave Δ_W. As the label accuracy θ falls from θ_h to θ_l, we choose two constants, μ_L and μ_R, as shown in Figure A.4. We set μ_L to be the highest μ such that $\Delta_W\left(\frac{1-\theta_l}{\gamma}\,\mu\right) \leq \Delta_W\left(\frac{\theta_h}{1-\gamma}\,\mu\right)$, and set μ_R to be the highest μ such that $\Delta_W\left(\frac{1-\theta_h}{\gamma}\,\mu\right) \leq \Delta_W\left(\frac{\theta_l}{1-\gamma}\,\mu\right)$. As Δ_W is quasiconcave and $\lim_{\beta\downarrow 0}\Delta_W(\beta) = \lim_{\beta\uparrow\infty}\Delta_W(\beta) = 0$, $0 < \mu_L < \mu_R < \infty$. For illustration only, Figure A.4 plots the case where Δ_W is continuous so that $\Delta_W\left(\frac{1-\theta_l}{\gamma}\,\mu_L\right) = \Delta_W\left(\frac{\theta_h}{1-\gamma}\,\mu_L\right)$ (line $\overline{BD}$ in Figure A.4a) and $\Delta_W\left(\frac{1-\theta_h}{\gamma}\,\mu_R\right) = \Delta_W\left(\frac{\theta_l}{1-\gamma}\,\mu_R\right)$ (line $\overline{IG}$ in Figure A.4b). The proof works whether or not Δ_W is continuous. We show that *(i)* the change in welfare (A.2) is strictly positive for all $\mu < \mu_L$ and strictly negative for



(a) Small mass of informed traders $\mu = \mu_L$

(b) Large $\mu = \mu_R > \mu_L$

Figure A.4: Quasiconcave Δ_W

all $\mu > \mu_R$; and *(ii)* (A.2) is strictly decreasing between μ_L and μ_R. Together, *(i)* and *(ii)* establish the existence of a single cutoff.

To prove *(i)*, we set $\mu = \mu_L$. A strict upper bound of the second integral in (A.2) ("Gross Loss/γ" in red) is

$$\frac{\theta_h - \theta_l}{\gamma}\,\mu_L \cdot \Delta_W\left(\frac{1-\theta_l}{\gamma}\,\mu_L\right) = ||\overline{AB}|| \cdot ||\overline{BE}|| = ||ABEJ||,$$

A strict lower bound of the first integral in (A.2) ("Gross Gain/$(1-\gamma)$" in red) is $||CDFH||$. As $||\overline{BE}|| \leq ||\overline{DF}||$, the Gross Gain in welfare is strictly larger than the Gross Loss. The same argument applies to any $\mu < \mu_L$, so that (A.2) is strictly positive for all $\mu \leq \mu_L$. Similarly, (A.2) is strictly negative for all $\mu \geq \mu_R$.

To prove *(ii)* that (A.2) is strictly decreasing over $\mu \in (\mu_L, \mu_R)$, we take the derivative of (A.2) with respect to μ,

$$\underbrace{\left(\theta_h \cdot ||\overline{DF}|| - \theta_l \cdot ||\overline{GH}||\right)}_{\text{Derivative of the gross welfare gain}} - \underbrace{\left((1-\theta_l) \cdot ||\overline{BE}|| - (1-\theta_h) \cdot ||\overline{IJ}||\right)}_{\text{Derivative of the gross welfare loss}}. \tag{A.5}$$

Due to Δ_W being quasiconcave and how μ_L and μ_R are chosen, both $||\overline{BE}||$ and $||\overline{GH}||$ are strictly greater than $||\overline{DF}||$ and $||\overline{IJ}||$ when $\mu_L < \mu < \mu_R$. Then, (A.5) is strictly negative. In sum, as θ decreases from θ_h to θ_l, the change in welfare is strictly positive if $\mu \leq \mu_L$, strictly negative if $\mu \geq \mu_R$, and strictly decreasing across $\mu \in (\mu_L, \mu_R)$, which together imply that a single cutoff exists. The above continues to hold in the limit where $\theta_l \uparrow \theta_h$, which corresponds to marginally lowering the label accuracy θ.

We follow similar steps when the label accuracy γ falls from γ_h to γ_l. We choose two constants, $\tilde{\mu}_L < \tilde{\mu}_R$, as shown in Figure A.5. We set $\tilde{\mu}_L$ to be the highest μ such that $\Delta_W\left(\frac{\theta}{1-\gamma_l}\,\mu\right) \leq \Delta_W\left(\frac{\theta}{1-\gamma_h}\,\mu\right)$ (geometrically, $||\overline{GH}|| \leq ||\overline{CD}||$ in the left panel of Figure A.5), and set $\tilde{\mu}_R$ to be the highest μ such that $\Delta_W\left(\frac{1-\theta}{\gamma_h}\,\mu\right) \leq \Delta_W\left(\frac{1-\theta}{\gamma_l}\,\mu\right)$ ($||\overline{AB}|| \leq ||\overline{EF}||$ in the right panel).

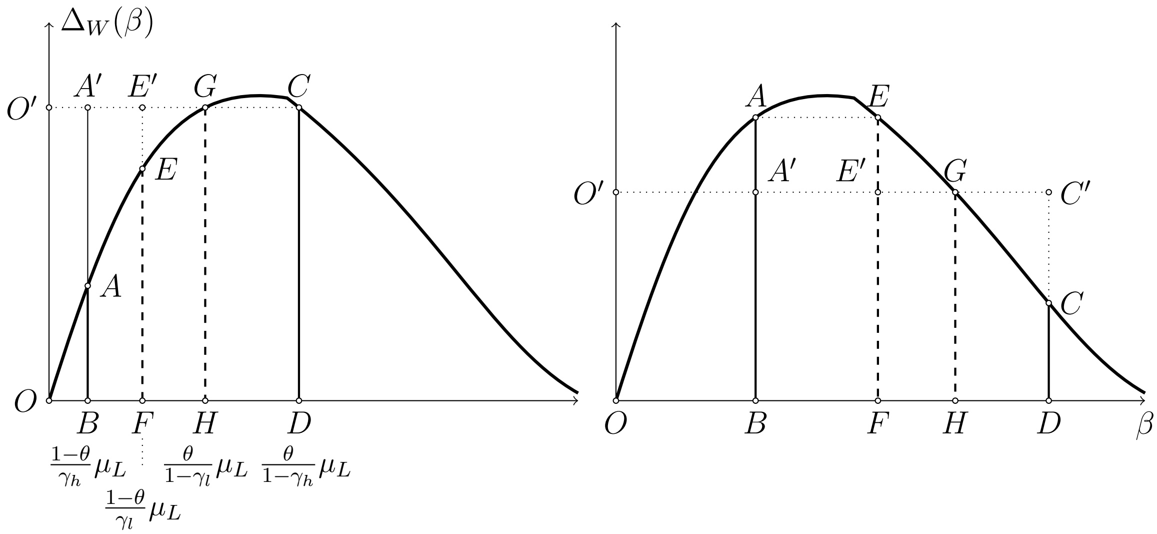


Figure A.5: Decrease in label accuracy γ

For $\mu \leq \tilde{\mu}_L$, the change in welfare (A.4) is strictly more than

$$(\gamma_h - \gamma_l)\left(||O'GHO|| - ||O'E'FO|| - ||EGHF||\right) > 0.$$

For $\mu \geq \tilde{\mu}_R$, (A.4) is strictly less than

$$(\gamma_h - \gamma_l)\left(||O'GHO|| - ||O'E'FO|| - ||EGHF||\right) < 0.$$

For $\mu \in (\tilde{\mu}_L, \tilde{\mu}_R)$, the derivative of (A.4) is

$$(1-\theta) \cdot ||\overline{AB}|| + \theta \cdot ||\overline{CD}|| - (1-\theta) \cdot ||\overline{EF}|| - \theta \cdot ||\overline{GH}|| < 0.$$

Therefore, a single cutoff exists. The above continues to hold in the limit where $\gamma_l \uparrow \gamma_h$, which corresponds to marginally lowering the label accuracy γ.

Parts (e): Figure A.6 plots a decreasing Δ_V. As the label accuracy θ falls from θ_h to θ_l, the second term of (A.1) has a strict lower bound

$$(\theta_h - \theta_l)\,\mu \cdot ||\overline{BE}||,$$

which is larger than the first term's strict upper bound

$$(\theta_h - \theta_l)\,\mu \cdot ||\overline{DF}||,$$

and thus the change (A.1) in aggregate volume V is strictly negative. The above continues to hold in the limit where $\theta_l \uparrow \theta_h$, which corresponds to marginally lowering the label accuracy θ.

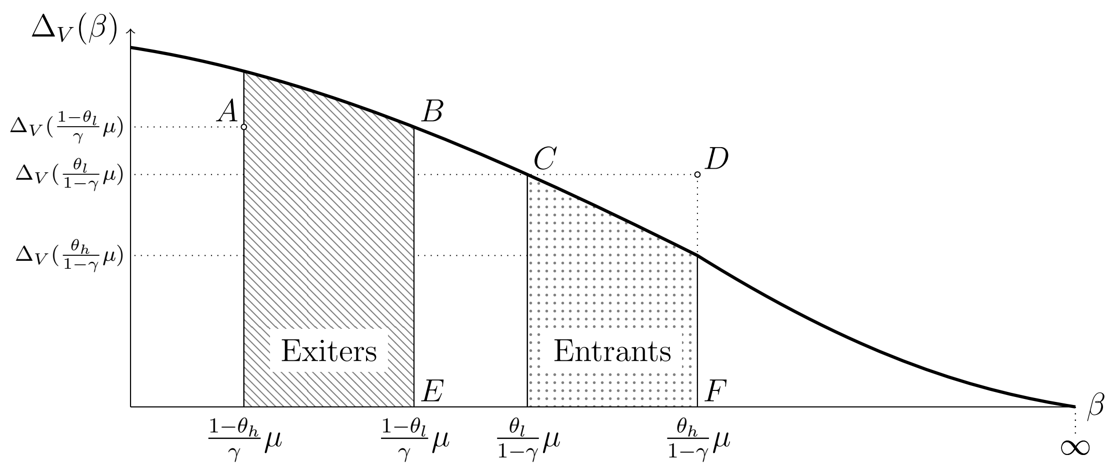


Figure A.6: Decreasing Δ_V

As the label accuracy γ falls from γ_h to γ_l, the change (A.3) in aggregate volume V is strictly less than

$$(\gamma_h - \gamma_l)(||O'GHO|| - ||O'E'FO|| - ||EGHF||) < 0,$$

marked in Figure A.7. The above continues to hold in the limit where $\gamma_l \uparrow \gamma_h$, which corresponds to marginally lowering the label accuracy γ. $\square$

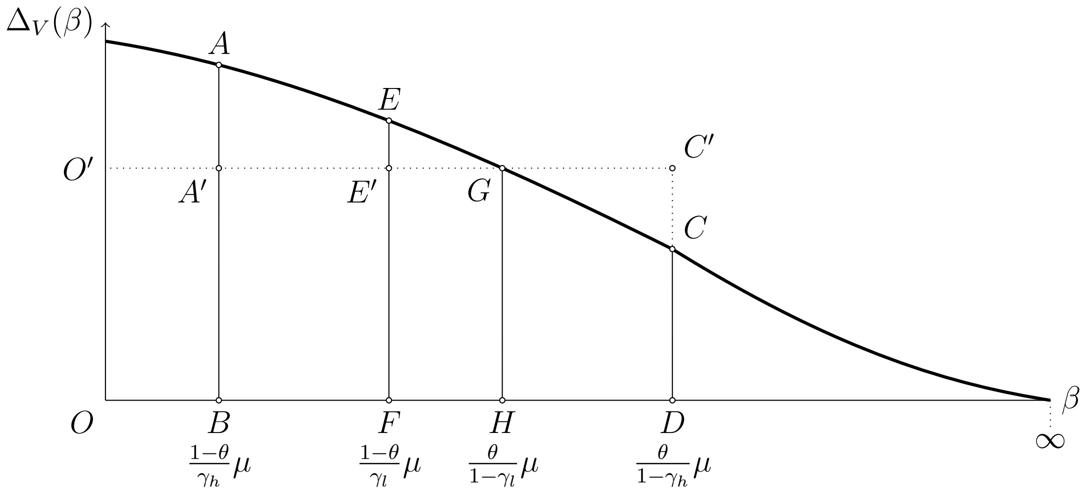


Figure A.7: Decrease in label accuracy γ

Proof of Proposition 2. The marginal welfare Δ_W can be written as

$$\Delta_W(\beta) = S'(\beta)S(\beta)f(S(\beta)) = \frac{S(\beta)f(S(\beta))}{\beta'(S(\beta))}.$$

Because the spread function $S(\beta)$ is strictly increasing in β, then $\Delta_W(\beta)$ is quasiconcave if and only if $\beta'(x)/(xf(x))$ is quasiconvex in $x \in (0, 2\alpha - 1)$. Differentiating (2) with respect to x yields

$$1 - F(x) + \beta - xf(x) = (2\alpha - 1 - x)\beta'(x),$$

which can be rearranged to

$$\beta'(x) = \frac{1 - F(x) + \beta - xf(x)}{2\alpha - 1 - x}. \tag{A.6}$$

From (2), we can express $\beta(x)$ as function of x:

$$\beta(x) = \frac{(1 - F(x))x}{2\alpha - 1 - x}. \tag{A.7}$$

Then, marginal welfare $\Delta_W(\beta)$ is quasiconcave if and only if condition (5) holds. Likewise, marginal volume $\Delta_V(\beta)$ is decreasing if and only if (6) is true. $\square$

Proof of Proposition 3. Part (a): The ratio $V_{\mathrm{E}}/V_{\mathrm{O}}$ equals

$$\frac{V_{\mathrm{E}}}{V_{\mathrm{O}}} := \frac{(1-\gamma)[1-F(S(\beta_{\mathrm{E}}))]+\theta\mu}{\gamma[1-F(S(\beta_{\mathrm{O}}))]+(1-\theta)\mu}.$$

The derivative of $V_{\mathrm{E}}/V_{\mathrm{O}}$ with respect to μ is strictly positive if and only if

$$\frac{-f(S(\beta_{\mathrm{O}}))S'(\beta_{\mathrm{O}})+1}{\frac{1-F(S(\beta_{\mathrm{O}}))}{\beta_{\mathrm{O}}}+1} < \frac{-f(S(\beta_{\mathrm{E}}))S'(\beta_{\mathrm{E}})+1}{\frac{1-F(S(\beta_{\mathrm{E}}))}{\beta_{\mathrm{E}}}+1}.$$

The above inequality holds for every μ, θ, and γ if and only if

$$\frac{-f(S(\beta))S'(\beta)+1}{\frac{1-F(S(\beta))}{\beta}+1}$$

is strictly increasing in β. After a change of variable using (A.6) and (A.7), the above is true if and only if

$$\frac{-\frac{(2\alpha-1-x)f(x)}{\frac{(2\alpha-1)[1-F(x)]}{2\alpha-1-x}-xf(x)}+1}{\frac{2\alpha-1-x}{x}+1}$$

is strictly increasing in $x \in (0, 2\alpha-1)$. After simplifying, the above is true if and only if

$$x\left(\frac{1}{2\alpha-1-x}-\frac{f(x)}{1-F(x)}\right)$$

is strictly increasing in $x \in (0, 2\alpha-1)$.

When $F = \mathbb{U}[0,1]$, the above expression equals

$$2(1-\alpha)\,\frac{s}{(2\alpha-1-s)(1-s)}$$

which is strictly increasing in $x \in (0, 2\alpha-1)$.

Parts (b) and (c) are corollaries of part (a) and Proposition 1. $\square$

3. *Proofs for Section III*

Proof of Propositions 4 and 5. We prove Propositions 4 and 5 together. First, imposing some Pigouvian tax T obtains a strictly higher welfare than closing the OTC market (Proposition 4 (a)). To show why, we proceed in three steps.

Step 1: We show that when $T = \overline{T}$, both the OTC and the exchange spreads are equal to the No-OTC spread S_{N}, $S_{\mathrm{O}}(\overline{T}) = S_{\mathrm{E}}(\overline{T}) = S_{\mathrm{N}}$. The definition of $\overline{T}$ implies $S_{\mathrm{O}}(\overline{T}) = S_{\mathrm{E}}(\overline{T})$, wherein the OTC dealer's and the exchange dealer's zero-profit conditions are

$$S(\overline{T})\left[1 - F\left(S(\overline{T})\right) + \beta_{\mathrm{O}}\right]\gamma = (2\alpha - 1)\beta_{\mathrm{O}}\gamma + \overline{T},$$

$$S(\overline{T})\left[1 - F\left(S(\overline{T})\right) + \beta_{\mathrm{E}}\right]\cdot(1-\gamma) = (2\alpha - 1)\beta_{\mathrm{E}}\cdot(1-\gamma) - \overline{T}.$$

Summing the two equations gives

$$S(\overline{T})\left[1 - F\left(S(\overline{T})\right) + \mu\right] = (2\alpha - 1)\mu.$$

Since the zero-OTC tax $\overline{T}$ is unique, it must be that $S_{\mathrm{O}}(\overline{T}) = S_{\mathrm{E}}(\overline{T}) = S_{\mathrm{N}}$.

Step 2: We show that a marginal tax cut $-dT$ from the tax $\overline{T}$ increases the aggregate trade volume V. The volume V increases if and only if $|dV_{\mathrm{O}}(\overline{T})/dV_{\mathrm{E}}(\overline{T})| > 1$, where

$$dV_{\mathrm{O}}(\overline{T}) = f(S_{\mathrm{N}})S'_{\mathrm{O}}(\overline{T})\gamma\, dT \text{ and } dV_{\mathrm{E}}(\overline{T}) = f(S_{\mathrm{N}})S'_{\mathrm{E}}(\overline{T})\cdot(1-\gamma)\, dT.$$

Taking the derivative of the zero-profit conditions (8)-(9) with respect to T:

$$S'_{\mathrm{O}}(T)\left[1 - F\left(S_{\mathrm{O}}(T)\right) - f\left(S_{\mathrm{O}}(T)\right)S_{\mathrm{O}}(T) + \beta_{\mathrm{O}}\right]\gamma = 1,$$

$$S'_{\mathrm{E}}(T)\left[1 - F\left(S_{\mathrm{E}}(T)\right) - f\left(S_{\mathrm{E}}(T)\right)S_{\mathrm{E}}(T) + \beta_{\mathrm{E}}\right]\cdot(1-\gamma) = -1.$$

Then, using that $S_{\mathrm{O}}(\overline{T}) = S_{\mathrm{E}}(\overline{T}) = S_N$,

$$\left|\frac{dV_{\mathrm{O}}(\overline{T})}{dV_{\mathrm{E}}(\overline{T})}\right| = \left|\frac{S'_{\mathrm{O}}(\overline{T})\gamma}{S'_{\mathrm{E}}(\overline{T}) \cdot (1-\gamma)}\right| = \frac{1 - F(S_{\mathrm{N}}) - f(S_{\mathrm{N}})S_{\mathrm{N}} + \beta_{\mathrm{E}}}{1 - F(S_{\mathrm{N}}) - f(S_{\mathrm{N}})S_{\mathrm{N}} + \beta_{\mathrm{O}}}.$$

The ratio $|dV_{\mathrm{O}}(\overline{T})/dV_{\mathrm{E}}(\overline{T})| > 1$ because $\beta_{\mathrm{E}} > \beta_{\mathrm{O}}$.

Step 3: We show that a strictly larger volume V implies a strictly higher welfare W. The entrants and the exiters upon a marginal change in the tax around $\overline{T}$ have the same hedging benefit. Since the entrants outnumber the exiters, welfare W is strictly higher.

Altogether, Steps 1-3 imply that the tax cut $-dT$ from $\overline{T}$ strictly raises welfare.

Second, we establish Proposition 5 (a) and (c). Upon a marginal increase dT in the tax $T < \overline{T}$, the gross welfare loss among traders over the counter is $|S_{\mathrm{O}}(T)dV_{\mathrm{O}}(T)|$. The gross welfare gain among traders on the exchange is $|S_{\mathrm{E}}(T)dV_{\mathrm{E}}(T)|$. Thus, welfare W increases ($dW/dT > 0$) if and only if $\mathrm{WSR}(T) > 1$ (Proposition 5 (a)).

If $F = \mathbb{U}[0,1]$, from the first half of this proof,

$$|S_{\mathrm{O}}(T)dV_{\mathrm{O}}(T)| = \frac{S_{\mathrm{O}}(T)}{1 - 2S_{\mathrm{O}}(T) + \beta_{\mathrm{O}}}\, dT = \frac{1}{\frac{1+\beta_{\mathrm{O}}}{S_{\mathrm{O}}(T)} - 2}\, dT,$$

and

$$|S_{\mathrm{E}}(T)dV_{\mathrm{E}}(T)| = \frac{S_{\mathrm{E}}(T)}{1 - 2S_{\mathrm{E}}(T) + \beta_{\mathrm{E}}}\, dT = \frac{1}{\frac{1+\beta_{\mathrm{E}}}{S_{\mathrm{E}}(T)} - 2}\, dT.$$

Because $S_O(T)$ is strictly increasing in T while $S_{\mathrm{E}}(T)$ is strictly decreasing, $|S_O(T)dV_O(T)|$ is strictly increasing in T and $|S_E(T)dV_E(T)|$ is strictly decreasing. Hence,

$$\mathrm{WSR}(T) = \left|\frac{S_{\mathrm{E}}(T)dV_{\mathrm{E}}(T)}{S_{\mathrm{O}}(T)dV_{\mathrm{O}}(T)}\right| = \frac{\frac{1+\beta_{\mathrm{O}}}{S_{\mathrm{O}}(T)} - 2}{\frac{1+\beta_E}{S_E(T)} - 2}$$

is strictly decreasing in T (Proposition 5 (c)).

Therefore, there exists a unique $T^* \in [0, \overline{T})$ such that $\mathrm{WSR}(T) > 1$ for $T \in [0, T^*)$ and $\mathrm{WSR}(T) < 1$ for $T > (T^*, \overline{T})$. Proposition 5 (a) then implies that $dW/dT > 0$ for $T < T^*$

and $dW/dT < 0$ for $T > T^*$. That is, T^* is the unique optimal tax that maximizes welfare W. Proposition 4 (b) and Proposition 5 (b) follow. $\square$

References

Admati, A. R. and P. Pfleiderer (1991): "Sunshine Trading and Financial Market Equilibrium," *The Review of Financial Studies*, 4, 443–481.

Aguirre, I., S. Cowan, and J. Vickers (2010): "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, 100, 1601–1615.

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488–500.

Allen, S. P. (1987): "Taxes, Redistribution, and the Minimum Wage: A Theoretical Analysis," *The Quarterly Journal of Economics*, 102, 477–489.

Babus, A. and C. Parlatore (2024): "Strategic Fragmented Markets," *Journal of Financial Economics*, 145, 876–908.

Baldauf, M. and J. Mollner (2021): "Trading in Fragmented Markets," *Journal of Financial and Quantitative Analysis*, 56, 93–121.

Benveniste, L. M., A. J. Marcus, and W. J. Wilhelm (1992): "What's Special about the Specialist?" *Journal of Financial Economics*, 32, 61–86.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

Biais, B., T. Foucault, and F. Salanié (1998): "Floors, Dealer Markets and Limit Order Markets," *Journal of Financial Markets*, 1, 253–284.

Boadway, R. and K. Cuff (2001): "A Minimum Wage Can Be Welfare-Improving and Employment-Enhancing," *European Economic Review*, 45, 553–576.

Bolton, P., T. Santos, and J. A. Scheinkman (2016): "Cream-Skimming in Financial Markets," *The Journal of Finance*, 71, 709–736.

Budish, E., P. Cramton, and J. Shim (2015): "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 130, 1547–1621.

Chao, Y., C. Yao, and M. Ye (2019): "Why Discrete Price Fragments U.S. Stock Exchanges and Disperses Their Fee Structures," *The Review of Financial Studies*, 32, 1068–1101.

Chen, D. and D. Duffie (2021): "Market Fragmentation," *American Economic Review*, 111, 2247–2274.

Cheng, I.-H. and W. Xiong (2014): "Why Do Hedgers Trade so Much?" *The Journal of Legal Studies*, 43, S183–S207.

Clemens, J. (2021): "How Do Firms Respond to Minimum Wage Increases? Understanding the Relevance of Non-Employment Margins," *Journal of Economic Perspectives*, 35, 51–72.

Collin-Dufresne, P., B. Junge, and A. B. Trolle (2020): "Market Structure and Transaction Costs of Index CDSs," *The Journal of Finance*, 75, 2719–2763.

de Roure, C., E. Moench, L. Pelizzon, and M. Schneider (2021): "OTC Discount," .

Desgranges, G. and T. Foucault (2005): "Reputation-Based Pricing and Price Improvements," *Journal of Economics and Business*, 57, 493–527.

Dugast, J., S. Üslü, and P.-O. Weill (2022): "A Theory of Participation in OTC and Centralized Markets," *The Review of Economic Studies*, 89, 3223–3266.

Exton, G. and R. Healey (2022): "Liquidity Landscape H1 2022," Tech. rep., Liquidnet.

Fishman, M. J. and F. A. Longstaff (1992): "Dual Trading in Futures Markets," *The Journal of Finance*, 47, 643–671.

Forster, M. M. and T. J. George (1992): "Anonymity in Securities Markets," *Journal of Financial Intermediation*, 2, 168–206.

Foucault, T., S. Moinas, and E. Theissen (2007): "Does Anonymity Matter in Electronic Limit Order Markets?" *Review of Financial Studies*, 20, 1707–1747.

Glode, V. and C. C. Opp (2019): "Over-the-Counter vs. Limit-Order Markets: The Role of Traders' Expertise," *Review of Financial Studies*, 33, 866–915.

Glosten, L. R. and P. R. Milgrom (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

Goldstein, I. and A. Guembel (2008): "Manipulation and the Allocational Role of Prices," *The Review of Economic Studies*, 75, 133–164.

Guesnerie, R. and K. Roberts (1987): "Minimum Wage Legislation as a Second Best Policy," *European Economic Review*, 31, 490–498.

Hau, H., P. Hoffmann, S. Langfield, and Y. Timmer (2021): "Discriminatory Pricing of Over-the-Counter Derivatives," *Management Science*.

Hendershott, T. and A. Madhavan (2015): "Click or Call? Auction versus Search in the Over-the-Counter Market," *Journal of Finance*, 70, 419–447.

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

Lee, D. and E. Saez (2012): "Optimal Minimum Wage Policy in Competitive Labor Markets," *Journal of Public Economics*, 96, 739–749.

Lee, T. (2019): "Latency in Fragmented Markets," *Review of Economic Dynamics*, 33, 128–153.

Malamud, S. and M. Rostek (2017): "Decentralized Exchange," *American Economic Review*, 107, 3320–3362.

Malinova, K. and A. Park (2013): "Liquidity, Volume and Price Efficiency: The Impact of Order vs. Quote Driven Trading," *Journal of Financial Markets*, 16, 104–126.

Manning, A. (2021): "The Elusive Employment Effect of the Minimum Wage," *Journal of Economic Perspectives*, 35, 3–26.

O'Hara, M. and X. A. Zhou (2021): "The Electronic Evolution of Corporate Bond Dealers," *Journal of Financial Economics*, 140, 368–390.

Pagano, M. (1989): "Trading Volume and Asset Liquidity," *The Quarterly Journal of Economics*, 104, 255–274.

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

Pagnotta, E. S. and T. Philippon (2018): "Competing on Speed," *Econometrica*, 86, 1067–1115.

Pigou, A. (1920): *The Economics of Welfare*, St. Martin's Street, London: MacMillan and Co., Limited.

Riggs, L., E. Onur, D. Reiffen, and H. Zhu (2020): "Swap Trading after Dodd-Frank: Evidence from Index CDS," *Journal of Financial Economics*, 137, 857–886.

Rindi, B. (2008): "Informed Traders as Liquidity Providers: Anonymity, Liquidity and Price Formation," *Review of Finance*, 12, 497–532.

Röell, A. (1990): "Dual-Capacity Trading and the Quality of the Market," *Journal of Financial Intermediation*, 1, 105–124.

Rust, J. and G. Hall (2003): "Middlemen versus Market Makers: A Theory of Competitive Exchange," *Journal of Political Economy*, 111, 353–403.

Securities Industry and Financial Markets Association (2018): "Re: Post-Trade Name Give-Up on Swap Execution Facilities; Request for Comment - RIN 3038-AE79, 83 Fed. Reg. 61751," .

Selten, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

Seppi, D. J. (1990): "Equilibrium Block Trading and Asymmetric Information," *The Journal of Finance*, 45, 73–94.

Vives, X. (2011): "Strategic Supply Function Competition With Private Information," *Econometrica*, 79, 1919–1966.

Vogel, S. (2019): "When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?" *Working Paper*.

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

Internet Appendix for "Why Trade Over the Counter?"

Tomy Lee Chaojun Wang*

June 4, 2024

IA.A. Implications for Permissioned Blockchain

Blockchain is an electronic recordkeeping procedure for a network of members, called "nodes." Each transaction between the nodes is broadcast to the other nodes in the network, and each node maintains a ledger of all transactions. An algorithm periodically reconciles these ledgers, and the sequence of reconciled ledgers forms the 'official' record. Transactions on a blockchain include pseudonymized identifiers of the counterparties involved in the trade. The reconciled record provides the complete trade history of every account on the blockchain. Since pseudonymity violates anti-money-laundering rules (Elwell, Murphy, and Seitzinger, 2013), most proposed blockchains for financial markets fully reveal traders' identities to their nodes.

The Depository Trust & Clearing Corporation (DTCC)[1] proposed a plan to move the ownership records of most credit derivatives onto a blockchain (Irrera, 2017). The DTCC envisions a "permissioned" blockchain, wherein 15 major dealers act as the nodes (DTCC, 2018). Soon thereafter, ICAP (the dominant interdealer broker) initiated a pilot of a permissioned blockchain for foreign exchange, and the DTCC proposed a permissioned blockchain for equities (DTCC, 2020).

All three proposals would reveal each trader's transaction history to the major dealers. The trade histories may aid the dealers to more accurately separate traders by their trading motives. If so, the blockchain proposals would increase the label accuracy, θ or γ, in our

*Lee is at the Central European University. Wang is at the Wharton School, University of Pennsylvania. Emails: leeso@ceu.edu; wangchj@wharton.upenn.edu. *Citation format: Lee, Tomy and Chaojun Wang, Internet Appendix to "Why Trade Over the Counter?," Journal of Finance [DOI STRING].

[1]DTCC is the dominant clearinghouse for most securities.

model. Then, our model predicts a decrease in welfare for the assets with low adverse selection risk and a widening of the bid-ask spreads on exchanges.

IA.B. Request for Quote

We microfound the zero-profit condition and the trading protocol in the OTC market. We add two features to the model of Section I: There are $n \geq 2$ OTC dealers and $m \geq 2$ exchange dealers ($n + m$ dealers), and traders must request quotes from the OTC dealers to trade over the counter. We show that these additions do not change the equilibrium allocation.

Strategies. In Stage 1, each trader chooses (i) whether to simultaneously request two-way quotes and (ii) from whom among the n OTC dealers (possibly all n). (The equilibrium allocation is the same whether the traders' choices (i) and (ii) are disclosed to the intermediaries or not.) In Stage 2, each exchange dealer posts one bid and one ask available to all traders and, for each quote request received, each OTC dealer offers a bid and an ask conditional on the requester's label. In Stage 3, each trader observes her prices then makes two decisions: (a) buy, sell, or exit and (b) against which price to trade. We impose the same tie-breaking rule Assumption 2 as in Section I.A. Figure IA.B.1 summarizes the timing and Figure IA.B.2 the traders' choices.

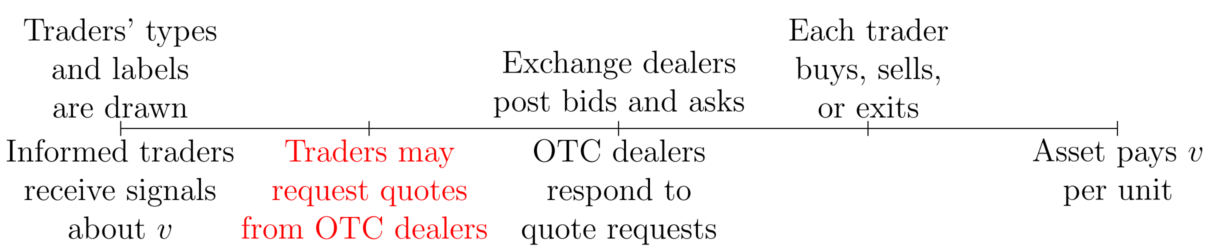


Figure IA.B.1: Timeline

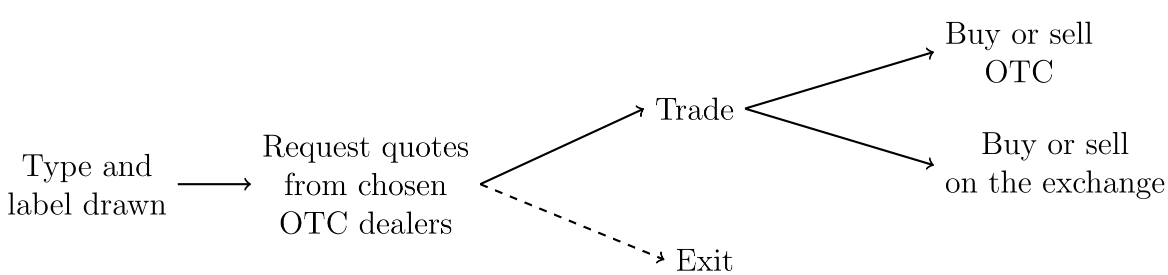


Figure IA.B.2: Traders' Choices

Equilibrium. We study the properties of an extensive-form trembling-hand perfect (THP) equilibrium (Selten, 1975). The next proposition shows that any extensive-form THP equilibrium must induce an identical allocation as the equilibrium of the original model.

Proposition IA.B.1. *Given any extensive-form trembling-hand perfect equilibrium,*[2] *with probability* 1*:*

- *In Stage 1, each trader sends a quote request to all n OTC dealers.*
- *In Stage 2, every exchange dealer posts the exchange spread S_E, and every OTC dealer offers the OTC spread S_O to a Likely Uninformed trader and S_E to a Likely Informed trader. (The spreads S_O and S_E are given in Proposition 0.)*
- *In Stage 3, every LI trader chooses the exchange and receives S_E, and every LU trader chooses the OTC market and receives S_O.*

Proof. We first show that in Stage 1, each trader sends an RFQ to all the n OTC dealers[3] with probability 1 under a THP equilibrium E. We fix a converging sequence of completely mixed strategy profiles E_n that gives rise to the equilibrium E. Each strategy profile E_n is a Nash equilibrium of a perturbed game. In Stage 1, each trader optimally submits an RFQ to every OTC dealer with the maximum probability that is allowed in the perturbed game for two reasons. First, sending an RFQ is cost-free. Second, every dealer follows a completely mixed pricing strategy under the Nash equilibrium E_n, so any OTC dealer offers the overall best price with a strictly positive probability. Hence, each trader submits an RFQ to all the OTC dealers with probability 1 under the limiting equilibrium E.

Since every trader submits an RFQ to all the OTC dealers with probability 1, the extended trading game reduces to the original game with price posting instead of RFQ. Therefore, a THP equilbrium of the extended trading game induces an identical equilibrium allocation as the equilibrium of Section I.C. □

IA.C. Informed Traders with Hedging Benefits

We modify the model of Section II.B to endow the informed traders with hedging benefits. Precisely, each trader's payoff V_i from owning an asset (or $-V_i$ from shorting an asset) is the

[2] There may be multiple THP equilibria. The proof of Proposition IA.B.1 establishes that all such equilibria must have the same allocation. For example, the THP equilibria may differ in how an OTC dealer responds at an information set where the dealer does not receive quote requests from all traders. Such an information set is reached with probability 0 under any THP equilibrium.

[3] Wang (2023) shows that a trader submits an RFQ to only two dealers, who quote noncompetitive prices, when a dealer must incur a cost to respond to an RFQ. Adding such an OTC friction to our model would strengthen the benefit of restricting the OTC dealer. We abstract away from any OTC friction to focus on the effect of cheap substitution.

sum of a private value $\pm b_i$ (a hedging need) and the common value v of the asset, $V_i = \pm b_i + v$. The trader's private value is equally likely to be b_i and $-b_i$. The hedging benefits $b_i \overset{iid}{\sim} F$, whose pdf f is analytic in some neighborhoods of 0 and $(2\alpha - 1)$ with support $[0, 1]$.

We normalize the total mass of traders to 1. A fraction μ of the traders are informed in that they each receive a private binary signal about v, as do the informed traders of the baseline model. This setup guarantees that the total attainable hedging benefit is invariant to μ. A trader's private value $\pm b_i$ and whether she is informed are her private information.

The remaining setup is identical to the model in Section II.B. In this extension, the informed traders have "dual intent" of earning speculative profit and attaining their hedging benefits.

Proposition IA.C.1. *Given any α, we let each uninformed trader's hedging benefit $b_i \overset{iid}{\sim} F$ and its pdf f be analytic in some neighborhoods of 0 and $(2\alpha - 1)$ with support $[0, 1]$. Given any pairs $(\theta_l, \gamma_l) < (\theta_h, \gamma_h) < (1, 1)$ that satisfy $\theta_h + \gamma_l > 1$ and $\theta_l + \gamma_h > 1$, there exists a cutoff $\mu^d > 0$ such that:*

- *(a) (1) Marginally lowering traders' label accuracy θ from θ_h to $\theta_h - d\theta$ or γ from γ_h to $\gamma_h - d\gamma$, (2) lowering the label accuracy θ from θ_h to θ_l or γ from γ_h to γ_l, or (3) closing the OTC market strictly raises welfare W for all $\mu < \mu^d$.*
- *(b) For $F = \mathbb{U}[0, 1]$, (1), (2), or (3) strictly reduces the aggregate trade volume V and widens the average bid-ask spread $\bar{S}$.*

Restricting the OTC dealer raises welfare if the mass of informed traders μ is sufficiently small. Welfare may rise or decline where μ is large.

The intuition is that cheap substitution remains dominant with small μ while, with large μ, other effects with ambiguous signs may dominate. In the base model, cheap substitution arises among (A) the uninformed traders, because the uninformed LI traders enter and the uninformed LU traders exit. Adding dual intent, two further instances of cheap substitution arise. Among (B) the informed traders with large hedging benefits, who only trade in the direction of their benefits, cheap substitution takes the same form as for (A). Among (C) the informed traders with small benefits, who only trade against the direction of their benefits to profit on their private signals, cheap substitution is "double flipped." (First flip) Trading by (C) destroys welfare as each such trade is in the *opposite* direction of its underlying hedging benefit. (Second flip) Each exiter among (C) has a *larger* hedging benefit than each entrant, since a trader in (C) trades only if her hedging benefit is *small*. As a result of the double flip, cheap substitution in (C) raises welfare because each exiter would otherwise inflict a larger welfare loss than each entrant. All three cheap substitutions raise welfare and dominate if μ is low. When μ is high, all cheap substitutions vanish, which leaves the volume effect (present in the base model) and several other effects (absent in the base model), whose net effect depends on distributional assumptions.

Proof of Proposition IA.C.1. If a dealer charges a spread s, the dealer earns zero profit in expectation if and only if

$$
\begin{aligned}
s \cdot [\underbrace{1 - F(s)}_{(A)}] = \Bigg[\quad & (2\alpha - 1 - s)\left(\frac{1}{2} + \underbrace{\frac{F(2\alpha - 1 - s)}{2}}_{(C)}\right) \\
& - (2\alpha - 1 + s)\underbrace{\frac{1 - F(2\alpha - 1 + s)}{2}}_{(B)}\Bigg] \cdot \beta.
\end{aligned}
\tag{IA.C.1}
$$

The left-hand side is the dealer's expected profit when facing an uninformed trader, and is the same as the LHS of (2). On the right-hand side, a dealer who faces an informed trader j expects either (i) a loss of $2\alpha-1-s$ if j trades in the direction of her signal, or (ii) a gain of $2\alpha-1+s$ if j trades against her signal. Thus, any zero-profit spread must be less than $2\alpha-1$. Given a spread $s < 2\alpha - 1$, event (i) occurs if and only if the private value $\pm b_j$ of informed trader j and her signal are in the same direction, or (C) $\pm b_j$ is in the opposite direction of her signal and $b_j < 2\alpha-1-s$. Event (i) occurs with probability $[1+F(2\alpha-1-s)]/2$. Event (ii) occurs if and only if (B) $\pm b_j$ is in the opposite direction of her signal and $b_j > 2\alpha-1+s$. Event (ii) occurs with probability $[1 - F(2\alpha - 1 + s)]/2$. Last, β is the mass of informed traders per unit mass of uninformed traders.

The Intermediate Value Theorem implies that the zero-profit condition (2) admits at least one solution in $(0, 2\alpha - 1)$. Its smallest solution is the equilibrium spread, denoted as $\tilde{S}(\beta)$.

Part (a): Welfare is equal to

$$
\begin{aligned}
(1-\mu)\Bigg(\quad (1-\gamma)\Bigg[\quad & \int_{\tilde{S}(\beta_{\mathrm{E}})}^{1} sf(s)\,\mathrm{d}s + \\
& \frac{\beta_{\mathrm{E}}}{2}\left(\int_{2\alpha-1-\tilde{S}(\beta_{\mathrm{E}})}^{1} sf(s)\,\mathrm{d}s + \int_{2\alpha+1+\tilde{S}(\beta_{\mathrm{E}})}^{1} sf(s)\,\mathrm{d}s\right)\Bigg] \\
+ \gamma\Bigg[\quad & \int_{\tilde{S}(\beta_{\mathrm{O}})}^{1} sf(s)\,\mathrm{d}s + \\
& \frac{\beta_{\mathrm{O}}}{2}\left(\int_{2\alpha-1-\tilde{S}(\beta_{\mathrm{O}})}^{1} sf(s)\,\mathrm{d}s + \int_{2\alpha+1+\tilde{S}(\beta_{\mathrm{O}})}^{1} sf(s)\,\mathrm{d}s\right)\Bigg]\Bigg),
\end{aligned}
$$

where

$$
\beta_{\mathrm{E}} = \frac{\theta\mu}{(1-\gamma)(1-\mu)}, \qquad \beta_{\mathrm{O}} = \frac{(1-\theta)\mu}{\gamma(1-\mu)}.
$$

Expanding the terms, for $M \in \{\text{O,E}\}$, we have

$$\begin{aligned}
&\frac{\beta_M}{2}\int_{2\alpha-1-\tilde{S}(\beta_M)}^{1} sf(s)\,\mathrm{d}s \\
=\ &\frac{1}{2}\int_0^{\beta_M}\Bigg[\int_{2\alpha-1-\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s + \\
&\qquad \beta\tilde{S}'(\beta)\left(2\alpha-1-\tilde{S}(\beta)\right) f\left(2\alpha-1-\tilde{S}(\beta)\right)\Bigg]\mathrm{d}\beta,
\end{aligned}$$

$$\begin{aligned}
&\frac{\beta_M}{2}\int_{2\alpha-1+\tilde{S}(\beta_M)}^{1} sf(s)\,\mathrm{d}s \\
=\ &\frac{1}{2}\int_0^{\beta_M}\Bigg[\int_{2\alpha-1+\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s - \\
&\qquad \beta\tilde{S}'(\beta)\left(2\alpha-1+\tilde{S}(\beta)\right) f\left(2\alpha-1+\tilde{S}(\beta)\right)\Bigg]\mathrm{d}\beta.
\end{aligned}$$

As the label accuracy θ falls from θ_h to θ_l, the change in welfare W is

$$(1-\mu)\left[(1-\gamma)\int_{\frac{\theta_l\mu}{(1-\gamma)(1-\mu)}}^{\frac{\theta_h\mu}{(1-\gamma)(1-\mu)}} \Delta_W^d(\beta)\,\mathrm{d}\beta - \gamma\int_{\frac{(1-\theta_h)\mu}{\gamma(1-\mu)}}^{\frac{(1-\theta_l)\mu}{\gamma(1-\mu)}} \Delta_W^d(\beta)\,\mathrm{d}\beta\right],$$

where

$$\begin{aligned}
\Delta_W^d(\beta) =\ &\tilde{S}(\beta)\tilde{S}'(\beta)f(\tilde{S}(\beta)) \\
&-\frac{1}{2}\Bigg[\int_{2\alpha-1-\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s + \beta\tilde{S}'(\beta)\left(2\alpha-1-\tilde{S}(\beta)\right) f\left(2\alpha-1-\tilde{S}(\beta)\right) + \\
&\qquad \int_{2\alpha-1+\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s - \beta\tilde{S}'(\beta)\left(2\alpha-1+\tilde{S}(\beta)\right) f\left(2\alpha-1+\tilde{S}(\beta)\right)\Bigg].
\end{aligned}$$

As the label accuracy γ falls from γ_h to γ_l, the change in welfare W is

$$\begin{aligned}
(1-\mu)\Bigg[&(1-\gamma_h)\int_{\frac{\theta\mu}{(1-\gamma_l)(1-\mu)}}^{\frac{\theta\mu}{(1-\gamma_h)(1-\mu)}} \Delta_W^d(\beta)\,\mathrm{d}\beta - \gamma_h\int_{\frac{(1-\theta)\mu}{\gamma_h(1-\mu)}}^{\frac{(1-\theta)\mu}{\gamma_l(1-\mu)}} \Delta_W^d(\beta)\,\mathrm{d}\beta \\
&- (\gamma_h-\gamma_l)\int_{\frac{(1-\theta)\mu}{\gamma_l(1-\mu)}}^{\frac{\theta\mu}{(1-\gamma_l)(1-\mu)}} \Delta_W^d(\beta)\,\mathrm{d}\beta\Bigg].
\end{aligned}$$

Next, we show that Δ_W^d is strictly increasing in a neighborhood of $\beta = 0$. Then, Proposition IA.C.1 Part (a) follows from identical steps as in the proof of Proposition 1 (a)–(c).

First, one can show that $\lim_{\beta \downarrow 0}[\tilde{S}(\beta)\tilde{S}'(\beta)f(\tilde{S}(\beta))]'(\beta)$ exists and is strictly positive. Thus, $\tilde{S}(\beta)\tilde{S}'(\beta)f(\tilde{S}(\beta))$ is increasing in a neighborhood of $\beta = 0$ with a strictly positive slope. Second,

$$\begin{aligned} & \left(\int_{2\alpha-1-\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s + \int_{2\alpha-1+\tilde{S}(\beta)}^{1} sf(s)\,\mathrm{d}s\right)'(\beta) \\ = \quad & \tilde{S}'(\beta)\big[\ (2\alpha-1-\tilde{S}(\beta))f(2\alpha-1-\tilde{S}(\beta)) \\ & \qquad -(2\alpha-1+\tilde{S}(\beta))f(2\alpha-1+\tilde{S}(\beta))\big] \\ \xrightarrow{\beta\downarrow 0} \quad & 0 \qquad\qquad\qquad \left(\text{since } \tilde{S}(\beta) \xrightarrow{\beta\downarrow 0} 0 \text{ and } f \text{ is analytic in a neighborhood of } 2\alpha-1\right) \end{aligned}$$

Third,

$$\begin{aligned} \frac{1}{\beta}\Big[\ & \beta\tilde{S}'(\beta)(2\alpha-1-\tilde{S}(\beta))f(2\alpha-1-\tilde{S}(\beta)) - \\ & \beta\tilde{S}'(\beta)(2\alpha-1+\tilde{S}(\beta))f(2\alpha-1+\tilde{S}(\beta))\Big] \xrightarrow{\beta\downarrow 0} \quad 0 \end{aligned}$$

Therefore, Δ_W^d is strictly increasing in a neighborhood of $\beta = 0$.

Part (b): Under the uniform distribution $F = \mathbb{U}[0,1]$, as θ falls from θ_h to θ_l, the change in aggregate volume V is

$$(1-\mu)\left[(1-\gamma)\int_{\frac{\theta_l\mu}{(1-\gamma)(1-\mu)}}^{\frac{\theta_h\mu}{(1-\gamma)(1-\mu)}} \Delta_V^d(\beta)\,\mathrm{d}\beta - \gamma\int_{\frac{(1-\theta_h)\mu}{\gamma(1-\mu)}}^{\frac{(1-\theta_l)\mu}{\gamma(1-\mu)}} \Delta_V^d(\beta)\,\mathrm{d}\beta\right],$$

where

$$\Delta_V^d(\beta) = \tilde{S}'(\beta) - 2\,(1-\alpha)\,.$$

As γ falls from γ_h to γ_l, the change in V is

$$\begin{aligned} (1-\mu)\Bigg[(1-\gamma_h)\int_{\frac{\theta\mu}{(1-\gamma_l)(1-\mu)}}^{\frac{\theta\mu}{(1-\gamma_h)(1-\mu)}} \Delta_V^d(\beta)\,\mathrm{d}\beta - \gamma_h\int_{\frac{(1-\theta)\mu}{\gamma_h(1-\mu)}}^{\frac{(1-\theta)\mu}{\gamma_l(1-\mu)}} \Delta_V^d(\beta)\,\mathrm{d}\beta \\ - (\gamma_h-\gamma_l)\int_{\frac{(1-\theta)\mu}{\gamma_l(1-\mu)}}^{\frac{\theta\mu}{(1-\gamma_l)(1-\mu)}} \Delta_V^d(\beta)\,\mathrm{d}\beta\Bigg]. \end{aligned}$$

One can show that $\tilde{S}'$ is strictly decreasing in $\beta \geq 0$ and so is Δ_V^d. Then, Proposition IA.C.1 Part (b) follows from identical steps as in the proof of Proposition 1 (e). □

IA.D. Relation to Akerlof (1970)

The setup of Akerlof (1970) differs from ours in two ways: (I) the private values are perfectly correlated with the common value, and (II) each uninformed trader does not know her own private value. In the presence of assumption (I), assumption (II) is necessary to preserve information asymmetry. This section adds imperfect labels to Akerlof (1970) under every coherent combination of (I) and (II). First, assumptions (I) and (II) together *reverse* the effect of cheap substitution on welfare. Second, assumption (II) alone turns off cheap substitution and is equivalent to a setting with homogeneous hedging benefits. Under the first two settings, pooling can raise social welfare only if it also raises the aggregate trade volume. Third, removing both assumptions is necessary and sufficient to obtain our main results.

The baseline model. We recall the seminal Akerlof (1970) model. Each of a unit mass of sellers owns one unit of an asset with value $x_i \overset{\text{iid}}{\sim} \mathbb{U}[0,2]$, which is private information to seller i. One buyer values an asset at ax_i, with $a > 1$ (in Akerlof (1970), $a = 3/2$), and has no information about the realization of x_i. The buyer has Y dollars to purchase assets. In a pooled market with price p, the sellers supply $S(p) = p/2$ of assets, and the expected value of those assets is $\mu = p/2$. The demand of the buyer equals $D(p) = Y/p$ if $a\mu > p$ (or equivalently, $a > 2$) and $D(p) = 0$ if $a\mu < p$ (i.e., $a < 2$). If $a < 2$, no trade can exist in the pooled market. If $a > 2$, then $D(p^*) = S(p^*)$ in equilibrium. Therefore, $p^* = \sqrt{2Y}$, and only the assets with common values below p^* get traded.

Adding imperfect labels. The first step of our analysis extends Akerlof (1970) with imperfect labels to compare pooling and separating outcomes. The buyer receives an imperfect label of x_i, which indicates $x_i \sim G_1$ with probability $\beta > 0$ and $x_i \sim G_2$ with probability $1-\beta$ without revealing the realization of x_i. The distributions satisfy $G_1 \neq G_2$ and $\beta G_1 + (1-\beta)G_2 = \mathbb{U}[0,2]$. One can similarly solve for the market clearing prices p_1^* and p_2^* in each of the two markets. If market m breaks down ($m = 1, 2$), we set $p_m^* = 0$ by convention. We suppose that $p_1^* < p_2^*$ without loss of generality. Then, one must have $p_1^* < p^* < p_2^*$.[4] In market m, only the assets with common values below p_m^* get traded. Once pooled together at the price p^*, those market-1 assets with values between $[p_1^*, p^*]$ "enter," while those market-2 assets with values between $[p^*, p_2^*]$ "exit." Each entering asset attains a private value between $[(a-1)p_1^*, (a-1)p^*]$, and each exiting asset destroys a private value between $[(a-1)p_1^*, (a-1)p_2^*]$. Thus, every entrant has a lower private value than every exiter, which is precisely the opposite of cheap substitution. Therefore, pooling the two markets

[4] If $p^* \leq p_1^* < p_2^*$, then

$$D(p^*) = D_1(p^*) + D_2(p^*) > D_1(p_1^*) + D_2(p_2^*) = S_1(p_1^*) + S_2(p_2^*) > S_1(p^*) + S_2(p^*) = S(p^*),$$

which contradicts $D(p^*) = S(p^*)$. Likewise, one can rule out $p_1^* < p_2^* \leq p^*$.

can raise social welfare only if it also raises the aggregate trade volume.

The reversal of cheap substitution is caused by assumption (I), the perfect correlation between private and common values. Intuitively, because only the assets with relatively lower common values get traded, only lower private values get realized under the perfect correlation.

Removing the correlation between private and common values. Second, we break assumption (I), the perfect correlation between common and private values, while maintaining assumption (II) that the buyer does not know her own private value. To do so, we make one modification to the first extension: The buyer now values each asset at $x_i + b_i$, where $b_i \overset{\text{iid}}{\sim} F$. The private values b_i and the common values x_i are independently distributed. The buyer remains uninformed of the private values b_i. This extension is equivalent to the one with a homogeneous private value in that the private values can be replaced by their expectation $\mathbb{E}b_i$, because not knowing the private values makes $\mathbb{E}b_i$ a sufficient statistic for the buyer.

We denote the pooling equilibrium price as $\tilde{p}$. The separating equilibrium prices are $\tilde{p}_1$ and $\tilde{p}_2$, and we suppose $\tilde{p}_1 < \tilde{p}_2$ without loss of generality. By the same arguments as before, $\tilde{p}_1 < \tilde{p} < \tilde{p}_2$. Upon pooling, the market-1 assets with common values in $[\tilde{p}_1, \tilde{p}]$ "enter," while those market-2 assets valued between $[\tilde{p}, \tilde{p}_2]$ "exit." The aggregate volume changes by $\Delta V = \beta[G_1(\tilde{p}) - G_1(\tilde{p}_1)] - (1-\beta)[G_2(\tilde{p}_2) - G_2(\tilde{p})]$. Each entering asset attains the expected private value $\mathbb{E}b_i$, and each exiting asset destroys $\mathbb{E}b_i$. Thus, every entrant has the same expected private value as any exiter. The net change in welfare is $\mathbb{E}b_i \Delta V$. As such, pooling raises welfare if and only if the aggregate volume also rises.

Revealing private value to the buyer. Third, we reveal the private value b_i to the buyer. We denote the market-clearing prices as $\hat{p}$, $\hat{p}_1$, and $\hat{p}_2$. In the pooled market, an asset gets traded if and only if $\hat{p}/2 + b_i > \hat{p}$, or $b_i > \hat{p}/2$. That is, only private values *above* $\underline{b} \coloneqq \hat{p}/2$ get traded. Likewise, when separated, only the assets with private values *above* $\underline{b}_m \coloneqq \hat{p}_m - \mathbb{E}_{G_m}(x \,|\, x < \hat{p}_m)$ get traded in market m ($m = 1, 2$). We suppose $\underline{b}_1 < \underline{b}_2$ without loss of generality. Then, one must have $\underline{b}_1 < \underline{b} < \underline{b}_2$.[5] Upon pooling, the market-1 assets with private values between $[\underline{b}_1, \underline{b}]$ "exit," and the market-2 assets with private values

[5] One can show that $\underline{b}$ is some weighted average between $\underline{b}_1$ and $\underline{b}_2$:

$$
\begin{aligned}
\underline{b}_m &= \hat{p}_m - \mathbb{E}_{G_m}(x \,|\, x < \hat{p}_m) \\
&= \mathbb{E}_{G_m}(\hat{p}_m - x \,|\, x < \hat{p}_m) \\
&= \mathbb{E}_{\widetilde{G}_m}(\Delta_m \,|\, \Delta_m > 0) \;\text{ where } \Delta_m \coloneqq \hat{p}_m - x \text{ and } \widetilde{G}_m \text{ is the distribution of } \Delta_m \\
&= \frac{\mathbb{E}_{\widetilde{G}_m}(\Delta_m^+)}{\mathbb{P}_{\widetilde{G}_m}(\Delta_m > 0)} \\
&= \frac{\mathbb{E}_{\widehat{G}_m}(\Delta_m^+)}{\mathbb{P}_{\widetilde{G}_m}(\Delta_m > 0)} \text{ where } \widehat{G}_m \text{ is the distribution of } \Delta_m^+.
\end{aligned}
$$

between $[\underline{b}, \underline{b}_2]$ "enter." Therefore, every entrant has a higher private value than any exiter, which is precisely cheap substitution. This extension generates our result that pooling can raise welfare while reducing the aggregate trade volume.

IA.E. Evidence

OTC discount

A prediction of our model is that the bid-ask spread over the counter is lower than the spread on the exchange. Recent evidence in de Roure, Moench, Pelizzon, and Schneider (2021) is consistent with this prediction. They examine data on German Bunds, which are traded both in the OTC market and on exchanges. Their headline result is that 87% of OTC trades execute at a better price (for the initiator of the trade) than the contemporaneous price on the exchange. A typical OTC discount is substantial: The median OTC trade saves 65.7% of the half spread it would have cost on the exchange. Moreover, the OTC discount is not an artifact of larger trade sizes. The median trade size over the counter (EUR 5.4 million) is nearly identical to that on the exchange (EUR 5.0M). On average, the OTC trades are larger (EUR 12.74M vs 7.88M), solely because the biggest 5 percentile of trades are far larger over the counter (EUR 50M vs 13M). Likewise, Collin-Dufresne, Junge, and Trolle (2020) finds that both dealer-to-client (D2C) and dealer-to-dealer (D2D) trades of index CDSs incur smaller trading costs than the contemporaneous bid-ask spread on the largest limit order book.[6]

Analogous evidence is found in the literature on the upstairs market for equities. (Upstairs trades are the OTC trades of stocks.) Rose (2014) compares upstairs and limit order book (LOB) trades on the Australian Stock Exchange. While OTC trades make a loss on average, LOB trades earn a profit. Further, traders who consistently make losses are more likely to trade upstairs and receive a discount relative to the price on the LOB. The opposite holds for consistently profit-earning traders. Others find that upstairs trades are less informed and less costly (Madhavan and Cheng, 1997; Smith, Turnbull, and White, 2001;

We define G, Δ, $\widetilde{G}$, and $\widehat{G}$ in a similar manner in the pooled market: $G := \mathbb{U}[0,2]$, $\Delta := \hat{p} - x$, $\widetilde{G}$ is the distribution of Δ, and $\widehat{G}$ is the distribution of Δ^+. Since $\beta G_1 + (1-\beta)G_2 = G$, then

$$\beta\widetilde{G}_1 + (1-\beta)\widetilde{G}_2 = \widetilde{G}\,, \qquad \beta\widehat{G}_1 + (1-\beta)\widehat{G}_2 = \widehat{G}.$$

Therefore,

$$\begin{aligned}
&\beta\mathbb{E}_{\widehat{G}_1}\left(\Delta_1^+\right) + (1-\beta)\mathbb{E}_{\widehat{G}_2}\left(\Delta_2^+\right) = \mathbb{E}_{\widehat{G}}\left(\Delta^+\right), \\
&\beta\mathbb{P}_{\widetilde{G}_1}\left(\Delta_1 > 0\right) + (1-\beta)\mathbb{P}_{\widetilde{G}_2}\left(\Delta_2 > 0\right) = \mathbb{P}_{\widetilde{G}}\left(\Delta > 0\right).
\end{aligned}$$

[6] Collin-Dufresne et al. (2020) focuses on one limit order book, GFI Swaps Exchange ("D2D SEF" in their parlance). GFI Swaps Exchange provides non-discriminatory access to both dealers and buyside firms (GFI LLC, 2022, Sections 301, 302).

Booth, Lin, Martikainen, and Tse, 2002; Bessembinder and Venkataraman, 2004; Bernhardt, Dvoracek, Hughson, and Werner, 2005; Westerholm, 2009). Large trades do not drive the upstairs discount. In fact, the discount is minimized among large orders (Bernhardt et al., 2005) and can be monotonically shrinking in trade size (Westerholm, 2009). The findings echo our mechanism that dealers offer a discount to the less informed traders and cream skim them into the OTC market.

Migration of corporate bonds to the OTC market

Cream skimming in our model pulls a larger share of trades into the OTC market when a smaller share of traders are informed. The Depression-era migration of corporate bond trades is an episode consistent with this prediction. Most corporate bond trades moved from exchanges to the OTC market in the 1930s and 40s (Biais and Green, 2007). This movement coincided with a sudden, dramatic rise in the proportion of corporate bonds held by insurance companies and pension funds. Homer (1975) delivers the author's first-hand account of this transition as a brokerage president in the 1930s: Pre-1930s, the corporate bond traders primarily traded on NYSE and were "small, country investors or big-city investors" (p. 379). Particularly active were sophisticated traders who "grabbed up" new issues, which "if well priced they sold at quick premiums" (p. 379). During the Great Depression, distressed investors sold out to dealers, who in turn resold the bonds to life insurance companies (p. 381). The life insurers preferred to trade over the counter although "[t]he exchange tried hard to retain its bond business" (p. 381). Because insurers typically trade with a hedging motive, the history of corporate bonds mirrors our prediction that OTC trading dominates where most traders are uninformed.

IA.F. Empirics

We use US equities data to document (1) substantial OTC market share in US-listed equities and (2) a positive correlation between the exchange market share and the exchange quoted spread, which Proposition IA.F.1 below predicts.

Proposition IA.F.1. *For $F = \mathbb{U}[0,1]$ and given any triple (α, θ, γ), the exchange spread S_E and the exchange market share V_E/V are each strictly increasing in the mass of informed traders μ.*

1. Data, Variables, and Summary Statistics

We combine millisecond Trade and Quote (TAQ) data with weekly OTC trade volumes from the Financial Industry Regulatory Agency (FINRA).[7] In the FINRA data, OTC volumes are separated into Alternative Trading Systems (ATS) versus Non-ATS OTC volumes.

[7] FINRA data used here is available publicly at http://www.finra.org/industry/otc-transparency.

The ATS consists of dark pools, batch auctions, and limit order books that are not designated as "national securities exchanges" by the US Securities and Exchange Commission. Because trading on the ATS is anonymous, the ATS corresponds to the exchange in our model. The Non-ATS OTC refers to traditional bilateral and request-for-quote trades. Therefore, only the Non-ATS trades are counted as over the counter in our analysis.

The sample period is January 2, 2017–March 5, 2021, the available range of FINRA OTC data at the time of this analysis. We exclude all trades outside of market hours. Our sample consists of 3,210 US-listed non-ETF tickers that exist in both TAQ and FINRA data on both the first and the last weeks of the 218 weeks in the sample period. Only the trades during market hours are included to avoid an upward bias for OTC market share. (Results remain nearly identical if all trades are included).

Exchange market share for ticker i and week w is 1 minus the ratio of week w OTC dollar volume of trades from FINRA to the week w aggregate dollar volume from TAQ. To compute *percent quoted spread* for ticker i and week w, we use the millisecond TAQ quotes to calculate the time-weighted best quoted spread in percentage, *(best ask - best bid) /midpoint*, for each day, and then we take the simple average across the number of days observed for ticker i in week w. The total *number of trades* and total *dollar volume* of trades are computed for each ticker i and week w from millisecond TAQ trade data. Table IA.F.1 provides the summary statistics, first for all observations, and then by average weekly dollar volume of trades in quintiles.

2. *Exchange Market Share for US-listed Equities*

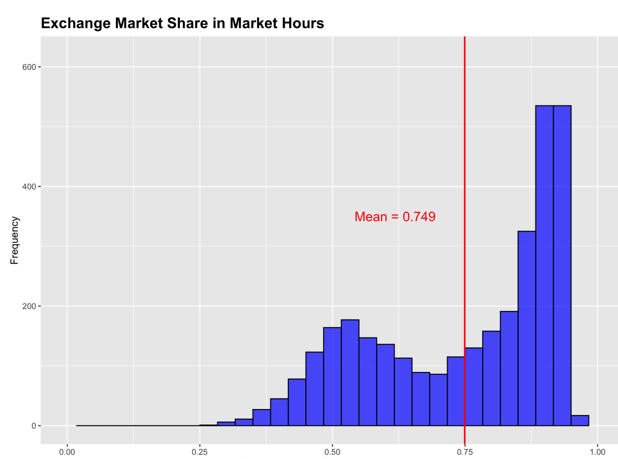


Figure IA.F.1: Exchange Market Share of Dollar Value of Trades in US-listed Equities

Does OTC trading dominate only for historically OTC-traded assets such as bonds? Figure IA.F.1 plots the average weekly exchange market shares of tickers in our sample.

Table IA.F.1: Summary Statistics

Each observation is one ticker for one week.

	Obs	Mean	SD	Min	25%	50%	75%	Max
	All Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	694,305	0.750	0.198	0.000	0.597	0.828	0.914	1.000
Percent Quoted Spread	694,305	0.027	0.053	0.000	0.004	0.010	0.026	6.995
Dollar Volume of Trades	694,305	384.4M	2,752.6M	0.0M	3.3M	27.5M	187.2M	376,000.0M
Number of Trades	694,305	38,401	107,820	2	1,714	9,434	36,594	13,458,361
	By Average Weekly Dollar Volume of Trades:							
	Quintile 1 Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	137,455	0.615	0.185	0.000	0.484	0.609	0.758	1.000
Percent Quoted Spread	137,455	0.067	0.086	0.001	0.016	0.042	0.089	6.995
Dollar Volume of Trades	137,455	1.6M	3.7M	0.0M	0.4M	0.9M	1.9M	543.8M
Number of Trades	137,455	997	3,495	2	189	474	1,027	491,416
	Quintile 2 Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	139,184	0.659	0.193	0.001	0.506	0.652	0.839	1.000
Percent Quoted Spread	139,184	0.037	0.055	0.001	0.011	0.022	0.044	4.943
Dollar Volume of Trades	139,184	8.3M	18.8M	0.0M	2.7M	5.4M	9.8M	1,570.1M
Number of Trades	139,184	5,240	14,348	7	1,345	2,815	5,745	950,784
	Quintile 3 Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	139,010	0.775	0.194	0.005	0.667	0.859	0.922	1.000
Percent Quoted Spread	139,010	0.018	0.033	0.000	0.007	0.012	0.019	3.687
Dollar Volume of Trades	139,010	37.6M	50.9M	0.0M	16.2M	27.8M	46.3M	4,236.7M
Number of Trades	139,010	14,577	25,721	5	5,120	9,755	17,342	1,986,649
	Quintile 4 Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	139,273	0.830	0.164	0.008	0.800	0.896	0.935	1.000
Percent Quoted Spread	139,273	0.008	0.012	0.000	0.003	0.005	0.009	0.487
Dollar Volume of Trades	139,273	153.5M	153.4M	0.0M	72.2M	119.7M	194.3M	14,100.0M
Number of Trades	139,273	32,783	36,820	59	14,249	25,524	41,738	3,485,211
	Quintile 5 Observations							
	Obs	Mean	SD	Min	25%	50%	75%	Max
Exchange Market Share	139,383	0.867	0.108	0.013	0.851	0.902	0.931	1.000
Percent Quoted Spread	139,383	0.003	0.005	0.000	0.001	0.002	0.004	0.226
Dollar Volume of Trades	139,383	1,714.1M	5,957.4M	0.3M	427.1M	726.2M	1,409.3M	376,000.0M
Number of Trades	139,383	137,773	206,702	1,161	51,324	86,649	153,925	13,458,361

Table IA.F.2: Summary Statistics by Exchange Market Share

Each observation is one ticker for one week. We present simple averages across observations with exchange market share less than 50% on the left column, and across observations with exchange share more than 50% on the right.

	Exchange Market Share $< 50\%$	Exchange Market Share $> 50\%$
Observations	80,746	613,559
Exchange Market Share	0.443	0.79
Percent Quoted Spread	0.031	0.026
Dollar Volume of Trades	40,283,547	429,702,482
Number of Trades	6,227	42,635

It shows that even among US-listed equities, many are OTC-dominated, and much of the remainder exhibit large OTC market shares.

3. *Correlation between Exchange Market Share and Spread*

Proposition IA.F.1 predicts a positive correlation between exchange market share and the exchange quoted spread *if* the observations vary substantively in the mass of informed traders μ but not in signal or label accuracy (α, θ, γ). We estimate the correlation within a narrow asset class, controlling for time (week) fixed effects. Specifically, we partition US exchange-listed equities into quintiles by average weekly dollar volume and estimate the correlation within each quintile. Our assumption is that across tickers within each dollar volume quintile in a given week, (i) similar amounts of information about individual traders (e.g., type of firm, reputation, past disclosures) are available, so the tickers have similar label accuracies (θ, γ), (ii) similar speculators are present, thus similar signal accuracies α, and (iii) there is significant variation in adverse selection risk among the tickers (due to, say, ticker-specific news).

Table IA.F.3 presents the regression estimates for log exchange market share on log quoted spreads.[8] All regressions control for week fixed effects, hence our estimates capture cross-sectional variation. Standard errors are clustered at the ticker level. Under each quintile, the left-most regression has no controls. Consistent with Proposition IA.F.1, the correlation between log quoted spread and log exchange market share is positive within every quintile. COVID or other time variation common across tickers do not drive our results, as we control for week fixed effects.

[8] Bogousslavsky and Collin-Dufresne (2022) reports a positive correlation between aggregate turnover and bid-ask spread. We examine the intensive margin of *where* trades occur and necessarily distinguish between the trades on exchanges versus over the counter. They examine the extensive margin of *how much* trading occurs in the aggregate.

Table IA.F.3: Dependent Variable: log(Exchange Market Share)

All regressions control for week fixed effects. Observations are weekly and include 3,210 non-ETF US-listed tickers that exist in both the first and last weeks of the 218-week sample, from January 2, 2017 to March 5, 2021. Trades outside of market hours are excluded. Standard errors are clustered at the ticker level. Corresponding t-statistics are shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Independent Variables	Quintile 1			Quintile 2			Quintile 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(percent quoted spread)	0.074***	0.076***	0.072***	0.046***	0.144***	0.069***	0.061***	0.143***	0.077***
	(11.61)	(11.05)	(11.66)	(4.32)	(15.57)	(7.52)	(3.62)	(10.10)	(5.71)
log(dollar volume)		0.005			0.129***			0.125***	
		(0.89)			(24.39)			(18.90)	
log(number of trades)			0.029***			0.123***			0.150***
			(5.74)			(19.27)			(14.49)
R^2	0.059	0.059	0.068	0.027	0.168	0.182	0.029	0.140	0.208
N		137,453			139,184			139,010	

Independent Variables	Quintile 4			Quintile 5		
	(10)	(11)	(12)	(13)	(14)	(15)
log(percent quoted spread)	0.087***	0.131***	0.095***	0.028***	0.038***	0.044***
	(5.74)	(9.65)	(8.16)	(2.61)	(2.98)	(3.20)
log(dollar volume)		0.095***			0.014	
		(10.30)			(1.57)	
log(number of trades)			0.146***			0.034***
			(13.08)			(2.83)
R^2	0.074	0.146	0.273	0.051	0.056	0.078
N		139,273			139,383	

A typical explanation for OTC trading is that illiquidity renders trading certain assets on exchanges impractical. We evaluate whether our results are genuinely separate from this liquidity mechanism and, if so, the relative magnitudes of the two mechanisms on the exchange market share. To this end, Table IA.F.3 presents regressions whose controls include one of two proxies for liquidity, log weekly total dollar volume ("dollar volume") or log weekly total number of trades. It shows that the quoted spread is positively correlated with the exchange market share after controlling for liquidity—in fact, the coefficient estimates seemingly become larger. Therefore, our findings are not explained by more liquid tickers being easier to trade on exchanges. Further, the coefficient estimates for the quoted spread and for the liquidity proxies are in the same magnitude, which suggests that our mechanism is not second-order to the illiquidity explanation.

References

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488–500.

Bernhardt, D., V. Dvoracek, E. Hughson, and I. M. Werner (2005): "Why Do Larger Orders Receive Discounts on the London Stock Exchange?" *Review of Financial Studies*, 18, 1343–1368.

Bessembinder, H. and K. Venkataraman (2004): "Does an Electronic Stock Exchange Need an Upstairs Market?" *Journal of Financial Economics*, 73, 3–36.

Biais, B. and R. C. Green (2007): "The Microstructure of the Bond Market in the 20th Century," *GSIA Working Papers.*

Bogousslavsky, V. and P. Collin-Dufresne (2022): "Liquidity, Volume, and Order Imbalance Volatility," *Journal of Finance, Forthcoming.*

Booth, G. G., J.-C. Lin, T. Martikainen, and Y. Tse (2002): "Trading and Pricing in Upstairs and Downstairs Stock Markets," *The Review of Financial Studies*, 15, 1111–1135.

Collin-Dufresne, P., B. Junge, and A. B. Trolle (2020): "Market Structure and Transaction Costs of Index CDSs," *The Journal of Finance*, 75, 2719–2763.

de Roure, C., E. Moench, L. Pelizzon, and M. Schneider (2021): "OTC Discount," .

DTCC (2018): "DTCC Enters Test Phase on Distributed Ledger Project for Credit Derivatives with MarkitSERV & 15 Leading Global Banks," Tech. rep., Depository Trust & Clearing Corporation.

——— (2020): "Project ION Case Study," Tech. rep., Depository Trust & Clearing Corporation.

Elwell, C. K., M. M. Murphy, and M. V. Seitzinger (2013): "Bitcoin: Questions, Answers, and Analysis of Legal Issues," Tech. Rep. 7-5700, Congressional Research Service.

GFI LLC (2022): "GFI Swaps Exchange LLC Rulebook," Tech. rep., GFI Swaps Exchange LLC.

Homer, S. (1975): "The Historical Evolution of Today's Bond Market," in *Explorations in Economic Research*, Cambridge, MA: National Bureau of Economic Research, Inc, vol. 2 of *NBER Books*, 378–389.

Irrera, A. (2017): "DTCC Completes Blockchain Repo Test," https://www.reuters.com/article/us-dtcc-blockchain-repos/dtcc-completes-blockchain-repo-test-idUSKBN1661L9.

Madhavan, A. and M. Cheng (1997): "In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets," *The Review of Financial Studies*, 10, 175–203.

Rose, A. (2014): "The Informational Effect and Market Quality Impact of Upstairs Trading and Fleeting Orders on the Australian Securities Exchange," *Journal of Empirical Finance*, 28, 171–184.

Selten, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

Smith, B. F., D. A. S. Turnbull, and R. W. White (2001): "Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects," *The Journal of Finance*, 56, 1723–1746.

Wang, C. (2023): "The Limits of Multi-Dealer Platforms," *Journal of Financial Economics*, 149, 434–450.

Westerholm, P. J. (2009): "Do Uninformed Crossed and Internalized Trades Tap into Unexpressed Liquidity? The Case of Nokia," *Accounting & Finance*, 49, 407–424.