Mixology: Order flow segmentation design*

Joshua Mollner†

November 14, 2024

Click here for latest version

Abstract

I analyze the welfare consequences of segmentation in financial markets. Venues vary in their mixture of information-motivated versus liquidity-motivated order flow. In a simple model, I consider the combinations of information-motivated investor welfare and liquidity-motivated investor welfare that can be achieved by some segmentation. This set's Pareto frontier can (under certain conditions) be implemented by a simple class of segmentations, in which a fraction of information-motivated flow is segregated, while the remainder pools with liquidity-motivated flow. These results call into question the wisdom of the current regulatory framework as it applies to segmentation, e.g., in U.S. equities.

Keywords: adverse selection, cream-skimming, segmentation, third-degree price discrimination

JEL: D47, D82, D83, G14, G18

1 Introduction

In modern financial markets, a great deal of infrastructure is geared toward segmenting different types of traders. In U.S. equities, many dark pools and other alternative trading systems (ATSs)—being exempt from the fair-access requirements that apply to exchanges—prohibit or limit the access of certain types of traders.[1] Retail brokers route the vast majority of their orders directly to market makers who, knowing that these orders originate from retail traders, execute them at better prices than those on-exchange. In fact, exchanges themselves operate retail liquidity programs, whereby certain limit orders can be made accessible only to market orders from retail traders. Recently, the SEC has proposed order-by-order auctions for retail orders, which would maintain retail-order segmentation, albeit via a different mechanism.

Segmentation also arises when trading occurs outside anonymous exchanges, as for asset classes that trade over-the-counter (bonds, swaps, etc.) or for equity blocks traded in the so-called "upstairs" market. In such settings, different traders might obtain different prices for the same trade,

First version: September 11, 2024. For helpful comments, I thank Nemanja Antić, Markus Baldauf, Ian Ball, Robert Battalio, Jean-Edouard Colliard, Darrell Duffie, Piotr Dworczak, Timothy Feddersen, Pete Kyle, Marzena Rostek, Alex Smolin, Chaojun Wang, and Bart Zhou Yueshen. I thank Elena Stella for research assistance. No financial relationships or other potential conflicts of interest apply.

†Kellogg School of Management, Northwestern University. E-mail: joshua.mollner@kellogg.northwestern.edu.

[1] Reg NMS Rule 610 requires exchanges to provide fair and non-discriminatory access to quotations. In contrast, Reg ATS Rule 301(b)(5) stipulates that an ATS need not do the same, provided that it accounts for no more than 5% of trading volume in any stock.

based on their reputation for being informed or uninformed, based on historical trading relationships, or based on other factors.

One question that springs forth from these facts is: which form of segmentation would a social planner prefer? That question is not only theoretically interesting but also practically relevant. The form and extent of segmentation that prevails in a market is subject, at least in part, to regulatory control. Indeed, many jurisdictions permit significantly less segmentation than the U.S. For example, Canadian and Australian equities markets have "trade-at" rules, which curtail off-exchange trading by prohibiting it not only at prices worse than those quoted on-exchange but also at prices that are not meaningfully better. In the U.S., the SEC once tested a trade-at rule on a subset of stocks, as a component of their Tick Size Pilot. Several other policy levers are, in principle, available. Regulators could strengthen or weaken the conditions under which trading venues must provide fair access. They could limit the allowable forms of discriminatory access. They could regulate the frequency with which brokers are allowed to route the orders of various customers to various venues. They could also vary what trader characteristics can or cannot be made observable and hence potential bases for price discrimination.

In short, policymakers have many available levers with which they can influence segmentation. But determining which of those levers to throw requires first answering the more basic question about which segmentations are desirable in financial market settings. This paper's main contribution is to provide, within a simple model, a sharp answer to that question.

Model. I consider these issues within a simple model that is in the spirit of Glosten and Milgrom (1985). A single investor with unit demand will arrive and will be either information-motivated (with probability α) or liquidity-motivated (with probability $1-\alpha$). If he is information-motivated, then he knows the asset's common-value component, whose magnitude is drawn from a distribution F_I. If he is liquidity-motivated, then he has a private-value component, whose magnitude is drawn from a distribution F_L.

Segmentation is modeled as by exogenous probabilities that specify, for each of these two investor types, the probability the investor is assigned to each of a set of venues. The interpretation is that there is an (unmodeled) planner who can control these probabilities, and the goal is to understand what welfare outcomes can be feasibly generated by the different choices that this planner might make. Competitive market makers, understanding these probabilities, set a bid-ask spread on each venue. A sufficient statistic for a venue's spread is the 'toxicity' of its order flow: the ratio of information-motivated flow to liquidity-motivated flow. Spreads are larger when order flow is more toxic.

Any given segmentation determines the expected welfare of the investor *(i)* conditional on being information-motivated, denoted w_I, and *(ii)* conditional on being liquidity-motivated, denoted w_L. It is straightforward to derive the welfare outcomes (w_I, w_L) produced by two extreme forms of segmentation:

- First, suppose that liquidity-motivated and information-motivated orders are fully separated. A no-trade theorem would apply to the venues featuring only information-motivated order flow, implying $w_I = 0$. On the other hand, liquidity-motivated investors would trade at a spread of zero, meaning that they obtain all available gains from trade, implying $w_L = \int v \, dF_L(v)$.

- Second, suppose that liquidity-motivated and information-motivated orders are fully pooled. In this case, both investor types will trade at the same spread, denoted s_{pool}. It follows that $w_I = \int [v - s_{pool}]^+ \, dF_I(v)$ and $w_L = \int [v - s_{pool}]^+ \, dF_L(v)$.

I then characterize the welfare pairs (w_I, w_L) that might arise from *any* form of segmentation—not only from the two extremes discussed above. Clearly, intermediate forms of segmentation interpolate between these extremes. What is potentially more surprising is that some forms of segmentation are Pareto-efficient, while others are Pareto-inefficient.

Results (the uniform case). It is useful to begin by considering the case in which F_I and F_L both correspond to the uniform distribution on $[0, 1]$. Assuming $F_I = F_L$ means that information-motivated and liquidity-motivated investors have identical elasticities with respect to the spread, which is a natural benchmark case. A uniform distribution makes the analysis simpler and more transparent.

When segmentation occurs in practice, it often takes a particular form. Often, some liquidity-motivated order flow are segmented into various non-exchange venues (or secondary exchanges), leaving relatively more information-motivated flow on the exchanges (or the primary exchange). See, for example, Easley, Kiefer and O'Hara (1996); Bessembinder and Kaufman (1997); Goldstein, Shkilko, Van Ness and Van Ness (2008); Hatheway, Kwan and Zheng (2017); Comerton-Forde, Malinova and Park (2018). Because liquidity-motivated flow is more profitable for a market maker (at a fixed spread), this form of segmentation is sometimes referred to as "cream-skimming" by these non-exchange venues. To capture cream-skimming in the model, assume that all information-motivated orders go to a certain venue, while only a β_L-fraction of liquidity-motivated orders go to that venue, with the remainder going elsewhere. Increasing β_L from 0 to 1 interpolates between full-separation and full-pooling. Despite the prevalence of cream-skimming segmentation in practice, the analysis indicates that it is an undesirable form a segmentation: these cream-skimming outcomes form the Pareto-inferior portion of the feasible set.

What are the Pareto-dominant forms of segmentation? They constitute the opposite—what I call "reverse cream-skimming"—whereby all liquidity-motivated orders go to a certain venue, while only a β_I-fraction of information-motivated orders go to that venue, with the remainder going elsewhere. Increasing β_I from 0 to 1 also interpolates between full-separation and full-pooling. Yet, it is a different interpolation than the "cream-skimming interpolation" discussed above, and it implements different welfare outcomes. All Pareto-dominant welfare outcomes are implemented by reverse cream-skimming segmentation.

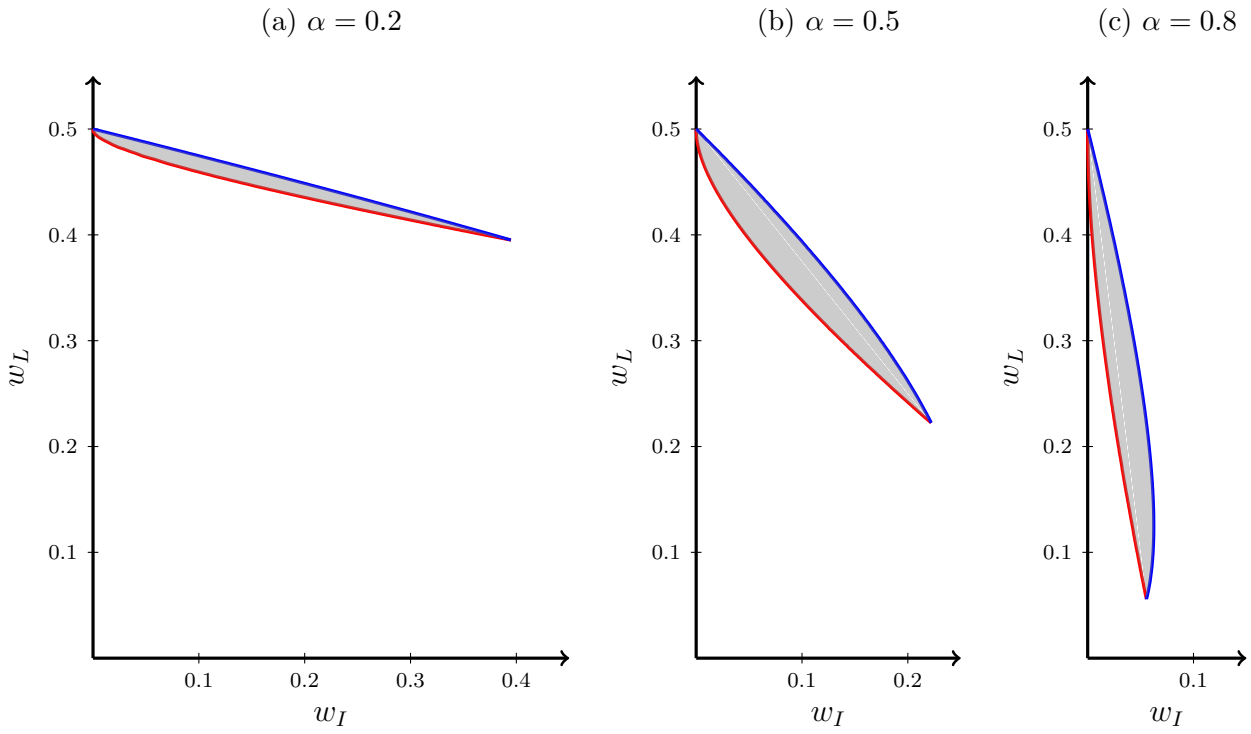
The welfare outcomes that can be feasibly implemented form a convex set, consisting of all points between the cream-skimming outcomes and the reverse cream-skimming outcomes. Figure 1 illustrates.

To state some of these results more formally, given a set $\mathcal{X} \subset \mathbb{R}^2$, define $\mathcal{X}^{pareto}$ as its Pareto frontier (i.e., the elements that are not Pareto-dominated by any other element of $\mathcal{X}$). Likewise, define $\mathcal{X}^{anti\text{-}pareto}$ as its Pareto anti-frontier (i.e., the elements that do not Pareto-dominate any other element of $\mathcal{X}$). Finally, let $\mathcal{CS}$ denote the set of cream-skimming outcomes, $\mathcal{RCS}$ the set of reverse cream-skimming outcomes, and $\mathcal{F}$ the entire set of feasible outcomes. In the uniform case, *(i)* $\mathcal{F}$ is the area between $\mathcal{CS}$ and $\mathcal{RCS}$, *(ii)* $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$, and *(iii)* $\mathcal{F}^{anti\text{-}pareto} = \mathcal{CS}$.[2] Because a planner need only consider how to implement Pareto-efficient segmentations, the main result from the perspective of practical relevance is clearly *(ii)*. The other results, *(i)* and *(iii)*, are primarily of theoretical interest.

Robustness and extensions. Versions of the results obtained for the uniform case hold for other distributional assumptions. For example, Appendix B considers the case in which F_I is a point

[2] It is not necessarily true that $\mathcal{F}^{pareto} = \mathcal{RCS}$, because some elements of $\mathcal{RCS}$ may be Pareto-dominated. See, for example, Figure 1(c).

Figure 1: The feasible welfare set



The feasible sets for $F_I = F_L = U[0,1]$ and $\alpha \in \{0.2, 0.5, 0.8\}$. In each panel, the red (blue) curve depicts the welfare pairs implementable by cream-skimming (reverse cream-skimming), and the gray area represents the welfare pairs implementable by all forms of segmentation.

mass on some $\sigma \in (0,1]$, while F_L remains $U[0,1]$. All the same qualitative conclusions obtain: *(i)*, *(ii)*, and *(iii)*.

Section 4 relaxes the specific distributional assumptions on F_I and F_L. I provide conditions under which the main result, *(ii)* extends.

Section 3.4 extends the model in a different direction, by weakening the assumption that segmentations can be based directly on the investor's trading motive. Concretely, suppose there are two imperfectly-informative labels: likely-information (`LI`) and likely-liquidity (`LL`). Conditional on being information-motivated (liquidity-motivated), the investor is labeled `LI` with probability ρ_I (ρ_L) and `LL` otherwise, where it is natural to assume $\rho_I \geq \rho_L$. Segmentations can be based on these labels, although not on the investor's type directly. For this analysis, I retain the assumption that $F_I = F_L = U[0,1]$. Relative to the baseline model, fewer segmentations are now possible—indeed, if $\rho_I = \rho_L$, then the labels are completely uninformative, and full-pooling is the only feasible outcome. With less scope for segmentation, the feasible set shrinks. Yet, versions of all the same qualitative conclusions obtain—in particular, *(i)*, *(ii)*, and a weakened version of *(iii)*: $\mathcal{F}^{\textit{anti-pareto}} \subseteq \mathcal{CS}$.[3]

Implications. Taken at face value, these results have considerable implications for policy. In U.S. equities, for example, some forms of segmentation are allowed, while other forms are not.

[3]For this analysis, I adjust the definitions in the natural way: a cream-skimming (reverse cream-skimming) segmentation is one where all `LI` (`LL`) orders go to a certain venue, while only a fraction of `LL` (`LI`) go there, with the remainder going elsewhere.

What is allowed, broadly speaking, is advantageous treatment for small fractions of orders (where a primary reason to provide advantageous treatment is because such orders are particularly likely to be liquidity-motivated). One example is the price improvement that retail orders tend to receive. Indeed, retail orders account for a relatively small fraction of overall market volume (depending on the calculation, about 20%). Another example is that ATSs are allowed to discriminate in terms of whom they provide access to, provided they are small (in the sense of accounting for less than 5% of trading volume in a stock). If an ATS grows bigger, then it is required to provide fair access.

What is not allowed, then, is the opposite: *dis*advantageous treatment for small fractions of orders (where a primary reason to provide disadvantageous treatment is because such orders are particularly likely to be information-motivated). For example, an exchange like Nasdaq or NYSE cannot block or charge additional fees to traders with a history of informed trading. The Access Rule (Reg NMS Rule 610) forbids it.

Effectively, what this means is that under the current regulatory framework, segmentation is geared more toward cream-skimming and less toward reverse cream-skimming. This approach, however, is precisely the opposite of what the model recommends.

Related literature. Conceptually, my approach resembles that of Bergemann, Brooks and Morris (2015). They characterize the combinations of consumer surplus and producer surplus that are feasible, depending on the information a monopolist might have about consumers. Each information structure corresponds to a segmentation, given which the monopolist will do third-degree price discrimination. Both their paper and mine characterize a two-dimensional feasible set. A difference is what is on the axes. Another difference is the setting: they consider a canonical product-market setting, whereas I consider a canonical microstructure setting, in which adverse selection plays an important role. A final difference is interpretation. They have in mind that segmentations are determined by the monopolist's information. My primary interpretation is that segmentations are determined by regulatory actions that affect how orders from investors of different types are routed to various venues. That said, my analysis could be reinterpreted through the lens of information: suppose there is just a single venue, but that market makers can price discriminate based on their information about the investor (as, e.g., in over-the-counter markets). Characterizing the feasible welfare set for that regime is the exact same problem. The analysis therefore also connects to Bayesian persuasion (Kamenica and Gentzkow, 2011), the literature on which is surveyed by Kamenica (2019) and Bergemann and Morris (2019).

The analysis also connects to theoretical analyses of segmentation in financial markets. Most closely related are models in which segmentation occurs because some market makers can condition access or pricing on observable characteristics (Easley, Kiefer and O'Hara, 1996; Battalio and Holden, 2001; Baldauf, Mollner and Yueshen, 2024; Lee and Wang, 2023). In other models, segmentation occurs because trader types differ in their need for immediacy (Admati and Pfleiderer, 1988, 1989; Hendershott and Mendelson, 2000; Brolley and Cimon, 2020), in the market access frictions that they face (Hoffmann, 2016), or because they have different perspectives on the tradeoffs between price and volume (Easley and O'Hara, 1987; Seppi, 1990) or between price and execution probability (Zhu, 2014). In still other models, segmentation occurs via reputation: some market makers can condition their current pricing on previous trading outcomes (Benveniste, Marcus and Wilhelm, 1992; Desgranges and Foucault, 2005). Many of these models compare the equilibrium with segmentation to a single counterfactual without any segmentation. I contribute to this literature by analyzing what is feasible given *all* possible segmentations.

Other papers model ways in which a change in the number of venues can affect the aggregate toxicity of order flow, although not differentially across venues—hence, without introducing

segmentation (Chowdhry and Nanda, 1991; Dennert, 1993; Baldauf and Mollner, 2021; Chen and Duffie, 2021).

Finally, my analysis also relates to empirical studies of segmentation in financial markets. These studies analyze segmentation between primary and secondary markets (Easley, Kiefer and O'Hara, 1996; Bessembinder and Kaufman, 1997; Barclay, Hendershott and McCormick, 2003; Goldstein, Shkilko, Van Ness and Van Ness, 2008), between lit and dark markets (Comerton-Forde and Putniņš, 2015; Degryse, De Jong and Kervel, 2015; Hatheway, Kwan and Zheng, 2017; Comerton-Forde, Malinova and Park, 2018), or between upstairs and downstairs markets (Madhavan and Cheng, 1997). All of these papers document ways in which different venues tend to attract different mixtures of information-motivated versus liquidity-motivated traders.

2 Model

2.1 Setup

A single risky asset will be traded by a single, unit-demand investor at a single point in time. The investor might be either information-motivated or liquidity-motivated. He will be assigned to one of several trading venues, where bid-ask spreads are set by competitive market makers. Segmentation occurs when the venues differ in their ratios of information- to liquidity-motivated order flow.

Asset. There is a single risky asset. Its common-value component is equally likely to be positive or negative, and its magnitude is an independent draw from a distribution F_I with support contained in $[0, 1]$.

Investor. A single risk-neutral investor will arrive and might place a market order to trade one share. With exogenous probability $\alpha \in (0, 1)$, he is information-motivated, in which case he privately knows the common-value component. With complementary probability $1 - \alpha$, he is liquidity-motivated, in which case he has a private-value component for trading one unit (and knows nothing about the common-value value).[4] This private-value component is equally likely to be positive or negative, and its magnitude is an independent draw from a distribution F_L with support contained in $[0, 1]$. For brevity, I will sometimes say that the investor's type is $k \in \{I, L\}$, where type-I means information-motivated, and type-L means liquidity motivated.

Venues. Venues are indexed $j \in \{1, \ldots, J\}$. Before the investor arrives, risk-neutral market makers compete to quote a bid and an ask at each of these venues. For each venue j, let s_j denote its half-spread (henceforth, simply 'spread'), which will be endogenously determined in equilibrium.

Segmentation. Conditional on being type $k \in \{I, L\}$, the investor will be assigned to venue j with exogenous and commonly-known probability δ_{kj}. Define $\boldsymbol{\delta}_k = (\delta_{k1}, \ldots, \delta_{kJ})$. By definition, $\boldsymbol{\delta}_k \in \Delta^J$. The interpretation is that an unmodeled planner can control $(\boldsymbol{\delta}_I, \boldsymbol{\delta}_L)$.

Timing. First, market makers post their quotes. Second, Nature determines the investor's type, the investor's venue assignment, the asset's common-value component, and (if the investor is

[4] Thus, a liquidity-motivated investor who buys $x \in \{-1, 0, 1\}$ shares at the price p when the common-value component is v_I and his private-value component is v_L receives utility $(v_I + v_L - p)x$. An information-motivated investor who buys x shares at the price p when the common-value component is v_I receives utility $(v_I - p)x$.

liquidity-motivated) his private-value component. Third, the investor chooses whether to buy, sell, or hold.

2.2 Spreads, welfare, and the feasible set

By symmetry, market makers quote symmetrically about zero, meaning that the ask (bid) on venue j will be s_j ($-s_j$). If assigned to venue j, and if information-motivated (liquidity-motivated), the investor will choose to trade if the magnitude of the common-value component (his private-value component) exceeds s_j.

To avoid needing to specify how liquidity-motivated investors break their indifference, I assume throughout that F_L is atomless. It does not matter for the analysis how information-motivated investors break their indifference. Yet, to avoid degenerate solutions, I assume that F_I is not simply a point mass on zero.

Spreads. This implies the following zero-profit condition for venue j, which, owing to Bertrand competition among the market makers, the half-spread s_j must solve

$$\underbrace{(1-\alpha)\delta_{Lj}\big[1-F_L(s_j)\big]s_j}_{\text{gains from type-}L\text{ investors}} = \underbrace{\alpha\delta_{Ij}\int_{s_j}^{1}(v-s_j)\,dF_I(v)}_{\text{losses to type-}I\text{ investors}}. \tag{1}$$

The left-hand side of (1) represents the market maker's gains from trading with liquidity-motivated investors on venue j. The investor is liquidity-motivated with probability $1-\alpha$, and conditional on that is assigned to venue j with probability δ_{Lj}. He then trades only if his private-value component exceeds the half-spread, which occurs with probability $1-F_L(s_j)$, and conditional on trading the market maker earns the half-spread s_j. The right-hand side represents the market maker's losses from trading with information-motivated investors on venue j. The investor is information-motivated with probability α, and conditional on that is assigned to venue j with probability δ_{Ij}. When he trades, the market maker loses the difference between the magnitude of the common-value component and the half-spread, and because he trades only when this difference is positive, the market maker expects to lose $\int_{s_j}^{1}(v-s_j)\,dF_I(v)$.

By standard arguments, s_j must in equilibrium be the smallest solution to (1).

Welfare. Conditional on being information-motived and assigned to venue j, the investor's expected surplus coincides with the market maker's expected loss: $\int_{s_j}^{1}(v-s_j)\,dF_I(v)$. Conditional on being liquidity-motived and assigned to venue j, the investor's expected surplus has a similar form: $\int_{s_j}^{1}(v-s_j)\,dF_L(v)$. It follows that the expected surplus of an information-motivated investor is

$$w_I := \sum_{j=1}^{J}\delta_{Ij}\int_{s_j}^{1}(v-s_j)\,dF_I(v). \tag{2a}$$

Likewise, the expected surplus of a liquidity-motivated investor is

$$w_L := \sum_{j=1}^{J}\delta_{Lj}\int_{s_j}^{1}(v-s_j)\,dF_L(v). \tag{2b}$$

The feasible set. Define the feasible welfare set as the set of pairs (w_I, w_L) that can be achieved by some segmentation. Formally,

$$\mathcal{F} := \left\{(w_I, w_L) \middle| \boldsymbol{\delta}_I \in \Delta^J, \boldsymbol{\delta}_L \in \Delta^J\right\}.$$

An initial observation is that the feasible set must be convex.

Proposition 1. *$\mathcal{F}$ is convex.*

To prove this result, suppose that the welfare pairs (w_I', w_L') and (w_I'', w_L'') are respectively implemented by segmentations $(\boldsymbol{\delta}_I', \boldsymbol{\delta}_L') \in \Delta^{J'}$ and $(\boldsymbol{\delta}_I'', \boldsymbol{\delta}_L'') \in \Delta^{J''}$. It is then straightforward to construct, for any $t \in (0,1)$, a segmentation $(\boldsymbol{\delta}_I''', \boldsymbol{\delta}_L''') \in \Delta^{J'+J''}$ that implements the welfare pair $t(w_I', w_L') + (1-t)(w_I'', w_L'')$.[5] For full details, see the proof in Appendix A.

2.3 Remarks

One motivation for considering w_I and w_L separately—as opposed to total welfare $w_T := \alpha w_I + (1-\alpha) w_L$—is because a social planner might put different Pareto weights on different trader types. On the one hand, a social planner might overweight liquidity-motivated investor welfare. Indeed, much of Gary Gensler's rulemaking as the current SEC Chair has indicated a disproportionate emphasis on minimizing the transaction costs of retail traders, who tend to be liquidity-motivated rather than information-motivated.[6] On the other hand, a different social planner with different priorities might overweight information-motivated investor welfare, as information-motivated traders contribute to price discovery. Although information acquisition is exogenous in this model, if it were to be endogenized, then a desire to enhance price informativeness would rationalize a higher Pareto weight on information-motivated traders: directing more surplus toward information-motivated traders would incentivize further information acquisition, hence more informative prices.

I do not consider market maker welfare—instead focusing only on investor welfare—because market makers are assumed to be competitive throughout.

I described the model in terms of a single investor who will be either information-motivated (with probability α) or liquidity-motivated (with probability $1-\alpha$). There is, however, an alternative interpretation involving a continuum of investors: a mass α of information-motivated investors, and a mass $1-\alpha$ of liquidity-motivated investors. All results apply equally to this alternative interpretation, and when convenient, I will speak in terms of this alternative interpretation instead.

For the single-investor interpretation, welfare is evaluated from an ex-interim perspective: after the investor knows his type $k \in \{I, L\}$ but before he knows his draw from F_k and before he knows his venue assignment. For the alternative continuum-of-investors interpretation mentioned in the previous paragraph, welfare is evaluated at the group-level rather than at the individual-level. If, for

[5] The proof, therefore, leverages the assumption that the number of venues can be any natural number $J \in \mathbb{N}$. Results from Doval and Smolin (2024), however, imply that at most $J = 3$ venues are needed to implement any point in $\mathcal{F}$. Numerical experimentation suggests that, in fact, only $J = 2$ venues are needed.

[6] For example, Gary Gensler said, "I asked staff to take a holistic, cross-market view of how we could update our rules and drive greater efficiencies in our equity markets, *particularly for retail investors*" (Gensler, 2022, emphasis added). Later in the same speech, he proposed order-by-order competition as a way to reduce transaction costs for retail investors. This idea later materialized as the SEC's proposed Order Competition Rule (SEC, 2022), and it specifically pertains to some of the issues of segmentation that this paper addresses:

> the vast majority of retail marketable orders are flowing to wholesalers that pay for this order flow. [...] This segmentation—which isolates retail orders—may not benefit the retail public as much as orders being exposed to order-by-order competition. [...] Thus, I've asked staff to make recommendations for the Commission's consideration around how to enhance order-by-order competition. (Gensler, 2022)

instance, one segmentation Pareto-dominates another, what that means is that both information-motivated investors (in aggregate) and liquidity-motivated investors (in aggregate) are better off under the former segmentation; it does not necessarily mean that every individual investor will be better off.

In practice, regulators do have a great deal of influence over the form and amount of segmentation that prevails—in language of the model, over the venue-assignment probabilities $(\boldsymbol{\delta}_I, \boldsymbol{\delta}_L)$. I do not, however, formally model the process by which different regulations induce different values for $(\boldsymbol{\delta}_I, \boldsymbol{\delta}_L)$. Rather, I simply assume directly that a planner can implement any $(\boldsymbol{\delta}_I, \boldsymbol{\delta}_L)$ she would like. Admittedly, this assumption is quite strong: in practice, regulators might be constrained in terms of the segmentations they can induce.[7] In that case, one interpretation of my analysis is that $\mathcal{F}$ is an outer bound on the what a planner can actually achieve. In comparison, many other papers make an assumption that is quite strong in the other direction, assuming that the planner is limited to a handful of specific segmentations, often just the full-pooling and full-separation segmentations, or versions thereof (e.g., Benveniste, Marcus and Wilhelm, 1992; Hendershott and Mendelson, 2000; Battalio and Holden, 2001; Desgranges and Foucault, 2005; Zhu, 2014; Lee and Wang, 2023; Baldauf, Mollner and Yueshen, 2024). An interesting direction for future research would be to formulate a suitable middle ground.

3 The uniform case

I begin by considering the case in which the distributions F_I and F_L are both $U[0,1]$.

3.1 Spreads and welfare

Provided the spread s_j lies in the interval $[0,1]$, which will be the case in equilibrium, the zero-profit condition (1) becomes

$$(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}\frac{(1-s_j)^2}{2}. \tag{3}$$

The spread must be the smallest solution to (3), namely

$$s_j = \frac{\alpha\delta_{Ij}}{\alpha\delta_{Ij} + 2(1-\alpha)\delta_{Lj}}. \tag{4}$$

As an aside, this could alternatively be expressed $s_j = \frac{\tau_j}{\tau_j+2}$, where $\tau_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$ represents the toxicity of venue j's order flow.

The welfare expressions (2a) and (2b) become

$$w_I = \sum_{j=1}^{J} \delta_{Ij}\frac{(1-s_j)^2}{2} \qquad \text{and} \qquad w_L = \sum_{j=1}^{J} \delta_{Lj}\frac{(1-s_j)^2}{2},$$

where each s_j is given by (4).

3.2 Cream-skimming and reverse cream-skimming

Before characterizing what is feasible for all possible segmentations, I begin by discussing two specific forms of segmentation.

[7] One class of constraints that I do formally address is the possibility that segmentations can be based only on an imperfectly informative signal of the investor's trading motive. See Appendix 3.4 for details.

Cream-skimming. First, consider the class of segmentations in which some liquidity-motivated orders are pooled with information-motivated orders, while the remaining liquidity-motivated orders are separated. Formally, suppose there are $J = 2$ venues, with all information-motivated orders routed to the first venue, and only a fraction $\beta_L \in [0, 1]$ of the liquidity-motivated orders routed to the first venue, with the remainder routed to the second. That is,

$$\boldsymbol{\delta}_I = (1, 0) \qquad \text{and} \qquad \boldsymbol{\delta}_L = (\beta_L, 1 - \beta_L). \tag{5}$$

This class roughly connects to some forms of segmentation that occur in practice. When segmentation occurs in practice, it is often geared toward segmenting certain liquidity-motivated orders into various non-exchange venues, leaving relatively more informed order flow on exchanges. Because liquidity-motivated orders are more profitable for market makers to trade with than information-motivated orders (at a fixed spread), segmentation of this type is commonly called "cream-skimming."

For example, retail orders tend to be uninformed. Hence, all the various practices by which retail orders receive special treatment constitute this sort of cream-skimming segmentation: off-exchange internalization of retail orders, on-exchange retail liquidity programs, the SEC's recently-proposed order-by-order auctions for retail orders, and so on. Another set of examples include ATSs that provide access only to certain institutional traders with a reputation for liquidity-motivated trading.

Accordingly, I say that a *cream-skimming segmentation* is one of the form (5). Such a segmentation leads to spreads

$$s_1 = \frac{\alpha}{\alpha + 2(1-\alpha)\beta_L} \qquad \text{and} \qquad s_2 = 0,$$

implying welfares

$$w_I = \frac{2(1-\alpha)^2\beta_L^2}{\big(\alpha + 2(1-\alpha)\beta_L\big)^2} \qquad \text{and} \qquad w_L = \frac{2(1-\alpha)^2\beta_L^3}{\big(\alpha + 2(1-\alpha)\beta_L\big)^2} + \frac{1-\beta_L}{2}.$$

For example, $\beta_L = 0$ implements full-separation: all information-motivated investors go to venue 1, and all liquidity-motivated investors to venue 2. On venue 1, a no-trade theorem (Milgrom and Stokey, 1982) applies: its spread must then be large enough to crowd out all information-motivated trade, so we have $s_1 = 1$ and $w_I = 0$. Venue 2, in contrast, has no adverse selection at all, so we have $s_2 = 0$ and $w_L = \frac{1}{2}$. In summary, full-separation leads to the welfare pair $(w_I, w_L) = (0, \frac{1}{2})$.

At the other extreme, $\beta_L = 1$ implements full-pooling: both investor types face the same spread. Because we have assumed $F_I = F_L$, the full-pooling outcome entails a welfare pair on the 45-degree line, namely $(w_I, w_L) = \frac{2(1-\alpha)^2}{(2-\alpha)^2}$.

Intermediate values of β_L interpolate between these extremes, in a way that is characterized by the following lemma.

Lemma 2. *In the uniform case, cream-skimming segmentations implement welfare outcomes* (w_I, w_L) *in the set*

$$\mathcal{CS} = \left\{ (w_I, w_L) \,\middle|\, w_I \in \left[0, \frac{2(1-\alpha)^2}{(2-\alpha)^2}\right] \text{ and } w_L = \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\big(\sqrt{2w_I} + 2w_I\big) \right\}.$$

Reverse cream-skimming. Second, consider the reverse situation: suppose there are $J = 2$ venues, with all liquidity-motivated orders routed to the first venue, and only a fraction $\beta_I \in [0, 1]$

of the information-motivated orders routed to the first venue, with the remainder routed elsewhere. That is,

$$\boldsymbol{\delta}_I = (\beta_I, 1-\beta_I) \qquad \text{and} \qquad \boldsymbol{\delta}_L = (1,0). \tag{6}$$

A *reverse cream-skimming segmentation* is one of the form (6). Such a segmentation leads to spreads

$$s_1 = \frac{\alpha\beta_I}{\alpha\beta_I + 2(1-\alpha)} \qquad \text{and} \qquad s_2 = 1,$$

implying welfares

$$w_I = \beta_I \frac{2(1-\alpha)^2}{\big(\alpha\beta_I + 2(1-\alpha)\big)^2} \qquad \text{and} \qquad w_L = \frac{2(1-\alpha)^2}{\big(\alpha\beta_I + 2(1-\alpha)\big)^2}.$$

As above, $\beta_I = 0$ implements full-separation and the welfare pair $(w_I, w_L) = (0, \frac{1}{2})$, while $\beta_I = 1$ implements full-pooling and $(w_I, w_L) = \frac{2(1-\alpha)^2}{(2-\alpha)^2}$. Intermediate values of β_I interpolate between these extremes—although the interpolation differs from the one above. The following lemma characterizes the outcomes induced by these segmentations.

Lemma 3. *In the uniform case, reverse cream-skimming segmentations implement welfare outcomes (w_I, w_L) in the set*

$$\mathcal{RCS} = \left\{ (w_I, w_L) \,\middle|\, w_L \in \left[\frac{2(1-\alpha)^2}{(2-\alpha)^2}, \frac{1}{2}\right] \text{ and } w_I = \frac{1-\alpha}{\alpha}\big(\sqrt{2w_L} - 2w_L\big) \right\}.$$

Cream-skimming vs. reverse cream-skimming. It follows from the characterizations in Lemmas 2 and 3 that reverse cream-skimming Pareto dominates cream-skimming, in the following sense.

Proposition 4. *For all $(w_I, w_L) \in \mathcal{CS}$, there exists a $(\hat{w}_I, \hat{w}_L) \in \mathcal{RCS}$ for which $\hat{w}_I \geq w_I$ and $\hat{w}_L \geq w_L$. Moreover, both inequalities can be made strict, with the possible exceptions of the boundary points $w_I = 0$ and $w_I = \frac{2(1-\alpha)^2}{(2-\alpha)^2}$.*

Here is a partial intuition. I provide a more complete intuition when discussing a stronger result in Section 3.4. As a baseline, consider the full-separation outcome with all information-motivated investors on venue 1 and all liquidity-motivated investors on venue 2. This outcome features spreads $s_1 = 1$ and $s_2 = 0$. Then consider a small change involving either cream-skimming or reverse cream-skimming.

- Cream-skimming means moving a small fraction of liquidity-motivated flow from venue 2 to venue 1. This creates three effects: *(i)* the liquidity-motivated investors who move suffer, as they face a larger spread; *(ii)* the liquidity-motivated investors who stay are unaffected, as the spread at venue 2 remains 0; and *(iii)* the information-motivated investors benefit, as the spread at venue 1 decreases. Crucially, however, the benefit to the information-motivated investors is only second-order. This is because, in general, a drop in the spread entails a first-order benefit to inframarginal investors who were already trading at the larger spread, and only a second-order benefit to marginal investors who begin to trade only at the smaller spread. However, in this case, there are no such inframarginal investors, because the spread at venue 1 was initially $s_1 = 1$, large enough to crowd out all trade.

- Reverse cream-skimming means moving a small fraction of information-motivated flow from venue 1 to venue 2. This creates three effects: *(i)* the information-motivated investors who move benefit, as they face a smaller spread; *(ii)* the information-motivated investors who stay are unaffected, as the spread at venue 1 remains 1; and *(iii)* the liquidity-motivated investors suffer, as the spread at venue 2 increases. And because the spread at venue 2 was initially $s_2 = 0$, there are inframarginal liquidity-motivated investors, and so the harm to the liquidity-motivated investors is first-order.

In summary, for small changes from the full-separation baseline, reverse cream-skimming entails a first-order benefit for information-motivated investors, at the cost of first-order harm for liquidity-motivated investors. In contrast, cream-skimming entails only a second-order benefit for information-motivated investors, yet still at the cost of first-order harm for liquidity-motivated investors. It follows that the reverse cream-skimming scenario Pareto dominates. By Proposition 4, reverse cream-skimming outcomes in fact Pareto dominate cream-skimming outcomes in general—not merely for small changes from the full-separation baseline.

3.3 Characterizing the feasible set

The next result characterizes the feasible welfare set as the area between the cream-skimming and reverse cream-skimming outcomes.

Proposition 5. *In the uniform case, the feasible welfare set is*

$$\mathcal{F} = \left\{ (w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right) \text{ and } w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right) \right\}.$$

One direction is immediate. I have already shown, by construction, that $\mathcal{CS}$ and $\mathcal{RCS}$ are feasible. Moreover, $\mathcal{F}$ is convex (by Proposition 1). It follows that $\mathcal{F}$ must contain the area between $\mathcal{CS}$ and $\mathcal{RCS}$. The other direction—that *only* the area between $\mathcal{CS}$ and $\mathcal{RCS}$ is feasible—is less straightforward. The proof (in Appendix A) provides a mathematical argument. I postpone a discussion of the intuition until Section 3.4, which contains a generalization of this result.

Figure 1 (in the introduction) depicts the feasible set for various choices of α. Full separation occurs when $\delta_{Ij}\delta_{Lj} = 0$ for all venues j. In the figure, in each panel, this is the top-left-most point: the best outcome for liquidity-motivated investors and the worst outcome for information-motivated investors. The intuition is that under full separation, a no-trade theorem applies to the venues on which information-motivated investors trade, so that they receive zero surplus. In contrast, liquidity-motivated investors trade at a spread of zero, and so receive their highest possible surplus.

Full pooling occurs when $\boldsymbol{\delta}_I = \boldsymbol{\delta}_L$. In the figure, in each panel, this is the bottom-most point: the worst outcome for liquidity-motivated investors. Interestingly, the figure illustrates that for sufficiently high values of α, the full pooling outcome is not the best outcome for information-motivated investors, and is consequently Pareto dominated by other points in the feasible set.[8]

The intuition for why full pooling can be Pareto dominated is actually slightly clearer in the alternative parametrization of the model considered in Appendix B.2, where F_I is a point mass

[8] In particular, this is the case if $\alpha > \frac{2}{3}$. This follows from the characterization of $\mathcal{RCS}$, since

$$\frac{d}{dw_L}\left(\frac{1-a}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right)\Big|_{w_L = \frac{2(1-\alpha)^2}{(2-\alpha)^2}} = \frac{3\alpha - 2}{2\alpha},$$

which is positive iff $\alpha > \frac{2}{3}$.

on 1. There, if α is sufficiently large, full pooling entails so much adverse selection that it causes trading to break down entirely, leading to $w_I = w_L = 0$. Because welfare has nowhere to go but up, a Pareto improvement is easy to find: if some information-motivated investors are shunted elsewhere, then adverse selection ameliorates, which allows trading to open, making both investor types better off. A similar intuition applies to the parametrization considered in this section, where $F_I = F_L = U[0, 1]$. Here, if α is large, full pooling entails so much adverse selection that—although trading does not break down entirely—it comes close enough that a Pareto improvement can be found in a similar way.

Figure 1 also indicates that the Pareto frontier of $\mathcal{F}$ consists of reverse cream-skimming outcomes: $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$. In contrast, the Pareto anti-frontier of $\mathcal{F}$ is exactly the set of cream-skimming outcomes: $\mathcal{F}^{anti\text{-}pareto} = \mathcal{CS}$. These results, in fact, follow as a corollary to the characterization of $\mathcal{F}$ given by Proposition 5.

Corollary 6. *In the uniform case,* $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$ *and* $\mathcal{F}^{anti\text{-}pareto} = \mathcal{CS}$.

Remark 1. As mentioned in Section 2.3, a potentially focal scenario is that of a social planner who—rather than putting different Pareto weights on different investor types—seeks to maximize total welfare $w_T = \alpha w_I + (1 - \alpha) w_L$. Such a planner would prefer full separation. The intuition is that under full separation, every liquidity-motivated investor faces a spread of zero, and hence all available gains from trade are realized. Under other segmentations, some gains from trade go unrealized, so total welfare is lower. This argument, in fact, applies for any F_I and F_L, not only for the uniform case. The argument does, however, rely on the assumption that segmentations can be based directly on the investor's trading motive. Without it, a corresponding result is not guaranteed. Indeed, the next section considers an extension in which segmentations can be based only on imperfectly informative labels: likely-information (`LI`) and likely-liquidity (`LL`). Full separation—that is, assigning all investors labeled `LI` to one venue and all labeled `LL` to another—might no longer maximize total welfare. This observation relates to a key result of Lee and Wang (2023).

3.4 An extension: imperfect binary signal of investor type

For this section only, I weaken the assumption that segmentations can be based directly on the investor's trading motive. Versions of all previous results extend.

Labels. Concretely, suppose there are two imperfectly informative labels: likely-information (`LI`) and likely-liquidity (`LL`). Conditional on being information-motivated (liquidity-motivated), the investor is labeled `LI` with probability ρ_I (ρ_L) and `LL` otherwise. Without loss of generality, assume $\rho_I \geq \rho_L$, so that the labels can be interpreted at face value.

The parameters (ρ_I, ρ_L) capture the informativeness of these labels. Labels are perfectly informative if $\rho_I = 1$ and $\rho_L = 0$; this corresponds to the baseline case analyzed before. On the other hand, labels are completely uninformative if $\rho_I = \rho_L$.

It is, of course, limiting to assume that there are only two labels. A more general model would allow for an unrestricted number of labels and an arbitrary joint distribution between labels and the investor's motive. However, allowing for only two labels lends tractability. Several other papers take a related approach (e.g., Benveniste, Marcus and Wilhelm, 1992; Easley, Kiefer and O'Hara, 1996; Battalio and Holden, 2001; Lee and Wang, 2023).

Segmentations. Segmentations can be based on the investor's label, although not on his type directly. Conditional on the label $\ell \in \{\texttt{LI}, \texttt{LL}\}$, the investor will be assigned to venue j with

exogenous and commonly-known probability $\delta_{\ell j}$. Define $\boldsymbol{\delta}_\ell = (\delta_{\ell 1}, \ldots, \delta_{\ell J})$. By definition, $\boldsymbol{\delta}_\ell \in \Delta^J$. The interpretation is that a planner who knows the labels can control $(\boldsymbol{\delta}_{\texttt{LI}}, \boldsymbol{\delta}_{\texttt{LL}})$.

For this analysis, adjust the definitions of cream-skimming and reverse cream-skimming in the natural way. A cream-skimming segmentation is one of the form $\boldsymbol{\delta}_{\texttt{LI}} = (1, 0)$ and $\boldsymbol{\delta}_{\texttt{LL}} = (\beta_{\texttt{LL}}, 1-\beta_{\texttt{LL}})$. Likewise, a reverse cream-skimming segmentation is one of the form $\boldsymbol{\delta}_{\texttt{LI}} = (\beta_{\texttt{LI}}, 1 - \beta_{\texttt{LI}})$ and $\boldsymbol{\delta}_{\texttt{LL}} = (1, 0)$.

Spreads. Provided the spread s_j lies in the interval $[0, 1]$, which will be the case in equilibrium, the zero-profit condition (3) generalizes to

$$(1-\alpha)\big[\rho_L \delta_{\texttt{LI}j} + (1-\rho_L)\delta_{\texttt{LL}j}\big](1-s_j)s_j = \alpha\big[\rho_I \delta_{\texttt{LI}j} + (1-\rho_I)\delta_{\texttt{LL}j}\big]\frac{(1-s_j)^2}{2}. \tag{7}$$

Define

$$\tau_j = \frac{\alpha}{1-\alpha}\frac{\rho_I \delta_{\texttt{LI}j} + (1-\rho_I)\delta_{\texttt{LL}j}}{\rho_L \delta_{\texttt{LI}j} + (1-\rho_L)\delta_{\texttt{LL}j}}$$

as the toxicity of venue j's order flow. A key difference from the case analyzed before is that, unless the labels are perfectly informative, these toxicities are bounded. Indeed, because $\delta_{\texttt{LI}j}$ and $\delta_{\texttt{LL}j}$ both lie in the unit interval, we must have $\tau_j \in [\tau_{\min}, \tau_{\max}]$, where

$$\tau_{\min} := \frac{\alpha}{1-\alpha}\frac{1-\rho_I}{1-\rho_L} \qquad \text{and} \qquad \tau_{\max} := \frac{\alpha}{1-\alpha}\frac{\rho_I}{\rho_L}.$$

The spread must be the smallest solution to (7), which, just as before, is given by

$$s_j = \frac{\tau_j}{\tau_j + 2}.$$

Welfare. The welfare expressions now become

$$w_I = \sum_{j=1}^{J} \big[\rho_I \delta_{\texttt{LI}j} + (1-\rho_I)\delta_{\texttt{LL}j}\big]\frac{(1-s_j)^2}{2},$$
$$w_L = \sum_{j=1}^{J} \big[\rho_L \delta_{\texttt{LI}j} + (1-\rho_L)\delta_{\texttt{LL}j}\big]\frac{(1-s_j)^2}{2}.$$

It will be convenient to use the zero-profit condition (7) to alternatively express w_I as

$$w_I = \frac{1-\alpha}{\alpha}\sum_{j=1}^{J} \big[\rho_L \delta_{\texttt{LI}j} + (1-\rho_L)\delta_{\texttt{LL}j}\big](1-s_j)s_j.$$

Characterizing the feasible set. Next, I perform a change of variables. In the original formulation, the planner chose vectors $(\boldsymbol{\delta}_{\texttt{LI}}, \boldsymbol{\delta}_{\texttt{LL}})$. Consider, instead, a formulation in which the planner chooses vectors $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$, where for each venue j,

$$\theta_j = \rho_L \delta_{\texttt{LI}j} + (1-\rho_L)\delta_{\texttt{LL}j}.$$

Because $\delta_{\texttt{LI}j}$ and $\delta_{\texttt{LL}j}$ must both lie in the unit interval, θ_j must also lie in the unit interval. In addition, $\sum_{j=1}^J \delta_{\texttt{LI}j} = \sum_{j=1}^J \delta_{\texttt{LL}j} = 1$ implies that $\sum_{j=1}^J \theta_j = 1$ and $\sum_{j=1}^J \theta_j \tau_j = \frac{\alpha}{1-\alpha}$.

Next, use $s_j = \frac{\tau_j}{\tau_j+2}$ to write $\frac{(1-s_j)^2}{2} = \frac{2}{(2+\tau_j)^2}$ and $(1-s_j)s_j = \frac{2\tau_j}{(2+\tau_j)^2}$. Collecting all the above, $(w_I, w_L) \in \mathcal{F}$ if and only if we can write

$$w_I = \frac{1-\alpha}{\alpha}\sum_{j=1}^{J}\theta_j\frac{2\tau_j}{(2+\tau_j)^2} \qquad \text{and} \qquad w_L = \sum_{j=1}^{J}\theta_j\frac{2}{(2+\tau_j)^2}.$$

for some $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ satisfying the constraints

$$\sum_{j=1}^{J}\theta_j = 1 \qquad \text{and} \qquad \theta_j \in [0,1] \text{ for all } j,$$

$$\sum_{j=1}^{J}\theta_j\tau_j = \frac{\alpha}{1-\alpha} \qquad \text{and} \qquad \tau_j \in [\tau_{\min}, \tau_{\max}] \text{ for all } j.$$

Formulated in this way, the planner's problem is to choose a distribution over toxicities in the interval $[\tau_{\min}, \tau_{\max}]$, where θ_j is the weight on toxicity τ_j, subject to a constraint that the mean of the distribution must be $\frac{\alpha}{1-\alpha}$. In these terms, w_I and w_L can each be thought of as an expected utility from this distribution over toxicities. The distribution is the same for both w_I and w_L; only the "utility function" is different. For w_I, it is $\frac{1-\alpha}{\alpha}\frac{2\tau}{(2+\tau)^2}$. For w_L, it is $\frac{1}{(2+\tau)^2}$.

Now, because $\frac{2}{(2+\tau)^2}$ is convex in τ, mean-preserving spreads in the distribution raise w_L. But because $\frac{2\tau}{(2+\tau)^2}$ has both convex and concave regions, mean-preserving spreads in the distribution might either raise or lower w_I.

A key quantity will be the "bang-for-the-buck" of a mean-preserving spread, which I define as the effect on w_I (possibly positive or negative) relative to the effect on w_L (necessarily positive). In general, the effect of a mean-preserving spread is captured by the second derivative of the utility function. Hence, the bang-for-the-buck of a (vanishingly small) mean-preserving spread in the region of a toxicity level τ is captured by the ratio of second derivatives:

$$\left[\frac{d^2}{d\tau^2}\frac{2\tau}{(2+\tau)^2}\right] \Big/ \left[\frac{d^2}{d\tau^2}\frac{2}{(2+\tau)^2}\right] = \frac{\tau-4}{3}.$$

Note that this is increasing in τ.

Hence, w_I can be increased while w_L is held fixed if it is possible to combine a mean-preserving spread in a high-τ region with a mean-preserving contraction in a low-τ region. Only when all such opportunities are exhausted can w_I be maximized subject to a particular value for w_L. All such opportunities are exhausted if and only if the distribution is supported on at most two toxicity levels: $\tau_{\max}$ and some $\tau \leq \frac{\alpha}{1-\alpha}$. These, however, are precisely the distributions over toxicities achieved by reverse cream-skimming segmentations.

Applying precisely the same logic in reverse, w_I is minimized subject to a particular value for w_L if and only if the distribution is supported on at most two toxicity levels: $\tau_{\min}$ and some $\tau \geq \frac{\alpha}{1-\alpha}$. These, however, are precisely the distributions over toxicities achieved by cream-skimming segmentations.

Stating this formally, we have the following result, which can be viewed as a generalization of Proposition 5 to the case of imperfect binary signals.

Proposition 5′. *Define $\underline{w}_L$ and $\overline{w}_L$ as the lowest and highest feasible values for w_L:*

$$\underline{w}_L = \min\{w_L : (w_I, w_L) \in \mathcal{F}\} \qquad \textit{and} \qquad \overline{w}_L = \max\{w_L : (w_I, w_L) \in \mathcal{F}\}.$$

For $w_L \in [\underline{w}_L, \overline{w}_L]$, define $\underline{w}_I(w_L)$ and $\overline{w}_I(w_L)$ as the lowest and highest feasible values for w_I that are consistent with that w_L:

$$\underline{w}_I(w_L) = \min\{w_I : (w_I, w_L) \in \mathcal{F}\} \qquad \text{and} \qquad \overline{w}_I(w_L) = \max\{w_I : (w_I, w_L) \in \mathcal{F}\}.$$

Then $\mathcal{F} = \big\{(w_I, w_L) \,\big|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I \in [\underline{w}_I(w_L), \overline{w}_I(w_L)]\big\}$. Moreover, in the uniform case with imperfect binary signals,

$$\mathcal{CS} = \big\{(w_I, w_L) \,\big|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I = \underline{w}_I(w_L)\big\},$$
$$\mathcal{RCS} = \big\{(w_I, w_L) \,\big|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I = \overline{w}_I(w_L)\big\}.$$

Proposition 5′, moreover, immediately implies the following analogue of Corollary 6.

Corollary 6′. *In the uniform case with imperfect binary signals, $\mathcal{F}^{\text{pareto}} \subseteq \mathcal{RCS}$ and $\mathcal{F}^{\text{anti-pareto}} \subseteq \mathcal{CS}$.*

One might wonder if a stronger result than Corollary 6′ would be true, entailing the claim $\mathcal{F}^{\text{anti-pareto}} = \mathcal{CS}$. Indeed, this stronger result is true if segmentations can be based directly on the investor's trading motive (*cf.* Corollary 6). Appendix C.1, however, provides an example to illustrate that this stronger result is not in fact guaranteed in this extension.

4 The general case

I now drop the assumption that F_I and F_L are both $U[0,1]$. I do, however, retain the assumption that segmentations can be based directly on the investor's trading motive (in contrast to Section 3.4). The goal of this section is to provide conditions under which the main result of the previous section extends: $\mathcal{F}^{\text{pareto}} \subseteq \mathcal{RCS}$.

4.1 Preliminaries

Re-expressing w_I. For the arguments below, it will be useful to work with an alternative expression for w_I. The idea is that, because market making is competitive, market maker revenue from liquidity-motivated traders equals market maker losses to information-motivated traders—which in turn equals αw_I. Hence,

$$w_I = \sum_{j=1}^{J} \delta_{Ij} \int_{s_j}^{1} (v - s_j)\, dF_I(v) = \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj}[1 - F_L(s_j)]s_j, \tag{8}$$

where the first equality repeats the definition of w_I from (2a), and the second equality follows from (1).

Defining $s(\tau)$. Provided that $\delta_{Lj} > 0$, it follows from (1) that the spread on a venue j depends on $(\alpha, \delta_{Ij}, \delta_{Lj})$ only through the toxicity $\tau_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$. In particular, $s_j = s(\tau_j)$, where $s(\tau)$ is defined as the function that delivers the smallest solution to

$$\big[1 - F_L(s)\big]s = \tau \int_{s}^{1} (v - s)\, dF_I(v). \tag{9}$$

To encompass $\delta_{Lj} = 0$, we say that $\tau_j = \infty$ in this case.[9] It follows from (1) that, in this case, the spread must be the largest element of the support of F_I. We therefore define $s(\infty) = \max \operatorname{supp} F_I$.

Note that a toxicity of zero implies a spread of zero: $s(0) = 0$. It is also useful to define s_{pool} as the half-spread that prevails when $\delta_{Ij} = \delta_{Lj}$, i.e., $s_{pool} = s\left(\frac{\alpha}{1-\alpha}\right)$. Furthermore, $s(\tau)$ is monotonically increasing:

Lemma 7. *$s(\tau)$ is weakly increasing.*

Defining $R(s)$. Define

$$R(s) := [1 - F_L(s)]s.$$

In words, $R(s)$ is the revenue that a market maker would earn from a liquidity-motivated investor when the spread is s. The investor trades with probability $1 - F_L(s)$, and conditional on trading, the market maker earns the spread. This allows us to write (8) somewhat more compactly as

$$w_I = \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj} R(s_j). \tag{10}$$

Next is a result about the supremum of the composition of the functions $R(s)$ and $s(\tau)$.

Lemma 8. *Define $R^* := \sup_\tau R(s(\tau))$.*

(i) If $R(s(\tau))$ achieves R^* on the domain $\tau \in [0, \infty)$, then there exists a smallest τ that does this.

(ii) Otherwise, $R(s(\tau))$ approaches R^* only in the limit, i.e., $R(s(\infty)) = R^*$ and $(\forall \bar{\tau})(\exists \varepsilon > 0)(\forall \tau < \bar{\tau}) : R(s(\tau)) < R^* - \varepsilon$.

Defining τ^* and s^*. In light of Lemma 8, there are two cases. If $R(s(\tau))$ achieves its supremum, R^*, then define τ^* as the smallest value at which R^* is achieved. Otherwise, say that $\tau^* = \infty$. Define also $s^* = s(\tau^*)$.

Characterizing $\mathcal{F}$. The alternative expression for w_I given by (8), together with the definition of $s(\tau)$ implies that $(w_I, w_L) \in \mathcal{F}$ if and only if

$$w_I = \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj} \underbrace{[1 - F_L(s(\tau_j))]s(\tau_j)}_{R(s(\tau_j))} \qquad \text{and} \qquad w_L = \sum_{j=1}^{J} \delta_{Lj} \int_{s(\tau_j)}^{1} (v - s(\tau_j))\, dF_L(v). \tag{11}$$

for some $\boldsymbol{\delta}_I \in \Delta^J$ and $\boldsymbol{\delta}_L \in \Delta^J$, where for all j,

$$\tau_j = \begin{cases} \frac{\alpha \delta_{Ij}}{(1-\alpha)\delta_{Lj}} & \text{if } \delta_{Lj} > 0, \\ \infty & \text{if } \delta_{Lj} = 0. \end{cases}$$

[9] A subtlety involves the case in which not only $\delta_{Lj} = 0$ but also $\delta_{Ij} = 0$. Although I had written that Bertrand competition implies s_j must be the smallest solution to (1), strictly speaking, this argument does not apply to the case of $\delta_{Ij} = \delta_{Lj} = 0$. Rather, Bertrand competition does nothing to pin down s_j in that case. For the following arguments, it is not relevant how τ_j and s_j are defined in those cases.

Equivalently, we can eliminate $\boldsymbol{\delta}_I$ as a choice variable, and instead think of the planner choosing $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)$ directly. To that end, $(w_I, w_L) \in \mathcal{F}$ if and only if (11) holds for some $\boldsymbol{\delta}_L$ and $\boldsymbol{\tau}$ satisfying the constraints

$$\sum_{j=1}^{J} \delta_{Lj} = 1 \qquad \text{and} \qquad \delta_{Lj} \in [0,1] \text{ for all } j$$

$$\sum_{j=1}^{J} \delta_{Lj}\tau_j \leq \frac{\alpha}{1-\alpha} \qquad \text{and} \qquad \tau_j \in [0,\infty) \text{ for all } j$$

In short, the planner can be thought of as choosing a distribution over toxicities, where δ_{Lj} is the weight on toxicity level τ_j. Moreover, the constraint $\sum_{j=1}^{J} \delta_{Lj}\tau_j \leq \frac{\alpha}{1-\alpha}$ means that she is restricted to distributions with expectation at most $\frac{\alpha}{1-\alpha}$.

Characterizing $\mathcal{RCS}$. A reverse cream-skimming segmentation is one in which $\boldsymbol{\delta}_L = (1,0)$. Plugging that into the characterization of $\mathcal{F}$ given above, $(w_I, w_L) \in \mathcal{RCS}$ if and only if

$$w_I = \frac{1-\alpha}{\alpha} \underbrace{[1 - F_L(s(\tau))]s(\tau)}_{R(s(\tau))} \qquad \text{and} \qquad w_L = \int_{s(\tau)}^{1} (v - s(\tau))\, dF_L(v). \tag{12}$$

for some $\tau \in \left[0, \frac{\alpha}{1-\alpha}\right]$.

4.2 A necessary and sufficient condition

Here is the main result of this section. It provides a necessary and sufficient condition on the distributions F_I and F_L under which the conclusion $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$ extends.

Proposition 9. *Fix F_I and F_L. $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$ for all $\alpha \in (0,1)$ if and only if both*

(i) $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$, and

(ii) the following parametrically-defined curve is concave on $s \in [0, s^*]$:

$$w_I(s) = [1 - F_L(s)]s \qquad \text{and} \qquad w_L(s) = \int_{s}^{1} (v - s)\, dF_L(v). \tag{13}$$

The rest of this section provides a sketch of the proof. Condition (i) of the proposition requires $R(s(\tau))$ to be concave. This condition is necessary and sufficient for ensuring that $\mathcal{RCS}$ contains a point that is w_I-optimal in $\mathcal{F}$—something that must be true in order to have $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$. To see this, note that by (11) we have

$$w_I = \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj} R(s(\tau_j)).$$

If $R(s(\tau))$ is concave, then a planner who chooses a distribution over toxicities (subject to the constraint that the distribution has expectation at most $\frac{\alpha}{1-\alpha}$) with the objective of maximizing w_I must put positive weight $\delta_{Lj} > 0$ on only a single toxicity level τ_j. Any such distribution over toxicities corresponds to a reverse cream-skimming segmentation. In fact, because $R(s(\tau))$ is maximized at τ^*, the argument requires concavity only on the domain $\tau \in [0, \tau^*]$. If, on the other

hand, $R(s(\tau))$ were not concave on $\tau \in [0, \tau^*]$, then w_I might be improved by mean-preserving spreads in the distribution over toxicities. Indeed, it would be possible to find an $\alpha \in (0, 1)$ such that the w_I-optimal distribution entails putting positive weight on two toxicity levels. Such a distribution does not correspond to any reverse cream-skimming segmentation.

Condition (ii) of the proposition requires concavity of the parametrically-defined curve (13) on the domain $s \in [0, s^*]$. Because $w_L(s)$ is monotonically decreasing in s and because $w_I(s)$ is maximized at s^*, this is the domain between the w_L-optimal and the w_I-optimal points on the curve. To see why concavity on this domain is necessary, note that by (12), $\mathcal{RCS}$ corresponds to a segment of a rescaling of that curve (i.e., the rescaling in which the w_I coordinate is multiplied by $\frac{1-\alpha}{\alpha}$, and the segment corresponding to $s \in [0, s_{pool}]$). So, if the curve (13) fails to be concave on that domain, then it would be possible to find an $\alpha \in (0, 1)$ such that there are points in the convex hull of $\mathcal{RCS}$ that are neither equal to nor Pareto dominated by any element of $\mathcal{RCS}$. And, by Proposition 1, the feasible set is convex, so those points are in $\mathcal{F}$, meaning that $\mathcal{F}^{pareto} \not\subseteq \mathcal{RCS}$.

Finally, to see why conditions (i) and (ii) are jointly sufficient, note that by (11), each point in $\mathcal{F}$ is a convex combination of points on the aforementioned rescaling of the curve (13). By condition (ii), the rescaled curve is concave, so each point in $\mathcal{F}$ is Pareto dominated by some point on the rescaled curve. In fact, we can be more specific. Because $w_L(s)$ is monotonically decreasing in s and because $w_I(s)$ is maximized at s^*, each point in $\mathcal{F}$ is Pareto dominated by some point on the segment of the rescaled curve corresponding to the domain $s \in [0, s^*]$. Consider now two cases.

- In the first case, α is such that $s_{pool} \geq s^*$. $\mathcal{RCS}$ then contains the segment of the rescaled curve corresponding to $s \in [0, s^*]$. So it immediately follows that $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$. This case is illustrated in Figure 2(b).

- In the second case, α is such that $s_{pool} < s^*$. Here, we recall that condition (i) implies that $\mathcal{RCS}$ contains a point that is w_I-optimal in $\mathcal{F}$. Incorporating this, we can again conclude that $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$. This case is illustrated in Figure 2(a).

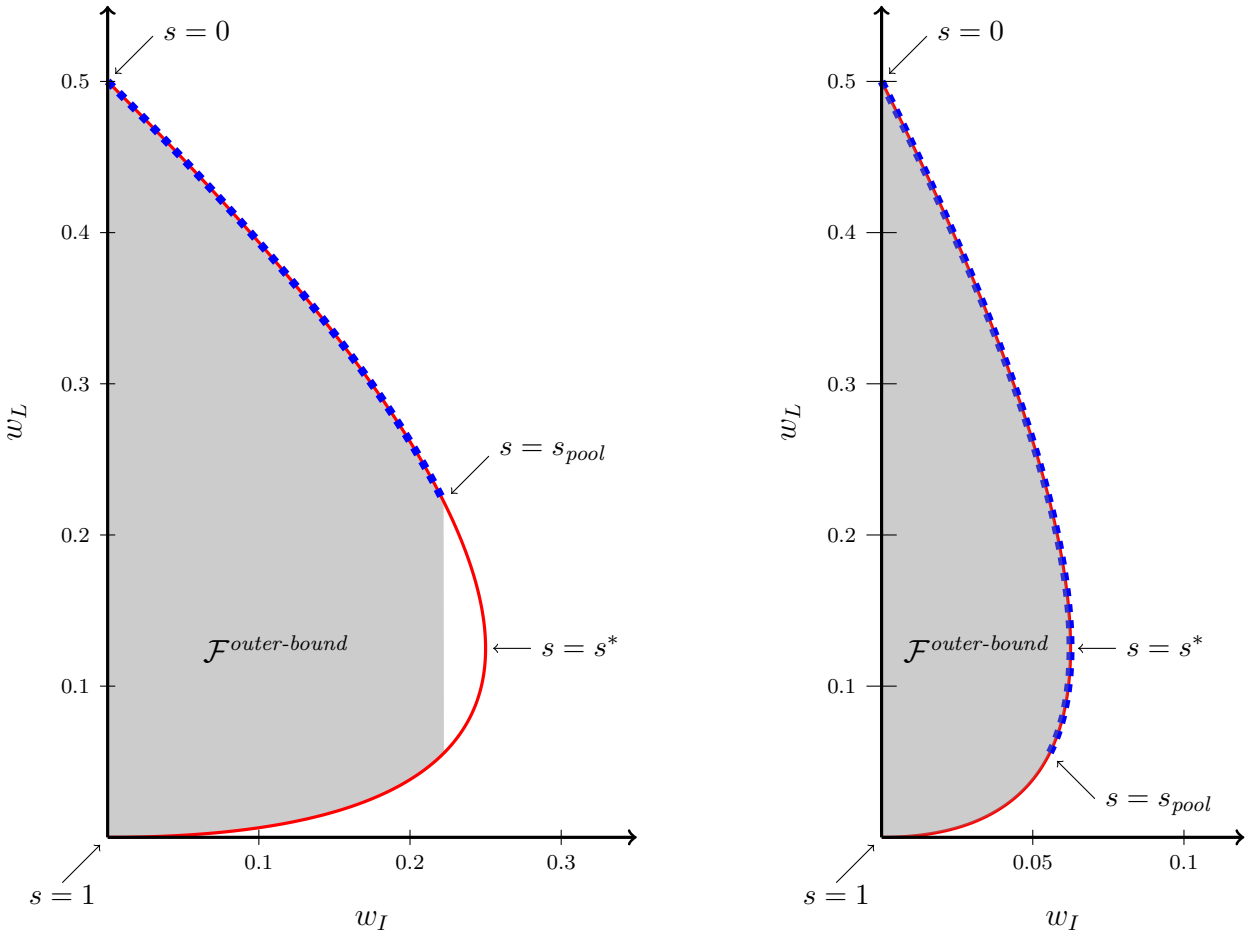
4.3 Distributions that satisfy the conditions of Proposition 9

The necessary and sufficient conditions given by Proposition 9 are satisfied by many common distributions.

[TODO: This section is in progress.]

Figure 2: Proof sketch of the sufficiency argument

(a) $\alpha = 0.5$ (b) $\alpha = 0.8$



An illustration of why conditions (i) and (ii) of Proposition 9 are sufficient for obtaining $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$, where the illustrations are drawn for the cases of $F_I = F_L = U[0,1]$ and $\alpha \in \{0.5, 0.8\}$. In each panel, the red curve depicts a rescaling of the parametrically-defined curve (13) (i.e., the rescaling in which the w_I coordinate is multiplied by $\frac{1-\alpha}{\alpha}$). As the parameter s increases from 0 to 1, the curve moves downward, from $(w_I, w_L) = (0, \frac{1}{2})$ to $(w_I, w_L) = (0, 0)$. The quantity s^* is defined as the parameter value at which the curve attains its w_I-maximal point. Condition (ii) ensures that this curve is convex on the domain $s \in [0, s^*]$. The blue dashed curve depicts $\mathcal{RCS}$, which is the segment of the red curve corresponding to the domain $s \in [0, s_{pool}]$. In the left panel, $s_{pool} < s^*$, while in the right panel, $s_{pool} > s^*$. The gray region is an outer bound on the feasible set, constructed using two observations: *(i)* $\mathcal{F}$ lies within the convex hull of the red curve, by equation (12), and *(ii)* $\mathcal{RCS}$ contains a point that is w_I-optimal in $\mathcal{F}$, from condition (i). In both panels, it holds that $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$.

5 Conclusion

One takeaway is that order flow segmentation has welfare consequences. Within that, an obvious point is that some segmentations favor information-motivated traders, while other segmentations favor liquidity-motivated traders. What is less obvious—and, to my knowledge, novel—is the finding that some segmentations are Pareto-efficient, while other segmentations are Pareto-inefficient.

Moreover, under certain conditions, the Pareto-efficient segmentations have a very simple structure. In particular, any point on the Pareto frontier can be achieved with what I've called a reverse cream-skimming segmentation. These are the segmentations in which all liquidity-motivated investors are assigned to one of two venues. The frontier is traced out by varying the allocation of information-motivated investors across the two venues.

Policy implications. The model is extremely simple, abstracting away from many considerations that might be important in practice. As such, its results should be interpreted with caution. But, taken at face value, these results have considerable implications for policy.

Consider, for example, the regulation of segmentation in U.S. equities. The prevailing regulations effectively encourage small venues that target only the most attractive (i.e., least informed) orders. Examples of such venues include *(i)* the off-exchange internalization of retail orders, and *(ii)* ATSs small enough to be exempt from fair access requirements. What regulations *prohibit* are large venues that might try to exclude only the least attractive (i.e., most informed) orders. For example, exchanges and large ATSs must provide fair access. Effectively, what this means is that under the current regulatory framework, segmentation is geared more toward cream-skimming and less toward reverse cream-skimming. This approach, however, is precisely the opposite of what the model recommends.

References

Admati, Anat R. and Paul Pfleiderer, "A theory of intraday patterns: Volume and price variability," *The Review of Financial Studies*, 1988, *1* (1), 3–40.

_ and _ , "Divide and conquer: A theory of intraday and day-of-the-week mean effects," *The Review of Financial Studies*, 1989, *2* (2), 189–223.

Baldauf, Markus and Joshua Mollner, "Trading in fragmented markets," *Journal of Financial and Quantitative Analysis*, 2021, *56* (1), 93–121.

_ , _ , and Bart Zhou Yueshen, "Siphoned apart: A portfolio perspective on order flow segmentation," *Journal of Financial Economics*, 2024, *154*, 103807.

Barclay, Michael J., Terrence Hendershott, and D. Timothy McCormick, "Competition among trading venues: Information and trading on electronic communications networks," *The Journal of Finance*, 2003, *58* (6), 2637–2665.

Battalio, Robert and Craig W. Holden, "A simple model of payment for order flow, internalization, and total trading cost," *Journal of Financial Markets*, 2001, *4* (1), 33–71.

Benveniste, Lawrence M., Alan J. Marcus, and William J. Wilhelm, "What's special about the specialist?," *Journal of Financial Economics*, 1992, *32* (1), 61–86.

Bergemann, Dirk and Stephen Morris, "Information design: A unified perspective," *Journal of Economic Literature*, 2019, *57* (1), 44–95.

_ , Benjamin Brooks, and Stephen Morris, "The limits of price discrimination," *American Economic Review*, 2015, *105* (3), 921–957.

Bessembinder, Hendrik and Herbert M. Kaufman, "A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks," *Journal of Financial Economics*, 1997, *46* (3), 293–319.

Brolley, Michael and David A. Cimon, "Order-flow segmentation, liquidity, and price discovery: The role of latency delays," *Journal of Financial and Quantitative Analysis*, 2020, *55* (8), 2555–2587.

Chen, Daniel and Darrell Duffie, "Market fragmentation," *American Economic Review*, 2021, *111* (7), 2247–2274.

Chowdhry, Bhagwan and Vikram Nanda, "Multimarket trading and market liquidity," *The Review of Financial Studies*, 1991, *4* (3), 483–511.

Comerton-Forde, Carole and Tālis J. Putniņš, "Dark trading and price discovery," *Journal of Financial Economics*, 2015, *118* (1), 70–92.

_ , Katya Malinova, and Andreas Park, "Regulating dark trading: Order flow segmentation and market quality," *Journal of Financial Economics*, 2018, *130* (2), 347–366.

Degryse, Hans, Frank De Jong, and Vincent van Kervel, "The impact of dark trading and visible fragmentation on market quality," *Review of Finance*, 2015, *19* (4), 1587–1622.

Dennert, Jürgen, "Price competition between market makers," *The Review of Economic Studies*, 1993, *60* (3), 735–751.

Desgranges, Gabriel and Thierry Foucault, "Reputation-based pricing and price improvements," *Journal of Economics and Business*, 2005, *57* (6), 493–527.

Doval, Laura and Alex Smolin, "Persuasion and Welfare," *Journal of Political Economy*, 2024, *132* (7), 000–000.

Easley, David and Maureen O'Hara, "Price, trade size, and information in securities markets," *Journal of Financial Economics*, 1987, *19* (1), 69–90.

_ , Nicholas M. Kiefer, and Maureen O'Hara, "Cream-skimming or profit-sharing? The curious role of purchased order flow," *The Journal of Finance*, 1996, *51* (3), 811–833.

Gensler, Gary, ""Market Structure and the Retail Investor:" Remarks Before the Piper Sandler Global Exchange Conference," June 2022. https://www.sec.gov/news/speech/gensler-remarks-piper-sandler-global-exchange-conference-060822.

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

Goldstein, Michael A., Andriy V. Shkilko, Bonnie F. Van Ness, and Robert A. Van Ness, "Competition in the market for NASDAQ securities," *Journal of Financial Markets*, 2008, *11* (2), 113–143.

Hatheway, Frank, Amy Kwan, and Hui Zheng, "An empirical analysis of market segmentation on US equity markets," *Journal of Financial and Quantitative Analysis*, 2017, *52* (6), 2399–2427.

Hendershott, Terrence and Haim Mendelson, "Crossing networks and dealer markets: Competition and performance," *The Journal of Finance*, 2000, *55* (5), 2071–2115.

Hoffmann, Peter, "Adverse selection, market access, and inter-market competition," *Journal of Banking & Finance*, 2016, *65*, 108–119.

Kamenica, Emir, "Bayesian persuasion and information design," *Annual Review of Economics*, 2019, *11*, 249–272.

_ **and Matthew Gentzkow**, "Bayesian persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Lee, Tomy and Chaojun Wang, "Why Trade Over-the-Counter?," *Working Paper*, 2023. https://finance.wharton.upenn.edu/~wangchj/papers/OTC_EX.pdf.

Madhavan, Ananth and Minder Cheng, "In search of liquidity: Block trades in the upstairs and downstairs markets," *The Review of Financial Studies*, 1997, *10* (1), 175–203.

Milgrom, Paul and Nancy Stokey, "Information, trade and common knowledge," *Journal of Economic Theory*, 1982, *26* (1), 17–27.

Securities and Exchange Commission, "Order Competition Rule," 2022. https://www.sec.gov/files/rules/proposed/2022/34-96495.pdf.

Seppi, Duane J., "Equilibrium block trading and asymmetric information," *the Journal of Finance*, 1990, *45* (1), 73–94.

Zhu, Haoxiang, "Do dark pools harm price discovery?," *The Review of Financial Studies*, 2014, *27* (3), 747–789.

A Proofs

A.1 Proofs corresponding to Sections 2–3

Proof of Proposition 1. Suppose that the welfare pairs (w_I', w_L') and (w_I'', w_L'') are respectively implemented by segmentations $(\boldsymbol{\delta}_I', \boldsymbol{\delta}_L') \in \Delta^{J'}$ and $(\boldsymbol{\delta}_I'', \boldsymbol{\delta}_L'') \in \Delta^{J''}$. For any $t \in (0,1)$, the welfare pair $t(w_I', w_L') + (1-t)(w_I'', w_L'')$ can be implemented by the segmentation $(\boldsymbol{\delta}_I''', \boldsymbol{\delta}_L''') \in \Delta^{J'+J''}$ defined by $\boldsymbol{\delta}_k''' = \big(t\delta_{k1}', t\delta_{k2}', \ldots, t\delta_{kJ'}', (1-t)\delta_{k1}'', (1-t)\delta_{k2}'', \ldots, (1-t)\delta_{kJ''}''\big)$ for $k \in \{I, L\}$.

Indeed, suppose that the segmentations $(\boldsymbol{\delta}_I', \boldsymbol{\delta}_L')$ and $(\boldsymbol{\delta}_I'', \boldsymbol{\delta}_L'')$ respectively induce the spreads $(s_1', \ldots, s_{J'}')$ and $(s_1'', \ldots, s_{J''}'')$. By (1), the spread depends on the segmentation only through the ratio of information-motivated to liquidity-motivated order flow. Hence, the segmentation $(\boldsymbol{\delta}_I''', \boldsymbol{\delta}_L''')$ induces the spreads $(s_1', \ldots, s_{J'}', s_1'', \ldots, s_{J''}'')$. Letting (w_I''', w_L''') be the welfare pair implemented by $(\boldsymbol{\delta}_I''', \boldsymbol{\delta}_L''')$, we have for $k \in \{I, L\}$

$$\begin{aligned} w_k''' &= \sum_{j=1}^{J'} t\delta_{kj}' \int_{s_j'}^{1} (v - s_j')\, dF_k + \sum_{j=1}^{J''} (1-t)\delta_{kj}'' \int_{s_j''}^{1} (v - s_j'')\, dF_k \\ &= tw_k' + (1-t)w_k''. \end{aligned}$$

□

Proof of Lemma 2. Define $\tau_1 = \frac{\alpha}{(1-\alpha)\beta_L}$. Because $\beta_L \in [0,1]$, we have $\tau_1 \in [\frac{\alpha}{1-\alpha}, \infty]$. Using $\beta_L = \frac{\alpha}{(1-\alpha)\tau_1}$, we can write

$$w_I = \frac{2}{(2+\tau_1)^2}, \tag{14a}$$

$$w_L = \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{4+\tau_1}{2(2+\tau_1)^2}. \tag{14b}$$

By (14a), w_I is continuous and decreasing in τ_1. Because $\tau_1 \in [\frac{\alpha}{1-\alpha}, \infty]$, we therefore have

$$w_I \in \left[0, \frac{2(1-\alpha)^2}{(2-\alpha)^2}\right]. \tag{15}$$

Moreover, solving (14a) for τ_1 (using the fact that τ_1 must be the nonnegative solution), we can write $\tau_1 = \sqrt{\frac{2}{w_I}} - 2$. Plugging that into (14b), we obtain

$$w_L = \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\big(\sqrt{2w_I} + 2w_I\big). \tag{16}$$

Together, (15) and (16) constitute the desired characterization. □

Proof of Lemma 3. Define $\tau_1 = \frac{\alpha\beta_I}{1-\alpha}$. Because $\beta_I \in [0,1]$, we have $\tau_1 \in [0, \frac{\alpha}{1-\alpha}]$. Using $\beta_I = \frac{1-\alpha}{\alpha}\tau_1$, we can write

$$w_I = \frac{1-\alpha}{\alpha}\frac{2\tau_1}{(2+\tau_1)^2}, \tag{17a}$$

$$w_L = \frac{2}{(2+\tau_1)^2}. \tag{17b}$$

By (17b), w_L is continuous and decreasing in τ_1. Because $\tau_1 \in [0, \frac{\alpha}{1-\alpha}]$, we therefore have

$$w_L \in \left[\frac{2(1-\alpha)^2}{(2-\alpha)^2}, \frac{1}{2}\right]. \tag{18}$$

Moreover, solving (17b) for τ_1 (using the fact that τ_1 must be the nonnegative solution), we can write $\tau_1 = \sqrt{\frac{2}{w_L}} - 2$. Plugging that into (17a), we obtain

$$w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right). \tag{19}$$

Together, (18) and (19) constitute the desired characterization. □

Proof of Proposition 4. Let $(w_I, w_L) \in \mathcal{CS}$. Let $\hat{w}_L = w_L$ and let $\hat{w}_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)$. The claim will follow if we can show $\hat{w}_I \geq w_I$.

$$\begin{aligned}
\hat{w}_I \geq w_I &\iff \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right) \geq w_I \\
&\iff \sqrt{2w_L} \geq \frac{\alpha}{1-\alpha} w_I + 2w_L \\
&\iff 2w_L \geq \frac{\alpha^2}{(1-\alpha)^2} w_I^2 + \frac{4\alpha}{1-\alpha} w_I w_L + 4w_L^2 \\
&\iff 2\left[\frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right)\right] \geq \frac{\alpha^2}{(1-\alpha)^2} w_I^2 \\
&\qquad + \frac{4\alpha}{1-\alpha} w_I \left[\frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right)\right] + 4\left[\frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right)\right]^2 \\
&\iff \frac{\alpha}{2(1-\alpha)}\sqrt{2w_I} \geq \frac{\alpha(2-\alpha)}{2(1-\alpha)^2} w_I \\
&\iff 2w_I \geq \frac{(2-\alpha)^2}{(1-\alpha)^2} w_I^2,
\end{aligned}$$

The first step simply uses the definition of $\hat{w}_I$. The second step rearranges, using $\alpha \in (0,1)$. The third step squares each side, which preserves the inequality because each side is nonnegative. The fourth step uses the fact that, because $(w_I, w_L) \in \mathcal{CS}$, we must have $w_L = \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right)$. The fifth step is algebra. For the sixth step, multiply through by $\frac{2(1-\alpha)}{\alpha}$ then square each side; this preserves the inequality because $\alpha \in (0,1)$ and because each side is nonnegative. Furthermore, the final condition is guaranteed to hold, because $w_I \in \left[0, \frac{2(1-\alpha)^2}{(2-\alpha)^2}\right]$, by the definition of $\mathcal{CS}$.

By the same arguments, we would in fact have $\hat{w}_I > w_I$ if $w_I \in \left(0, \frac{2(1-\alpha)^2}{(2-\alpha)^2}\right)$. It follows that, in such a case, there exists some $\varepsilon > 0$ such that we could have instead defined $\hat{w}_L = w_L + \varepsilon$ and $\hat{w}_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2\hat{w}_L} - 2\hat{w}_L\right)$, and we would still have had $\hat{w}_I > w_I$. □

Proof of Proposition 5. As observed in the text, the fact that $\mathcal{F}$ is convex (by Proposition 1) immediately implies that every point in the set

$$\left\{(w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right) \text{ and } w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}$$

can be implemented by some segmentation. It remains to show that every segmentation implements some point in the set. Consider an arbitrary segmentation, with $J \in \mathbb{N}$, $\boldsymbol{\delta}_I \in \Delta^J$, and $\boldsymbol{\delta}_L \in \Delta^J$. The conclusion follows from establishing the following two claims.

Claim 1: $w_L \geq \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right)$.

Proof of Claim 1: We argue as follows.

$$\begin{aligned}
w_L = \frac{1}{2} - \left[\frac{1}{2} - w_L\right] &= \frac{1}{2} - \left[\frac{1}{2} - \sum_{j=1}^{J} \delta_{Lj} \frac{(1-s_j)^2}{2}\right] \\
&= \frac{1}{2} - \frac{1}{2}\left[1 - \sum_{j=1}^{J} \delta_{Lj}\big(1 - 2s_j + s_j^2\big)\right] \\
&= \frac{1}{2} - \frac{1}{2}\sum_{j=1}^{J} \delta_{Lj} s_j (2 - s_j) \\
&= \frac{1}{2} - \frac{\alpha}{4(1-\alpha)} \sum_{j=1}^{J} \delta_{Ij} (1 - s_j)(2 - s_j) \\
&= \frac{1}{2} - \frac{\alpha}{4(1-\alpha)} \sum_{j=1}^{J} \delta_{Ij} \left[\sqrt{(1-s_j)^2} + (1-s_j)^2\right] \\
&\geq \frac{1}{2} - \frac{\alpha}{4(1-\alpha)} \left[\sqrt{\sum_{j=1}^{J} \delta_{Ij}(1-s_j)^2} + \sum_{j=1}^{J} \delta_{Ij}(1-s_j)^2\right] \\
&= \frac{1}{2} - \frac{\alpha}{4(1-\alpha)} \left[\sqrt{2w_I} + 2w_I\right]
\end{aligned}$$

In the above, the second and last steps use the definitions of w_L and w_I, respectively. The fourth step uses $\sum_{j=1}^{J} \delta_{Lj} = 1$. The fifth step uses the market maker's zero-profit condition: $(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}\frac{(1-s_j)^2}{2}$. The seventh step uses the fact that $f(x) = \sqrt{2x} + 2x$ is a concave function and that $\boldsymbol{\delta}_I \in \Delta^J$. The remaining steps are algebra.

Claim 2: $w_I \leq \dfrac{1-\alpha}{\alpha}\big(\sqrt{2w_L} - 2w_L\big)$.

Proof of Claim 2: We argue as follows.

$$\begin{aligned}
w_I &= \sum_{j=1}^{J} \delta_{Ij} \frac{(1-s_j)^2}{2} \\
&= \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj}(1-s_j)s_j \\
&= \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj} \left[\sqrt{(1-s_j)^2} - (1-s_j)^2\right] \\
&\leq \frac{1-\alpha}{\alpha} \left[\sqrt{\sum_{j=1}^{J} \delta_{Lj}(1-s_j)^2} - \sum_{j=1}^{J} \delta_{Lj}(1-s_j)^2\right] \\
&= \frac{1-\alpha}{\alpha}\left[\sqrt{2w_L} - 2w_L\right]
\end{aligned}$$

In the above, the first and last steps use the definitions of w_I and w_L, respectively. The second step uses the market maker's zero-profit condition: $(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}\frac{(1-s_j)^2}{2}$. The third step is algebra. The fourth step uses the fact that $f(x) = \sqrt{2x} - 2x$ is a concave function and that $\boldsymbol{\delta}_L \in \Delta^J$. □

Proof of Corollary 6. The characterization of $\mathcal{F}$ given by Proposition 5, together with the characterizations of $\mathcal{CS}$ and $\mathcal{RCS}$ given by Lemmas 2 and 3, respectively, immediately imply $\mathcal{F}^{\textit{anti-pareto}} \subseteq \mathcal{CS}$ and $\mathcal{F}^{\textit{pareto}} \subseteq \mathcal{RCS}$. It remains only to strengthen the former to $\mathcal{F}^{\textit{anti-pareto}} = \mathcal{CS}$. This would follow if it were true that $\mathcal{CS}^{\textit{anti-pareto}} = \mathcal{CS}$. To establish that, recall the characterization of $\mathcal{CS}$ from Lemma 2:

$$\mathcal{CS} = \left\{ (w_I, w_L) \,\middle|\, w_I \in \left[0, \frac{2(1-\alpha)^2}{(2-\alpha)^2}\right] \text{ and } w_L = \frac{1}{2} - \frac{\alpha}{4(1-\alpha)}\left(\sqrt{2w_I} + 2w_I\right) \right\}.$$

The slope of the curve that defines this set is

$$\frac{dw_L}{dw_I} = -\frac{\alpha}{4(1-\alpha)}\left(2 + \frac{1}{\sqrt{2w_I}}\right),$$

which never becomes positive. □

Proof of Proposition 5′. In light of the fact that $\mathcal{F}$ is convex (by Proposition 1), the characterization of $\mathcal{F}$ as

$$\left\{ (w_I, w_L) \,\middle|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I \in [\underline{w}_I(w_L), \overline{w}_I(w_L)] \right\}$$

is tautological. The primary substance of the proposition are the claims

$$\mathcal{CS} = \left\{ (w_I, w_L) \,\middle|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I = \underline{w}_I(w_L) \right\},$$
$$\mathcal{RCS} = \left\{ (w_I, w_L) \,\middle|\, w_L \in [\underline{w}_L, \overline{w}_L] \text{ and } w_I = \overline{w}_I(w_L) \right\}.$$

These, however, were fully established by the discussion in the text. □

Proof of Corollary 6′. This follows immediately from Proposition 5′. □

A.2 Proofs corresponding to Section 4

Proof of Lemma 7. Let $\tau' < \tau''$. Define $s' = s(\tau')$ and $s'' = s(\tau'')$. We show $s' \leq s''$. Define $\pi(s;\tau) := \big[1 - F_L(s)\big]s - \tau \int_s^1 (v - s)\, dF_I(v)$. In words, $\pi(s;\tau)$ are market making profits from a spread s given a toxicity τ. Given the assumption that F_L is atomless, π is continuous in s. We have

$$\begin{aligned}
\pi(s'';\tau') &= \big[1 - F_L(s'')\big]s'' - \tau' \int_{s''}^1 (v - s'')\, dF_I(v) \\
&\geq \big[1 - F_L(s'')\big]s'' - \tau'' \int_{s''}^1 (v - s'')\, dF_I(v),
\end{aligned}$$

which is equal to zero by definition. We also have $\pi(0;\tau') = -\tau' \int_0^1 v\, dF_I(v) \leq 0$. By the intermediate value theorem, there exists some $s \in [0, s'']$ for which $\pi(s,\tau') = 0$. Meanwhile, s' is defined as the *smallest* value of s for which $\pi(s,\tau') = 0$. It follows that $s' \leq s''$. □

Proof of Lemma 8. We begin by establishing three preliminary claims. Then we prove the result.

Preliminary Claim 1: $s(\tau)$ is left-continuous.

Proof of Preliminary Claim 1: Suppose we have an increasing sequence $\tau_k \to \tau_\circ$. Define $s_k = s(\tau_k)$. By monotonicity of $s(\tau)$, s_k is also an increasing sequence. This sequence is also bounded (by 1). We can therefore define $s_\circ$ as its limit. We wish to show that $s(\tau_\circ) = s_\circ$. Define $R(s) := [1 - F_L(s)]s$ and $C(s) := \int_s^1 (v - s)\, dF_I(v)$. Note that R and C are both nonnegative. In addition, C is continuous;

in fact, it is differentiable, with derivative $-[1-F_I(s)]$. R is also continuous, given that we have assumed F_L to be atomless. We also have $R(0)=0<C(0)$. Because $s(\tau)$ is the smallest solution to $R(s)=\tau C(s)$, it follows that *(i)* $R(s_k)=\tau_k C(s_k)$, and *(ii)* $R(s)<\tau_k C(s)$ for all $s<s_k$. By continuity of R and C, it follows that *(i)* $R(s_\circ)=\tau_\circ C(s_\circ)$, and *(ii)* $R(s)<\tau_\circ C(s)$ for all $s<s_\circ$. Hence, $s(\tau_\circ)=s_\circ$, as desired.

Preliminary Claim 2: $R(s(\tau))$ is left-continuous, i.e., $\lim_{\tau\to\tau_\circ^-} R(s(\tau))=R(s(\tau_\circ))$.

Proof of Preliminary Claim 2: That $R(\tau)$ is left-continuous follows from *(i)* the assumption that F_L is atomless, and *(ii)* the fact that, by Preliminary Claim 1, $s(\tau)$ is left-continuous.

Preliminary Claim 3: $R(s(\tau))$ can be discontinuous only at downward jumps, i.e., $R(s(\tau_\circ))\geq \lim_{\tau\to\tau_\circ^+} R(s(\tau))$ for all $\tau_\circ$.

Proof of Preliminary Claim 3: Suppose $R(\tau)$ is discontinuous at $\tau_\circ$. Given the assumption that F_L is atomless, such a discontinuity can happen only if $s(\tau)$ is discontinuous at $\tau_\circ$. Because $s(\tau)$ is weakly increasing (by Lemma 7), we must in particular have $s(\tau_\circ)<\lim_{\tau\to\tau_\circ^+} s(\tau)$.

As above, define $R(s):=[1-F_L(s)]s$ and $C(s):=\int_s^1 (v-s)\,dF_I(v)$. Note that $s(\tau)$ is the smallest solution to $R(s)=\tau C(s)$. Note also that C is decreasing; in fact, it is differentiable, with derivative $-[1-F_L(s)]$. Additionally, R is continuous, given that we have assumed F_L to be continuous. We argue as follows:

$$\lim_{\tau\to\tau_\circ^+} R(s(\tau))=R\left(\lim_{\tau\to\tau_\circ^+} s(\tau)\right)=\tau_\circ C\left(\lim_{\tau\to\tau_\circ^+} s(\tau)\right)<\tau_\circ C(s(\tau_\circ))=R(s(\tau_\circ)).$$

The first step follows because R is continuous. The second and fourth steps follow because $s(\tau)$ solves $R(s)=\tau C(s)$. The third step is because $s(\tau_\circ)<\lim_{\tau\to\tau_\circ^+} s(\tau)$ and because C is decreasing.

Proof of part (i): To prove part (i), suppose we have a sequence $\tau_k\to\tau_\circ$ where $\tau_k>\tau_\circ$ and $R(\tau_k)=R^*$ for all k. We cannot have $R(\tau_\circ)>R^*$, for R^* is the supremum of $R(\tau)$. We also cannot have $R(\tau_\circ)<R^*$, by Preliminary Claim 3. It follows that $R(\tau_\circ)=R^*$.

Proof of part (ii): To prove part (ii), suppose by way of contradiction that $R(s(\tau))<R^*$ for all $\tau\in[0,\infty)$ and moreover that $(\exists\bar\tau)(\forall\varepsilon>0)(\exists\tau<\bar\tau):R(s(\tau))\geq R^*-\varepsilon$. We can then construct a sequence of elements $\tau_k<\bar\tau$ such that $R(s(\tau_k))\to R^*$. Let $\tau_\circ$ be a limit point of that sequence. By Preliminary Claim 2, $\lim_{\tau\to\tau_\circ^-} R(s(\tau))=R(s(\tau_\circ))$, and by Preliminary Claim 3, $\lim_{\tau\to\tau_\circ^+} R(s(\tau))\leq R(s(\tau_\circ))$. Hence, $R(s(\tau_\circ))\geq R^*$. But this is a contradiction.

We have therefore established $(\forall\bar\tau)(\exists\varepsilon>0)(\forall\tau<\bar\tau):R(s(\tau))<R^*-\varepsilon$. In this case, $R(s(\infty))=R^*$ follows because it is the only way to have $R^*=\sup_\tau R(s(\tau))$. □

Proof of Proposition 9. I establish the result via a series of claims. Claims 1–3 together establish the necessity of (i) and (ii). Claim 5 establishes sufficiency.

Claim 1: $\mathcal{F}^{pareto}\subseteq\mathcal{RCS}$ for all $\alpha\in(0,1)$ only if $R(s(\tau))$ is concave on $\tau\in[0,\tau^*]$.

Proof of Claim 1: Define $\bar R(s(\tau))$ as the upper concave envelope of $R(s(\tau))$. By the definition of τ^*, $\bar R(s(\tau))$ is strictly increasing on $[0,\tau^*]$. Now, suppose $R(s(\tau))$ is not concave on $[0,\tau^*]$. It follows that there is some $\tau_\circ\in[0,\tau^*]$ for which $\bar R(s(\tau_\circ))>R(s(\tau_\circ))$. Note also that, by monotonicity of $\bar R(s(\tau))$, for all $\tau<\tau_\circ$ we have $\bar R(s(\tau_\circ))>\bar R(s(\tau))\geq R(s(\tau))$. Combining the previous two sentences, for all $\tau\leq\tau_\circ$ we have $\bar R(s(\tau_\circ))>R(s(\tau))$.

Define $\alpha_\circ=\frac{\tau_\circ}{1+\tau_\circ}$. It follows from the characterization of $\mathcal{F}$ in Section 4.1 that, when $\alpha=\alpha_\circ$, the highest feasible value for w_I is

$$\frac{1-\alpha_\circ}{\alpha_\circ}\bar R\left(s\left(\frac{\alpha_\circ}{1-\alpha_\circ}\right)\right)=\frac{1-\alpha_\circ}{\alpha_\circ}\bar R(s(\tau_\circ)).$$

Likewise, it follows from the characterization of $\mathcal{RCS}$ in Section 4.1 that, when $\alpha = \alpha_\circ$, the values of w_I that are implemented by reverse cream-skimming segmentations are those of the form

$$\frac{1-\alpha_\circ}{\alpha_\circ} R(s(\tau)) \quad \text{for } \tau \leq \tau_\circ.$$

As argued above, for all $\tau \leq \tau_\circ$ we have $\bar{R}(s(\tau_\circ)) > R(s(\tau))$. It follows that, when $\alpha = \alpha_\circ$, the highest feasible w_I is strictly greater than any w_I that can be implemented by reverse cream-skimming. Consequently, $\mathcal{F}^{pareto} \not\subseteq \mathcal{RCS}$.

Claim 2: $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$ only if $s(\tau)$ is continuous on $\tau \in [0, \tau^*]$.

Proof of Claim 2: Suppose $s(\tau)$ is discontinuous at some $\tau \in [0, \tau^*]$. Because $s(\tau)$ is increasing (by Lemma 7), any discontinuity must be an upward jump. Recall $R(s) = [1 - F_L(s)]s$. Define also $C(s) = \int_s^1 (v - s)\, dF_I(v)$. An upward jump in $s(\tau)$ must translate into a downward jump in $C(s(\tau))$. And because $R(s(\tau)) = \tau C(s(\tau))$, a downward jump in $C(s(\tau))$ implies a corresponding downward jump in $R(s(\tau))$. Any such discontinuity in $R(s(\tau))$ means that $R(s(\tau))$ is not concave on $[0, \tau^*]$.

Claim 3: Suppose $s(\tau)$ is continuous on $\tau \in [0, \tau^*]$. Then $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$ for all $\alpha \in (0, 1)$ only if the parametrically-defined curve (13) is concave on $s \in [0, s^*]$.

Proof of Claim 3: Suppose $s(\tau)$ is continuous on $\tau \in [0, \tau^*]$. Because $s(\tau)$ is also increasing (by Lemma 7), we have $s = s(\tau)$ for some $\tau \in [0, \tau^*]$ iff $s \in [0, s^*]$. The characterization of $\mathcal{RCS}$ in Section 4.1 can then be simplified as follows:

$$\mathcal{RCS} = \left\{ (w_I, w_L) \,\middle|\, w_I = \frac{1-\alpha}{\alpha}[1 - F_L(s)]s \text{ and } w_L = \int_s^1 (v - s)\, dF_L(v) \text{ for } s \in \left[0, s\Big(\frac{\alpha}{1-\alpha}\Big)\right] \right\}.$$

In short, $\mathcal{RCS}$ is a rescaled version of the graph of the parametrically-defined curve (13) on the domain $s \in \left[0, s(\frac{\alpha}{1-\alpha})\right]$. In particular, the w_I coordinate is scaled by a factor of $\frac{1-\alpha}{\alpha}$. Now, suppose the parametrically-defined curve (13) is not concave on $s \in [0, s^*]$. There are two cases.

- First, suppose $\tau^* < \infty$. Then define $\alpha_\circ = \frac{\tau^*}{1+\tau^*}$. Because $s(\frac{\alpha_\circ}{1-\alpha_\circ}) = s(\tau^*) = s^*$, with this choice for $\alpha_\circ$,

  $$\mathcal{RCS} = \left\{ (w_I, w_L) \,\middle|\, w_I = \frac{1-\alpha_\circ}{\alpha_\circ}[1 - F_L(s)]s \text{ and } w_L = \int_s^1 (v - s)\, dF_L(v) \text{ for } s \in [0, s^*] \right\}.$$

  Notice that $w_L(s) = \int_s^1 (v - s)\, dF_L(v)$ is continuous in s. It is also monotonically decreasing on the interval $s \in [0, s^*]$ from $w_L^{\max} := \int_0^1 v\, dF_L(v)$ to $w_L^{\min} := \int_{s^*}^1 (v - s^*)\, dF_L(v)$. Therefore, there exists some function $w_I^{RCS}(w_L)$ for which we can write

  $$\mathcal{RCS} = \left\{ (w_I, w_L) \,\middle|\, w_I = w_I^{RCS}(w_L) \text{ and } w_L \in \left[w_L^{\min}, w_L^{\max}\right] \right\}.$$

  Furthermore, because the parametrically-defined curve (13) is not concave on $s \in [0, s^*]$, $w_I^{RCS}(w_L)$ is likewise not concave on the interval $w_L \in \left[w_L^{\min}, w_L^{\max}\right]$. Next, define $\bar{w}_I^{RCS}(w_L)$ as the upper concave envelope of $w_I^{RCS}(w_L)$ over the interval $w_L \in \left[w_L^{\min}, w_L^{\max}\right]$. Because $w_I^{RCS}(w_L)$ is not concave on $w_L \in \left[w_L^{\min}, w_L^{\max}\right]$, there exists a $w_L^\circ \in \left[w_L^{\min}, w_L^{\max}\right]$ for which $\bar{w}_I^{RCS}(w_L^\circ) > w_I^{RCS}(w_L^\circ)$. Because τ^* is defined as the smallest value at which $[1 - F_L(s(\tau))]s(\tau)$ achieves its maximum, because $s^* = s(\tau)$, and because $s(\tau)$ is monotonically increasing, it follows that

$w_I^{RCS}(w_L^{\min}) > w_I^{RCS}(w_L)$ for all $w_L \in (w_L^{\min}, w_L^{\max}]$. It then further follows that $\bar{w}_I^{RCS}(w_L)$ is strictly decreasing on $w_L \in [w_L^{\min}, w_L^{\max}]$.

Clearly, every element of $\mathcal{RCS}$ is feasible. And because $\mathcal{F}$ is a convex set (by Proposition 1), the point $(w_I, w_L) = (\bar{w}_I^{RCS}(w_L^\circ), w_L^\circ)$ is feasible also. I claim that this feasible point is neither equal to nor Pareto dominated by any element of $\mathcal{RCS}$. It will follow that $\mathcal{F}^{pareto} \not\subseteq \mathcal{RCS}$. Indeed, suppose $(w_I, w_L) \in \mathcal{RCS}$ with $w_L = w_L^\circ$. Then $w_I = w_I^{RCS}(w_L^\circ)$. But by definition, $w_I^{RCS}(w_L^\circ) < \bar{w}_I^{RCS}(w_L^\circ)$. Alternatively, suppose $(w_I, w_L) \in \mathcal{RCS}$ with $w_L > w_L^\circ$. Then $w_I = w_I^{RCS}(w_L) \leq \bar{w}_I^{RCS}(w_L) < \bar{w}_I^{RCS}(w_L^\circ)$, where the first inequality follows from the definition of $\bar{w}_I^{RCS}(\cdot)$ as the upper concave envelope of $w_I^{RCS}(\cdot)$, and the second inequality follows from the fact that $\bar{w}_I^{RCS}(\cdot)$ is strictly decreasing.

- Second, suppose $\tau^* = \infty$. In this case, no $\alpha \in (0,1)$ is large enough to deliver $s(\frac{\alpha}{1-\alpha}) = s^*$, but for α sufficiently close to one, we can come arbitrarily close. It follows that for all $\alpha_\circ$ sufficiently close to one, both of the following will be true. First, $w_I^{RCS}(w_L)$ will fail to be concave somewhere on the interval $w_L \in [w_L^{\min}, w_L^{\max}]$. Second, $\bar{w}_I^{RCS}(w_L)$ will be strictly decreasing on the same interval, where all quantities are defined as in the previous case. Given this, all the previous arguments carry over, and we again conclude that $\mathcal{F}^{pareto} \not\subseteq \mathcal{RCS}$.

Claim 4: Suppose $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$. Then for all $\alpha \in (0,1)$, $\mathcal{RCS}$ contains a w_I-optimal point, i.e.,

$$\sup_{(w_I, w_L) \in \mathcal{F}} w_I = \max_{(w_I, w_L) \in \mathcal{RCS}} w_I.$$

Proof of Claim 4: There are two cases.

- *Case 1:* $\tau^* \leq \frac{\alpha}{1-\alpha}$. Following (10), we have the upper bound

$$w_I \leq \frac{1-\alpha}{\alpha} R(s^*). \tag{20}$$

Defining $\beta_I = \tau^* \frac{1-\alpha}{\alpha}$, we have that $\beta_I \in [0,1]$ and moreover that the reverse cream-skimming segmentation $\boldsymbol{\delta}_I^{RCS} = (\beta_I, 1-\beta_I)$ and $\boldsymbol{\delta}_L^{RCS} = (1,0)$ implements $\tau_1 = \tau^*$, hence $s_1 = s^*$. This segmentation therefore achieves the upper bound (20) and so must be w_I-optimal.

- *Case 2:* $\tau^* > \frac{\alpha}{1-\alpha}$. Given an arbitrary segmentation $(\boldsymbol{\delta}_I, \boldsymbol{\delta}_L)$, we can find a reverse cream-skimming segmentation that does at least as well in terms of w_I. Without loss, we can make some assumptions about the given segmentation.

  - Without loss, we can assume that, for the given segmentation, there exists at least one venue j for which $\delta_{Lj} = 0$. Indeed, if there were not, then we could modify the segmentation by adding an additional venue that receives no order flow, without affecting the resulting (w_I, w_L).
  - Without loss, we can also assume that, for the given segmentation, there exists a $J' < J$ such that *(i)* $\delta_{Lj} > 0$ for venues $j = 1, \ldots, J'$, and *(ii)* $\delta_{Lj} = 0$ for venues $j = J'+1, \ldots, J$. Indeed, if it were otherwise, then we could simply relabel venues to make it so.
  - Without loss, we can also assume that, for the given segmentation, $\tau_j \leq \tau^*$ for venues $j = 1, \ldots, J'$. Indeed, if it were otherwise, then we could modify the segmentation in the following way. For all venues $j = 1, \ldots, J'$ at which $\tau_j > \tau^*$, reduce δ_{Ij} until $\tau_j = \tau^*$; this reduces the spread to $s_j = s^*$. Commensurately increase δ_{Ij} at some other venue $j > J'$, where $\delta_{Lj} = 0$. Following (10), this modification can only improve w_I.

Define $\beta_I = \sum_{j=1}^{J'} \delta_{Ij}$. Using (10), this segmentation leads to

$$w_I = \frac{1-\alpha}{\alpha}\sum_{j=1}^{J}\delta_{Lj}R(s(\tau_j)) = \frac{1-\alpha}{\alpha}\sum_{j=1}^{J'}\delta_{Lj}R(s(\tau_j)) \leq \frac{1-\alpha}{\alpha}R\left(s\left(\sum_{j=1}^{J'}\delta_{Lj}\tau_j\right)\right)$$
$$= \frac{1-\alpha}{\alpha}R\left(s\left(\sum_{j=1}^{J'}\frac{\alpha\delta_{Ij}}{1-\alpha}\right)\right) = \frac{1-\alpha}{\alpha}R\left(s\left(\frac{\alpha\beta_I}{1-\alpha}\right)\right),$$

where the step involving the inequality follows from the assumption that $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$. Note, however, that the final term is, by (10), precisely the information-motivated investor welfare that prevails under the reverse cream-skimming segmentation $\boldsymbol{\delta}_I^{RCS} = (\beta_I, 1-\beta_I)$ and $\boldsymbol{\delta}_L^{RCS} = (1, 0)$.

If there exists a segmentation that is w_I-optimal in the class of reverse cream-skimming segmentations, then it follows from the above that this segmentation must be w_I-optimal overall. That this is indeed the case follows from results we have already shown above: because $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$, $s(\tau)$ is continuous on $\tau \in [0, \tau^*]$, hence we can argue as in the proof of Claim 3 that

$$\mathcal{RCS} = \left\{(w_I, w_L)\,\middle|\, w_I = \frac{1-\alpha}{\alpha}[1-F_L(s)]s \text{ and } w_L = \int_s^1 (v-s)\,dF_L(v) \text{ for } s \in \left[0, s\left(\frac{\alpha}{1-\alpha}\right)\right]\right\},$$

which is a compact set, and so must contain a w_I-optimal element.

Claim 5: $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$ for all $\alpha \in (0,1)$ if both *(i)* $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$, and *(ii)* the parametrically-defined curve (13) is concave on $s \in [0, s^*]$.

Proof of Claim 5: Suppose $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$, and suppose the parametrically-defined curve (13) is concave on $s \in [0, s^*]$. Fix some $\alpha \in (0,1)$. Choose some segmentation $(\hat{\boldsymbol{\delta}}_I, \hat{\boldsymbol{\delta}}_L)$. Applying (10) and (2b), this segmentation produces welfare outcomes

$$\hat{w}_I = \sum_{j=1}^{J}\hat{\delta}_{Lj}\frac{1-\alpha}{\alpha}s_j[1-F_L(s_j)]$$
$$\hat{w}_L = \sum_{j=1}^{J}\hat{\delta}_{Lj}\int_{s_j}^1 (v-s_j)\,dF_L(v)$$

In short, $(\hat{w}_I, \hat{w}_L)$ is a convex combination of points on the parametrically-defined curve

$$w_I(s) = \frac{1-\alpha}{\alpha}[1-F_L(s)]s \qquad \text{and} \qquad w_L(s) = \int_s^1 (v-s)\,dF_L(v) \tag{21}$$

Note that the only difference between this parametrically-defined curve and the one given by (13) is that the w_I coordinate has been rescaled by a factor of $\frac{1-\alpha}{\alpha}$. Hence, this curve is also concave on the domain $s \in [0, s^*]$. I establish the result by constructing some $(w_I^{\mathcal{RCS}}, w_L^{\mathcal{RCS}}) \in \mathcal{RCS}$ such that $w_I^{\mathcal{RCS}} \geq \hat{w}_I$ and $w_L^{\mathcal{RCS}} \geq \hat{w}_L$.

Without loss, we can make some assumptions about the given segmentation $(\hat{\boldsymbol{\delta}}_I, \hat{\boldsymbol{\delta}}_L)$.

- Without loss, we can assume that, for the given segmentation, there exists at least one venue j for which $\hat{\delta}_{Lj} = 0$. Indeed, if there were not, then we could modify the segmentation by adding an additional venue that receives no order flow, without affecting the resulting $(\hat{w}_I, \hat{w}_L)$.

- Without loss, we can also assume that, for the given segmentation, there exists a $J' < J$ such that *(i)* $\hat{\delta}_{Lj} > 0$ for venues $j = 1, \ldots, J'$, and *(ii)* $\hat{\delta}_{Lj} = 0$ for venues $j = J'+1, \ldots, J$. Indeed, if it were otherwise, then we could simply relabel venues to make it so.

- Without loss, we can also assume that, for the given segmentation, $s_j \leq s^*$ for venues $j = 1, \ldots, J'$. Indeed, if it were otherwise, then we could modify the segmentation in the following way. For all venues $j = 1, \ldots, J'$ at which $s_j > s^*$, reduce $\hat{\delta}_{Ij}$ until $s_j = s^*$. Commensurately increase $\hat{\delta}_{Ij}$ at some other venue $j > J'$, where $\delta_{Lj} = 0$. This modification can only improve both $\hat{w}_I$ and $\hat{w}_L$.

Based on these arguments, we can in fact assume that $(\hat{w}_I, \hat{w}_L)$ is a convex combination of points on graph of the parametrically-defined curve (21) restricted to the domain $s \in \big[0, s^*\big]$.

Now, because this curve is concave on this domain, there exists a point on the curve that equals or Pareto dominates $(\hat{w}_I, \hat{w}_L)$. If there are multiple such points, then take the one corresponding to the smallest value of s, which I define $\hat{s}$. We will be done if we can show that this point, $(w_I(\hat{s}), w_L(\hat{s}))$ is itself an element of $\mathcal{RCS}$.

Next, because $R(s(\tau))$ is concave on $\tau \in [0, \tau^*]$, Claim 2 implies that $s(\tau)$ is continuous on $\tau \in [0, \tau^*]$. Therefore, we can argue as in the proof of Claim 3 that $\mathcal{RCS}$ is defined as the graph of the parametrically-defined curve (21) restricted to the domain $s \in \big[0, s\big(\frac{\alpha}{1-\alpha}\big)\big]$. Thus, what we need to show is that $\hat{s} \leq s\big(\frac{\alpha}{1-\alpha}\big)$.

Because $w_L(s)$ is monotonically decreasing in s, we have that for all $s \leq \hat{s}$, $w_L(s) \geq w_L(\hat{s}) \geq w_L$. We must therefore have that, for all $s < \hat{s}$, $w_I(s) < \hat{w}_I$, for otherwise that would contradict the definition of $\hat{s}$ as the smallest value of s for which the corresponding point on the curve (21) equals or Pareto dominates $(\hat{w}_I, \hat{w}_L)$. Suppose then, by way of contradiction, that $\hat{s} > s\big(\frac{\alpha}{1-\alpha}\big)$. It would follow that every element of $\mathcal{RCS}$ entails a value of w_I that is strictly less than $\hat{w}_I$. But this would contradict Claim 4, which shows that $\mathcal{RCS}$ contains a w_I-optimal element of $\mathcal{F}$. □

B Point-mass distribution for F_I

To probe robustness, this appendix considers an alternate parametrization of the model, in which F_I is a point mass on σ, for some $\sigma \in (0,1]$. In contrast, Section 3 assumed that F_I was $U[0,1]$. As in Section 3, this appendix assumes that F_L is $U[0,1]$. Lee and Wang (2023), for example, make a similar combination of distributional assumptions.

In all other regards, the model is precisely as described in Section 2. Appendix B.1 treats the case of $\sigma \in (0,1)$, while Appendix B.2 treats the case of $\sigma = 1$.

B.1 Case 1: $F_I(v) = \mathbb{1}\{v \geq \sigma\}$, where $\sigma \in (0,1)$

With this parametrization, the spread on venue j must be the smallest solution to

$$(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}(\sigma - s_j),$$

Define, as before, $\tau_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$ as the toxicity of venue j's order flow. Define also the function

$$s(\tau) := \frac{1}{2}\left[\tau + 1 - \sqrt{(\tau+1)^2 - 4\tau\sigma}\right].$$

We then have $s_j = s(\tau_j)$. From the spreads, the welfares of the two trader types can be computed as

$$w_I = \sum_{j=1}^{J} \delta_{Ij}(\sigma - s_j) \qquad \text{and} \qquad w_L = \sum_{j=1}^{J} \delta_{Lj}\frac{(1-s_j)^2}{2}.$$

B.1.1 Cream-skimming and reverse cream-skimming

Cream-skimming. In cream-skimming scenarios, we have $\boldsymbol{\delta}_L = (\beta_L, 1-\beta_L)$ and $\boldsymbol{\delta}_I = (1,0)$. This gives $s_1 = s\left(\frac{\alpha}{(1-\alpha)\beta_L}\right)$ and $s_2 = s(0) = 0$. Note that because $\beta_L \in [0,1]$, we have $s_1 \in \left[s\left(\frac{\alpha}{1-\alpha}\right), s(\infty)\right] = \left[\frac{1-\sqrt{1-4\alpha(1-\alpha)\sigma}}{2(1-\alpha)}, \sigma\right]$.

We have $w_I = \sigma - s_1$. Because $s_1 \in \left[\frac{1-\sqrt{1-4\alpha(1-\alpha)\sigma}}{2(1-\alpha)}, \sigma\right]$, we have

$$w_I \in \left[0, \sigma - \frac{1}{2(1-\alpha)}(1 - \sqrt{1-4\alpha(1-\alpha)\sigma})\right].$$

Next, note that we can invert the equation for $s(\tau)$ to obtain $\tau = \frac{s(1-s)}{\sigma - s}$. It follows that $\beta_L = \frac{\alpha}{1-\alpha}\frac{\sigma - s_1}{s_1(1-s_1)}$. We also have

$$\begin{aligned} w_L = \beta_L\frac{(1-s_1)^2}{2} + \frac{1-\beta_L}{2} &= \frac{1}{2} - \frac{\beta_L}{2}s_1(2-s_1) \\ &= \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{\sigma - s_1}{2s_1(1-s_1)}s_1(2-s_1) \\ &= \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{(\sigma - s_1)(2-s_1)}{2(1-s_1)}. \end{aligned}$$

Using $s_1 = \sigma - w_I$, we can write $w_L = \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I(2-\sigma+w_I)}{2(1-\sigma+w_I)}$. In summary,

$$\mathcal{CS} = \left\{(w_I, w_L) \,\middle|\, w_I \in \left[0, \sigma - \frac{1-\sqrt{1-4\alpha(1-\alpha)\sigma}}{2(1-\alpha)}\right] \text{ and } w_L = \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I(2-\sigma+w_I)}{2(1-\sigma+w_I)}\right\}.$$

Reverse cream-skimming. In reverse cream-skimming scenarios, we have $\boldsymbol{\delta}_L = (1, 0)$ and $\boldsymbol{\delta}_I = (\beta_I, 1-\beta_I)$. This gives $s_1 = s\left(\frac{\alpha\beta_I}{1-\alpha}\right)$ and $s_2 = s(\infty) = \sigma$. Note that because $\beta_I \in [0,1]$, we have $s_1 \in \left[s(0), s\left(\frac{\alpha}{1-\alpha}\right)\right] = \left[0, \frac{1-\sqrt{1-4\alpha(1-\alpha)\sigma}}{2(1-\alpha)}\right]$.

We have $w_L = \frac{(1-s_1)^2}{2}$. Because $s_1 \in \left[0, \frac{1-\sqrt{1-4\alpha(1-\alpha)\sigma}}{2(1-\alpha)}\right]$, we have

$$w_L \in \left[\frac{\left(1-2\alpha+\sqrt{1-4\alpha(1-\alpha)\sigma}\right)^2}{8(1-\alpha)^2}, \frac{1}{2}\right].$$

Next, as above, we can invert to obtain $\tau = \frac{s(1-s)}{\sigma-s}$. It follows that $\beta_I = \frac{1-\alpha}{\alpha}\frac{s_1(1-s_1)}{\sigma-s_1}$. We also have $w_I = \beta_I(\sigma - s_1) + (1-\beta_I)(\sigma - s_2) = \beta_I(\sigma - s_1) = \frac{1-\alpha}{\alpha}s_1(1-s_1)$. Inverting $w_L = \frac{(1-s_1)^2}{2}$, we obtain $s_1 = 1 - \sqrt{2w_L}$. Hence, we can write $w_I = \frac{1-\alpha}{\alpha}\left(1-\sqrt{2w_L}\right)\sqrt{2w_L} = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)$. In summary,

$$\mathcal{RCS} = \left\{(w_I, w_L) \,\middle|\, w_L \in \left[\frac{\left(1-2\alpha+\sqrt{1-4\alpha(1-\alpha)\sigma}\right)^2}{8(1-\alpha)^2}, \frac{1}{2}\right] \text{ and } w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}.$$

B.1.2 Characterizing the feasible set

As in Section 3, the feasible set consists of the area between the cream-skimming and reverse cream-skimming outcomes. See an illustration in Figure 3.

Proposition 10. *For the version of the model considered in this appendix, the feasible welfare set is*

$$\mathcal{F} = \left\{(w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I(2-\sigma+w_I)}{2(1-\sigma+w_I)} \text{ and } w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}.$$

The proof is similar to that of Proposition 5.

Proof of Proposition 10. We have already shown, by construction, that the boundary points of the set

$$\left\{(w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I(2-\sigma+w_I)}{2(1-\sigma+w_I)} \text{ and } w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}$$

can be implemented, (either by a cream-skimming segmentation or by a reverse cream-skimming segmentation). Moreover, $\mathcal{F}$ is convex (by Proposition 1). It follows that every point in the set can be implemented by some segmentation. It remains to show that every segmentation implements some point in the set. Consider an arbitrary segmentation, with $J \in \mathbb{N}$, $\boldsymbol{\delta}_I \in \Delta^J$, and $\boldsymbol{\delta}_L \in \Delta^J$. The conclusion follows from establishing the following two claims.

Claim 1: $w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I(2-\sigma+w_I)}{2(1-\sigma+w_I)}$.

Proof of Claim 1: We argue as follows.

$$w_L = \frac{1}{2} - \left[\frac{1}{2} - w_L\right] = \frac{1}{2} - \left[\frac{1}{2} - \sum_{j=1}^{J}\delta_{Lj}\frac{(1-s_j)^2}{2}\right]$$

$$
\begin{aligned}
&= \frac{1}{2} - \frac{1}{2}\left[1 - \sum_{j=1}^{J} \delta_{Lj}\big(1 - 2s_j + s_j^2\big)\right] \\
&= \frac{1}{2} - \frac{1}{2}\sum_{j=1}^{J} \delta_{Lj} s_j (2 - s_j) \\
&= \frac{1}{2} - \frac{\alpha}{1-\alpha}\sum_{j=1}^{J} \delta_{Ij} \frac{(\sigma - s_j)(2 - s_j)}{2(1 - s_j)} \\
&= \frac{1}{2} - \frac{\alpha}{1-\alpha}\sum_{j=1}^{J} \delta_{Ij} \frac{(\sigma - s_j)\big(2 - \sigma + (\sigma - s_j)\big)}{2\big(1 - \sigma + (\sigma - s_j)\big)} \\
&\geq \frac{1}{2} - \frac{\alpha}{1-\alpha} \frac{\left(\sum_{j=1}^{J} \delta_{Ij}\,(\sigma - s_j)\right)\left(2 - \sigma + \sum_{j=1}^{J} \delta_{Ij}(\sigma - s_j)\right)}{2\left(1 - \sigma + \sum_{j=1}^{J} \delta_{Ij}(\sigma - s_j)\right)} \\
&= \frac{1}{2} - \frac{\alpha}{1-\alpha} \frac{w_I(2 - \sigma + w_I)}{2(1 - \sigma + w_I)}
\end{aligned}
$$

In the above, the second and last steps use the definitions of w_L and w_I, respectively. The fourth step uses $\sum_{j=1}^{J} \delta_{Lj} = 1$. The fifth step uses the market maker's zero-profit condition: $(1-\alpha)\delta_{Lj}(1 - s_j)s_j = \alpha\delta_{Ij}(\sigma - s_j)$. The seventh step uses the fact that $f(x) = \frac{x(2-\sigma+x)}{2(1-\sigma+x)}$ is a concave function on the domain $x \geq 0$ (thanks to $\sigma < 1$) and that $\boldsymbol{\delta}_I \in \Delta^J$. The remaining steps are algebra.

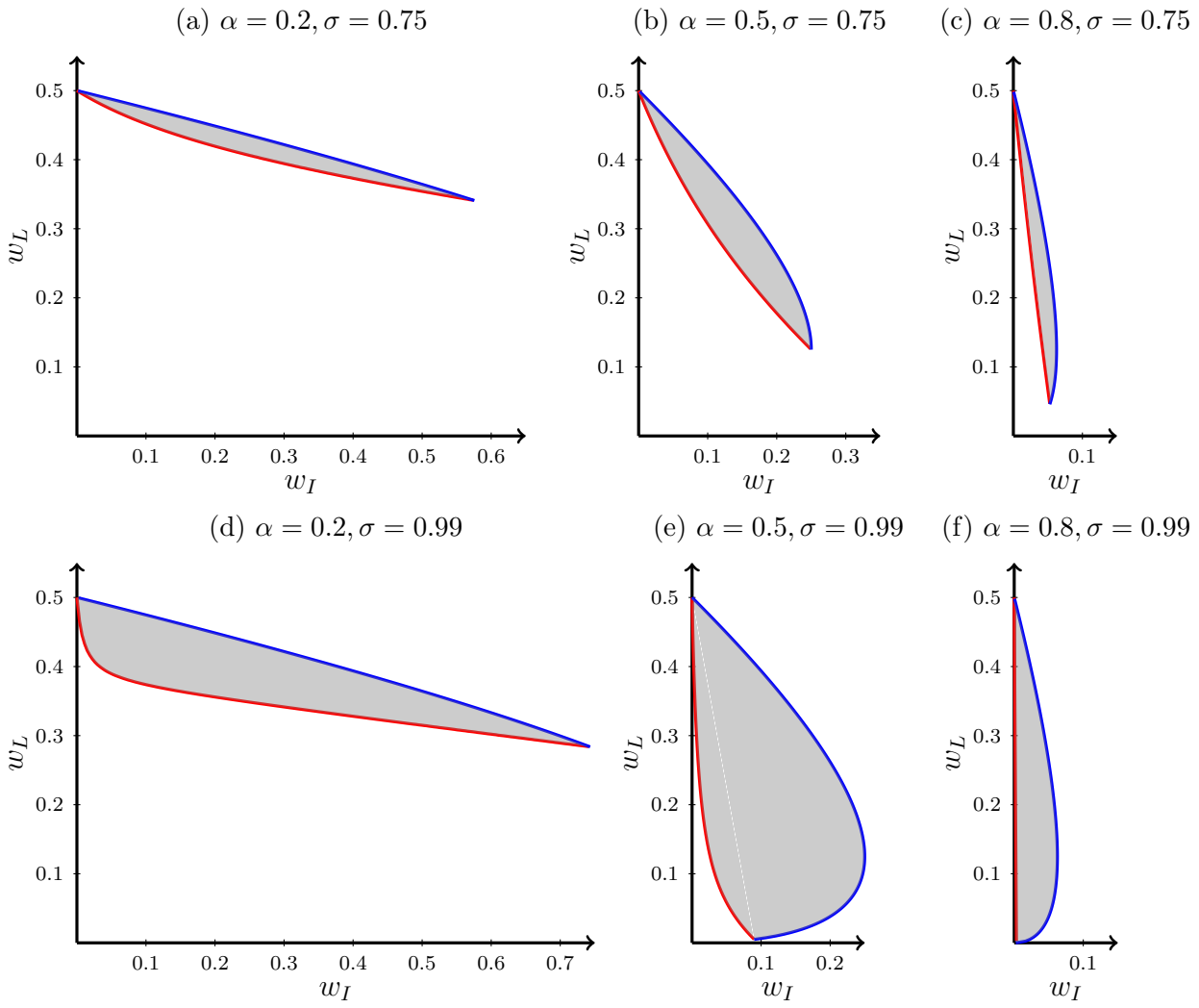
Claim 2: $w_I \leq \dfrac{1-\alpha}{\alpha}\big(\sqrt{2w_L} - 2w_L\big)$.

Proof of Claim 2: We argue as follows.

$$
\begin{aligned}
w_I &= \sum_{j=1}^{J} \delta_{Ij}(\sigma - s_j) \\
&= \frac{1-\alpha}{\alpha}\sum_{j=1}^{J} \delta_{Lj}(1 - s_j)s_j \\
&= \frac{1-\alpha}{\alpha}\sum_{j=1}^{J} \delta_{Lj}\left[\sqrt{(1 - s_j)^2} - (1 - s_j)^2\right] \\
&\leq \frac{1-\alpha}{\alpha}\left[\sqrt{\sum_{j=1}^{J} \delta_{Lj}(1 - s_j)^2} - \sum_{j=1}^{J} \delta_{Lj}(1 - s_j)^2\right] \\
&= \frac{1-\alpha}{\alpha}\big[\sqrt{2w_L} - 2w_L\big]
\end{aligned}
$$

In the above, the first and last steps use the definitions of w_I and w_L, respectively. The second step uses the market maker's zero-profit condition: $(1-\alpha)\delta_{Lj}(1 - s_j)s_j = \alpha\delta_{Ij}(\sigma - s_j)$. The third step is algebra. The fourth step uses the fact that $f(x) = \sqrt{2x} - 2x$ is a concave function and that $\boldsymbol{\delta}_L \in \Delta^J$. □

Figure 3: The feasible welfare set



The feasible sets for $F_I(v) = \mathbb{1}\{v \geq \sigma\}$, with $\sigma \in \{0.75, 0.99\}$, $F_L = U[0,1]$, and $\alpha \in \{0.2, 0.5, 0.8\}$. In each panel, the red (blue) curve depicts the welfare pairs implementable by cream-skimming (reverse cream-skimming), and the gray area represents the welfare pairs implementable by all forms of segmentation.

B.2 Case 2: $F_I(v) = \mathbb{1}\{v \geq 1\}$

With this parametrization, the spread on venue j must be the smallest solution to

$$(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}(1-s_j),$$

Define, as before, $\tau_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$ as the toxicity of venue j's order flow. We then have

$$s_j = \begin{cases} \tau_j & \text{if } \tau_j \leq 1, \\ 1 & \text{if } \tau_j > 1. \end{cases}$$

From the spreads, the welfares of the two trader types can be computed as

$$w_I = \sum_{j=1}^{J} \delta_{Ij}(1-s_j) \qquad \text{and} \qquad w_L = \sum_{j=1}^{J} \delta_{Lj}\frac{(1-s_j)^2}{2}.$$

B.2.1 Cream-skimming and reverse cream-skimming

Cream-skimming. In cream-skimming scenarios, we have $\boldsymbol{\delta}_L = (\beta_L, 1-\beta_L)$ and $\boldsymbol{\delta}_I = (1,0)$. This gives $\tau_1 = \frac{\alpha}{(1-\alpha)\beta_L}$ and $s_2 = 0$. There are two cases.

- First, suppose $\alpha \geq \frac{1}{2}$. In that case, for all $\beta_L \in [0,1]$, we have $\tau_1 \geq 1$ and therefore $s_1 = 1$. It follows that $w_I = 0$ and $w_L = \frac{1-\beta_L}{2}$. Because $\beta_L \in [0,1]$, we have $w_L \in \left[0, \frac{1}{2}\right]$. In summary,

$$\mathcal{CS} = \left\{ (w_I, w_L) \,\middle|\, w_I = 0 \text{ and } w_L \in \left[0, \frac{1}{2}\right] \right\}.$$

- Second, suppose $\alpha < \frac{1}{2}$. There are two sub-cases.

  In the first sub-case, $\beta_L \in \left[0, \frac{\alpha}{1-\alpha}\right]$. In this case, we have $\tau_1 \geq 1$ and therefore $s_1 = 1$. It follows that $w_I = 0$ and $w_L = \frac{1-\beta_L}{2}$. Because $\beta_L \in \left[0, \frac{\alpha}{1-\alpha}\right]$, we have $w_L \in \left[\frac{1-2\alpha}{2-2\alpha}, \frac{1}{2}\right]$.

  In the second sub-case, $\beta_L \in \left(\frac{\alpha}{1-\alpha}, 1\right]$. In this case, we have $\tau_1 \in \left[\frac{\alpha}{1-\alpha}, 1\right)$ and therefore $s_1 = \tau_1$. It follows that $w_I = 1-\tau_1$. Because $\tau_1 \in \left[\frac{\alpha}{1-\alpha}, 1\right)$, we have $w_I \in \left(0, \frac{1-2\alpha}{1-\alpha}\right]$. Rearranging this, we have $\tau_1 = 1 - w_I$. It also follows that $w_L = \beta_L\frac{(1-\tau_1)^2}{2} + \frac{1-\beta_L}{2}$. Plugging in $\tau_1 = 1-w_I$ and $\beta_L = \frac{\alpha}{(1-\alpha)\tau_1} = \frac{\alpha}{(1-\alpha)(1-w_I)}$, we get $w_L = \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I+1}{2}$.

  In summary,

$$\begin{aligned} \mathcal{CS} = &\left\{ (w_I, w_L) \,\middle|\, w_I = 0 \text{ and } w_L \in \left[\frac{1-2\alpha}{2-2\alpha}, \frac{1}{2}\right] \right\} \\ &\cup \left\{ (w_I, w_L) \,\middle|\, w_I \in \left(0, \frac{1-2\alpha}{1-\alpha}\right] \text{ and } w_L = \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{w_I+1}{2} \right\}. \end{aligned}$$

Reverse cream-skimming. In reverse cream-skimming scenarios, we have $\boldsymbol{\delta}_L = (1, 0)$ and $\boldsymbol{\delta}_I = (\beta_I, 1 - \beta_I)$. This gives $\tau_1 = \frac{\alpha\beta_I}{1-\alpha}$ and $s_2 = 1$. There are two cases.

- First, suppose $\alpha < \frac{1}{2}$. Because $\beta_I \in [0, 1]$, we have $\tau_1 \in \left[0, \frac{\alpha}{1-\alpha}\right]$. Because $\alpha < \frac{1}{2}$, we have $\tau_1 < 1$ and therefore $s_1 = \tau_1$. It follows that $w_L = \frac{(1-\tau_1)^2}{2}$. Because $\tau_1 \in \left[0, \frac{\alpha}{1-\alpha}\right]$, we have $w_L \in \left[\frac{(1-2\alpha)^2}{2(1-\alpha)^2}, \frac{1}{2}\right]$. Inverting $w_L = \frac{(1-\tau_1)^2}{2}$, we obtain $\tau_1 = 1 - \sqrt{2w_L}$. It also follows that $w_I = \beta_I(1 - \tau_1)$. Plugging in $\tau_1 = 1 - \sqrt{2w_L}$ and $\beta_I = \frac{(1-\alpha)\tau_1}{\alpha} = \frac{(1-\alpha)(1-\sqrt{2w_L})}{\alpha}$, we get $w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)$. In summary,

$$\mathcal{RCS} = \left\{(w_I, w_L) \,\middle|\, w_L \in \left[\frac{(1-2\alpha)^2}{2(1-\alpha)^2}, \frac{1}{2}\right] \text{ and } w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}.$$

- Second, suppose $\alpha \geq \frac{1}{2}$. There are two sub-cases.

  In the first sub-case, $\beta_I \in \left[0, \frac{1-\alpha}{\alpha}\right]$. In this case, we have $\tau_1 \in [0, 1]$ and therefore $s_1 = \tau_1$. It follows that $w_L = \frac{(1-\tau_1)^2}{2}$. Because $\tau_1 \in [0, 1]$, we have $w_L \in \left[0, \frac{1}{2}\right]$. We can also argue as above that $w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)$.

  In the second sub-case, $\beta_I \in \left(\frac{1-\alpha}{\alpha}, 1\right]$. In this case, we have $\tau_1 > 1$ and therefore $s_1 = 1$. It follows that $w_I = w_L = 0$.

  In summary,

$$\mathcal{RCS} = \left\{(w_I, w_L) \,\middle|\, w_L \in \left[0, \frac{1}{2}\right] \text{ and } w_I = \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}.$$

B.2.2 Characterizing the feasible set

As in Section 3, the feasible set consists of the area between the cream-skimming and reverse cream-skimming outcomes. See an illustration in Figure 4.

Proposition 11. *For the version of the model considered in this appendix, the feasible welfare set is*

$$\mathcal{F} = \left\{(w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{1+w_I}{2} \text{ and } 0 \leq w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\},$$

which, in the case of $\alpha \geq \frac{1}{2}$ can be written more simply as

$$\mathcal{F} = \left\{(w_I, w_L) \,\middle|\, 0 \leq w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}.$$

The proof is similar to that of Propositions 5 and 10.

Proof of Proposition 11. We have already shown, by construction, that the boundary points of the set

$$\left\{(w_I, w_L) \,\middle|\, w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{1+w_I}{2} \text{ and } 0 \leq w_I \leq \frac{1-\alpha}{\alpha}\left(\sqrt{2w_L} - 2w_L\right)\right\}$$

can be implemented, (either by a cream-skimming segmentation or by a reverse cream-skimming segmentation). Moreover, $\mathcal{F}$ is convex (by Proposition 1). It follows that every point in the set can

be implemented by some segmentation. It remains to show that every segmentation implements some point in the set. Consider an arbitrary segmentation, with $J \in \mathbb{N}$, $\boldsymbol{\delta}_I \in \Delta^J$, and $\boldsymbol{\delta}_L \in \Delta^J$. It is clear that we must have $w_I \geq 0$. The conclusion therefore follows from establishing the following two claims.

Claim 1: $w_L \geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{1+w_I}{2}$.

Proof of Claim 1: We argue as follows.

$$\begin{aligned}
w_L = \frac{1}{2} - \left[\frac{1}{2} - w_L\right] &= \frac{1}{2} - \left[\frac{1}{2} - \sum_{j=1}^{J} \delta_{Lj} \frac{(1-s_j)^2}{2}\right] \\
&= \frac{1}{2} - \frac{1}{2}\left[1 - \sum_{j=1}^{J} \delta_{Lj}\big(1 - 2s_j + s_j^2\big)\right] \\
&= \frac{1}{2} - \frac{\sum_{j=1}^{J} \delta_{Lj} s_j (2 - s_j)}{2} \\
&\geq \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{\sum_{j=1}^{J} \delta_{Ij}(2 - s_j)}{2} \\
&= \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{1 + \sum_{j=1}^{J} \delta_{Ij}(1 - s_j)}{2} \\
&= \frac{1}{2} - \frac{\alpha}{1-\alpha}\frac{1 + w_I}{2}
\end{aligned}$$

In the above, the second and last steps use the definitions of w_L and w_I, respectively. The fourth step uses $\sum_{j=1}^{J} \delta_{Lj} = 1$. The fifth step uses the fact that s_j must be the smallest solution to the market maker's zero-profit condition $(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}(1-s_j)$. To explain further, there are two cases: *(i)* if $s_j < 1$, then we must have $s_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$, and the step holds with equality, or *(ii)* if $s_j = 1$, then we must have $s_j \leq \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$, for it were otherwise, then a smaller solution to the zero-profit condition would exist. The sixth step uses $\sum_{j=1}^{J} \delta_{Ij} = 1$. The remaining steps are algebra.

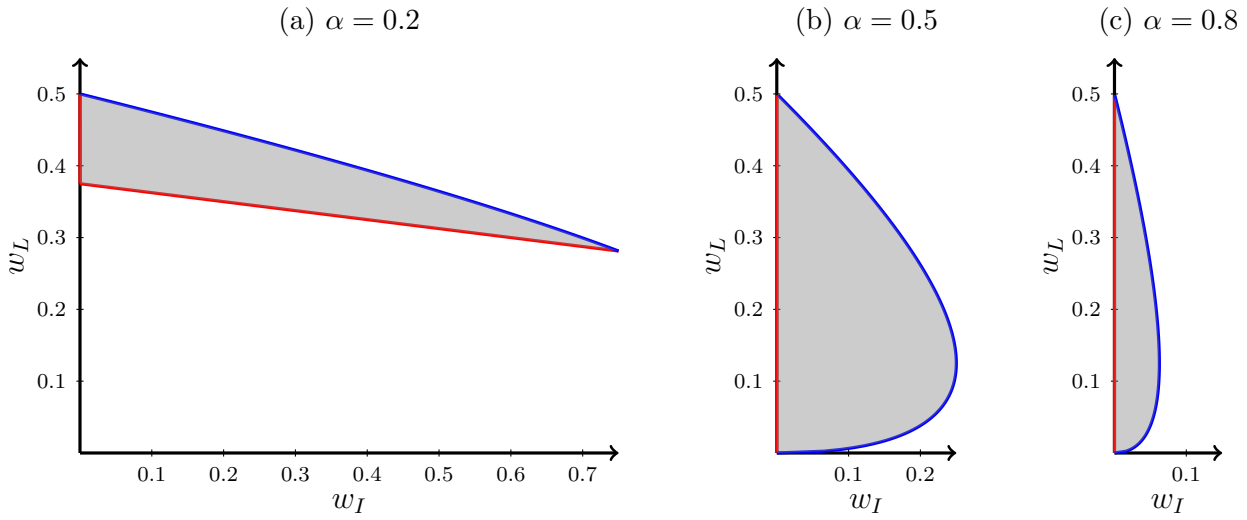
Claim 2: $w_I \leq \frac{1-\alpha}{\alpha}\big(\sqrt{2w_L} - 2w_L\big)$.

Proof of Claim 2: We argue as follows.

$$\begin{aligned}
w_I &= \sum_{j=1}^{J} \delta_{Ij}(1 - s_j) \\
&= \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj}(1 - s_j)s_j \\
&= \frac{1-\alpha}{\alpha} \sum_{j=1}^{J} \delta_{Lj}\left[\sqrt{(1-s_j)^2} - (1-s_j)^2\right] \\
&\leq \frac{1-\alpha}{\alpha}\left[\sqrt{\sum_{j=1}^{J} \delta_{Lj}(1-s_j)^2} - \sum_{j=1}^{J} \delta_{Lj}(1-s_j)^2\right] \\
&= \frac{1-\alpha}{\alpha}\big[\sqrt{2w_L} - 2w_L\big]
\end{aligned}$$

In the above, the first and last steps use the definitions of w_I and w_L, respectively. The second step uses the market maker's zero-profit condition: $(1-\alpha)\delta_{Lj}(1-s_j)s_j = \alpha\delta_{Ij}(1-s_j)$. The third step is algebra. The fourth step uses the fact that $f(x) = \sqrt{2x} - 2x$ is a concave function and that $\boldsymbol{\delta}_L \in \Delta^J$. □

Figure 4: The feasible welfare set



The feasible sets for $F_I(v) = \mathbb{1}\{v \geq 1\}$, $F_L = U[0,1]$, and $\alpha \in \{0.2, 0.5, 0.8\}$. In each panel, the red (blue) curve depicts the welfare pairs implementable by cream-skimming (reverse cream-skimming), and the gray area represents the welfare pairs implementable by all forms of segmentation.

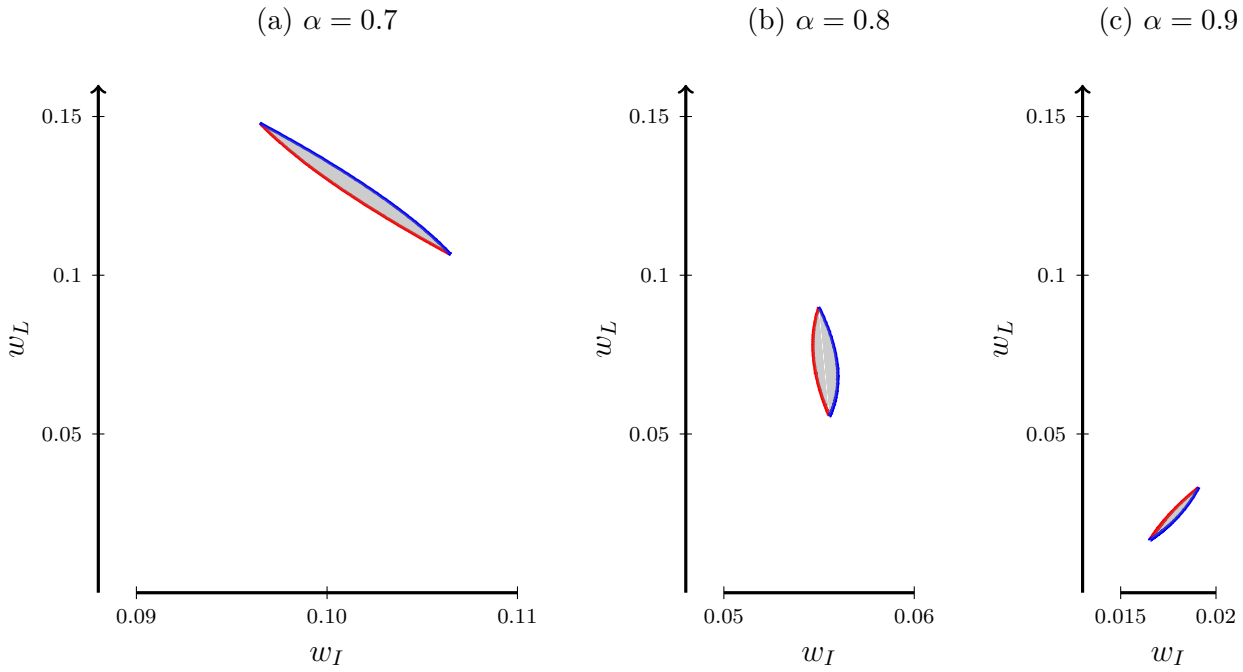
C Examples

C.1 With an imperfect binary signal, might have $\mathcal{F}^{\textit{anti-pareto}} \subsetneq \mathcal{CS}$

Recall that, in the uniform case, if segmentations can be based directly on the investor's trading motive, then we have $\mathcal{F}^{\textit{anti-pareto}} = \mathcal{CS}$. Section 3.4 considers an extension of the uniform case in which segmentations can be based only on potentially imperfect signals about the investor's trading motive. There, I argued only that $\mathcal{F}^{\textit{anti-pareto}} \subseteq \mathcal{CS}$. This appendix provides an example to illustrate that the stronger result ($\mathcal{F}^{\textit{anti-pareto}} = \mathcal{CS}$) does not carry over to this extension.

In particular, suppose that $\rho_I = \frac{2}{3}$ and $\rho_L = \frac{1}{3}$. If α is sufficiently large, then $\mathcal{CS} \subseteq \mathcal{F}^{\textit{anti-pareto}}$ does not obtain. See panels (b) and (c) of Figure 5.

Figure 5: The feasible welfare set



The feasible sets for $F_I = F_L = U[0,1]$, $\rho_I = \frac{2}{3}$, and $\rho_L = \frac{1}{3}$, and $\alpha \in \{0.7, 0.8, 0.9\}$. In each panel, the red (blue) curve depicts the welfare pairs implementable by cream-skimming (reverse cream-skimming), and the gray area represents the welfare pairs implementable by all forms of segmentation.

D Robustness of secondary results

D.1 The conditions under which the primary results extend do not suffice to ensure that the secondary results also extend

Three features of the uniform case analyzed in Section 3 were *(i)* $\mathcal{F}$ is the area between $\mathcal{CS}$ and $\mathcal{RCS}$, *(ii)* $\mathcal{F}^{pareto} \subseteq \mathcal{RCS}$, and *(iii)* $\mathcal{F}^{anti\text{-}pareto} = \mathcal{CS}$. As seen in Section 4, feature *(ii)* extends to a wide class of more general distributions. One might wonder if—under the same conditions—features *(i)* and *(iii)* also extend. They do not.

To illustrate, assume, as in Appendix B.2, that $F_I(v) = \mathbb{1}\{v \geq 1\}$. Now, however, assume that $F_L(v) = v^2$. This parametrization satisfies the conditions of Proposition 9, meaning that feature *(ii)* extends to this case.[10] Yet, we shall see that the same is not true for the other features.

With this parametrization, the spread on venue j must be the smallest solution to

$$(1-\alpha)\delta_{Lj}(1-s_j^2)s_j = \alpha\delta_{Ij}(1-s_j). \tag{22}$$

Define, as before, $\tau_j = \frac{\alpha\delta_{Ij}}{(1-\alpha)\delta_{Lj}}$ as the toxicity of venue j's order flow. Define also the function

$$s(\tau) = \begin{cases} \sqrt{\tau + \frac{1}{4}} - \frac{1}{2} & \text{if } \tau \leq 2, \\ 1 & \text{if } \tau > 2, \end{cases} \tag{23}$$

We then have $s_j = s(\tau_j)$. From the spreads, the welfares of the two trader types can be computed as

$$w_I = \sum_{j=1}^{J} \delta_{Ij}(1-s_j) \qquad \text{and} \qquad w_L = \sum_{j=1}^{J} \delta_{Lj}\frac{(1-s_j)^2(2+s_j)}{3}.$$

Cream-skimming. In cream-skimming scenarios, we have $\boldsymbol{\delta}_L = (\beta_L, 1-\beta_L)$ and $\boldsymbol{\delta}_I = (1,0)$. Here, we have $s_2 = 0$ and $s_1 = s(\tau_1)$, where $\tau_1 = \frac{\alpha}{(1-\alpha)\beta_L}$. There are two cases.

If $\beta_L < \frac{\alpha}{2(1-\alpha)}$, then $\tau_1 > 2$, meaning that the spread on venue 1 is $s_1 = 1$, and hence

$$w_I = 0 \qquad \text{and} \qquad w_L = (1-\beta_L)\frac{2}{3}.$$

If $\alpha \geq \frac{2}{3}$, then all cream-skimming outcomes take this form, so that

$$\mathcal{CS} = \left\{ (w_I, w_L) \,\middle|\, w_I = 0 \text{ and } w_L \in \left[0, \frac{2}{3}\right] \right\}.$$

On the other hand, if $\alpha < \frac{2}{3}$, then it is possible to have $\beta_L \geq \frac{\alpha}{2(1-\alpha)}$, in which case $\tau_1 \leq 2$, meaning that $s_1 = \sqrt{\tau_1 + \frac{1}{4}} - \frac{1}{2}$. It follows that have $w_I = 1 - s_1 = \frac{3}{2} - \sqrt{\tau_1 + \frac{1}{4}}$. Because $\beta_L \in \left[\frac{\alpha}{2(1-\alpha)}, 1\right]$, we have $\tau_1 \in [\frac{\alpha}{1-\alpha}, 2]$, and therefore

$$w_I \in \left[0, \frac{3}{2} - \sqrt{\frac{\alpha}{1-\alpha} + \frac{1}{4}}\right].$$

[10] Indeed, we have $R(s) = (1-s^2)s$, and on the domain $\tau \in [0, \tau^*]$, we have $s(\tau) = \sqrt{\tau + \frac{1}{4}} - \frac{1}{2}$. Hence, $R(s(\tau)) = \frac{3\tau}{2} - \frac{\tau\sqrt{1+4\tau}}{2}$, which is indeed concave, verifying Condition (i) of Proposition 9. Third, $F_L(v) = v^2$ is the Beta$(2,1)$ CDF, so Condition (ii) of Proposition 9 is satisfied (since it can be shown that all beta distributions satisfy this condition).

It also follows that

$$w_L = \beta_L \frac{(1-s_1)^2(2+s_1)}{3} + (1-\beta_L)\frac{2}{3} = \frac{2}{3} - \beta_L \frac{s_1(3-s_1^2)}{3}.$$

Next, invert $s_1 = \sqrt{\tau_1 + \frac{1}{4}} - \frac{1}{2}$ to obtain $\tau_1 = s_1(1+s_1)$, hence $\beta_L = \frac{\alpha}{1-\alpha}\frac{1}{s_1(1+s_1)}$. Plugging this in yields

$$w_L = \frac{2}{3} - \frac{\alpha}{1-\alpha}\frac{3-s_1^2}{3(1+s_1)}.$$

Next, plugging in $s_1 = 1 - w_I$ yields

$$w_L = \frac{2}{3} - \frac{\alpha}{1-\alpha}\frac{2+2w_I - w_I^2}{3(2-w_I)}.$$

In summary, if $\alpha < \frac{2}{3}$, then

$$\begin{aligned}\mathcal{CS} = &\left\{(w_I, w_L) \,\middle|\, w_I = 0 \text{ and } w_L \in \left[\frac{2-3\alpha}{3-3\alpha}, \frac{2}{3}\right]\right\} \\ &\cup \left\{(w_I, w_L) \,\middle|\, w_I \in \left[0, \frac{3}{2} - \sqrt{\frac{\alpha}{1-\alpha} + \frac{1}{4}}\right] \text{ and } w_L = \frac{2}{3} - \frac{\alpha}{1-\alpha}\frac{2+2w_I - w_I^2}{3(2-w_I)}\right\}.\end{aligned}$$

Another feasible outcome. Next, consider the segmentation $\hat{\boldsymbol{\delta}}_I = (\frac{3}{4}, \frac{1}{4})$ and $\hat{\boldsymbol{\delta}}_L = (\frac{2}{5}, \frac{3}{5})$. This gives rise to $\hat{s}_1 = \frac{\sqrt{34}}{4} - \frac{1}{2} \approx 0.9577$ and $\hat{s}_2 = \frac{\sqrt{6}}{3} - \frac{1}{2} \approx 0.3165$. Hence,

$$\begin{aligned}\hat{w}_I &= \frac{3}{4}(1-s_1) + \frac{1}{4}(1-s_2) \approx 0.2026 \\ \hat{w}_L &= \frac{2}{5}\frac{(1-s_1)^2(2+s_1)}{3} + \frac{3}{5}\frac{(1-s_2)^2(2+s_2)}{3} \approx 0.2171\end{aligned}$$

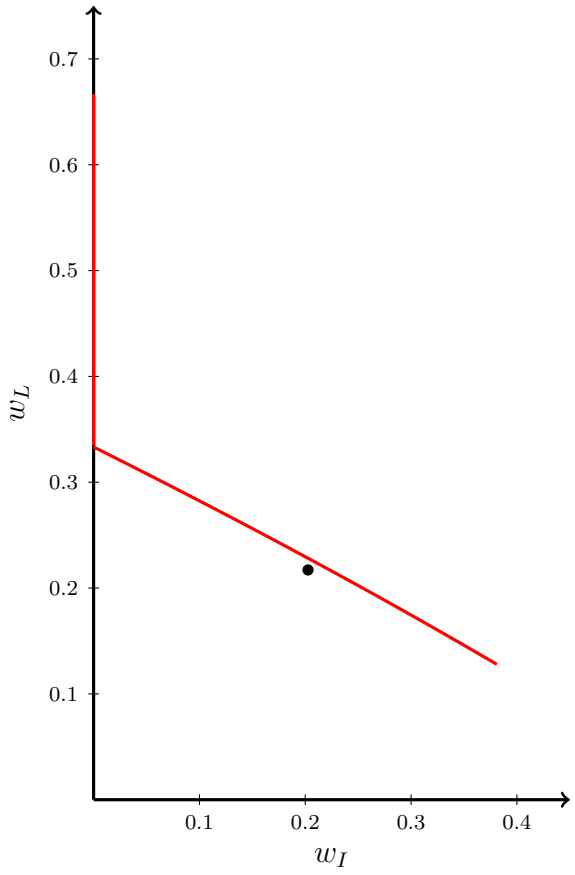
This outcome is Pareto-dominated by a cream-skimming outcome. Indeed, with $\alpha = 0.5$, there is a different outcome, $(w_I, w_L) \approx (0.2026, 0.2282)$, which lies in $\mathcal{CS}$.

It follows that, with this parametrization, it is not true that $\mathcal{F}^{\textit{anti-pareto}} = \mathcal{CS}$, and it is not true that $\mathcal{F}$ is the area between $\mathcal{CS}$ and $\mathcal{RCS}$.

D.2 Additional conditions under which the secondary results also extend

[TODO: This appendix is in progress.]

Figure 6: Cream-skimming outcomes ($\alpha = 0.5$)



The red curve depicts the welfare pairs implementable by the cream-skimming for $F_I(v) = \mathbb{1}\{v \geq 1\}$, $F_L(v) = v^2$, and $\alpha = 0.5$. The black dot depicts the feasible outcome $(\hat{w}_I, \hat{w}_L) \approx (0.2026, 0.2171)$, which is Pareto-dominated by a cream-skimming outcome.