Regulating Over-the-Counter Markets

Tomy Lee Chaojun Wang*

November 8, 2024

Abstract

Over-the-counter (OTC) trading thrives despite competition from exchanges. We let OTC dealers cream skim from exchanges in an otherwise standard Glosten and Milgrom (1985) framework. Restricting the dealer's ability to cream skim induces "cheap substitution": Some traders exit while others with larger gains from trade enter. Cheap substitution implies trading costs, trade volumes, and market shares are poor indicators for policy. In a benchmark case, restricting the dealer raises welfare only if trading cost increases, volume falls, and OTC market share is high. By contrast, the restriction improves welfare *whenever* adverse selection risk is low. A simple procedure implements the optimal Pigouvian tax.

---

*This paper was previously titled "Why Trade Over the Counter?" 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.

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

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

Over-the-counter (OTC) markets are opaque and difficult to access.[1] Yet, they host the vast majority of financial trades, as most assets are seldom traded on centralized exchanges. Dealers argue against regulatory intervention, citing the high OTC market shares and evidence of lower trading costs compared to exchanges.[2] Do the 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, while strictly raising 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 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

[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, Onur, Reiffen, and Zhu (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).

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,

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

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.[3] 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 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

[3] Proposition A.1 provides that the volume effect $m(\text{exiters})/m(\text{entrants})$ is finite for any distribution of hedging benefits whose pdf f is continuous and strictly positive in $[0, 1]$.

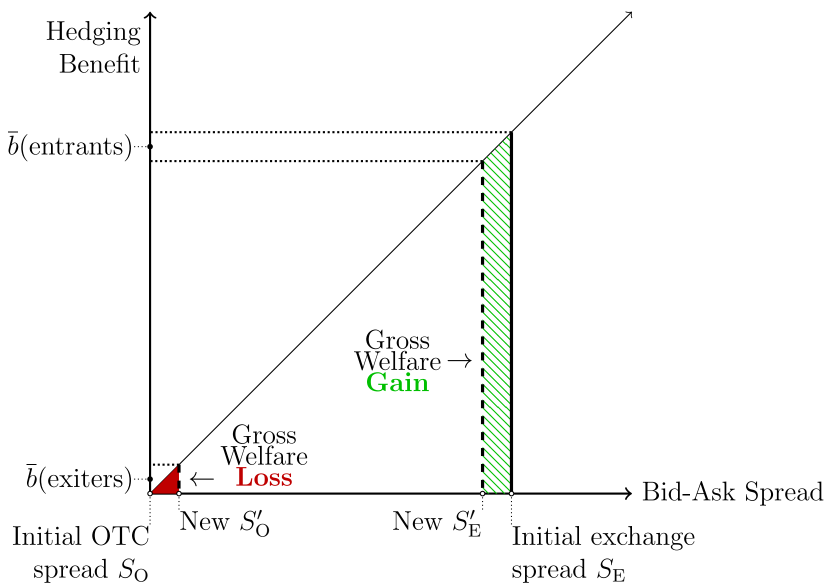


Figure 1: Restricting the OTC dealer can raise welfare

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).

intervention. Exchanges are indeed available for many OTC-dominated assets.[4] 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 $WSR := |S_{\text{E}}/S_{\text{O}} \div (dV_{\text{O}}/dV_{\text{E}})\,|$. The ratio of spreads $S_{\text{E}}/S_{\text{O}}$ captures the cheap substitution effect, and the ratio of changes in trade volumes $|dV_{\text{O}}/dV_{\text{E}}|$ 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 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

[4]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.

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. Electronic OTC trading platforms (e.g., MarktAxess, Bloomberg) have the capability to finetune the information revealed to dealers. Morover, several proposals to implement blockchain technology for recordkeeping of OTC trades are under consideration. Most such proposals would disseminate the 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 $\theta < 1$ or $\gamma < 1$, and that a trader's odds $\mathcal{O}_{LI}$ of being informed conditional on being labeled LI is strictly higher than the unconditional odds μ and the odds $\mathcal{O}_{LU}$ conditional on being LU. Precisely, $\mathcal{O}_{LI} = \frac{\theta\mu}{1-\gamma} > \mu > \frac{(1-\theta)\mu}{\gamma} = \mathcal{O}_{LU}$ or, equivalently, $\theta > 1 - \gamma$.

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.

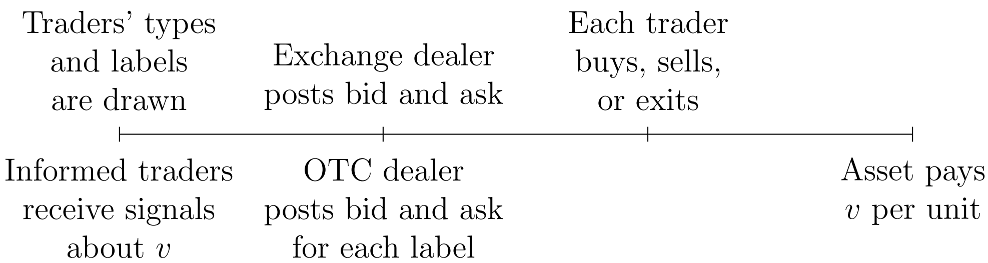


Figure 2: Timeline

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

Assumption 1. There exist neighborhoods of $x = 0$ and $x = 2\alpha - 1$ such that the hedging benefit distribution F has a pdf f that is analytic over these two neighborhoods.[5]

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.[6] The rule is equivalent to imposing a small cost

[5]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.

[6]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 $\rho > 0$.

on OTC trades, which can represent the inconvenience of soliciting prices that is absent on exchanges.

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 $V := V_{\text{O}} + V_{\text{E}}$ the aggregate trade volume.[7] *Average bid-ask spread* $\bar{S}$ is the volume-weighted average of bid-ask spreads in the OTC market and on the exchange,

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

We say "lower average spread $\bar{S}$" interchangeably with "higher aggregate volume V," because spread $\bar{S} \propto 1/V$ in equilibrium.[8]

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

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

[8]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.

a marginal increase in the ratio of informed to uninformed traders β:

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

Marginal welfare Δ_W is the decrease in welfare, $m(\text{exiters}) \times \bar{b}(\text{exiters})$, upon a marginal increase in β. It is equivalent to $\Delta_V(\beta)$ times the marginal exiters' hedging benefit $\bar{b}(\text{exiters})$:

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

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 and 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* $\bar{\mu} = \underline{\mu}$.[9]

(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

[9] 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."

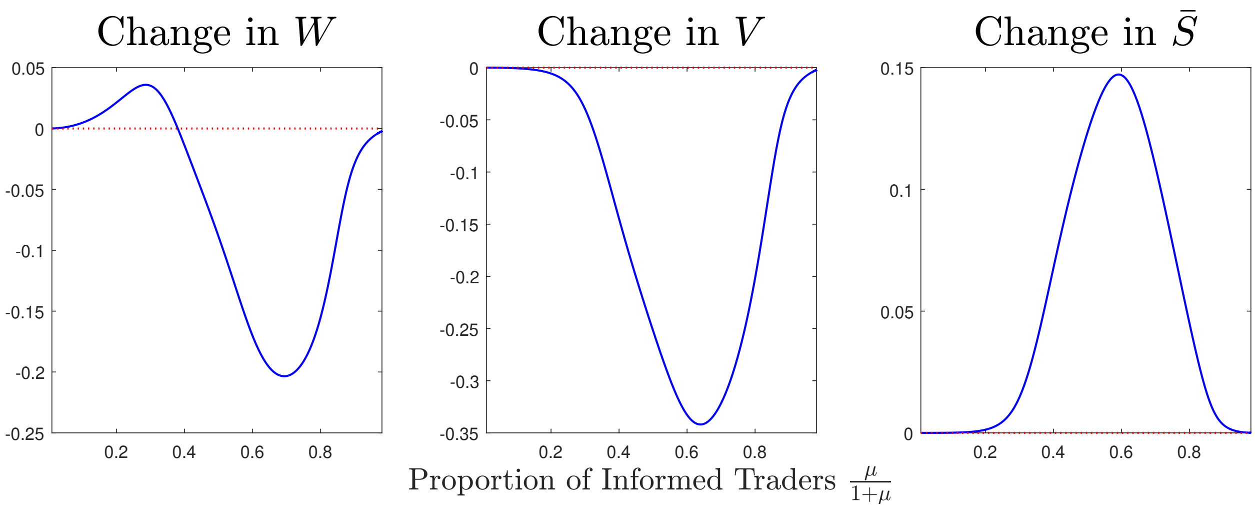


Figure 3: Effects of Restricting the OTC Dealer's Ability to Price Discriminate

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 $d\theta$ raises welfare when the mass μ of informed traders is low, because cheap substitution persists yet the volume effect vanishes. (The 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.[10] 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.

[10]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 $\beta_{\text{O}} = \frac{1-\theta}{\gamma}\mu$.

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 fraction of informed traders from the exchange to the OTC market,[11] which transfers the corresponding adverse selection cost from the exchange to the OTC dealer. The resulting change in the dealer's gross loss $V_{\text{M}}\,S_{\text{M}}$ is approximately the same in the two markets and equal to $V_{\text{M}}\,\mathrm{d}S_{\text{M}}$. Therefore, the masses of exiters and entrants are about equal at small μ: $m(\text{exiters}) = \gamma|\mathrm{d}S_{\text{O}}| \approx V_{\text{O}}|\mathrm{d}S_{\text{O}}| \approx V_{\text{E}}|\mathrm{d}S_{\text{E}}| \approx (1-\gamma)\mathrm{d}S_{\text{E}} = m(\text{entrants})$.[12] The volume effect $m(\text{exiters})/m(\text{entrants})$ is thus close to 1.

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 LU traders are equally likely to be informed, $\theta = 1-\gamma$. Figure 4 shows $\bar{b}(\text{entrants}) \in (S_{\text{N}}, S_{\text{E}})$ and $\bar{b}(\text{exiters}) \in (S_{\text{O}}, S_{\text{N}})$ 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.

---

[11] 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.

[12] This sentence uses the case of the uniform distribution. For analytical distributions whose $\lim_{x\to 0} f'(x) > 0$, the approximation is $m(\text{exiters}) = \lim_{x\to 0} f(x)\gamma|\mathrm{d}S_{\text{O}}| \approx \lim_{x\to 0} f(x)V_{\text{O}}|\mathrm{d}S_{\text{O}}| \approx \lim_{x\to 0} f(x)V_{\text{E}}|\mathrm{d}S_{\text{E}}| \approx \lim_{x\to 0} f(x)(1-\gamma)\mathrm{d}S_{\text{E}} = m(\text{entrants})$. 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.

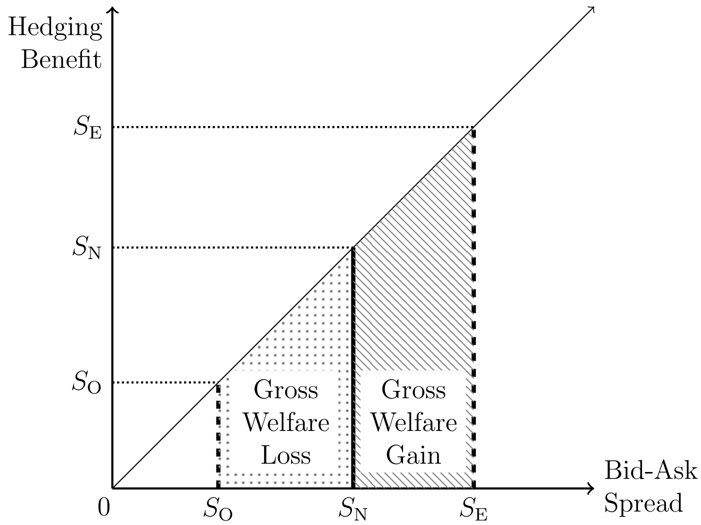


Figure 4: Effect of closing the OTC market on welfare

We mark the initial bid-ask spreads S_O and S_E, and the No-OTC spread S_N that prevails after the closure.

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.[13] 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 $\mathrm{Beta}(a, b)$ with parameters $a \leq 1, b \leq 1$. That a wide range of distributions satisfy (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 be such that*

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

[14]

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 $\frac{V_O}{V} < \widehat{M}$.

[13] The beta distribution $\mathrm{Beta}(a, b)$ (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 $[0, 1]$.

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

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 $\frac{V_{\mathrm{O}}}{V}$ to strictly decrease in μ (part (a)). Parts (b) and (c) are corollaries of part (a) and Proposition 1: 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.

[14] 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 $a, b > 0$.

A. Characterizing the Optimal Pigouvian Tax

Gross tax revenue T is transferred from the OTC dealer to the exchange dealer.[15] 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.[16] 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.

---

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

[16]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.

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:[17]

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

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

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

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^*.

---

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

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

Estimating the Weighted Spread Ratio

The volume term in the WSR, $|dV_{\text{O}}/dV_{\text{E}}|$, 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 $S_{\text{E}}/S_{\text{O}}$ 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 4 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.[18]

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;

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

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).[19] 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

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

[19] 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).

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.[20] 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 discovery 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.

[20]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.

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 $s \cdot [1 - F(s)] - (2\alpha - 1 - s)^+\beta$ over the spread s. 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.

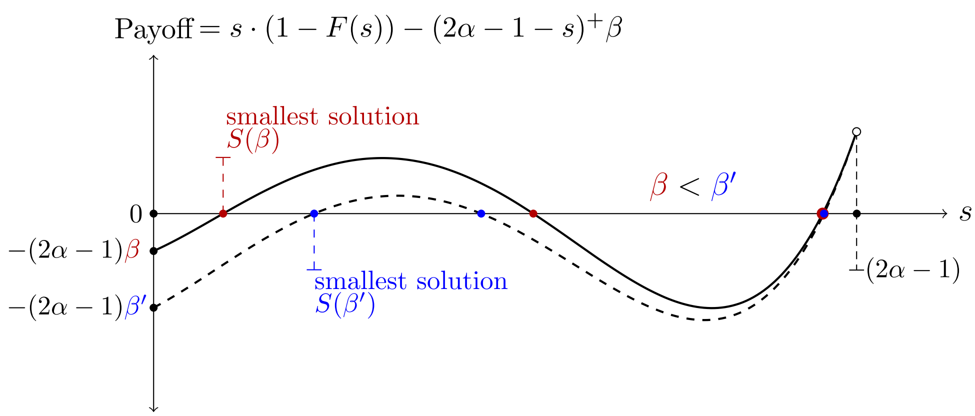


Figure A.1: Finding the equilibrium spread

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 $S(\beta_{\text{LU}})$, 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 $S_{\text{E}} \in (S(\beta_{\text{LU}}), S(\beta_{\text{LI}})]$, all LU traders choose the OTC market, whereas all LI traders choose the exchange. Then the informed ratio on the exchange $\beta_{\text{E}} = \beta_{\text{LI}}$. The exchange dealer thus earns zero profit if and only if she sets $S_{\text{E}} = S(\beta_{\text{LI}})$ in this case. If the exchange dealer sets $S_{\text{E}} \le S(\beta_{\text{LU}})$, then every trader chooses the exchange, implying $\beta_{\text{E}} = \mu > \beta_{\text{LU}}$, and thus the exchange dealer earns a non-zero profit. Therefore, in equilibrium, *(i)* the exchange spread is $S_{\text{E}} = S(\beta_{\text{LI}}) = S\left(\frac{\theta}{1-\gamma}\mu\right)$, 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 $\beta_{\text{O}} < \beta_{\text{E}}$ and the spread function $S(\beta)$ is strictly increasing in the informed ratio β, then $S_{\text{O}} < S_{\text{E}}$. □

Proposition A.1. *If the pdf f of the hedging benefit distribution F is continuous and strictly positive in $[0,1]$, then the volume effect $\frac{m(\textit{exiters})}{m(\textit{entrants})}$ associated with marginally reducing θ from $\theta = 1$ is finite.*

Proof of Proposition A.1. The equilibrium half bid-ask spread $S(\beta)$ is the smallest solution to the dealers' zero-profit condition,

$$s \cdot [1 - F(s)] - (2\alpha - 1 - s)^{+}\beta = 0, \tag{A.1}$$

which Figure A.1 illustrates.

The masses of entrants and exiters are:

$$\begin{aligned} m(\text{exiters}) &= f\left(S(\beta_{\text{O}})\right) S'\left(\beta_{\text{O}}^{+}\right) \mu \,\mathrm{d}\theta = f(0) S'\left(0^{+}\right) \mu \,\mathrm{d}\theta, \\ m(\text{entrants}) &= f\left(S(\beta_{\text{E}})\right) S'\left(\beta_{\text{E}}^{-}\right) \mu \,\mathrm{d}\theta, \end{aligned}$$

in which $S'\left(\beta^{\pm}\right)$ are the left and the right derivatives of $S(\beta)$ at β. Applying the Implicit

Function Theorem to (A.1), for any $\beta > 0$,

$$S'(\beta^-) = \frac{2\alpha - 1 - S(\beta)}{1 - F(S(\beta)) - S(\beta)f(S(\beta)) + \beta}.$$

Since a dealer's expected payoff, $s \cdot [1 - F(s)] - (2\alpha - 1 - s)^+\beta$, crosses zero for the first time at $s = S(\beta)$ from below, its derivative with respect to s at $s = S(\beta)$ must be non-negative,

$$1 - F(S(\beta)) - S(\beta)f(S(\beta)) + \beta \geq 0.$$

Then, $S'(\beta^-)$ is either infinite or strictly positive for any $\beta > 0$. Similarly, $S'(0^+) > 0$. Therefore, $m(\text{exiters})/m(\text{entrants})$ is finite. □

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

$$(1-\gamma)\int_{\frac{\theta_l}{1-\gamma}\mu}^{\frac{\theta_h}{1-\gamma}\mu} \Delta_V(\beta)\,\mathrm{d}\beta - \gamma \int_{\frac{1-\theta_h}{\gamma}\mu}^{\frac{1-\theta_l}{\gamma}\mu} \Delta_V(\beta)\,\mathrm{d}\beta. \tag{A.2}$$

Similarly, the change in welfare is

$$\underbrace{(1-\gamma)\int_{\frac{\theta_l}{1-\gamma}\mu}^{\frac{\theta_h}{1-\gamma}\mu} \Delta_W(\beta)\,\mathrm{d}\beta}_{\text{Gross welfare gain}} - \underbrace{\gamma\int_{\frac{1-\theta_h}{\gamma}\mu}^{\frac{1-\theta_l}{\gamma}\mu} \Delta_W(\beta)\,\mathrm{d}\beta}_{\text{Gross welfare loss}}. \tag{A.3}$$

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

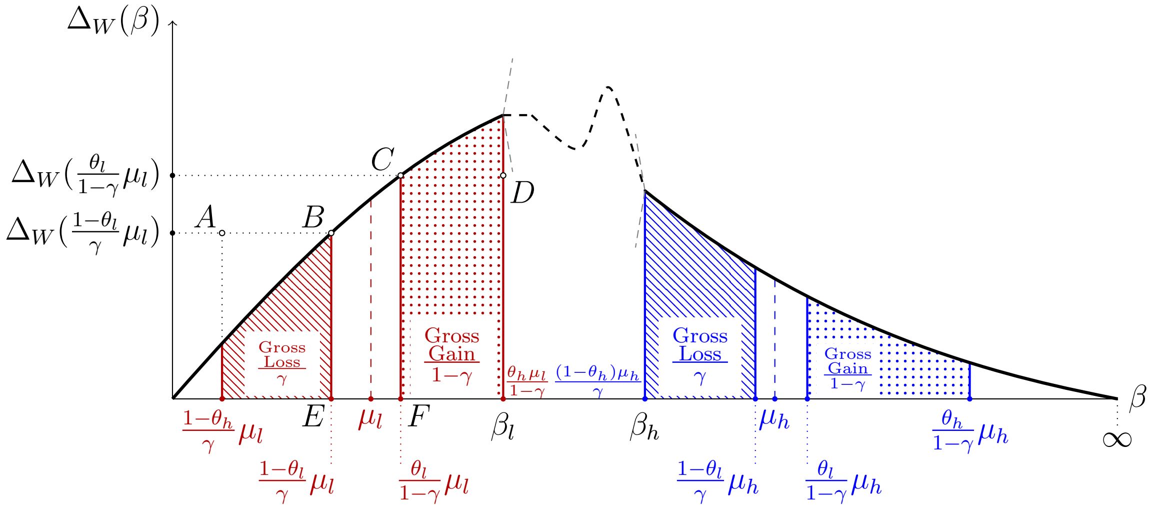


Figure A.2: Generic Δ_W

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.3) (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.3) (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

$$\underbrace{(1-\gamma_h)\int_{\frac{\theta}{1-\gamma_l}\mu}^{\frac{\theta}{1-\gamma_h}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\substack{\text{Entry by uninformed}\\ \text{LI-traders}}} - \underbrace{\gamma_h \int_{\frac{1-\theta}{\gamma_h}\mu}^{\frac{1-\theta}{\gamma_l}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\substack{\text{Exit by uninformed}\\ \text{LU traders}}} - \underbrace{(\gamma_h-\gamma_l)\int_{\frac{1-\theta}{\gamma_l}\mu}^{\frac{\theta}{1-\gamma_l}\mu} \Delta_V(\beta)\,\mathrm{d}\beta}_{\text{Exit by relabeled traders}}\,. \tag{A.4}$$

Similarly, the change in welfare is

$$(1-\gamma_h)\int_{\frac{\theta}{1-\gamma_l}\mu}^{\frac{\theta}{1-\gamma_h}\mu} \Delta_W(\beta)\,\mathrm{d}\beta - \gamma_h \int_{\frac{1-\theta}{\gamma_h}\mu}^{\frac{1-\theta}{\gamma_l}\mu} \Delta_W(\beta)\,\mathrm{d}\beta - (\gamma_h-\gamma_l)\int_{\frac{1-\theta}{\gamma_l}\mu}^{\frac{\theta}{1-\gamma_l}\mu} \Delta_W(\beta)\,\mathrm{d}\beta. \tag{A.5}$$

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

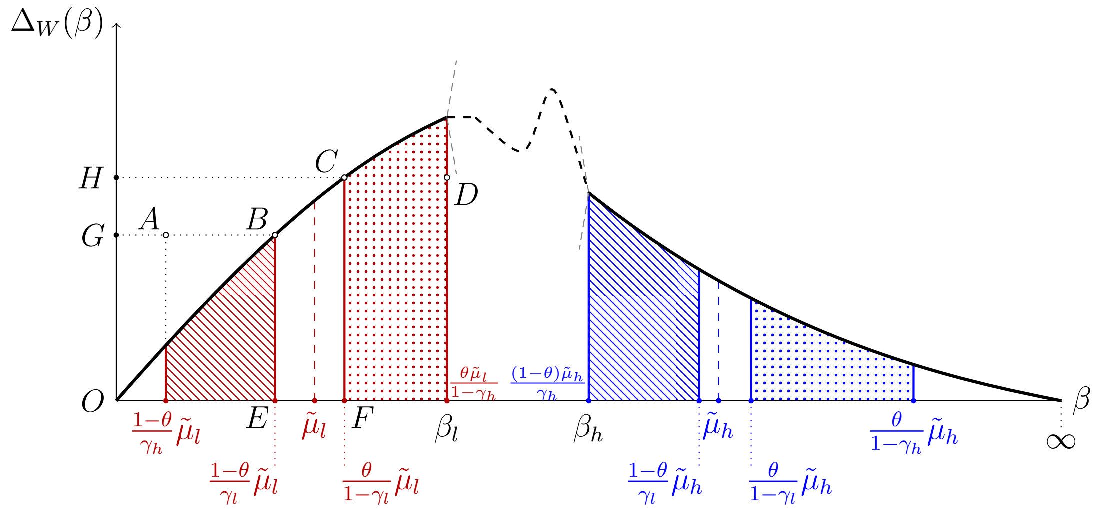


Figure A.3: Generic Δ_W

in (A.5) (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.5) has a strict lower bound

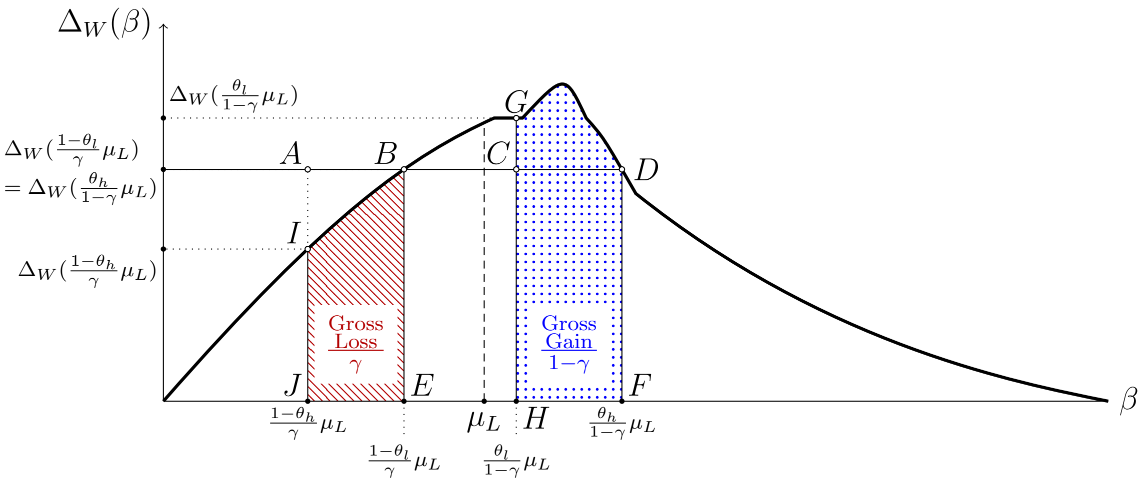
$$(\gamma_h-\gamma_l)\cdot||HCFO||,$$

The third term in (A.5) equals $(\gamma_h-\gamma_l)\cdot||BCFE||$. Then (A.5) is strictly more than

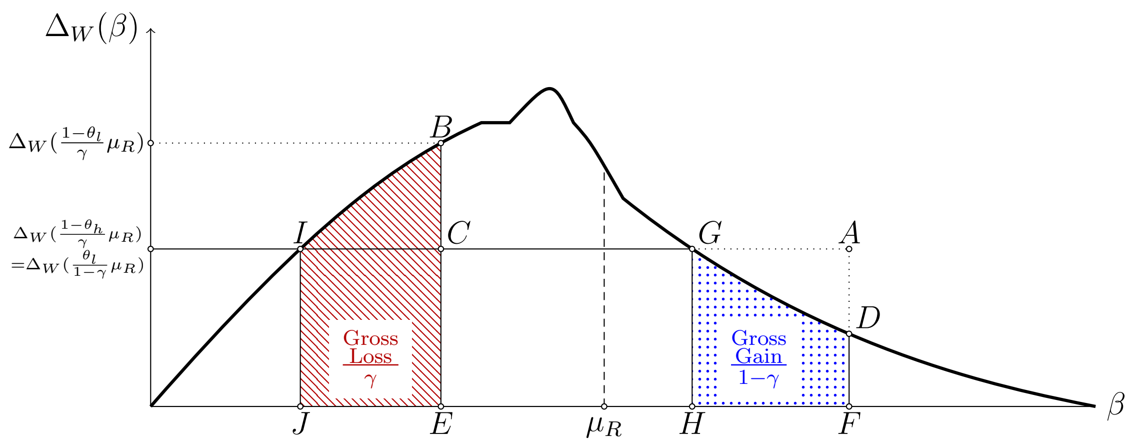
$$(\gamma_h-\gamma_l)\,(||HCFO||-||GBEO||-||BCFE||)>0.$$

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 $\lim_{\beta\downarrow0}\Delta_W(\beta)=\lim_{\beta\uparrow\infty}\Delta_W(\beta)=0$, $0<\mu_L<\mu_R<\infty$. 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.3) 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.3) 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.3) ("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.3) ("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.3) is strictly positive for all $\mu \leq \mu_L$. Similarly, (A.3) is strictly negative for all $\mu \geq \mu_R$.

To prove *(ii)* that (A.3) is strictly decreasing over $\mu \in (\mu_L, \mu_R)$, we take the derivative of (A.3) with respect to μ,

$$\underbrace{\left(\theta_h \cdot ||\overline{DF}|| - \theta_l \cdot ||\overline{GH}||\right)}_{\text{Derivative of the gross welfare gain}} - \underbrace{\left((1-\theta_l) \cdot ||\overline{BE}|| - (1-\theta_h) \cdot ||\overline{IJ}||\right)}_{\text{Derivative of the gross welfare loss}}. \tag{A.6}$$

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.6) 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).

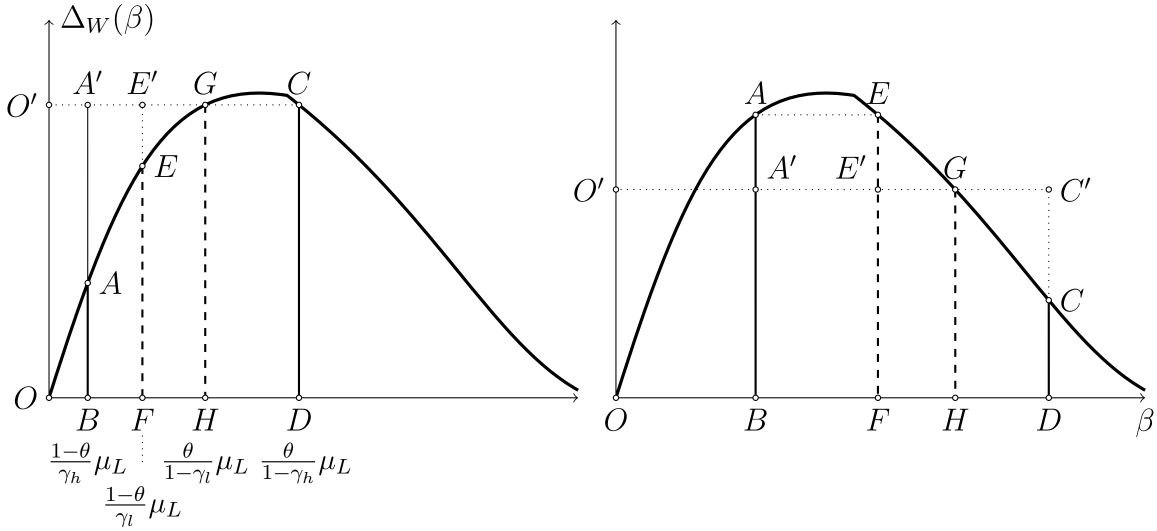


Figure A.5: Decrease in label accuracy γ

For $\mu \leq \tilde{\mu}_L$, the change in welfare (A.5) is strictly more than

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

For $\mu \geq \tilde{\mu}_R$, (A.5) is strictly less than

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

For $\mu \in (\tilde{\mu}_L, \tilde{\mu}_R)$, the derivative of (A.5) 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.2) 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.2) 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.

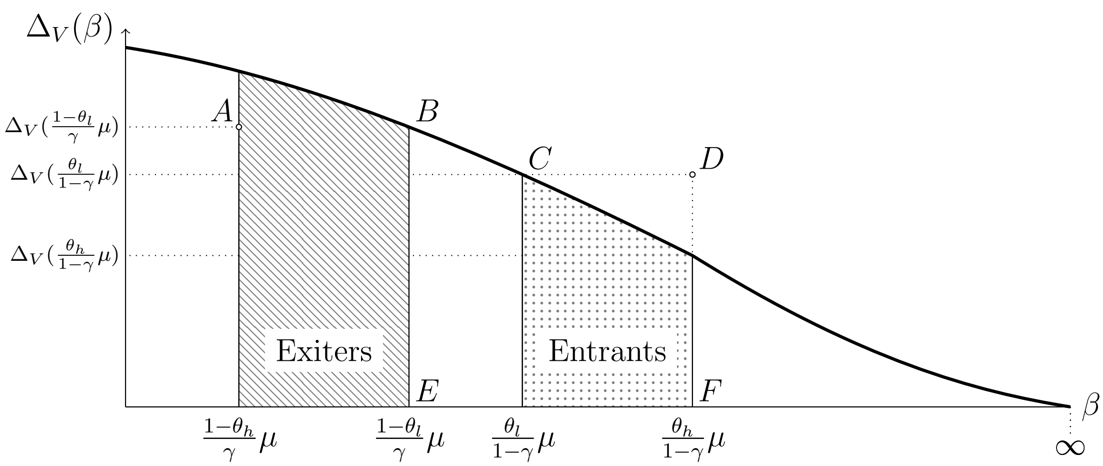


Figure A.6: Decreasing Δ_V

As the label accuracy γ falls from γ_h to γ_l, the change (A.4) 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. □

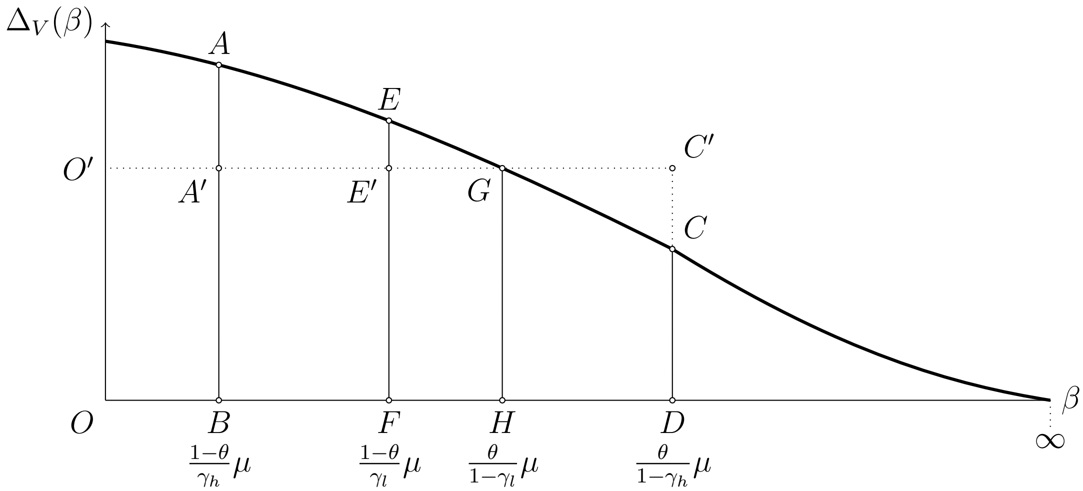


Figure A.7: Decrease in label accuracy γ

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

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

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

$$\beta(x) = \frac{(1 - F(x))x}{2\alpha - 1 - x}. \tag{A.8}$$

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.7) and (A.8), 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. □

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.

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.