Imperfect Competition and Moral Hazard in Financial Markets

Milena Wittwer*

December 1, 2025

Abstract

Moral hazard is traditionally analyzed as a bilateral principal–agent problem. In financial markets, however, intermediaries (agents) investing on behalf of clients (principals) must also compete with one another when trading, creating an interaction between incentives and market clearing. I develop a model to study how frictions from moral hazard and imperfect competition interact. The model shows that greater competition can reduce welfare by tightening incentive constraints, contradicting the standard prediction that more traders improve efficiency. Using trade-data from Canadian exchanges, I show how the model helps assess the relative importance of moral hazard and its interaction with competition.

1 Introduction

Most theoretical work on moral hazard builds on the principal–agent tradition, following Ross (1973), Mirrlees (1976), Holmström (1979), and Holstrom and Milgrom (1987), and analyzes incentives within a bilateral relationship. In finance, the intermediary asset-pricing literature, following He and Krishnamurthy, 2012, 2013, adopts this perspective, modeling clients who delegate investment decisions to an intermediary (or a continuum of intermediaries). Yet intermediaries do not act in isolation: they compete strategically in asset markets and internalize price impact (Kyle (1989); Vayanos (1999); Vives (2011)). Although each friction—moral hazard and imperfect competition—is well understood in isolation, we lack a framework for studying their interaction.

*Correspondence to Milena Wittwer, Columbia University, and NBER: mw3941@columbia.edu. I thank Connor Breck, Carel Chok, Costanza Didonna, Matthew Hagerty, Harrison Lynch, Brendan McLaughlin, Daniel Smith, and Joseph Wagner for excellent research assistance. I also thank Jason Allen and Adi Sunderam. Any errors are my own.

In this paper, I develop a model that unifies the two frictions and compares their welfare implications, motivated by suggestive evidence from Canadian exchange data indicating that both may arise in practice. The model shows that, because moral hazard and imperfect competition interact, stronger competition can tighten incentive constraints and ultimately reduce welfare. Interpreting the empirical patterns through the lens of the model suggests that this interaction could play a role in practice and should therefore be considered when designing policies.

To motivate the two frictions I focus on—moral hazard and imperfect competition—I use proprietary trade-level data from the Toronto Stock Exchange Group (TMX), which accounts for roughly 60% of Canadian equity trading. A key advantage of these data is that they contain firm-level dealer identifiers and account flags, allowing me to observe which dealer is trading, and whether the trade is executed for the dealer's own inventory or on behalf of a client—granularity that is rarely available in equity markets. Exploiting cases where the same dealer trades the identical security on the same day across exchanges with different levels of competition, I show that execution prices are systematically worse in the less competitive exchange. Moreover, comparing trades executed for clients with those executed for the dealer's own account, I find that trades executed on behalf of clients systematically receive worse prices, pointing to the presence of moral-hazard frictions.

Guided by these empirical patterns, I develop a model that incorporates both frictions in a unified framework. A continuum of clients seeks exposure to a risky asset with normally distributed returns but lacks direct market access. Clients are partitioned into N equal groups, and each group hires an intermediary—hereafter, a trader—compensated through a contract linear in realized wealth. Both clients and traders have CARA preferences, possibly with different risk tolerances. Given the contract, each trader chooses whether to accept and whether to exert effort or shirk when learning the client-specific expected return: effort yields a precise signal, while shirking produces a noisier one and confers a private benefit at the client's expense. Traders then compete in a centralized market (a uniform-price auction) by submitting (net) demand schedules; the market clears where aggregate demand meets supply, and all winning units transact at a single price.

I show that there exists an equilibrium in which clients choose the same contract, no trader shirks, and all submit identical demand functions conditional on observing the same signal, and I characterize this equilibrium. Two cases arise. If the traders' incentive constraint binds, the equilibrium contract depends on the number of traders and thus on

the degree of competition. If the incentive constraint is slack, the contract coincides with the first-best contract in a perfectly competitive, price-taking market.

To build an intuition for how the frictions affect market outcomes and welfare, I first examine each friction in isolation. Under imperfect competition, traders internalize how their orders move prices (Kyle, 1989) and strategically scale back demand to depress the clearing price and improve their terms of trade (Ausubel et al., 2014). This pushes prices below the first best. Moral hazard, in turn, affects contract choice: stronger private benefits from shirking require granting the trader a larger share of wealth to induce effort (Holmström and Tirole, 1997; He and Krishnamurthy, 2012). In my setting, this feedbacks into trading. A higher contractual share effectively raises the trader's risk aversion and increases the risky positions they are willing to take for a given expected return. This also pushes prices downward.

I then examine how the two frictions interact by asking whether stronger competition—measured by the number of competing traders—mitigates or exacerbates moral hazard, as reflected in the contract share a trader must receive to exert effort. Three effects emerge. First, as the number of traders increases, individual price impact declines, reducing the gains from strategic behavior. Unless the incentive constraint is slack, the contract share must rise to keep traders willing to exert effort. This price-impact channel is the manifestation of the broader idea that stronger competition compresses market-power rents, thereby requiring higher compensation to sustain effort.

Second, having more traders makes the clearing price more predictable. This affects the contract choice, because contracts are chosen in expectation, before prices realize. To highlight this, I temporarily shut down price impact and derive an equilibrium in which traders behave as naïve price takers. Even in this simplified environment, the optimal contract varies with the number of traders.

Third, traders must receive a larger contract share when shirking becomes more attractive, and the attractiveness of shirking depends on the information-risk ratio, which captures how informative a signal is relative to the risk borne when acting on it. To isolate this effect from the other two, I let the number of traders go to infinity, which renders prices perfectly predictable and eliminates price impact. In this limit, the trader's incentives to shirk depend on the precision of the signal obtained when shirking. With a highly precise signal, a shirker takes large positions and is exposed to substantial idiosyncratic risk, lowering the information-risk ratio and making shirking less appealing. With a noisy signal, the shirker scales back positions, reducing risk exposure enough to raise the information-risk ratio and thereby increasing the attractiveness of shirking.

Together, the price-impact, price-information, and information–risk-ratio effects give rise to two contract regimes. With high shirking noise, the contract (weakly) increases in the number of traders—an unfavorable case where competition amplifies moral hazard. With low noise, the contract instead (weakly) falls—the favorable case where competition mitigates moral hazard.

Then I analyze how imperfect competition and moral hazard affect welfare, measured by the expected allocative efficiency. Lowering the private benefit of shirking, and thus aligning incentives and mitigating moral hazard, always raises welfare. By contrast, increasing the number of traders may either raise or reduce welfare, depending on how the contract is adjusted in responds to the changing number of traders.

This welfare result is counterintuitive: in standard Kyle-type models, welfare typically increases as the number of traders grows. The key difference arises from an indirect effect which is absent in models without moral hazard. Changes in the number of traders modify the equilibrium contract. When the contract is unaffected by trader entry (which is the case when the incentive constraint is slack), welfare always rises because risk-sharing improves. But when the contract responds to the number of traders, an additional welfare channel opens.

Whether welfare rises or declines depends on the contract regime. When the contract declines with more traders (the favorable regime), this indirect effect reinforces the direct gains from risk-sharing. When instead the contract rises with more traders (the unfavorable regime), the negative indirect effect can outweigh the direct positive effect, so overall welfare falls. This happens when additional traders require a higher contract share to maintain incentives, raising the risk-bearing cost enough to outweigh the welfare gains from improved risk-sharing and information aggregation.

In the final part of the paper, I use the model as a rough, first-order approximation to interpret trade-level behavior on Canadian exchanges.[1] For this I introduce a few targeted adjustments that preserve tractability while allowing the model to speak directly to the empirical patterns that motivate it. I approximate continuous trading with frequent batch auctions (Budish et al. (2015)), treating the buy-, and sell side, client-account, and inventory-account trades on each exchange separately to capture the systematic price differences observed across trade types and exchanges.

I focus on learning about three parameters that shape the relative importance of the

[1] I do not conduct a full structural estimation of all model primitives. A fully structural exercise would require information that trade-level data cannot provide—for example, the costs and benefits of shirking, which leave no trace in equilibrium.

two frictions: traders' risk tolerance, the contract share intermediaries retain, and clients' risk tolerance (for which I can identify a lower bound). The moments used for identification mirror the earlier suggestive evidence. Price differences for inventory-account trades across exchanges with different levels of competition identify traders' risk tolerance, while price differences between client- and inventory-account trades identify the contract share (the portion of surplus accruing to the intermediary). Given these two estimates, the model's incentive-compatibility condition yields a lower bound on clients' risk tolerance consistent with the observed contract.

The resulting estimate of traders' risk tolerance is noisy but broadly in line with existing evidence (Wittwer and Allen (2023)). The implied lower bound on client risk tolerance is also relatively high, consistent with the fact that many active clients in my setting are large institutional investors rather than small retail traders. The estimated contract share is stable across all specifications I consider—between 0.6 and 0.8—implying that intermediaries retain well over half of the wealth generated on behalf of their clients.

A central implication of the model is that contract terms should vary with the number of traders, because moral hazard and imperfect competition interact. To evaluate this prediction, I estimate the contract separately across different levels of trader competition and find that it becomes more favorable to intermediaries as the number of active traders increases. This empirical pattern is consistent with the unfavorable contract regime highlighted by the theory, in which greater competition is associated with tighter moral-hazard frictions.

To quantify the associated welfare loss, I evaluate welfare in a representative client-account batch auction. Because the effects depend on unobserved primitives—such as client risk tolerance and the information–risk ratio—the absolute magnitudes are less informative than their relative size. What stands out is that adding one extra trader has a larger relative impact on welfare than eliminating moral hazard when clients are less risk-tolerant than dealers. This impact is negative: as competition intensifies, the contract tilts further toward intermediaries, and the resulting indirect effect outweighs the direct pro-competitive gain. This pattern indicates that analyses ignoring the interaction between moral hazard and trade competition can substantially understate the true costs of moral hazard.

Taken together, my findings highlight that the two frictions are deeply intertwined: the efficiency of asset-market outcomes depends jointly on the severity of incentive problems and on the number of traders competing for assets. Policies that promote com-

petition in asset markets can therefore have unintended consequences when incentive problems are severe, as additional entry may tighten incentive constraints and lower welfare. Effective regulation thus requires a coordinated approach.

Theoretical literature. My main theoretical contribution is to show how moral hazard and imperfect competition interact in a trading environment and jointly shape welfare. In doing so, I connect to several strands of literature.

At a high level, my paper relates to the principal–agent literature—following Ross (1973), Mirrlees (1976), Holmström (1979) and Holstrom and Milgrom (1987). In my setting, the clients are the principals and the traders the agents. In addition, because traders decide whether to acquire information, my model also connects to the costly information acquisition literature—dating back to Stigler (1961), and Grossman and Stiglitz (1980). Lastly, the paper relates to the financial contracting literature, which applies these ideas to settings such as delegated portfolio management (e.g., Bhattacharya and Pfleiderer (1985); Starks (1987); Carpenter (2000); Buffa et al. (2022)) and delegated asset management with moral hazard (e.g., DeMarzo and Fishman (2007); Di Tella and Sannikov (2021)).

A common feature of these three strands of literature is their micro-level focus. Within the principal–agent framework, this typically means analyzing incentives and contract design between a single principal and agent while abstracting from interactions among multiple agents or principals. Although some papers, such as Garcia and Vanden (2009), Buffa et al. (2022), and Kashyap et al. (2023), develop general-equilibrium models, they assume frictionless—and thus perfectly competitive—asset markets. My contribution is move away from that assumption. I explicitly incorporates strategic interactions among traders in the spirit of Wilson (1979); Kyle (1989), which allows me to analyze how imperfect competition and moral hazard interact.

Despite extensive work on moral hazard in lending and portfolio management, its role in shaping trading behavior and price formation remains largely unexplored.[2] Two exceptions are Kyle et al. (2011) and Baldauf et al. (2024). The former study a setting with one agent who exerts costly effort to acquire information before trading on behalf of an investor. This agent trades against competitive market makers, so competition among informed traders is absent. The latter analyze block trading as a contracting problem under moral hazard, but take price impact as exogenous. In contrast, my model embeds moral

[2]More remotely related are papers that study trading and price formation under different frictions but share modeling features with elements of my framework. Examples include Glode and Opp (2020), Chen et al. (2025), and Wittwer and Allen (2023).

hazard in a setting with multiple strategic traders who interact through endogenous market clearing. This framework is key to answering my research question—how imperfect trade competition and moral-hazard frictions interact to jointly shape contracts, price impact, and welfare.

By analyzing the interaction between imperfect competition and moral hazard, I contribute to a broader literature spanning banking and corporate finance, where this interaction has long been central. In banking, the traditional view holds that greater competition—among institutions providing loans, deposits, or credit—compresses margins and induces risk-taking or attracts riskier borrowers, undermining stability (e.g., Keeley (1990); Crawford et al. (2018)). Building on Stiglitz and Weiss (1981), Boyd and Nicolo (2005) and Martinez-Miera and Rapullo (2010) show that the link between competition and risk-taking is not necessarily monotonic, as competition can both discipline and destabilize banks depending on how it affects lending incentives. A similar insight emerges in corporate finance, where product-market competition shapes managerial effort and incentive design. Schmidt (1997), and related work such as Scharfstein (1988) and Raith (2003), show that competition can either mitigate or exacerbate moral hazard depending on how it influences rents and performance sensitivity.

The central welfare result—that competition can either mitigate or exacerbate moral hazard in a non-monotonic way—echoes findings from the banking and corporate finance literatures, but through distinct mechanisms. In these literatures, competition is typically modeled through Bertrand or Cournot interactions over prices or quantities. In contrast, traders in my model compete in demand (or supply) functions, a formulation better suited to trading environments with endogenous price formation. More importantly, welfare losses in my framework arise not from excessive risk-taking or weakened monitoring incentives, but from how moral hazard and imperfect competition jointly distort traders' information–risk trade-off. This makes the asset price less predictable, amplifies price impact, and leads to inefficient asset allocations.

Empirical literature. To my knowledge, neither the magnitude of moral-hazard frictions in financial trading nor the relative welfare consequences of moral hazard versus imperfect competition have been empirically quantified.

Quantifying moral hazard is difficult because the relevant incentives and contracts are rarely observed directly and, when observable, reflect equilibrium outcomes (see Einav et al., 2021). My idea is to use trade-level data to infer what fraction of wealth intermediaries must obtain to be willing to exert effort. Because contracts are private while

trading activity is observable to regulators, this approach provides a rough first pass—one that can be implemented with readily available data—to obtain better information for a subsequent, fully structural analysis.

The magnitude of the implied contract estimate is broadly consistent with evidence from asset management, where comparisons of fund performance to fee structures show that mutual fund managers capture most of the value they generate (Berk and van Binsbergen, 2015; Fama and French, 2010; French, 2008), hedge fund compensation schemes imply similarly large intermediary shares (Agarwal et al., 2009; Lan et al., 2013), and private-equity general partners also retain a substantial fraction of total surplus (Metrick and Yasuda, 2010).

2 Motivating empirical evidence

To motivate my focus on imperfect competition and moral hazard in a trading environment, I draw on evidence from Canadian stock exchanges. The data are unusually rich: they provide firm-level dealer identifiers across exchanges and indicate whether each trade is executed for a dealer's own account or on behalf of a client—granularity that is rarely available and useful for learning about both frictions.

Institutional environment and data. The Canadian equity markets provide an excellent setting for studying how moral hazard and imperfect competition may affect asset prices and allocations. Unlike in the United States, Canada does not enforce a strict National Best Bid and Offer (NBBO) regime, allowing execution prices to vary legitimately across venues according to local trading conditions. Moreover, off-exchange internalization is far less prevalent, so the vast majority of orders is executed on exchanges.

Canada has nine equity exchanges, which share the standard microstructure features of modern stock markets. Most trading occurs continuously throughout the day on centralized electronic limit order books, creating a relatively competitive trading environment (at least compared with illiquid over-the-counter markets). I focus on the three exchanges operated by TMX—TSX (the primary and most liquid market), TSXV, and Alpha—which together account for roughly 60% of trading activity.

As on most exchanges, only member firms—typically banks, large broker-dealers, and asset managers—are permitted to submit orders directly. All other investors, including large institutional investors, must trade through these members, so each firm executes both for its own account and on behalf of clients. Despite this not all institutional investors choose to become an exchange member, because membership is relatively costly:

firms must be affiliated with the Canadian Investment Regulatory Organization (CIRO), have access to the clearing system, and pay a C$65,000 entry fee plus monthly charges.

In handling client orders, member firms retain considerable discretion, creating scope for moral hazard. Although their execution decisions fall under CIRO'S principles-based best-execution requirements and the supervisory obligations of the Universal Market Integrity Rules (UMIR), these rules rely on firms to design, implement, and monitor their own policies and procedures.

I have trade-level data from all exchanges owned by the TMX Group between 2019 and 2022. For each trade, I observe the time of execution, the traded security—the 'symbol'—, trade size, price. I also see the firm's legal entity identifier, which I will, for simplicity, often refer to as the 'trader,' recognizing that multiple individuals may trade under the same identifier. In addition, I know whether the trade was executed for the exchange member's own account or on behalf of a client. Moreover, I observe whether the trade is active, meaning that it is added to the limit book and consumes liquidity, or passive. I restrict attention to active trades throughout the paper. From 2020 onward, I also see the national best bid and offer for each symbol immediately prior to trade execution.

Evidence of imperfect competition. To show that imperfect competition matters for asset prices, I exploit the fact that the same stocks are traded on two Canadian exchanges—TSX and Alpha—but market concentration is higher on Alpha than TSX, as shown in Figure 1.[3] Both exchanges have been well-established for over a decade and are owned by the same parent company (TMX). They operate under similar market rules and regulatory requirements, and list the same set of stocks. This common structure allows me to rule out many (yet not all) confounding factors, unrelated to competition, that might otherwise explain price differences.

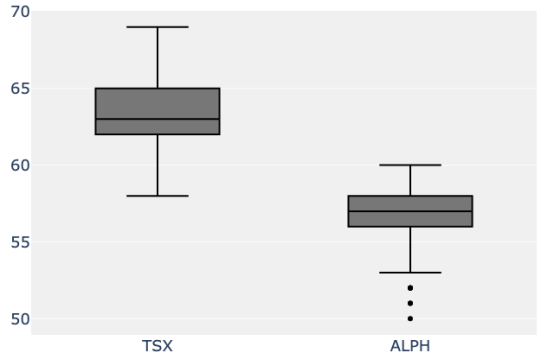
To reduce the sample size, I collapse the raw data to the level of trading venue (TSX/Alpha), day, symbol, trader, trade direction (buy/sell). I refer to each row in the collapsed dataset as τ, and compute the total quantity traded, quantity_τ, and the average price, price_τ, for each τ. In Appendix Table A1 I verify that the results are qualitatively robust when I treat each trade individually, using a subset of uncollapsed data.[4]

[3]One potential reason for fewer traders operating on Alpha compared to TSX is that the marginal benefits of joining Alpha—the smaller of the two exchanges—may be limited, conditional on already being a member of TSX.

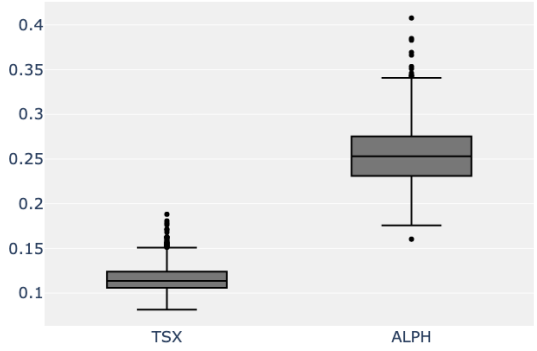
[4]The results also hold when distinguishing between account types (e.g., client, inventory, market-maker) and when replacing trader fixed effects with trader-by-account-type fixed effects.

Figure 1: Market Concentration is Higher on Alpha

(a) Number of active traders



(b) Daily HHI



Notes: Figure 1 visualizes that market concentration on Alpha is higher than on TSX. I choose two examples of variables that are related to market concentration, with all other measures pointing in the same direction. Figure 1a shows the distribution of the number of active traders on a day for the TSX and Alpha exchange. Figure 1b shows the distribution of daily HHI indices (which is between 0 and 1) on TSX and Alpha. The $HHI_t = \sum_i (s_{ti})^2$ is computed using the fraction of total daily trade value executed by a dealer i on day t as $s_{ti} \in [0, 1]$.

I regress buy- and sell-side trade prices separately on an Alpha-exchange indicator, trade quantity, plus day–symbol–trader effects ζ_{tsj}:

$$\text{price}_\tau = \alpha + \beta \text{alpha}_\tau + \gamma \text{quantity}_\tau + \zeta_{tsj} + \epsilon_\tau. \tag{1}$$

Standard errors are clustered by day to capture within-day correlations across dealers, securities, and trades.[5]

My findings, reported in Table 1, show that that same dealer trades the same symbol on the same day at worse prices on Alpha—where market concentration is greater—than on TSX: ca. C$0.005 worse, with corresponds to roughly 2 bps of the median stock price (C$22.9). The regression results further show that trading prices are higher when trading larger amounts, suggesting that both prices and quantities matter. This is in line with early empirical studies on the NYSE and regional U.S. exchanges (listed in Bernhardt et al. (2005)), according to which larger trades are executed at worse prices on the exchanges. It suggests that traders have price impact—a key feature of the model.

I view this evidence as suggestive that imperfect competition affects prices, while ac-

[5]Alternative clustering by dealer or symbol yields unreliable inference with few dealers and uneven cluster sizes (MacKinnon et al. (2023)).

Table 1: Prices are worse on the more concentrated exchange (Alpha)

	(Buy)	(Sell)
alpha	0.006***	-0.004***
	(0.001)	(0.001)
quantity	0.016***	-0.012***
	(0.002)	(0.002)
N	10,257,480	9,784,486
R^2	1.000	1.000
Within-R^2	0.000	0.000

Notes: Table 1 shows that prices are worse on the more concentrated exchange (Alpha) when using collapsed data. Appendix Table A1 reports qualitatively similar results using a subsample of the uncollapsed trade-level data. In column (Buy), I regress prices of active buy-side trades on an Alpha indicator (vs. TSX baseline), traded quantity, and date-symbol-trader fixed effects; column (Sell) uses sell-side trades. In all regressions I include date-symbol-trader fixed effects. Prices are in C$, and quantities in million C$. Standard errors are clustered at the daily level.

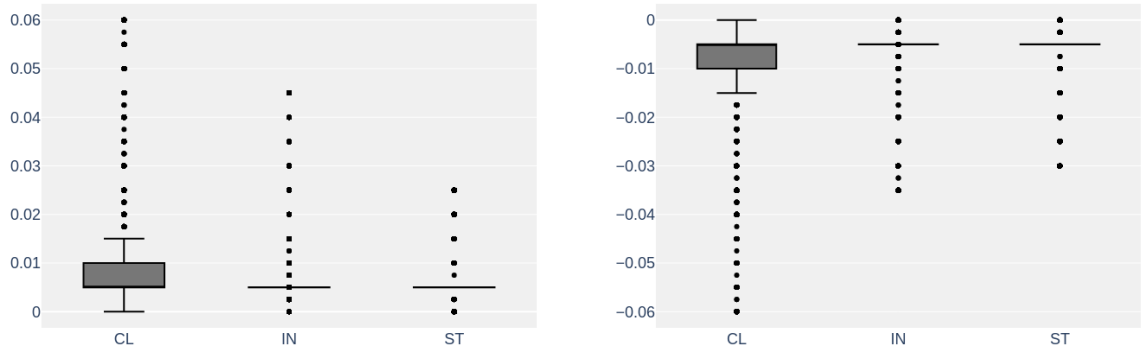
knowledging that fixed-effects cannot rule out all confounding factors. One concern is that dealers may selectively route harder-to-execute orders—which trade at worse prices—to Alpha. For instance, TSX's greater liquidity may make it the preferred venue for simple orders, with more complex or information-sensitive orders pushed to Alpha.

Evidence of moral hazard. Identifying moral hazard in financial settings is difficult because incentives and contractual arrangements are rarely observed. I offer suggestive evidence based on the idea that, when moral-hazard frictions exist between traders and their clients, client trades should execute at systematically worse prices than trades for a trader's own account. In practice, this pattern can arise for several reasons that I interpret as manifestations of moral hazard. For example, traders may simply pass through client orders without optimizing execution, or clients may trade at suboptimal times and fail to time the market effectively. In both cases, the trader—acting as the client's dealer—does not exert sufficient effort to improve the client's trading performance, either by timing trades more carefully or by helping the client adopt better trading practices.

Figure 2 shows that client (CL) trades consistently buy at higher prices and sell at lower prices than inventory (IN) trades, using TSX/Alpha/TSXV data for all active common-stock and ETF trades in July 2021. To quantify these pricing differences more systematically, I regress prices on whether the trade is on behalf of a client or for the trader's own account. In my main specification (Table 2), I collapse active trades by exchange, day, symbol, trader, trade direction, and account type, and regress ask and bid

Figure 2: Trade price over national mid-price

(a) Active side buys for the account (July 2021) (b) Active side sells for the account (July 2021)



Notes: These boxplots show the distribution of the difference between the trade-price and the mid-price before the trade executes, excluding upper and lower 2.5% quantiles, using data for common stocks and ETFs on TMX from July 2021 for CL, IN, ST accounts. The mid-price is the average of the national best bid and ask before the trade.

prices on account-type indicator, traded quantity, and exchange-day-symbol fixed effects:

$$\text{price}_\tau = \alpha + \sum_{k \in \{CL, IN\}} \beta_k \mathbb{I}(\text{account-type} = k) + \gamma \text{quantity}_\tau + \zeta_{ets} + \epsilon_\tau. \tag{2}$$

As robustness, I rely on the uncollapsed trade-level data, using a shorter sample period, but richer fixed effects (Appendix Table A2). Across all specifications, the inventory account serves as the benchmark. While magnitudes vary, the qualitative pattern is consistent: client-account trades systematically receive worse prices than trades executed by dealers for their own accounts. Because the regressions include day–symbol–exchange fixed effects, these differences cannot be explained by clients selecting different securities or trading on different exchanges.

In practice, several factors related to moral hazard may lead to worse execution for client accounts. My model captures this idea in a stylized way: prices are worse when the trader invests on behalf of the client because the trader captures part of the asset's surplus as compensation. This pushes up (down) the market price when investors are buying (selling).

Take away. The empirical patterns motivate a model that jointly captures imperfect competition in trading and moral-hazard frictions. The goal is to understand how these

Table 2: Clients trade at worse prices than dealers for their own account

	(Buy)	(Sell)
account_type: CL	0.005***	-0.010***
	(0.001)	(0.001)
quantity	0.009***	-0.006***
	(0.001)	(0.001)
N	23,761,003	22,780,098
R^2	1.000	1.000
Within-R^2	0.000	0.000

Notes: Table 2a regresses (average) prices from active buy-side trades the account-type indicator and traded quantity plus symbol-day-exchange fixed effects in column (Buy) using the collapsed data of all trades. Column (Sell) shows the analogous estimation results for the sell-side. The inventory account is the baseline. Standard errors are clustered at the daily-level.

frictions interact, how they shape welfare, and under what conditions one dominates the other.

To retain tractability, the model abstracts from several real-world complexities. For instance, traders are ex ante identical, so I omit heterogeneity across dealers such as firm size or type (e.g., banks versus brokers). Trading takes place in a centralized market that clears at a single price (following Kyle (1989) and related references). This setup abstracts from frictions arising from market fragmentation (e.g., between TSX and Alpha) and does not incorporate bid–ask spreads. I also abstract from differences across account types: while the model distinguishes between trading for one's own account and for a client, it does not model market clearing when traders operate across multiple accounts.

3 Model with imperfect competition and moral hazard

In my model, traders compete to buy assets on behalf of clients, consistent with the intermediary asset-pricing literature that focuses on asset investment. The market is therefore one-sided. However, modifying the model to allow for selling or trading is straightforward, and common in the finance literature[6] I present the model and characterize the equilibrium, and then discuss key tractability assumptions. Afterwards I analyze the interaction between moral hazard and imperfect competition. Proofs omitted from the main text are provided in Appendix A.3.

[6]See Rostek and Yoon (2025) for a comprehensive overview of two-sided uniform-price auctions, including Kyle-style models.

Setting. There are $N > 2$ clients who seek to invest in a risky asset of fixed supply $A > 0$. Clients have no direct access to the asset market. Therefore, each client hires a trader in period 1 of the game, which unfolds over 4 periods. For their investment service (described below), trader i receives a fixed amount K_i and a fraction, $\phi_i \geq 0$, of wealth, ω_i, that is generated from the investment at the end of the game.[7]

Wealth results from purchasing amount a_i at price p: $\omega_i = a_i(R_i - p)$. $R_i \in \mathbb{R}$ is the per-unit gross return that the asset pays to client i at the end of the game. Returns are normally distributed $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, with known covariance matrix $\Sigma > 0$, conditional on mean ν_i.

The expected return, ν_i, is client-specific, reflecting factors such as existing balance–sheet exposures that lower the appetite for additional holdings, different macro or market forecasts, hedging needs, or variations in funding costs and regulatory capital requirements that influence the valuation of the asset. The means are drawn independently from a common prior, $\nu_i \overset{iid}{\sim} \mathcal{N}(\nu, \Sigma_\nu)$, with $\nu \in \mathbb{R}$ and $\Sigma_\nu > 0$, implying that $R_i \sim \mathcal{N}(\nu_i, \Sigma + \Sigma_\nu)$.[8]

Traders must form beliefs about their expected return, ν_i, in period 2 of the game. They can choose to exert effort ($s_i = 0$), in which case they observe $\nu_i \in \mathbb{R}$, and therefore know the conditional return distribution. Alternatively, they can shirk ($s_i = 1$) which implies that they only observe the return distribution with noise. Summarizing, they see the following private return signal:

$$\mu_i(s_i) = \begin{cases} \nu_i & \text{when trader exerts effort, } s_i = 0 \\ \nu_i + \epsilon_i \text{ with } \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon) & \text{when trader shirks, } s_i = 1, \end{cases}$$

where $\sigma_\epsilon > 0$. Note that unlike in a standard principal-agent setting (dating back to Ross (1973), and Holstrom and Milgrom (1987)) exerting effort does not create larger output, but rather changes the information available to the trader in the subsequent trading stage. In this regard, my model is more similar to the literature on costly information acquisition (dating back to Stigler (1961), and Grossman and Stiglitz (1980)).

Shirking generates a private benefit to the trader, $b \geq 0$, and a loss for the client

[7] For tractability, I do not restrict $\phi_i \leq 1$. However, model parameters could, if desired, be constrained so that $0 \leq \phi_i \leq 1$. Thus their total compensation is: $K_i + \phi_i \omega_i$. When the trader declines the job offer, they receive a reservation utility of $\overline{U}$.

[8] Independence reflects the idea that each trader's private information stems from idiosyncratic forces—such as observing information from their specific client base, or their balance sheet—so that any common component is captured by the prior mean ν. While this assumption simplifies the analysis, it could be relaxed to allow correlated signals.

$\Delta \geq 0$. In practice, the benefit from shirking may reflect the value of leisure from working less, while the client's loss captures broader negative spillovers from traders neglecting other, non-modeled responsibilities. For example, incomplete follow-through on reporting, communication, or compliance duties can impose administrative or regulatory costs on the client. I assume that Δ is sufficiently high that it is never optimal for the client to allow their trader to shirk—an assumption I discuss in Section 3.

Given their private signal, traders compete against one another to purchase shares of the asset in period 3 of the game. The asset market clears through a uniform price auction—a common approach to modeling market clearing in trading environments (e.g., Wilson (1979); Klemperer and Meyer (1989); Kyle (1989); Vayanos (1999); Vives (2011); Rostek and Yoon (2021)). Each trader submits a decreasing and twice continuous demand function

$$a_i(\cdot, \mu_i(s_i)) : \mathbb{R}_+ \to \mathbb{R}$$

that maps prices to quantities, conditional on their private information and their shirking decision.[9] The submitted demand functions are aggregated, and the market clears at prices where aggregate demand equals supply.[10] From the perspective of trader i who takes choices of all other traders as given:

$$p^c(s_i) : \sum_i a_i(p^c(s_i), \mu_i(s_i)) = A. \tag{3}$$

There are two possible clearing prices, depending on whether the trader exerts effort or shirks. To highlight this, I label the clearing price as $p^c(s_i)$. Note that this notation omits the fact that the clearing price depends on all contracts and effort choices by all traders not just trader i.

Clients and traders obtain CARA utility over the wealth they hold at the end of the game. Clients have risk-tolerance $\rho_H > 0$, and traders $\rho_I > 0$. Specifically, a trader who chooses their demand function expects to obtain the following utility when choosing

[9] If we want to model two-sided trade-settings, it suffices to define the net-demand (demand minus supply).

[10] In case of excess demand—which never happens in equilibrium—traders are rationed pro-rata on the margin (see Kastl (2012) for details).

demand $a^c(s_i) = a_i(p, \mu_i(s_i))$ at clearing price $p = p^c(s_i)$:

$$\mathbb{E}\left[1 - \exp\left(-\frac{1}{\rho_I}\omega_I^c(s_i;\phi_i)\right)\middle|\mu_i(s_i)\right], \tag{4}$$

$$\text{with } \omega_I^c(s_i;\phi_i) = \phi_i\left\{a^c(s_i)(R_i - p^c(s_i))\right\} + s_i b + K_i. \tag{5}$$

Given that all random variables are normally distributed, asset returns conditional on return signals are also normally distributed:

$$R_i|\mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i)], Var[R_i \mid \mu_i(s_i)]). \tag{6}$$

Under normal return distributions (6), maximizing (4) is equivalent to maximizing the certainty equivalent $\mathcal{U}(s_i;\phi_i)$ point-wise at every clearing price $p^c(s_i)$:

$$\mathcal{U}(s_i;\phi_i) + K_i\ , \tag{7}$$

$$\text{where } \mathcal{U}(s_i;\phi_i) = \phi_i\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)a_i^c(s_i) - \phi_i^2\frac{1}{2\rho_I}a_i^c(s_i)Var[R_i \mid \mu_i(s_i)]a_i^c(s_i) + bs_i. \tag{8}$$

In period 2, when the trader chooses their effort-level, s_i, they do not know their private signal, $\mu_i(s_i)$. They take an expectation, integrating out the uncertainty over their signal—a feature discussed in Section 3. When

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] + K_i \geq \bar{U} \tag{IR}$$

the trader participates in the game. They choose to exert effort when

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] \geq \mathbb{E}[\mathcal{U}(1;\phi_i)], \tag{IC}$$

and shirk otherwise.

In period 1, each client chooses their contract, $\{\phi_i, K_i\}$. Each client maximizes their expected certainty equivalent, integrating out the uncertainty about the trader's signal, similarly to the trader in period 2.

The client's certainty equivalent is defined analogously to the trader's certainty equivalent; with four differences: (i) the client's risk tolerance is ρ_H rather than ρ_I, (ii) the client obtains fraction $1 - \phi_i$ rather than ϕ_i of wealth generated by the asset, (iii) they pay rather than earn K_i, (iv) and suffer a loss, Δ, when the trader shirks, rather than a

benefit, b. Formally, their certainty equivalent, conditional on $\mu_i(s_i)$, is:

$$\mathcal{W}(\phi_i; s_i) - K_i, \tag{9}$$

$$\begin{aligned}\text{where } \mathcal{W}(\phi_i; s_i) =& (1-\phi_i)\Big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\Big)a_i^c(s_i) \\ &- (1-\phi_i)^2 \frac{1}{2\rho_H} a_i^c(s_i) Var[R_i \mid \mu_i(s_i)] a_i^c(s_i) - \Delta s_i.\end{aligned} \tag{10}$$

Integrating out the uncertainty about the signal, the client chooses the contract $\{\phi_i, K_i\}$ to maximize:

$$\mathbb{E}[\mathcal{W}(\phi_i; 0] - K_i \text{ subject to (IR) and (IC).} \tag{11}$$

Summarizing, the sequence of events is as follows:

(1) N clients choose $\{K_i, \phi_i\}$ simultaneously, and independently. Each trader chooses to participate in the game or walk away, in which case the game ends.

(2) N traders decide simultaneously, and independently, whether to exert effort: $s_i \in \{0, 1\}$.

(3) Each trader observe their private expected return signal, depending on their chosen effort-level, $\mu_i(s_i)$, and compete to buy assets in a uniform-price auction.

(4) The assets pay their returns, and all transactions take place.

While I introduced the model in a buy-side setting, it is straightforward to extend it to selling or to a two-sided (double-auction) environment. One can simply allow the exogenous net supply to take any real value, $A \in \mathbb{R}$, and interpret traders' submitted schedules as net-demand curves $a_i : \mathbb{R} \to \mathbb{R}$, where positive quantities correspond to purchases and negative ones to sales.

Equilibrium. I focus on symmetric Bayesian Nash equilibria (BNE), in which traders and clients choose identical strategies when observing the same signal. The key equilibrium objects are the contracts K^*, ϕ^*, the trader's shirking decisions, $s_i \in \{0, 1\}$, and their demand functions, $a^*(\cdot, \mu_i(s_i))$ for $s_i \in \{0, 1\}$ for all i.

Proposition 1. *Denote* $\tilde{\Sigma} = Var(v_i|\mu_i(1))$.

In a symmetric equilibrium, all clients offer the following contract $\{\phi^*, K^*\}$*:*

$$\phi^* = \begin{cases} \Phi(N) = \frac{1}{|A|}\sqrt{\frac{\beta(N)+b\gamma(N)}{\alpha(N)}} & \text{when } \frac{\rho_I}{\rho_I+\rho_H} \leq \Phi(N) \\ \frac{\rho_I}{\rho_I+\rho_H} & \text{otherwise,} \end{cases} \tag{12}$$

and

$$K^* = \bar{U} + \frac{N-2}{2(N-1)}\frac{\rho_I}{\phi^*}\frac{\Sigma_\nu}{\Sigma} + \frac{A^2\Sigma}{2N(2N-1)}\frac{\phi^*}{\rho_I}, \tag{13}$$

where $\alpha(N) = (-1+N)\Sigma^2\left(\tilde{\Sigma}-\Sigma\right) > 0, \beta(N) = (-2+N)N\rho_I^2((-1+N)\Sigma\sigma_\epsilon + (-2+N)\Sigma\Sigma_\nu - (-2+N)\Sigma_\nu\tilde{\Sigma}) \in \mathbb{R}, \gamma(N) = 2(-1+N)N\rho_I\Sigma(2\Sigma+(-2+N)\tilde{\Sigma}) > 0$.

All traders exert effort, and submit the following demand function:

$$a^*(p, \mu_i(s_i)) = \rho_I(\Lambda^*\rho_I + \phi^* Var[R_i \mid \mu_i(s_i)])^{-1}(\mathbb{E}[R_i \mid \mu_i(s_i)] - p), \tag{14}$$

$$\text{with } \Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\Sigma. \tag{15}$$

This equilibrium is the unique symmetric outcome in which traders submit linear demand functions and exert effort. When the client's loss is sufficiently high, as specified in Appendix A.3, this equilibrium is the unique symmetric linear equilibrium—an equilibrium refinement that is standard in the related auction literature (Rostek and Yoon (2025)).

To derive an intuition for the equilibrium consider one client-trader pair, i, and fix the behavior of all clients and traders at the symmetric equilibrium. Consider first the contract choice. Since the trader's incentive constraint is independent of the fixed part of the contract, K_i, it is optimal to set K_i as high as possible, so as to still incentivize the trader to participate. This is the case when

$$K(\phi_i; 0) = \bar{U} - \mathbb{E}\left[\mathcal{U}(0; \phi_i)\right]. \tag{IR}$$

Note that $K(\phi_i; 0)$ depends on ϕ_i, because ϕ_i affects the trader's demand function in the trading game. Inserting the binding participation constraint (IR) into the client's objective function (16), the optimization problem simplifies to:

$$\max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i; 0] - K_i(\phi_i; 0)\right\} = \max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i; 0] + \mathbb{E}[\mathcal{U}(0; \phi_i] - \bar{U}\right\} \tag{16}$$

subject to (IC). There are two solutions: an unconstrained contract, $\phi_i = \frac{\rho_I}{\rho_I+\rho_H}$ that maximizes the objective function without constraints, and a constrained contract, $\phi_i = \Phi(N)$.

Under the unconstrained solution, the asset allocation would be first best-efficient if the market was perfectly competitive or if traders would be price-takers. With this contract, the trader perfectly internalizes that they invest for a client, and correctly chooses demand, depending on their differing degree of risk-tolerance, i.e., maximizing risk-sharing.

The constrained contract is pined down by the binding incentive constraint (IC): $\mathbb{E}[\mathcal{U}(0;\phi_i)] = \mathbb{E}[\mathcal{U}(1;\phi_i)]$. This gives a best response ϕ_i to all other clients choosing the equilibrium contract, ϕ. Hereby all clients correctly anticipate traders' reactions in periods 2 and 3 of the game, which determine the market price and allocations. In a symmetric equilibrium, each client i chooses the same contract as all others. Solving for ϕ then yields a unique positive equilibrium solution, denoted by $\Phi(N)$.

In both the constrained and unconstrained case, the trader exerts effort. They choose demand, $a_i = a_i(p, \mu_i(s_i))$, by point-wise maximizing their certainty equivalent (8) at each possible clearing price. The necessary condition for quantity a_i at the fixed market clearing price p and both $s_i \in \{0,1\}$ is

$$\mathbb{E}[R_i \mid \mu_i(s_i)] - \phi_i \frac{1}{\rho_I} Var[R_i \mid \mu_i(s_i)] a_i = p + \Lambda_i a_i, \text{ where } \Lambda_i = \left(\frac{dp}{da_i}\right). \tag{17}$$

Λ_i measures the price impact of the trader's own demand on the clearing price. It is the amount by which the market clearing price moves when trader i demands marginally more than quantity a_i. Since all other traders j but i submit the equilibrium guess (15), price impact is

$$\Lambda_i = \frac{\phi}{\rho_I (N-2)} Var[R_i \mid \mu_i(0)]. \tag{18}$$

Moreover, the necessary conditions are sufficient, since $Var[R_i \mid \mu_i(s_i)]$ is positive semi-definite, $N > 2$ and $\phi_i > 0$. Therefore, equation (17) pins down the trader's best reply $a_i(p, \mu_i(s_i))$ for $s_i \in \{0,1\}$.

Given the the contracts, the trader's shirking decisions, and their demand functions, the equilibrium price (19) and equilibrium allocations are then pinned down via market clearing.

Corollary 1. *In equilibrium, the market clears at price:*

$$p^* = \frac{1}{N} \sum_i \nu_i - \frac{\phi^*}{\rho_I} \Sigma \left(\frac{N-1}{N-2}\right) \frac{A}{N}, \tag{19}$$

and trader i purchases amount

$$a^*(p^*, v_i) = \frac{A}{N} + \frac{\rho_I}{\phi^*}\frac{N-2}{N-1}\left(v_i - \frac{1}{N}\sum_i v_i\right)\Sigma^{-1}, \tag{20}$$

where I used $\mu_i(0) = v_i$, $\mathbb{E}[R_i \mid \mu_i(0)] = v_i$ *and* $Var[R_i \mid \mu_i(0)] = \Sigma$.

Discussing tractability assumptions. The model relies on a set of assumptions that ensure tractability. Before analyzing the equilibrium and unpacking the underlying economic mechanisms, I discuss the key modeling choices that render the model tractable. The discussion focuses on features that are specific to this paper's contribution, rather than widely used assumptions in the related literatures (e.g., CARA utility, normally distributed returns, linear equilibria).

First, clients offer simple, linear compensation contracts, $\{K_i, \phi_i\}$, consisting of a fixed fee and a constant share of the future wealth generated by the asset.[11] At a high level, the contract resembles common institutional arrangements, combining a fixed base salary with a performance-based bonus tied to trading outcomes. The bonus ϕ_i can be viewed as an average share earned across repeated market-clearing events and does not depend on one-time market outcomes such as the clearing price. Allowing for more sophisticated, outcome-contingent contracts would complicate the analysis, as the trader's demand would no longer be linear in prices (a feature that is known to render uniform-price auctions intractable).

Second, I assume that traders and clients choose their effort levels and contracts by maximizing their expected certainty equivalents—that is, the utility they would obtain from receiving a given level of wealth with certainty, conditional on the realized signal. In doing so, they integrate out the uncertainty that arises from not knowing the signal realization at the time of the decision.[12] This modeling choice preserves tractability and reflects the idea that traders and clients do not make a single, one-shot decision based on a unique signal. Instead, trading occurs repeatedly across many rounds, and decisions

[11] In the classic principal–agent framework, where the agent's effort directly increases output, Holstrom and Milgrom (1987) show that such linear contracts are optimal when the agent has CARA utility and output noise is normally distributed. In my model the trader's (agent's) effort does not increase productivity, but affects their information when competing for the asset. Given the strategic complexity of the setting, establishing the optimality of linear contracts is beyond the scope of this paper.

[12] Formally, maximizing the expected certainty equivalent, where expectations are taken over signals, $\mathbb{E}[\mathcal{U}(s_i; \phi_i)]$, is not equivalent to maximizing expected CARA utility in (4), where expectations are taken jointly over wealth and signals.

are likely to reflect average outcomes across these rounds, which the model captures through expectations.

Third, I focus on equilibria in which clients choose contracts that incentivize effort, by including a large enough loss to the client when the trader shirks. This rules out cases in which one dealer would let its trader shirk while others induce effort and serves as a reduced-form way to capture the social or institutional pressures that deter shirking when peers exert effort.

Fourth, I abstract from common-value concerns, meaning that all traders and clients share the same beliefs about the market-wide sentiment for the asset (ν), so prices convey no information about the value of the asset. Allowing for common values would introduce additional strategic considerations in the trading stage: bidders would have an additional incentive to shade their bids to avoid overpaying when they are more optimistic than the market—the classic winner's curse. Crucially, however, effort and contract choices occur before trading and would only be affected insofar as certainty equivalents incorporate expectations over both the true value and traders' private signals. Thus, abstracting from common values helps isolate the two key forces that govern the interaction between competition and moral hazard: a price-impact effect, capturing how traders strategically adjust their quantities, and an information-anticipation effect, arising from the fact that contracts are determined before trading occurs. Both forces should also arise in a less tractable framework with common values, though their functional forms would differ.

Finally, the game features a single trading stage, abstracting from intertemporal considerations, and assumes ex-ante identical clients and traders, thereby ruling out heterogeneity arising from persistent differences in preferences or beliefs. Both assumptions could be relaxed in future work, though doing so would be challenging. A promising (yet difficult) direction would be to study a dynamic environment with common values, where prices reveal information about the asset's value and clients can revise their contracts after observing market outcomes.

3.1 How imperfect competition and moral hazard affects choices

My goal is to analyze how the two frictions—imperfect competition and moral hazard—affect asset allocations, prices, and welfare.

Imperfect competition. In an imperfectly competitive market, each trader has price impact (Kyle (1989)). Strategic traders take this into account when choosing how to

trade: they deliberately reduce their demand for additional units to lower the market-clearing price—this is known as 'strategic demand reduction'—, thereby improving the terms of their winning bids (Ausubel et al. (2014)). In contrast, price-taking traders do not strategically adjust their demand. They act as if their choice does not have price impact (i.e., set $\Lambda_i = 0$ in expression (17)), and therefore submit the following naive demand schedule:

$$a^o(p^o, \mu_i(s_i)) = \frac{\rho_I}{\phi^o} Var[R_i \mid \mu_i(s_i)]^{-1}(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^o). \tag{21}$$

The effects of demand reduction become apparent when comparing demand (21) to demand with strategic traders (15), which simplifies to

$$a^*(p, \mu_i(s_i)) = \frac{\rho_I}{\phi^*}\left(\frac{N-2}{N-1}\right) Var[R_i \mid \mu_i(s_i)]^{-1}(\mathbb{E}[R_i \mid \mu_i(s_i)] - p). \tag{22}$$

Imperfect competition introduces a multiplicative adjustment factor, $\left(\frac{N-2}{N-1}\right)$, which converges to 1 as the number of traders grows ($N \to \infty$) and the market becomes perfectly competitive.

Strategic demand reduction distorts both prices and allocations, leading to a welfare loss. It lowers prices ($p^* < p^o$) and causes traders with above-average return expectations to win too little, and those with below-average expectations to win too much compared to the efficient auction outcome. For example, when everyone exerts effort and contracts are identical, $\phi^* = \phi^o$:

$$a^o(p^o, \nu_i) - a(p^*, \nu_i) = \frac{\rho_I}{\phi^*}\frac{1}{N-1}\left(\nu_i - \frac{1}{N}\sum_i \nu_i\right)\Sigma^{-1}, \tag{23}$$

where I have inserted $\nu_i = \mathbb{E}[R_i \mid \mu_i(0)]$ and $\Sigma = Var(R_i \mid \mu_i(0))$. This expression is positive when ν_i exceeds the average and negative otherwise.

Moral hazard. Moral hazard frictions directly affect the contract choice, and through the flexible part of the contract, ϕ^*, the trading decisions. Since the fixed part of the contract, K^*, represents a transfer between the client and trader, and does not affect allocative efficiency, I focus on ϕ^*, throughout the paper.

When the benefit of shirking is sufficiently low so that the trader has no incentives to shirk, $\phi^* = \frac{\rho_I}{\rho_I + \rho_H}$. This achieves the first best-solution when the market is perfectly competitive ($N \to \infty$) or traders are price-takers. The unconstrained contract is independent of the degree of competition among traders since we assume that traders do not

switch dealers—they are loyal to the dealer with whom they have their contract with. When moral hazard is binding, which is the case when $\frac{\rho_I}{\rho_I+\rho_H}$ is too low, the contract is set to ensure that the trader exerts effort: $\Phi(N)$. While the expression is messy, the key property is that it depends non-trivially on the number of traders, and therefore the degree of competition.

The contract, ϕ^*, affects trading because it affects the trader's risk-return trade-off. Increasing the contract raises the trader's 'effective risk-aversion' $\frac{\phi^*}{\rho_I}$. Since traders weigh the extra profit they expect from the purchase against the additional risk it brings, traders are less comfortable taking on larger (risky) positions for a fixed expected return. As a result, the price (19) needs to decline to persuade them to absorb extra amounts of the asset, and price impact (15) increases.

This ultimately affects asset allocations, and allocative efficiency. To see how, compare the winning quantity of trader i when moral hazard is not binding ($\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$)—$a^*_{i,UC}$—to the winning quantity when the constrained is binding, $\Phi(N)$—$a^*_{i,C}$:

$$a^*_{i,UC} - a^*_{i,C} = \left[(\rho_I+\rho_H) - \frac{\rho_I}{\Phi(N)}\right]\frac{1}{N-1}\left(\nu_i - \frac{1}{N}\sum_i \nu_i\right)\Sigma^{-1} \geq 0. \tag{24}$$

Traders with above than average return expectations purchase too much, and traders with below than average expectations too little when the constrained binds. The higher the contract, i.e., the worse the moral hazard friction, the larger this distortion.

How the frictions interact. Intuitively, it is ambiguous whether stronger competition alleviates or amplifies moral hazard. Adding an additional trader can increase or decrease the contract, ϕ^*, depends on three forces.

(i) Price-impact effect: The first force is that greater competition reduces each trader's individual price impact, altering their incentive to strategically scale back demand. As the market becomes more competitive, each trader's impact on the price declines. To be willing to exert effort, traders need to obtain a larger fraction of wealth as soon as the incentive constraint binds. Formally, trader i's price impact—the amount by which the trader can move the market clearing price given demand of all other traders—declines in the number of traders, N, when keeping the contract fixed:

$$\Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\Sigma. \tag{15}$$

The price-impact effect is related to, yet distinct from, the idea that greater competi-

tion compresses the rents traders can extract from market power and thus increases the compensation needed to sustain effort. In our trading environment, market-power rents depend not only on price impact but also on the information–risk ratio, which measures how informative the signal is relative to the residual uncertainty the trader must bear, Σ_ν/Σ.[13] The influence of this ratio on market-power rents varies with the number of traders in a way not captured by the price-impact effect.

(ii) Price-information effect: Adding more traders makes prices more predictable, and thus affects both contract and effort choices. For illustration consider a client i who chooses their contract ϕ_i given all other clients choose ϕ. This client can better predict the clearing price as N increases, because the variance of clearing prices declines in N:

$$Var(p^c(0)) = \frac{\Sigma}{\big(2\phi + (N-2)\phi_i\big)^2}\left(\phi^2 + \frac{\big(\phi + (N-2)\phi_i\big)^2}{N-1}\right).$$

To isolate information effect from the price-impact effect, I shut off the latter by assuming that traders behave as price-takers (see Proposition 3 in the appendix). The chosen constrained contract, $\tilde{\Phi}(N)$, shares close similarities to the contract with strategic demand reduction, $\Phi(N)$. The constraint contract shares close similarities to the one that arises with strategic traders, highlighting the crucial role of the information effect.

(iii) Information-risk-ratio effect: The third force reflects how tempting it is for a trader to shirk, which depends on how noisy the signal becomes when shirking relative to exerting effort. To isolate this effect from the other two, I consider the limiting case as the number of traders N grows large—an environment approximating perfect competition, where prices are fully predictable and no trader strategically reduces demand.

Corollary 2. *Let $\gamma_\infty(\sigma_\epsilon) = \frac{\rho_I}{2}\left(\frac{\Sigma_\nu}{\Sigma} - \frac{\sigma_\epsilon+\Sigma_\nu}{\tilde{\Sigma}}\right)$; note that $\gamma'_\infty(\sigma_\epsilon) < 0$ for $\sigma_\epsilon > \Sigma_\nu\frac{\Sigma_\nu-\Sigma}{\Sigma+\Sigma_\nu}$.*

(i) When σ_ϵ is sufficiently high so that $b > \gamma_\infty(\sigma_\epsilon)$, there exists no finite contract that incentivizes traders to exert effort in a perfectly competitive market ($N \to \infty$).

(ii) When σ_ϵ is sufficiently low so that $b < \gamma_\infty(\sigma_\epsilon)$, the incentive constraint never binds in a perfectly competitive market ($N \to \infty$), so that $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$.

[13] One way to measure market-power rents from the perspective of an individual trader is to compare the expected payoff when acting strategically to the payoff when behaving as a price-taker while all other traders remain strategic: $\mathbb{E}[R_i(0)] = \Lambda^*\frac{A^2\phi^*}{2(N-1)^2N} + \frac{(N-2)\rho_I}{2(N-1)^3}\frac{\Sigma_\nu}{\Sigma}$. A more collective measure compares the expected payoff of a trader (who exerts effort) when all traders behave strategically with the payoff when all traders behave as price-takers, holding contracts fixed: $\mathbb{E}[R(0)] = \Lambda^*\phi^*\left(\frac{A}{N}\right)^2 - \frac{\rho_I}{2N(N-1)}\frac{\Sigma_\nu}{\Sigma}$.

There are two cases: a favorable and an unfavorable one. When the noise from shirking, σ_ϵ, is low relative to the private benefit from shirking, the contract is unconstrained and achieves the first-best; when the noise is large no finite contract can induce effort. To see why, note that in the competitive limit the trader's expected payoff take a familiar mean–variance structure and is proportional to the information–risk ratio, which is:

$$\frac{Var(\mathbb{E}[R_i|\mu_i(s_i)])}{Var(R_i|\mu_i(s_i))} = \frac{Var(\mu_i(s_i))}{Var(R_i|\mu_i(s_i))}. \tag{25}$$

A trader can be incentivized to exert effort if their expected payoff under effort is at least as high as their expected payoff when shirking. When the trader exerts effort, they expect the following payoff:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(0;\phi_i)] = \frac{\rho_I}{2}\frac{Var(\mu_i(0))}{Var(R_i|\mu_i(0))} = \frac{\rho_I}{2}\frac{\Sigma_\nu}{\Sigma}. \tag{26}$$

When they shirk, they expect to get:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(1;\phi_i)] = \frac{\rho_I}{2}\frac{Var(\mu_i(1))}{Var(R_i|\mu_i(1))} + b = \frac{\rho_I}{2}\frac{\sigma_\epsilon + \Sigma_\nu}{\tilde{\Sigma}} + b. \tag{27}$$

When signals under shirking are precise (low σ_ϵ) traders take aggressive positions on small signal differences, because they trust their information. This exposes them to a lot of conditional risk. The denominator $Var(R_i|\mu_i(1))$ is small, so the risk penalty is high. As a result, their expected payoff is lower. When signals are noisy (high σ_ϵ) each trader's information is less sharp. They scale back their positions, taking on less risk. The posterior variance $Var(R_i|\mu_i(1))$ is larger, so the risk penalty is lighter. Even though their information is worse, the risk-adjusted payoff is higher, because they're not as exposed to as much risk.

With finitely many traders, the three forces— (i) the price-impact effect, (ii) the price-information effect, and (iii) the information-risk ratio effect—jointly give rise to the non-trivial relationship between competition and moral hazard. To capture this relationship we analyze how the constraint contract, $\Phi(N)$, changes when we add an additional trader. When the trader needs to obtain a higher fraction of wealth to still exert effort, we say that moral hazard intensifies. When the opposite is true, moral hazard is relaxed.

Corollary 3. *Define*

$$\gamma(\sigma_\epsilon) = \frac{\rho_I\Sigma(-2(-1+N)^3\sigma_\epsilon + (4+N(-8+(7-2N)N))\Sigma_\nu) + (-2+N)(2+N(-3+2N))\rho_I\Sigma_\nu\tilde{\Sigma}}{4(-1+N)^2\Sigma(\Sigma+(-1+N)\tilde{\Sigma})}$$

(i) When σ_ϵ is sufficiently high so that $b > \gamma(\sigma_\epsilon)$, $\Phi(N)$ increases in N.

(ii) When σ_ϵ is sufficiently high so that $b < \gamma(\sigma_\epsilon)$, $\Phi(N)$ decreases in N.

Note that $\gamma'(\sigma_\epsilon) < 0$ for $\sigma_\epsilon > \overline{\sigma}_\epsilon = \frac{(2N\Sigma_\nu^3((-2+(-2+N)^2N)\Sigma+(-2+N)(-1+N)(2+N(-3+2N))\Sigma_\nu))^{1/2}-2(-1+N)^{3/2}N\Sigma\Sigma_\nu}{2(-1+N)^{3/2}(N(\Sigma+\Sigma_\nu)-\Sigma_\nu)}$.

Corollary 3 underscores the non-trivial interaction between moral hazard and imperfect competition, foreshadowed in Corollary 2. When noise from shirking is large, increasing the number of traders worsens the moral hazard problem—the fraction of wealth the trader must obtain to exert effort increases in N (statement (i)); otherwise it decreases (statement (ii)). Given that the cutoff, $\gamma(\sigma_\epsilon)$ varies in N, however, the relationship is non-trivial.

Figure 3 provides numerical examples of how the contract ϕ^* can behave as the number of traders increases. The examples are chosen to illustrate the two cases presented in Corollary 2, though these cases do not exhaust all possible shapes the contract may take. In the first example—corresponding to statement (i)— noise from shirking is high, $\sigma_\epsilon = 1$, and $\Phi(N)$ increases in N for large enough N. In the second example—corresponding to statement (ii), noise from shirking is low, $\sigma_\epsilon = 0.1$. Then, for N large enough the contract is unconstrained. In both settings, $\rho_I = 0.5, \rho_H = 0.5, \Sigma = 0.9, \Sigma_\nu = 1.5, b = 0.002$; $A = 1$ in the first example and $A = 5$ in the second example.

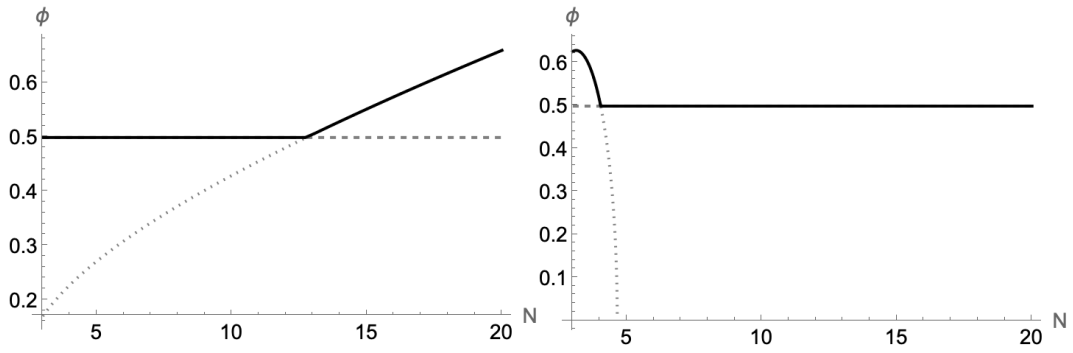
The finding that contracts vary in the number of traders suggests that policies designed to promote competition in asset markets, for example, lowering entry barriers for traders, may inadvertently aggravate moral hazard problems. It further clarifies the conditions under which this risk arises, namely when predicting asset returns requires effort so that shirking is beneficial. Such situations are more likely in markets where returns are information-sensitive, such as corporate bonds or certain derivatives, and less so in markets like short-term government debt, where returns are easier to forecast. The difficulty of forecasting returns also varies with economic conditions, becoming especially pronounced during periods of heightened uncertainty or market stress. The result therefore suggests that moral hazard problems could be more severe in times of distress.

3.2 How imperfect competition and moral hazard affect welfare

Having analyzed how moral hazard and imperfect competition distort trading behavior, the welfare implications follow directly. In the first best, traders take positions that reflect both their own and their clients' risk tolerance. With moral hazard and imperfect competition, however, allocations depend on the equilibrium contract—and the contract

Figure 3: Examples of how the flexible contract, ϕ^*, can change in N

(a) ϕ^*: High noise from shirking (b) ϕ^*: Low noise from shirking



Notes: Figure 3 plot the equilibrium contract ϕ^* (black line) as a function of the number of traders N for two numeric examples. The chosen contract is a piecewise function consisting of the unconstrained contract, $\frac{\rho_I}{\rho_I+\rho_H}$, the dashed gray line, and the constrained contract, $\Phi(N)$, the dotted gray line. In both examples, $\rho_I = 0.5, \rho_H = 0.5, \Sigma = .9, \Sigma_\nu = 1.5, b = 0.002, \bar{U} = 0.01$. In Figure 3a shirking-noise is high, $\sigma_\epsilon = 1.5$ and supply is large, $A = 5$, in Figure 3b it is low, $\sigma_\epsilon = 0.001$ and supply is small $A = 1$. Appendix Figure shows the corresponding fixed part of the contract, K^*.

itself varies with the number of traders. Thus, whether greater competition improves or reduces welfare should turn on how the contract responds to N: if tighter competition weakens incentives, welfare could fall even as markets become more competitive.

To make this precise, I evaluate allocative efficiency under the equilibrium allocation. Because traders receive a lump-sum transfer that leaves them exactly at their outside option, this welfare measure is equivalent to aggregate client utility. Formally,

$$W(\phi^*, N) = \sum_i W_i(\phi^*, N), \tag{28}$$

$$W_i(\phi^*, N) = a^*(p^*, \nu_i)\nu_i - \sum_{k\in\{h,i\}} \frac{1}{2}\phi_k^{*2}\frac{1}{\rho_k} a^*(p^*, \nu_i)\Sigma\, a^*(p^*, \nu_i), \tag{29}$$

where $\phi_h = 1 - \phi_i$. Following the market-design literature, I adopt an ex-ante perspective and evaluate $\mathbb{E}[W(\phi^*, N)]$, ensuring that welfare comparisons do not depend on a particular realization of return expectations. I refer to this expected allocative efficiency as 'welfare'.

Welfare is shaped by three forces: how wealth is split between traders and clients, their relative risk tolerances, and how much useful information traders have relative to the underlying asset's risk (the information–risk ratio). To build intuition, I begin with two simple benchmarks. In the first benchmark, a single trader ($N = 1$) purchases an

asset with returns $\mathcal{N}(\nu,\Sigma)$ and keeps a share ϕ of the payoff. Welfare is

$$\mathbb{E}[W^{B1}(\phi,N)] = A\nu - \frac{1}{2}\sum_{k\in\{i,h\}}\frac{1}{\rho_k}\phi_k^2\Sigma A^2 = A\nu - \beta(\phi,1)\Sigma A^2,$$

where $\beta(\phi,N)$ is given by expression (32) in Proposition 2. The first term reflects expected gains from holding the asset; the second captures risk-bearing costs, which depend on the wealth split and on risk tolerances. In the second benchmark, $N>2$ traders trade on heterogeneous expected returns and keep all assets ($\phi=1$). Welfare is

$$\mathbb{E}[W^{B2}(1,N)] = A\nu - \frac{1}{2N\rho_I}\Sigma A^2 + \frac{\rho_I N(N-2)}{2(N-1)}\frac{\Sigma_\nu}{\Sigma} = A\nu - \beta(1,N)\Sigma A^2 + \alpha(1,N)\beta(1,N)\frac{\Sigma_\nu}{\Sigma},$$

with $\alpha(\phi,N)$ given by expression (31) in Proposition 2. Relative to the first benchmark, an additional force appears: the information–risk ratio Σ_ν/Σ, which raises welfare when information is precise and traders are risk-tolerant. In the full model with moral hazard, welfare is driven by the same three forces.

Proposition 2 (a). *Welfare is given by*

$$\mathbb{E}[W(\phi^*,N)] = A\nu - \beta(\phi^*,N)\Sigma A^2 + \alpha(\phi^*,N)\beta(\phi^*,N)\frac{\Sigma_\nu}{\Sigma}, \tag{30}$$

with

$$\alpha(\phi^*,N) = \frac{(-2+N)N\rho_I^2((2+N(-2+\phi^*)-2\phi^*)\phi^*\rho_H + (-2+N)(-1+\phi)^{*2}\rho_I)}{-(-1+N)\phi^{*2}(\rho_I - 2\phi^*\rho_I + \phi^{*2}(\rho_H+\rho_I))}, \tag{31}$$

$$\beta(\phi^*,N) = \frac{1}{2N}\left(\frac{1}{\rho_I}\phi^{*2} + \frac{1}{\rho_H}(1-\phi^*)^2\right), \tag{32}$$

where $\alpha(\phi^*,N)\geq 0$ *for* $\phi^*<1$, *and* $\beta(\phi^*,N)\geq 0$. *Note that* $\frac{\partial\beta(\phi^*,N)}{\partial N}\leq 0$; $\frac{\partial\alpha(\phi^*,N)\beta(\phi^*,N)}{\partial N}\geq 0$ *for* $\phi^*<1$; $\frac{\partial\alpha(\phi^*,N)\beta(\phi^*,N)}{\partial\phi}\leq 0$; $\frac{\partial\beta(\phi^*,N)}{\partial\phi}\geq 0$.

Changes in moral hazard (measured by ϕ, since an increase in the private benefit of shirking eventually increases ϕ) and changes in the degree of competition (captured by N, since the market becomes perfectly competitive as $N\to\infty$) both affect welfare but in different ways.

Proposition 2 (b).

(i) Welfare increases when moral hazard becomes less severe, i.e., ϕ^* decreases, because the private benefit from shirking, b, decreases.

(ii) Welfare decreases when the number of traders (N) increases when

$$\underbrace{\frac{\partial \mathbb{E}[W(\phi^*_{UC}, N)]}{\partial N}}_{\text{Direct effect}} - \underbrace{\frac{\partial \beta(\Phi(N), N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\Sigma A^2 + \frac{\partial \alpha(\Phi(N), N)\beta(\Phi(N), N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\frac{\Sigma_\nu}{\Sigma}}_{\text{Indirect effect through } \Phi(N)} < 0, \tag{33}$$

and increases otherwise. Here $\phi^*_{UC} = \frac{\rho_I}{\rho_I + \rho_H}$ denotes the unconstrained contract.

Reducing moral hazard is always welfare enhancing. As the contract share ϕ declines, the loss from risk aversion falls ($\frac{\partial \beta(\phi^*, N)}{\partial \phi} \geq 0$) because, when the incentive constraint binds, the trader is effectively more risk-averse than the client ($\frac{\rho_I}{\rho_I + \rho_H} \leq \phi$), so shifting surplus from the trader to the client reduces the aggregate cost of bearing risk.[14] A second force reinforces this effect: when ϕ falls, the trader absorbs a smaller share of the risky position, which raises their effective willingness to take risk per unit of information. As a result, the trader reacts more aggressively to informative signals, improving welfare. Formally, this is captured by $\frac{\partial(\alpha(\phi^*, N)\beta(\phi^*, N))}{\partial \phi} \leq 0$, which shows that lowering ϕ strengthens the information-driven component of expected welfare.

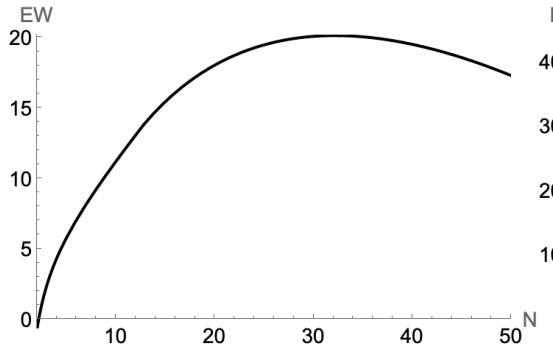
However, increasing competition by adding traders can either raise or reduce welfare. To build intuition, begin with the simpler case in which the contract does not adjust with N—that is, when incentive constraints are slack. In this benchmark, more competition unambiguously increases welfare, for two reasons. First, as the number of traders grows, each trader (and thus each client) holds a smaller share of the risky asset, A/N, so the utility loss from risk aversion falls. Formally, since $\frac{\partial \beta(\phi^*, N)}{\partial N} \leq 0$, the term $-\frac{\partial \beta(\phi^*, N)}{\partial N}\Sigma A^2 \geq 0$ contributes positively to welfare. Second, individual uncertainty about the asset's return matters less when more traders participate, because prices aggregate information across a larger set of signals. This improves welfare through the information–risk channel, captured by $\frac{\partial(\alpha(\phi^*, N)\beta(\phi^*, N))}{\partial N}\frac{\Sigma_\nu}{\Sigma} \geq 0$.

When the contract is constrained, welfare may decline as the number of traders increases because N affects welfare not only directly, but also indirectly through its effect on the equilibrium contract. A decline occurs precisely when the contract becomes more generous to traders as N rises—an unfavorable case in which additional competition forces the principal to reward effort more heavily. In this case, the negative indirect effect (the second line of expression 33) can outweigh the positive direct effect of competition (the first line of expression 33), leading to a net welfare loss.

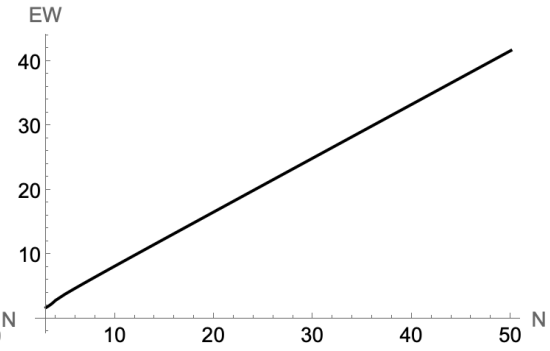
[14] For instance, if trader and client have equal risk tolerance ($\rho_H = \rho_I$), welfare rises when moving from a situation in which the trader receives more than half the surplus ($\phi \geq \frac{1}{2}$) toward a more balanced split.

Figure 4: Examples of how equilibrium welfare can change in N

(a) High noise from shirking



(b) Low noise from shirking



Notes: Figure 4 plots equilibrium welfare, $\mathbb{E}[W(\phi^*, N)]$ of Proposition 2, as a function of the number of traders N for two numeric examples of Figure 3.

To illustrate this, I return to the two numerical examples. In the high–σ_ϵ example of Figure 3a, the contract is constrained at $N = 35$ (with $\Phi(35) = 0.946$) and is *increasing* in N. At this point, the indirect marginal effect of competition on welfare is negative: $\frac{\partial \mathbb{E}[W(\Phi(35),35)]}{\partial N} = 0.479655 - 0.54074 < 0$. Although the direct effect $\frac{\partial \mathbb{E}[W(\phi^*,N)]}{\partial N}$ is positive, the contract's unfavorable response to competition leads to an overall decline in welfare. By contrast, in the low–σ_ϵ example of Figure 3b, the contract is constrained already at $N = 3$ (with $\Phi(3) = 0.62$) but *decreasing* in N. The indirect effect is therefore positive, reinforcing the direct effect, and the overall welfare effect of increasing N is positive.

Take away. The finding implies that policies aimed at promoting competition in asset trade settings may, paradoxically, reduce efficiency by exacerbating moral hazard problems. In such cases, an increase in the number of traders worsens incentive misalignments and thereby diminishes welfare. This highlights a potential negative side-effect of competition that runs counter to the standard intuition that increase competition promotes efficiency (especially in Kyle-models where traders have price impact and distort allocations because of it). Corollary 2 identifies when this is the case, for example, when forecasting returns requires costly effort—so that contracts become more generous in N to maintain traders' incentives—and when the expected per-unit return of assets is sufficiently high, additional entry of traders reduces welfare.

In Appendix A.2, I assess which friction—moral hazard or imperfect competition—has the larger welfare impact and under what conditions one dominates. I isolate the effect of imperfect competition by comparing outcomes when traders act as price-takers with those when they behave strategically and reduce demand. To isolate moral hazard, I hold

competition fixed by imposing price-taking behavior and compare welfare under unconstrained contracting with welfare when incentive constraints bind. This yields separate welfare losses for each friction while still allowing the number of traders to influence equilibrium contracts and prices. The resulting characterization helps determine whether policy should prioritize promoting competition—e.g., by lowering exchange membership requirements—or mitigating moral hazard through improved incentive contracts, recognizing that the two frictions interact and cannot be treated independently.

4 Calibration

For the remainder of the paper, I connect the theoretical framework back to the empirical evidence that motivated the two frictions—in particular, Tables 1 and 2. I use the model as a rough, first-order approximation to interpret the trade data and to get a sense of the relative importance of moral-hazard and imperfect-competition frictions, without attempting a full structural estimation of all model primitives (such as the costs and benefits from shirking). Such an estimation would require information that trade-level data cannot provide, since shirking does not occur in equilibrium and therefore leaves no trace in observed transactions. I therefore view my quantification as a useful first step that shows how far one can get using trade-level data to detect patterns that could point towards moral-hazard frictions, and its interactions with imperfect competition.

Model-data mapping. My model relies on several simplifying assumptions that may not hold exactly in practice. I therefore introduce a few targeted adjustments that preserve tractability while still allowing the model to capture the key empirical patterns that motivate it. Rather than cataloguing every potential discrepancy between the model and the data, I highlight the most consequential simplifications and the conditions under which the model provides a rough approximation to observed trading behavior.

The first main difference between my model and real-world exchange trading is that the model is static, whereas actual trading unfolds continuously throughout the day. To address this, I approximate continuous trading with frequent, periodically held batch auctions—effectively discretizing trading into a sequence of uniform-price auctions that clear at Γ periods during the day.[15] In each auction, I treat active orders as net demand and passive resting orders as fixed supply. This approach empathizes active trading decisions while abstracting from the more complex dynamics considerations underpinning

[15]This follows the spirit of Budish et al. (2015), who propose replacing continuous trading with frequent batch auctions.

orders that remain in the book.

The second key difference is that, in the model, all trades clear at a single uniform price, whereas in practice trading is subject to bid–ask spreads, so buys and sells occur at different prices. To address this, I model separate auctions for the (active) buy- versus sell-side. Treating the two sides separately avoids a mechanical bias from averaging transaction prices over short intervals—namely, higher averages when active buy orders exceed active sell orders, and lower ones when the reverse holds.

To reflect the observed price differences across accounts and exchanges, I also model separate auctions for client-account and inventory-account trades on each exchange. Institutionally, client and proprietary books within a dealer are typically ring-fenced by mandate, client-priority rules, and distinct risk budgets, making joint optimization across accounts operationally costly and often infeasible in real time. Conceptually, allowing for optimal bidding in an auction with multiple asymmetric account types would introduce substantial analytical complexity. It is therefore plausible—both operationally and theoretically—that traders themselves rely on a simpler, account-by-account decision rule.

In addition, I treat each security separately so that prices and trade sizes are not averaged across securities with different price levels, market capitalizations, or liquidity. Technically, auctions for a given security clear independently: demand and supply are aggregated only within that security, not across securities. This reflects how stock markets operate, but it implies that traders bid as if securities are unrelated, even though their returns may be correlated.

Data sample. I use all active trades executed in 2021 on TSX and Alpha for securities in the S&P/TSX 60 index, focusing on these highly liquid securities to ensure sufficient trading activity over short intervals. In my baseline specification, I divide the trading day into 15-minute intervals—the shortest window that still provides sufficient activity on both exchanges, with Alpha being the limiting factor. As robustness, I also estimate the parameters using 10-, 20-, and 30-minute intervals.

Given this discretization, I collapse the raw data per exchange $e \in \{\text{TSX, Alpha}\}$, day t, security s, account-type $k \in \{\text{CL, IN}\}$, side $l \in \{\text{buy, sell}\}$, period τ. I average over prices—these average prices reflect the hypothetical auction prices, $p^{el}_{tsk\tau}$, and aggregate traded quantities to obtain the total amount traded in each auction, $A^{el}_{tsk\tau}$. Appendix Figures A2a and A2b shows the distribution of prices and net-supply for each auction type.

For each day–security–account-type–side combination, I count the number of distinct

traders that execute a trade in that category and use that count as the number of potential bidders in the corresponding auctions for this security on that day, N^{el}_{tsk}. For example, if I observe 15 trader LEIs executing client-account buy trades for security s on TSX on day t, then all buy-side client auctions for that security on TSX on that day are among 15 potential traders.[16] Appendix Figure A2c shows the distribution of the number of traders auction type.

I approximate the return variance, Σ_{ts}, using the intraday variance of the national mid-quote for each security and day; and, as robustness, also construct weekly and monthly mid-quote variances. Quote-based volatility measures are standard in the related literature, because they avoid microstructure noise that inflates transaction-price-based estimates (see Andersen et al. (2001); Aït-Sahalia et al. (2005); Hendershott et al. (2011)). The median variance is approximately C$0.038, about 8 basis points of the median stock price (C$50.33); the full distribution of Σ_{ts} is shown in Appendix Figure A2d.

Moments. The goal is to learn about the trader's risk-tolerance, ρ_I, the contract size, ϕ^*, and the clients' risk-tolerance, ρ_H.

The first two parameters can be point-identified. The key idea is that the price difference between inventory-account trades on TSX and Alpha identifies the trader's risk tolerance, ρ_I (similar to Table 1). Conditional on this, the price difference between client- and inventory-account trades on a given exchange identifies the contract size, ϕ^* (similar to Table 2). I rely on TSX trades for this second step, as TSX is Canada's largest and most liquid exchange and therefore the natural venue for identifying contract terms.

To see why these moments identify ρ_I, ϕ^*, consider an l-side auction of security s on exchange $e \in \{\text{TSX, Alpha}\}$ on day t in period τ for account $k \in \{IN, CL\}$. This auction clears at price:

$$p^{el}_{t\tau sk} = \frac{1}{N^{el}_{tsk}} \sum_i v_{ti\tau s} - \frac{\phi^*_k}{\rho_I} \Sigma_{ts} \left(\frac{N^{el}_{tsk} - 1}{N^{el}_{tsk} - 2} \right) \frac{A^{el}_{t\tau sk}}{N^{el}_{tsk}}, \tag{19'}$$

[16] Two alternative approaches to measuring the number of potential bidders in each auction would be to count the number of trades executed in a given period or to count the number of distinct trader IDs that execute trades in that period. The first approach would imply that, if five inventory-account purchases occur in period τ on day t for security s, five traders participated in that buy-side auction, implicitly treating multiple trades submitted by the same trader as independent private signals and thereby overstating the amount of information in the auction. A more conceptually appropriate measure is the number of distinct trader IDs, but in practice I do not observe sufficiently many distinct traders (that not only bid but also win and execute trades) at the period level often enough to implement this reliably.

with $\phi^*_{IN} = 1$ for inventory-accounts, and $\phi^*_{CL} > \phi^* \geq 0$ for client-accounts. Since $\nu_{ti\tau s}$'s are drawn iid within a day/security from $N(\nu_{ts}, \Sigma_{ts})$, averaging over intraday periods yields the following average auction price:[17]

$$p^{el}_{tsk} = \frac{1}{\Gamma}\sum_{\tau} p^{el}_{t\tau sk} \approx \nu_{ts} - \frac{\phi^*_k}{\rho_I}\Sigma_{ts}\left(\frac{N^{el}_{tsk}-1}{N^{el}_{tsk}-2}\right)\frac{1}{N^{el}_{tsk}}\frac{1}{\Gamma}\sum_{\tau} A^{el}_{t\tau sk}. \tag{19"}$$

Since p^{el}_{tsk}, N^{el}_{tsk}, A^{el}_{tsk}, and Σ_{ts} are observed (or estimated), the only unknown parameters in expression (19")—after differencing prices across accounts or exchanges—are the trader's risk tolerance ρ_I and the contract size ϕ^*.

Then, having identified ρ_I and ϕ^*, I exploit the equilibrium condition from Proposition 1 that the contract must satisfy $\frac{\rho_I}{\rho_H+\rho_I} \leq \phi^*$. In the model, this inequality reflects the fact that clients must offer sufficiently strong incentives to deter traders from shirking, and tighter incentives are required when clients are more risk-averse. Rearranging therefore yields a lower bound on the client's risk-tolerance parameter, ρ^{lb}_H, consistent with the observed contract.

Operationally, I first estimate ρ_I and ϕ^* parameters sequentially using median-based moment conditions in a two-stage GMM procedure: the first stage recovers ρ_I from TSX–Alpha price differences for inventory accounts, and the second estimates ϕ^* from client–inventory price differences on TSX, pooling moments from both buy- and sell-side auctions.[18] With these two estimates, I compute the implied lower bound on client risk tolerance, thereby recovering the minimum degree of client risk tolerance that is consistent with the observed contract in the model's unconstrained benchmark. To ensure that taking daily averages provides a reasonable approximation to the expected auction price for a given account type and trade side, I drop observations when a security is not bought (or sold) by at least two distinct traders for the respective account on a given day.

For inference, I cluster standard errors at the daily level by resampling entire trading days with replacement and re-estimating both parameters in each bootstrap replication. I construct confidence intervals from the empirical percentiles of the bootstrap draws, which provide more reliable coverage for median-based estimators with skewed sampling distributions. For more details about the estimation procedure see Appendix B.1.

[17] In the limit, when either the number of periods or the number of traders grows to infinity, this expression holds with equality.

[18] I use a two-stage rather than a joint estimator because the first-stage estimate of ρ_I is relatively noisy, making sequential estimation more stable and numerically robust while remaining consistent (see Newey and McFadden, 1994). The median-based specification limits the influence of outliers and heavy-tailed ratios, which would otherwise distort mean-based GMM estimates.

Median risk-tolerance and contract. The median estimate of risk tolerance, ρ_I, is about 670 (see Table 3). Although imprecisely estimated,[19] the estimate is broadly consistent with Wittwer and Allen (2023), who infer low levels of risk aversion ($1/\rho_I \approx 0.004$) for Canadian primary dealers—those accounting for the largest trading volume in the exchange data. While not directly comparable in units or normalization, my estimate of $\rho_I \approx 670$ implies a similarly low degree of risk aversion (about 0.0015).

The median contract size, ϕ^*, between clients and their intermediaries is estimated with much greater precision and is consistently around 0.6 across specifications. This implies that intermediaries capture more than half of the surplus generated by a trade—a sizable share that indicates economically meaningful moral-hazard frictions.

Although I am not aware of any empirical study that directly measures the fraction of wealth intermediaries obtain through incentive contracts when trading on behalf of clients on exchanges (as ϕ^* does), related work in asset management provides benchmarks. That literature estimates intermediaries' surplus shares vis-à-vis investors through management fees and performance-based compensation (e.g., Berk and van Binsbergen (2015); Fama and French (2010); French (2008); Agarwal et al., 2009: Lan et al. (2013)). While not directly comparable to ϕ^*, and applicable only to specific types of intermediaries such as mutual funds or hedge funds, these estimates fall in a broadly similar range and generally suggest that intermediaries capture more than half of the total surplus.

Finally, the implied lower bound on clients' risk tolerance is around 385, which is about 60% of the trader's risk-tolerance. This is because, in the model, the observed median contract can arise only when clients are themselves relatively risk-tolerant, given that traders are risk-tolerant.[20] One possible explanation for this result is that many of the active clients in my setting are large institutional investors (such as pension funds, endowments, foundations, or fully managed investment funds) who trade through dealer client-accounts, because they do not themselves have exchange membership. These investors typically hold diversified portfolios, making them relatively risk-tolerant.

[19] This imprecision reflects limited data: only a small number of securities on Alpha are traded on inventory accounts by at least two dealers in every 15-minute interval, restricting usable mini-auctions and introducing noise.

[20] Formally, $\rho_H^{\text{lb}} = \rho_I(1/\phi^* - 1)$. For ρ_H^{lb} to be much smaller than ρ_I, the contract would need to be larger than the one we observe. For example, the client would be at least 10% risk-tolerant as the trader if the contract was about 0.9.

Table 3: Parameter estimates

Parameter	Estimate	SE (SD)	SE (Pct)	95% CI (lower)	95% CI (upper)
ρ_I	670.562	144.447	140.913	372.127	924.504
ρ_H^{lb}	385.241	86.735	86.676	213.174	552.946
ϕ^*	0.635	0.023	0.022	0.596	0.685

Notes: Table 3 reports parameter estimates from the main specification (2021, 15-minute intervals, with return variance measured daily at the security level). Standard errors and confidence intervals are computed using a cluster bootstrap with 999 draws, clustering at the daily level (228 clusters). I report both standard-deviation bootstrap SEs and percentile-based SEs, the latter being more reliable for median-based estimators. The table also reports the implied lower bound on client risk tolerance, ρ_H^{lb}, constructed jointly with the bootstrap distribution of $(\hat{\rho}_I, \hat{\phi}^*)$; see Appendix B.1 for details.

Robustness and limitations. Appendix B.2 reports a series of robustness checks and discusses limitations of the calibration, emphasizing that the estimated parameters may also reflect forces beyond pure risk tolerance or contracting behavior—forces that my tractable model necessarily abstracts from.

How the contract varies with the number of traders. My model predicts that moral hazard and imperfect competition interact, implying that the contract should vary with the number of traders, unless incentives constraints do not bind (Corollary 3).

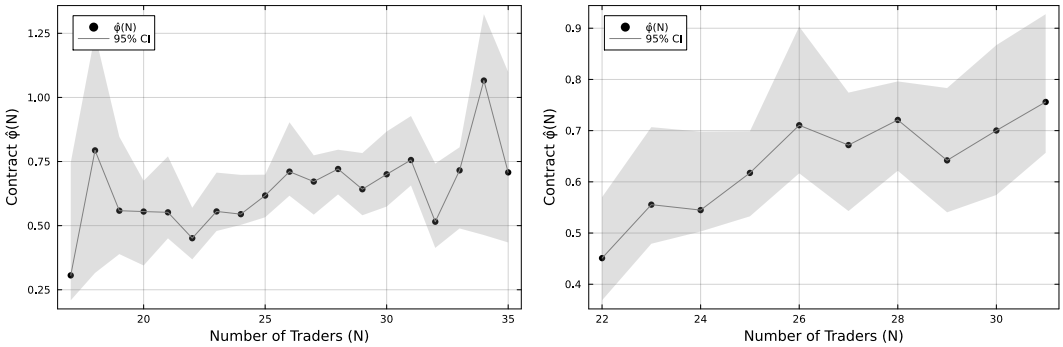
To visualize how the contract varies in the data, I estimate $\phi^*(N)$ separately for each number of active TSX client-account traders per security and day, $N_{ts}^{TSX,l} = N$ (details in Appendix B.1). Figure 5a shows an upward trend but with substantial noise at the lowest and highest trader counts, where data are sparse. After excluding those thinly populated groups in Figure 5b, the pattern is pronounced: the contract increases from about 0.45 to 0.76 as the number of active traders rises by 40% from 22 to 31.

This empirical pattern could suggest that moral-hazard frictions intensify as the number of traders increases—a concerning possibility that motivates further research to better account for confounding factors. In particular, the observed relationship could also reflect time-varying market conditions, which are averaged over in this reduced-form estimation. For example, the model predicts that $\Phi(N)$ declines in the total amount traded, $|A|$; thus, a systematic decrease in A as N rises could generate the observed pattern. Appendix Figure A3 shows that the opposite is true: $|A|$ tends to increase with N. We additionally examine whether return variance differs systematically across trader-count groups and find no evidence of such differences.

However, cleanly accounting for time-varying market conditions would require a fully structural approach that jointly estimates all model primitives and allows parame-

Figure 5: How ϕ^* changes in N

(a) No restrictions (b) At least 50 observations per group



Notes: Figure 5a reports the contract estimates and their 95% confidence intervals for each trader-count group using all available observations. Figure 5b repeats the exercise but excludes groups with fewer than 50 observations. This restriction narrows the support of trader counts from 17–35 to 22–31 and removes much of the noise in the extremes. Appendix B.1 provides further details on the estimation procedure.

ters—such as the noise and the private benefit from shirking, σ_ϵ and b—to vary across time and securities. This is not feasible with trade-level data alone, because key inputs are unobserved: actual contract terms and the behavioral primitives that determine the costs and benefits of shirking. Since shirking does not occur in equilibrium, these objects leave no direct imprint in transaction data and cannot be recovered without additional information.

Welfare implications. With this caveat in mind, and taking the model at face value, Proposition 2 offers a way to gauge the relative importance of moral hazard and imperfect competition. To illustrate this, I focus on a representative TSX client-account auction, constructed using the median return variance $\overline{\Sigma} = \Sigma_{ts}$ and the median daily average net supply (C$1,604) from TSX client-account auctions with exactly 25 traders. I take $N = 25$ because it lies near the center of the empirical distribution of active client-account traders (Figure 5b), and both N and $N+1$ have enough observations for the corresponding contracts to be estimated reliably.

To highlight how welfare depends on information quality (which I do not observe), I vary the trader's signal variance by setting $\Sigma_\nu = \kappa\overline{\Sigma}$ for different values of the information–risk ratio κ. Because the client's risk tolerance is also unknown, I let ρ_H vary from its lower bound up to values that exceed the trader's risk tolerance. Further, to express welfare changes in percentage terms, I set the trader's expected signal, ν, equal to the median mid-quote in the relevant auctions (C$57.16). Because this value is large

Table 4: Welfare effects for median client auction

ρ		ϕ			ΔW_ϕ		ΔW_N		$\Delta W_N^{\text{direct}}$		$\Delta W_{\phi(N)}^{\text{indirect}}$	
$\hat{\rho}_I$	ρ_H	ϕ_{UC}	$\hat{\phi}^*$	$\hat{\phi}^*_+$	C$	%	C$	%	C$	%	C$	%
670	405	0.62	0.62	0.71	-0.01	-0.01	-0.09	-0.09	0.54	0.51	-0.63	-0.6
670	425	0.61	0.62	0.71	0.01	0.01	-0.21	-0.2	0.55	0.52	-0.76	-0.72
670	450	0.6	0.62	0.71	0.06	0.06	-0.35	-0.33	0.56	0.53	-0.91	-0.86
670	500	0.57	0.62	0.71	0.24	0.23	-0.58	-0.55	0.58	0.55	-1.16	-1.1
670	600	0.53	0.62	0.71	0.8	0.76	-0.93	-0.87	0.61	0.57	-1.53	-1.45
670	670	0.5	0.62	0.71	1.31	1.23	-1.11	-1.04	0.62	0.59	-1.73	-1.63
670	700	0.49	0.62	0.71	1.55	1.45	-1.18	-1.1	0.63	0.59	-1.8	-1.69
670	800	0.46	0.62	0.71	2.4	2.25	-1.36	-1.27	0.64	0.6	-2.01	-1.88
670	900	0.43	0.62	0.71	3.34	3.12	-1.51	-1.41	0.66	0.61	-2.16	-2.02
670	1000	0.4	0.62	0.71	4.33	4.03	-1.62	-1.51	0.67	0.62	-2.29	-2.13

Notes: Table 4 reports welfare effects for different values of the client's risk tolerance ρ_H in a representative TSX client-account auction ($N = 25$, $A =$ C$1,604, $\nu =$ C$57.16, and $\Sigma_\nu = \overline{\Sigma} = 0.038$). Columns 1–2 list the estimated trader risk tolerance $\hat{\rho}_I$ and the ρ_H values considered. Columns 3–5 report, respectively, the unconstrained contract $\phi_{UC} = \rho_I/(\rho_I + \rho_H)$, the estimated contract $\hat{\phi}^*$ for $N = 25$, and the estimated contract $\hat{\phi}^*_+$ for $N+1 = 26$. Columns 6–13 present welfare changes, shown first in thousands of Canadian dollars (C$ k) and, directly below, as percentage changes relative to baseline welfare. ΔW_ϕ reports the welfare gain from relaxing moral hazard at $N = 25$ (moving from $\hat{\phi}^*(N)$ to ϕ_{UC}). ΔW_N gives the total effect of increasing competition from N to $N+1$ with the contract adjusting endogenously. $\Delta W_N^{\text{direct}}$ isolates the welfare effect of entry holding the contract fixed at $\hat{\phi}^*$, while the indirect effect captures the additional effect arising from the contract change from $\hat{\phi}^*$ to $\hat{\phi}^*_+$. Absolute and percentage welfare changes are numerically similar because baseline welfare is large (approximately C$105,000), dominated by the level term $A\nu$, whereas the risk-adjustment components are small. Since ν scales levels but not differences, relative welfare comparisons are insensitive to its specific value.

relative to the return variance (0.038), the resulting percentage changes are numerically small. For this reason, I focus on the relative size of welfare changes across scenarios rather than their absolute percentage levels. Reassuringly, the qualitative patterns are the same whether welfare is measured in absolute terms (which do not depend on ν) or in percentage terms.

First, I compare welfare under the observed contract to welfare under the unconstrained benchmark, $\phi_{UC} = \frac{\rho_I}{\rho_H+\rho_I}$, holding the number of traders fixed at $N = 25$. This isolates the welfare effect of removing moral hazard while keeping competition constant. I report results beginning at the lower bound for ρ_H implied by the observed contract at $N = 25$.[21] Across all feasible ρ_H, eliminating moral hazard increases welfare—as ex-

[21]This lower bound exceeds the median lower bound in Table 3 because the contract at $N = 25$ (0.617) is slightly smaller than the median estimate (0.635). For ρ_H below this threshold, the unconstrained contract would exceed the observed one, which cannot occur in equilibrium and

pected. The percentage gain rises in the client's risk tolerance and in the information–risk ratio, but declines in supply (Appendix Figure A4a). This is because the improvement from stronger incentives grows much more slowly with A than baseline welfare, which rises roughly linearly in Av. As a result, the absolute gains remain modest relative to total surplus, so the percentage gains shrink mechanically as auctions become larger.

Second, I evaluate the effect of increasing competition by adding one more trader. The welfare impact has a direct component—additional traders improve risk sharing—and an indirect component that operates through the adjustment of the contract when N changes. The direct effect is always positive, but the indirect effect can be either positive or negative, depending on how the contract responds to higher competition. In the data, the contract rises when moving from 25 to 26 traders, generating a negative indirect effect, which dominates the direct effect, leading to a net welfare loss from adding an extra trader.[22]

The key insight—robust across specifications I have considered—is that the welfare loss generated by the interaction between moral hazard and imperfect competition is sizable relative to the gains from eliminating moral hazard itself. This effect is strongest when clients are less risk-tolerant than traders, which is empirically plausible. Together, these findings imply that ignoring the adverse interaction with limited competition can lead to substantially understating the severity of moral-hazard frictions.

Take away. Taken together, these findings show that the mechanism emphasized by the model is not merely theoretical: the interaction between moral hazard and imperfect competition is empirically relevant, and inefficiency is driven by how the two frictions amplify one another rather than by either friction in isolation. A fuller understanding of this interaction—and how it responds to changing market conditions—will require richer data and more structural analyses, offering a natural direction for future research.

5 Conclusion

I introduce a model showing that imperfect competition and moral hazard interact in shaping trading outcomes. Inefficiency arises from this interaction—more competition is not always welfare-improving when incentives are weak. Using rich proprietary trade

would mechanically reduce welfare because excessively strong incentives are suboptimal.

[22] In the calibration, the indirect effect becomes positive whenever the contract falls—for example, from N=26 to N=27. Over wider ranges, such as N=22 to N=31, where the contract mostly rises, the indirect effect is negative and grows in magnitude.

data from Canadian exchanges, I illustrate how the model helps interpret observed price differences and assess the relative importance of the two frictions. Effective policy therefore needs to address both frictions jointly.

References

Agarwal, V., N. D. Daniel, and N. Y. Naik (2009). Role of managerial incentives and discretion in hedge fund performance. *Journal of Finance 64*(5), 2221–2256.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Journal of Finance 60*(2), 1207–1252.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

Ausubel, L. M., P. Cramton, M. Pycia, M. Rostek, and M. Weretka (2014). Demand reduction and inefficiency in multi-unit auctions. *Review of Economic Studies 81*, 1366–1400.

Baldauf, M., C. Frei, and J. Mollner (2024). Block trade contracting. *Journal of Financial Economics 160*, 1–29.

Berk, J. B. and J. H. van Binsbergen (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics 118*(1), 1–20.

Bernhardt, D., V. Dvoracek, E. Hughson, and I. M. Werner (2005). Why do larger orders receive discounts on the london stock exchange? *The Review of Financial Studies 18*(4), 1343–1368.

Bhattacharya, S. and P. Pfleiderer (1985). Delegated portfolio management. *Journal of Economic Theory 36*(1), 1–25.

Boyd, J. H. and G. D. Nicolo (2005). The theory of bank risk taking and competition revisited. *The Journal of Finance 60*(3), 1329–1343.

Budish, E., P. Cramton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics 130*(4), 1547–1621.

Buffa, A. M., D. Vayanos, and P. Woolley (2022). Asset management contract and equilibrium prices. *Journal of Political Economy 130*(12), 3025–3342.

Carpenter, J. N. (2000). Does option compensation increase managerial risk appetite? *Journal of Finance 55*(5), 2311–2311.

Chen, S., R. Kaniel, and C. C. Opp (2025). Market power in the securities lending market. Working paper.

Crawford, G., N. Pavanini, and F. Schivardi (2018). Asymmetric information and imperfect competition in lending markets. *American Economic Review 108*(7), 1659–1701.

DeMarzo, P. M. and M. J. Fishman (2007). Optimal long-term financial contracting. *Review of Economic Studies 20*(6), 2079–2128.

Di Tella, S. and Y. Sannikov (2021). Optimal asset management contracts with hidden savings. *Econometrica 89*(3), 1099–1139.

Einav, L., A. Fikelstein, and N. Mahoney (2021). Chapter 14: The IO of selection markets. In: Handbook of Industrial Organization, Volume 5, Issue 1: 389-429.

Fama, E. F. and K. R. French (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance 65*(5), 1915–1947.

French, K. R. (2008). Presidential address: The cost of active investing. *Journal of Finance 63*(4), 1537–1573.

Garcia, D. and J. M. Vanden (2009). Information acquisition and mutual funds. *Journal of Economic Theory 144*(5), 1965–1995.

Glode, V. and C. C. Opp (2020). Over-the-counter versus limit-order-markets: The role of traders' expertise. *The Review of Financial Studies 33*(2), 866–915.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), 393–408.

He, Z. and A. Krishnamurthy (2012). A model of capital and crises. *The Review of Economic Studies 79*(2), 735–777.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *The American Economic Review 103*(2), 732–770.

Hendershott, T., C. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*(1), 1–33.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics 10*(1), 74–91.

Holmström, B. and J. Tirole (1997). Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics 112*(3), 663–691.

Holstrom, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*(2), 303–328.

Kashyap, A. K., N. Kovrijnykh, J. Li, and A. Pavlova (2023). Is there too much benchmarking in asset managemetn? *American Economic Review 113*(4), 1112–1141.

Kastl, J. (2012). On the properties of equilibria in private value divisible good auctions with constrained bidding. *Journal of Mathematical Economics 48*, 339–352.

Keeley, M. (1990). Deposit insurance, risk, and market power in banking. *American Economic Review 80*(5), 1183–1200.

Klemperer, P. D. and M. A. Meyer (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica 57*(6), 1243–1277.

Kyle, A. S. (1989). Informed speculation with imperfect competition. *Review of Economic Studies 56*(3), 317–355.

Kyle, A. S., H. Ou-Yang, and B. Wei (2011). A model of portfolio delegation and strategic trading. *The Review of Financial Studies 24*(11), 3778–3812.

Lan, Y., C. Wang, and L. Yang (2013). The economics of hedge funds. *Journal of Financial Economics 110*(2), 300–323.

MacKinnon, J. G., M. O. Nielsen, and M. D. Webb (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics 232*, 272–299.

Martinez-Miera, D. and R. Rapullo (2010). Does competition reduce the risk of bank failure? *The Review of Financial Studies 23*(10), 3638–3664.

Metrick, A. and A. Yasuda (2010). The economics of private equity funds. *Review of Financial Studies 23*(6), 2303–2341.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics 7*(1), 105–131.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume IV*, Chapter 36, pp. 2111–2245. Elsevier.

Raith, M. (2003). Competition, risk, and managerial incentives. *American Economic Review 93*(4), 1425–1436.

Ross, S. (1973). The economic theory of agency: The principal's problem. *American Economic Review 63*(2), 134–139.

Rostek, M. and J. Yoon (2025). Equilibrium theory of financial markets: Recent developments. *The Journal of Economic Literature*, forthcoming.

Rostek, M. J. and J. H. Yoon (2021). Exchange design and efficiency. *Econometrica 89*(6), 2887–2928.

Scharfstein, D. S. (1988). Product-market competition and managerial slack. *RAND Journal of Economics 19*(1), 147–155.

Schmidt, K. M. (1997). Managerial incentives and product market competition. *Review of Economic Studies 64*(2), 191–213.

Starks, L. T. (1987). Performance incentive fees: An agency theoretic approach. *Journal of Financial and Quantitative Analysis 22*(1), 17–32.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy 69*(3), 213–225.

Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *American Economic Review 71*(3), 393–410.

Vayanos, D. (1999). Strategic trading and welfare in a dynamic market. *The Review of Economic Studies 66*(2), 219–254.

Vives, X. (2011). Strategic supply function competition with private information. *Econometrica 79*(6), 1919–1966.

Wilson, R. (1979). Auctions of shares. *The Quarterly Journal of Economics 93*(4), 675–689.

Wittwer, M. and J. Allen (2023). Market power and capital constraints. Working paper.

Wittwer, M. and A. Uthemann (2024). Market and product specialization in financial markets. Working paper.

ONLINE APPENDIX

Imperfect Competition and Moral Hazard in Financial Markets

by Milena Wittwer

Section A shows additional theoretical results, and proofs.
Section B presents estimation details and additional evidence.
Appendix Tables and Figures are displayed at the end.

A Additional theoretical evidence

Section A.1 solves for an equilibrium with price-taking traders but sophisticated dealers.
Section A.2 analyses when imperfect competition is worse than moral hazard.
Section A.3 presents mathematical auxiliary statements and all proofs.

A.1 Price takers

Here we analyze what happens when traders are naive and trade as if $\Lambda_i = 0$. Naive traders do not distort the trading outcome and simply offer what they are willing to pay. They are making a mistake since they behave as if their own demand did not have price impact, while it actually moves the price. Thus, they submit demand functions as in equation (15) but setting $\Lambda_i = 0$. Dealers anticipate this behavior and make different contract choices.

Proposition 3. *There exists and equilibrium with the following properties:*

(i) All dealers choose contract:

$$\phi^* = \begin{cases} \Phi_C(N) & \text{when the incentive constraint binds, i.e., when } \Phi_U(N) \leq \Phi_C(N) \\ \Phi_U(N) & \text{otherwise,} \end{cases}$$

where

$$\Phi_C(N) = \sqrt{\frac{\tilde{\beta}(N) + b\tilde{\gamma}(N)}{\tilde{\alpha}(N)}} \tag{34}$$

$$\Phi_U(N) = \frac{1}{2}\left(\frac{\sqrt{(\rho_H + \rho_I)((N-2)^2\rho_H + N^2\rho_I)}}{\rho_H + \rho_I} - (N-2) \right) > 0, \tag{35}$$

with $\tilde{\Sigma} \equiv Var[R_i|\mu_i(1)]$, *and*

$$\tilde{\alpha}(N) = A^2\,\Sigma^2\left(\Sigma-\tilde{\Sigma}\right)\left(\Sigma-(N-1)^2\,\tilde{\Sigma}_1\right) \geq 0$$
$$\tilde{\beta}(N) = -(N-1)N\rho_I^2\Big(\Sigma^2\Sigma_\nu - \Sigma\Big((N-1)N\,\sigma_\epsilon + \big(2+(N-2)N\big)\Sigma_\nu\Big)\,\tilde{\Sigma}_1 + (N-1)^2\,\Sigma_\nu\,\tilde{\Sigma}^2\Big),$$
$$\tilde{\gamma}(N) = 2N^2\rho_I\Sigma\left(\Sigma+(N-1)\tilde{\Sigma}\right)^2 \geq 0.$$

(ii) All traders exert effort, i.e., $s_i = 0$.

A trader with signal $\mu_i(s_i)$ *submits demand*

$$a^*(p,\mu_i(s_i)) = \rho_I(\phi^* Var[R_i|\mu_i(s_i)])^{-1}(\mathbb{E}[R_i \mid \mu_i(s_i)]-p). \tag{36}$$

When traders behave as price-takers despite having price impact, dealers anticipate this misperception and adjust their contracts accordingly—both when the incentive constraint binds and when it does not. However, because the contract is not auction–specific and, in particular, does not depend on the realized clearing prices, dealers cannot correct the trader's misperception. This is evident from the fact that the unconstrained contract $\Phi_U(N)$ depends on the number of traders and converges to the second–best benchmark $\frac{\rho_I}{\rho_I+\rho_H}$ only as $N \to \infty$, when traders truly have no price impact and thus no longer make a mistake.

A.2 When is imperfect competition worse than moral hazard?

In this Appendix examine which friction—when considered "in isolation"—has a larger impact on welfare and identify the conditions under which one dominates the other. Specifically, to assess the welfare loss from imperfect competition, which arises due to demand reduction, I remove frictions from moral hazard (by setting $\phi = \frac{\rho_I}{\rho_I+\rho_H}$). I compare the expected welfare when traders are price-takers (i.e., do not reduce their demand strategically) to the corresponding welfare when traders act strategically and reduce their demand (as in Proposition 1).

To analyze the effect of moral hazard, I, instead, remove the loss due to strategic demand reduction. I assume traders act as price-takers and compute the welfare loss from moral hazard as the difference between the expected welfare that would arise under unconstrained contracting—that is, with $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$ for all traders—and the welfare when moral hazard constraints bind for all traders, $\phi^* = \Phi(N)$. Note that this only removes the welfare loss that arises due to strategic demand reduction (i.e., the fact

that strategic traders internalize their own price impact when the market is imperfectly competitive), but it does not shut off the effect that a change in the number of traders has no the contract, and with that the actual price impact their own demand has on the price.

Proposition 4. *Consider the scenario where moral hazard constraints bind,* $\frac{\rho_I}{\rho_H+\rho_I} < \Phi(N)$.

The expected allocative welfare from imperfect competition is, for any Σ, A *and distribution of* v_i*, smaller than the expected allocative welfare loss from moral hazard if*

$$f(\phi) \leq N \text{ , where } f(\phi) = 1 + \frac{\phi\sqrt{\rho_H(\rho_H+\rho_I)}}{-\rho_I + \phi(\rho_H+\rho_I)} \text{ at } \phi = \Phi(N). \tag{37}$$

At $\phi = \Phi(N)$ *this expression rearranges to:*

$$\Phi(N) \geq \frac{\rho_I(N-1)}{(N-1)(\rho_H+\rho_I) - \sqrt{\rho_H(\rho_H+\rho_I)}} \geq \frac{\rho_I}{\rho_I+\rho_H}. \tag{38}$$

To build intuition for the mechanisms at work, let us temporarily shut down the indirect channel through which a change in N affects trading via its impact on the contract. Specifically, assume that ϕ is constant in N. While this does not hold in equilibrium, it allows us to separate the direct effect of changing the number of traders from the indirect effect operating through contract adjustments.

The corollary identifies the minimal number of traders necessary for imperfect competition to distort asset allocations less than moral hazard for a given contract ϕ.[23] As moral hazard becomes less sever (since b declines which decreases ϕ), $f(\phi)$ increases. This means the number of traders such that the loss from imperfect competition is smaller than the loss from moral hazard increases. When moral hazard disappears ($\phi \to \frac{\rho_I}{\rho_I+\rho_H}$), there is no finite number of traders such that that imperfect competition is less welfare harming than moral hazard.

When we account for the fact that the contract changes in the number of traders (as specified in Corollary 3), the relationship between moral hazard and the degree of competition becomes more complicated. The loss from imperfect competition is now worse than moral hazard when $\phi(N)$ is sufficiently high, as shown in expression (38).

As before, there are two cases. When $\Phi(N)$ is decreasing in N, the LHS of condition (38) decreases, while the RHS increases, so that this expression is violated for sufficiently high N: it is not the case that the loss from imperfect competition is smaller than from

[23] The relationship is visualized in Appendix Figure A5 for a simplified environment in which traders and clients have the same risk tolerance ($\rho_I = \rho_H$).

moral hazard. Intuitively, moral hazard frictions become worse as the number of traders increase, while the direct frictions from competition (due to demand reduction) decline—formally, price impact declines. When $\Phi(N)$ increases in N, condition (38) can be satisfied for sufficiently high N. In fact, we know from Corollary 2, that in the limit, when $b > \gamma(\sigma_\epsilon)$, the condition must hold, implying that the loss from imperfect competition is smaller than the loss from moral hazard.

Proposition 4 offers a tool for evaluating whether policy should primarily target imperfect competition or moral hazard, while recognizing that the two frictions interact. For instance, enhancing competition could involve lowering exchange membership fees to attract additional traders. Addressing moral hazard, by contrast, would require redesigning the contractual relationship between dealer institutions and traders—such as by structuring compensation to more directly reward effort that improves returns.

A.3 Proofs and auxiliary statements

Lemma 1. *Given the information structure, returns are normally distributed, conditional on the trader's signal, $\mu_i(s_i)$:*

$$R_i|\mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i)], Var[R_i \mid \mu_i(s_i)]) \tag{39}$$

$$\textit{where } Var[R_i \mid \mu_i(s_i)] = \begin{cases} \Sigma + \left(\Sigma_\nu^{-1} + \sigma_\epsilon^{-1}\right)^{-1} & \textit{if } s_i = 1 \\ \Sigma & \textit{if } s_i = 0. \end{cases} \tag{40}$$

and unconditionally: $\boldsymbol{R_i} \sim \mathcal{N}(\nu, \Sigma + \Sigma_\nu)$.

Proof of Lemma 1. Because $\nu_i \sim \mathcal{N}(\nu, \Sigma_\nu)$ and $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, the marginal distribution of R_i is given by: $R_i \sim \mathcal{N}(\nu, \Sigma_\nu + \Sigma)$, because $\mathbb{E}[R_i] = \mathbb{E}[\mathbb{E}[R_i|\nu_i]] = \mathbb{E}[\nu_i] = \nu$ and $Var(R_i) = \mathbb{E}[Var(R_i|\nu_i)] + Var(\mathbb{E}[R_i|\nu_i]) = \Sigma + \Sigma_\nu$. We now derive the conditional distribution of $R_i \mid \mu_i$. Using the law of total variance: $Var(R_i \mid \mu_i) = \mathbb{E}[Var(R_i \mid \nu_i) \mid \mu_i] + Var(\mathbb{E}[R_i \mid \nu_i] \mid \mu_i)$. Since $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, we have: $Var(R_i \mid \mu_i) = \Sigma + Var(\nu_i \mid \mu_i)$. To compute $Var(\nu_i \mid \mu_i)$, observe that this is a standard Bayesian updating problem with Gaussian prior and linear Gaussian signal. The prior is $\nu_i \sim \mathcal{N}(\nu, \Sigma_\nu)$, and the signal model is: $\mu_i = \nu_i + 1\epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon(s_i))$. Using the formula for the posterior covariance in Gaussian models with rank-one noise, we obtain: $Var(\nu_i \mid \mu_i) =$

$\left(\Sigma_\nu^{-1} + \frac{1}{\sigma_\epsilon(s_i)}\right)^{-1}$ when $s_i = 1$. Further,

$$\mathbb{E}[\nu_i \mid \mu_i] = \nu + \Sigma_\nu \left(\frac{1}{\sigma_\epsilon} + \Sigma_\nu\right)^{-1} (\mu_i - \nu) \text{ when } s_i = 1.$$

Therefore, the full conditional distribution of $R_i \mid \mu_i$ is:

$$R_i \mid \mu_i \sim \mathcal{N}\left(\mathbb{E}[\nu_i \mid \mu_i],\ \Sigma + \left(\Sigma_\nu^{-1} + \frac{1}{\sigma_\epsilon(1)}\right)^{-1}\right) \text{ when } s_i = 1$$

When $\sigma_\epsilon(s_i) = 0$, the signal is perfect and $\mu_i = \nu_i$. In that case, the posterior variance collapses to zero and: $R_i \mid (\mu_i = \nu_i) \sim \mathcal{N}(\nu_i, \Sigma)$, so the conditional variance simplifies to $Var(R_i \mid \mu_i) = \Sigma$, as in the classical setting. □

Proof of Proposition 1. I follow a guess-and verify strategy, and solve the game backwards.

Period 3. Starting in the last period, I consider the choice of trader i and assume all other traders play the equilibrium strategy. The trader observes signal $\mu_i(s_i)$, depending on their previous choice of effort, and chooses their demand schedule. Their objective is to maximize the expected utility from the money they will receive at the end of the game, which is:

$$\omega^c(s_i; \phi_i) = \phi_i \Big\{ a_i(p, \mu_i(s_i))(R_i - p) \Big\} + s_i b + K_i. \tag{41}$$

Fo this, the trader chooses quantities $a_i(p, \mu_i(s_i))$ for $s_i \in \{0, 1\}$, point-wise for each clearing price p to maximize their expected payoff from purchasing assets:

$$\mathbb{E}\left[1 - \exp\left(-\frac{1}{\rho_I}\omega^c(s_i; \phi_i)\right) \middle| \mu_i(s_i)\right]. \tag{4}$$

Since $R_i | \mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i)], Var[R_i \mid \mu_i(s_i)])$, point-wise maximizing (4) is equivalent to standard mean-variance optimization:

$$\phi_i(\mathbb{E}[R_i \mid \mu_i(s_i)] - p)a_i - \frac{1}{2}\frac{\phi_i^2}{\rho_I} a_i Var[R_i \mid \mu_i(s_i)] a_i + s_i b + K_i, \tag{42}$$

where $a_i = a_i(p, \mu_i(s_i))$. The necessary condition for quantity a_i at the fixed market clear-

ing price p and both $s_i \in \{0,1\}$ is

$$\mathbb{E}[R_i \mid \mu_i(s_i)] - \phi_i \frac{1}{\rho_I} Var[R_i \mid \mu_i(s_i)] a_i = p + \Lambda_i a_i, \text{ where } \Lambda_i = \left(\frac{dp}{da_i}\right). \tag{17}$$

Λ_i measures the price impact of the trader's own demand on the clearing price. It is the amount by which the market clearing price moves when trader i demands marginally more than quantity a_i. Since all other traders j but i submit the equilibrium guess (15), price impact is $\Lambda_i = \frac{\phi}{\rho_I(N-2)}\Sigma,$. Moreover, the necessary conditions are sufficient, since $Var[R_i \mid \mu_i(s_i)]$ is positive semi-definite, $N > 2$ and $\phi_i > 0$.

Therefore, equation (17) pins down the trader's best reply $a_i(p, \mu_i(s_i))$ for $s_i \in \{0,1\}$. When trader i submits this demand curve and everyone else the equilibrium demand, the market clears at

$$p^c(s_i) : A = a_i^c(s_i) + \sum_{j \neq i} a^*(p^c, \mu_j(0)),$$

where $a_i^c(s_i)$ abbreviates the clearing quantity, $a_i(p^c, \mu_i(s_i))$ and $a^*(p^c, \mu_j(0))$ is the equilibrium candidate (15).

Period 2. Before trading and observing their private signal, the dealer chooses their effort level in period 2. They exert effort when

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] + K_i \geq \mathbb{E}[\mathcal{U}(1;\phi_i)] + K_i, \tag{43}$$

where

$$\mathcal{U}(s_i;\phi_i) = \phi_i(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i))a_i^c(s_i) - \frac{1}{2}\frac{\phi_i^2}{\rho_I} a_i^c(s_i) Var[R_i \mid \mu_i(s_i)] a_i^c(s_i) + bs_i. \tag{8}$$

Period 1. In the first period client i choose contracts so as to maximize the expected certainty equivalent $\mathbb{E}[\mathcal{W}(\phi_i; s_i)] - K_i$, with

$$\begin{aligned} &\mathcal{W}(\phi_i;s_i) - K_i, \\ \text{where } \mathcal{W}(\phi_i;s_i) =& (1-\phi_i)\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)a_i^c(s_i) \\ &- (1-\phi_i)^2 \frac{1}{2\rho_H} a_i^c(s_i) Var[R_i \mid \mu_i(s_i)] a_i^c(s_i) - \Delta s_i. \end{aligned} \tag{10}$$

The client faces two constraints:

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] \geq \mathbb{E}[\mathcal{U}(1;\phi_i)] \tag{IC}$$

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] + K_i \geq \bar{U} \tag{IR}$$

Since the incentive constraint is independent of K_i, we can choose K_i to ensure participation:

$$K_i = K(\phi_i;0) = \bar{U} - \mathbb{E}[\mathcal{U}(0;\phi_i)]. \tag{IR }$$

Note that $K(\phi_i;0)$ depends on ϕ_i, because ϕ_i affects the trader's demand function in the trading game. Inserting the binding participation constraint (IR) into the client's objective function (16), the optimization problem simplifies to:

$$\max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;0] - K_i(\phi_i;0)\right\} = \max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;0] + \mathbb{E}[\mathcal{U}(0;\phi_i] - \bar{U}\right\} \tag{16}$$

subject to (IC), assuming Δ is sufficiently high that the client always wants to incentivize effort.

To preview, there are two possible solutions for ϕ_i: an unconstrained contract, $\phi_i = \frac{\rho_I}{\rho_I+\rho_H}$, and a constrained contract, $\phi_i = \Phi(N)$. The unconstrained contract maximizes $\mathbb{E}[\mathcal{W}(\phi_i;0)] + \mathbb{E}[\mathcal{U}(0;\phi_i)] - \bar{U}$. The constrained contract is pinned down by the binding incentive constraint: $\mathbb{E}[\mathcal{U}(0;\phi_i)] = \mathbb{E}[\mathcal{U}(1;\phi_i)]$. The solution is unconstrained when $\Phi(N) \leq \frac{\rho_I}{\rho_I+\rho_H}$ and constrained otherwise.

To see why this is the case, consider first the constrained scenario. The constrained contract is pined down by the binding incentive constraint, defined below, in expressions (44) and (45), respectively. Since all clients much choose the same contract in the symmetric equilibrium, we obtain the unique positive solution, $\Phi(N)$, by solving $\mathbb{E}[\mathcal{U}(0;\phi)] = \mathbb{E}[\mathcal{U}(1;\phi)]$ for ϕ.

Consider now the unconstrained scenario, which arises when $\Phi(N) \leq \frac{\rho_I}{\rho_I+\rho_H}$. First note that $\mathbb{E}[\mathcal{W}(\phi_i;0] + \mathbb{E}[\mathcal{U}(0;\phi_i] - \bar{U}$ is maximized at $\phi_i = \frac{\rho_I}{\rho_I+\rho_H}$, independently of other clients' contracts. Since clients are ex-ante identical, we know that other clients must be choosing the unconstrained contract as well if this is optimal for one client. Next, I verify that when the solution is unconstrained, the trader exerts effort, i.e., the incentive constraint is satisfied. To check this, compute the trader's expected certainty equivalents for an arbitrary ϕ_i for some given ϕ, relying on market clearing and equilibrium trading

behavior in period 3:

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] = \phi_i \frac{A^2(-1+N)\phi^2\Sigma^2 + (-2+N)^2 N\rho_I^2\Sigma_v}{2(-2+N)(-1+N)(2\phi+(-2+N)\phi_i)\rho_I\Sigma}, \tag{44}$$

$$\mathbb{E}[\mathcal{U}(1;\phi_i)] = b + \phi_i \frac{A^2(-1+N)\phi^2\Sigma^2 + (-2+N)^2\rho_I^2((-1+N)\sigma_\epsilon + N\Sigma_v)}{2(-2+N)(-1+N)(2\phi\Sigma+(-2+N)\phi_i\tilde{\Sigma})\rho_I}. \tag{45}$$

To ensure that the client has no incentive to deviate to shirking when the contract is unconstraint, $\phi_i = \frac{\rho_I}{\rho_H+\rho_I}$, I need to show that $\mathbb{E}[\mathcal{U}(0;\phi_i)] \geq \mathbb{E}[\mathcal{U}(1;\phi_i)]$ when all clients choose the unconstrained contract: $\phi_i = \phi = \frac{\rho_I}{\rho_H+\rho_I}$. To show this, let's first impose $\phi_i = \phi$ for some ϕ and simplify the incentive constraint:

$$\mathbb{E}[\mathcal{U}(0;\phi)] \geq \mathbb{E}[\mathcal{U}(1;\phi)] \Leftrightarrow b \leq \mathcal{C}(\phi) = \frac{A^2\phi^2\alpha(N) - \beta(N)}{\gamma(N)},$$

given $\alpha(N), \beta(N), \gamma(N)$ as defined in the proposition. When $b \leq \mathcal{C}\left(\frac{\rho_I}{\rho_I+\rho_H}\right)$ the incentive constraint is fulfilled. By construction, we know that $\mathcal{C}(\Phi(N)) = b$. Since we consider the case where $\Phi(N) \leq \frac{\rho_I}{\rho_I+\rho_H}$, and $\mathcal{C}'(\phi) \geq 0$, the incentive constraint is satisfied.

Finally, inserting the optimal ϕ^* into the expression for the fixed part of the contract (IR) completes the proof. □

How large much Δ be to rule out deviations to shirking. A client who is allowed to deviate to induce their own trader to shirk in responds to all other clients choosing to induce effort would choose the following contract:

$$\max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;1)] - K_i(\phi_i;1)\right\} = \max_{\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;1] + \mathbb{E}[\mathcal{U}(1;\phi_i] - \bar{U}\right\} \tag{46}$$

$$\text{subject to } \mathbb{E}[\mathcal{U}(1;\phi_i] \geq \mathbb{E}[\mathcal{U}(0;\phi_i] \tag{47}$$

In order to rule out this deviation, I must show that the optimal contract allowing for shirking, generates lower expected payoffs. There are two cases. Consider first the case where the incentive constraint, now incentivizing shirking, binds. Then the client's best reply to other clients choosing ϕ solves $\mathbb{E}[\mathcal{U}(1;\phi_i^S(\phi)] = \mathbb{E}[\mathcal{U}(0;\phi_i^S(\phi)]$ given in expres-

sions (44) and (45):

$$\begin{aligned}\phi_i^S(\phi) =& \Big(8b(-1+N)\phi^2\rho_I\Sigma^2\Big) \\ & \Big[-(-2+N)(-1+N)\phi\rho_I\Sigma\Big(\rho_I\sigma_\epsilon + 2b(\Sigma+\tilde{\Sigma}_1)\Big) \\ & \pm\Big\{\phi^2\rho_I\Sigma^2\Big((2-3N+N^2)^2\rho_I\Big(\rho_I\sigma_\epsilon+2b(\Sigma+\tilde{\Sigma}_1)\Big)^2 \\ & -8b(-1+N)\Big(A^2(-1+N)\phi^2\Sigma^2(\Sigma-\tilde{\Sigma}_1)+(-2+N)^2\rho_I \\ & \Big((-1+N)\rho_I\Sigma\sigma_\epsilon+N\rho_I\Sigma\Sigma_\nu+2b(-1+N)\Sigma\tilde{\Sigma}_1-N\rho_I\Sigma_\nu\tilde{\Sigma}_1\Big)\Big)\Big)\Big\}^{\frac{1}{2}}\Big]^{-1}. \end{aligned} \tag{48}$$

To show that that this is not a profitable deviation, the following must hold

$$\mathbb{E}[\mathcal{W}(\phi_i^S(\phi^*);1]+\mathbb{E}[\mathcal{U}(1;\phi_i^S(\phi^*)] \leq \mathbb{E}[\mathcal{W}(\phi^*;0)]+\mathbb{E}[\mathcal{U}(0;\phi^*)]$$

for $\phi^* = \frac{\rho_I}{\rho_H+\rho_I}$ and $\phi^* = \Phi(N)$ given $\phi_i^S(\phi) \leq \frac{\rho_I}{\rho_I+\rho_H}$. Next, consider the case when the incentive constraint does not bid, which is the case when $\phi_i^S(\phi) > \frac{\rho_I}{\rho_I+\rho_H}$. In that case, the best response is unconstrained. The optimal unconstrained shirking contract, in responds to other dealers choosing any contract ϕ is:

$$\phi_i^S = \frac{\rho_I}{\rho_I+\rho_H}.$$

To rule out this deviation, the following must hold:

$$\mathbb{E}[\mathcal{W}(\phi_i^S;1]+\mathbb{E}[\mathcal{U}(1;\phi_i^S] \leq \mathbb{E}[\mathcal{W}(\phi^*;0)]+\mathbb{E}[\mathcal{U}(0;\phi^*)]$$

for $\phi^* = \frac{\rho_I}{\rho_H+\rho_I}$ and $\phi^* = \Phi(N)$ given $\phi_i^S(\phi) > \frac{\rho_I}{\rho_I+\rho_H}$.

To summarize all conditions, let's compute the expected payoff under effort given all clients choose a contract ϕ:

$$\begin{aligned} f(\phi) &= \mathbb{E}[\mathcal{W}(\phi;0)]+\mathbb{E}[\mathcal{U}(0;\phi)] \\ &= \frac{(2\phi\rho_H+(-2+N)(-\rho_I-(-2+\phi)\phi(\rho_H+\rho_I)))(A^2(-1+N)\phi^2\Sigma^2+(-2+N)^2N\rho_I^2\Sigma_\nu)}{2(-2+N)(-1+N)N^2\phi^2\rho_H\rho_I\Sigma} \end{aligned} \tag{49}$$

In comparison, the payoff when allowing shirking and choosing some ϕ_i^S in responds to all other traders choosing contract ϕ is:

$$g(\phi_i^S,\phi) = \mathbb{E}[\mathcal{W}(\phi_i^S;1)] + \mathbb{E}[\mathcal{U}(1;\phi_i^S)]$$
$$= \frac{(A^2(-1+N)\phi^2\Sigma^2 + (N-2)^2\rho_I^2((-1+N)\sigma_\epsilon + N\Sigma_\nu))(2\phi\rho_H\Sigma - (-2+N)(\rho_I + (-2+\phi_i^S)\phi_i^S(\rho_H+\rho_I))\tilde{\Sigma})}{2(-2+N)(-1+N)\rho_H\rho_I(2\phi\Sigma + (-2+N)\phi_i^S\tilde{\Sigma})^2}$$
$$+ b - \Delta. \tag{50}$$

Taken together the conditions to rule out a client's deviation to induce shirking while everyone else induces effort are:

$$g(\phi_I^S,\phi^*) \leq f(\phi^*) \text{ for } \phi_i^S = \begin{cases} \phi_I^S(\phi^*) & \text{if } \frac{\rho_I}{\rho_I+\rho_H} > \phi_I^S(\phi^*) \\ \frac{\rho_I}{\rho_I+\rho_H} & \text{if } \frac{\rho_I}{\rho_I+\rho_H} \leq \phi_I^S(\phi^*) \end{cases} \text{ at } \phi^* \in \left\{\frac{\rho_I}{\rho_I+\rho_H}, \Phi(N)\right\}. \tag{51}$$

Proof of Proposition 2a. The statement follows form the definition of allocative efficiency. I take expectations over return expectations, and rely on the simplifying assumptions that these expectations are independent across traders i. Then to show that $\alpha(\phi,N)$ and $\beta(\phi,N)$ are non-negative for $\phi^* < 1$, I rely on the assumptions that $0 < \rho_H \leq \rho_I$, and $N > 2$, and the property that $\phi^* > 0$. □

Proof of Proposition 2b. To show statement (i) we take the partial derivative of expected allocative efficiency, characterize in Corollary 2, with respect to ϕ. Given $0 < \rho_H \leq \rho_I$, $N > 2$, and $\phi^* > 0$: $\frac{\partial\alpha(\phi^*,N)\beta(\phi^*,N)}{\partial\phi} \leq 0$, $\frac{\partial\beta(\phi^*,N)}{\partial N} \geq 0$. Therefore, expected allocative efficiency always decreases in ϕ, provided that $\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] \geq \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]$ and $A\Sigma A \geq 0$, which is always the case (in a one-asset setting). Similarly, taking the partial derivative with respect to the number of traders, N, shows statement (ii). □

Proof of Proposition 3. The proof is analogous to the proof of Proposition 1. The only difference is that we now set $\Lambda_i = 0$. □

Proof of Proposition 4. To assess the welfare loss from imperfect competition, which arises due to demand reduction, I remove frictions from moral hazard (by setting $\phi = \frac{\rho_I}{\rho_I+\rho_H}$). I compare the expected welfare—separately, either including or excluding expected payment—when traders are price-takers (i.e., do not reduce their demand strate-

gically) to the corresponding welfare when traders act strategically and reduce their demand (as in Proposition 1).

To analyze the effect of moral hazard, I assume traders act as price-takers and compute the loss as the difference between the expected welfare that would arise under unconstrained contracting—that is, with $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$ for all traders—and the welfare when moral hazard constraints bind for all traders, $\phi^* = \Phi(N)$. Note that this scenario would not arise in equilibrium, since Proposition 3 shows that dealers adjust their contracts when they anticipate traders' neglect of price impact. In other words, imperfect competition spills over to contract design. Instead, the comparison is designed to isolate the full effect of moral hazard—contrasting the constrained and unconstrained contracts of Proposition 1—when shutting off frictions stemming from imperfect competition in the trading game.

Expected allocative welfare loss due to imperfect competition (when shutting off frictions arising from moral hazard) is

$$LIC = +\underline{\gamma}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \underline{\delta}A\Sigma A, \text{ with} \tag{52}$$

$$\underline{\gamma} = \frac{\rho_H+\rho_I}{2(N-1)} \geq 0 \tag{53}$$

$$\underline{\delta} = 0. \tag{54}$$

As expected, the welfare loss converges to zero when $N \to \infty$.

Expected welfare loss due to moral hazard (when shutting off frictions arising from imperfect competition) is

$$LMH(N,\phi) = \tilde{\gamma}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \tilde{\delta}A\Sigma A, \text{ with} \tag{55}$$

$$\tilde{\gamma} = \frac{(\rho_I-\phi(\rho_H+\rho_I))^2(N-1)N\rho_I(\rho_H+\rho_I)}{2N\phi^2\rho_H\rho_I(\rho_H+\rho_I)} \geq 0 \tag{56}$$

$$\tilde{\delta} = -\frac{(\rho_I-\phi(\rho_H+\rho_I))^2\phi^2}{2N\phi^2\rho_H\rho_I(\rho_H+\rho_I)} \leq 0 \tag{57}$$

where $\phi = \Phi(N) \geq \frac{\rho_I}{\rho_I+\rho_H}$; $LMH(N, \frac{\rho_I}{\rho_H+\rho_I}) = 0$ when moral hazard constraints do not bind.

Imperfect competition leads to a smaller welfare loss (in allocative efficiency) for any A, Σ and distribution of ν_i when

$$\underline{\gamma} \leq \tilde{\gamma} \text{ and } \tilde{\delta} \leq \underline{\delta},$$

given $A\Sigma A \geq 0$ and $\left(\mathbb{E}[\nu_i \Sigma^{-1} \nu_i] - \mathbb{E}[\nu_i'] \Sigma^{-1} \mathbb{E}[\nu_i]\right) \geq 0$ by assumption. Since $N > 2$, $\frac{\rho_I}{\rho_I + \rho_H} < \phi$, this is the case when $f(\phi) \leq N$ with $f(\phi)$ given in Proposition 4.

To derive the condition when ϕ depends on N, we insert $\Phi(N)$ and rearrange the condition leveraging the fact that $\Phi(N) \geq \frac{\rho_I}{\rho_H + \rho_I}$ and $N \geq 2$:

$$1 + \frac{\Phi(N)\sqrt{\rho_H(\rho_H + \rho_I)}}{-\rho_I + \Phi(N)(\rho_H + \rho_I)} \leq N \iff \Phi(N)\Big((1-N)(\rho_H + \rho_I) + \sqrt{(\rho_H + \rho_I)\rho_H}\Big) + \rho_I N \leq \rho_I$$
$$\iff \Phi(N) \geq \frac{(N-1)\rho_I}{(N-1)(\rho_H + \rho_I) + \sqrt{(\rho_H + \rho_I)\rho_H}}. \quad \square$$

Proof of Corollary 1. To derive the market clearing price in equilibrium it suffices to insert the equilibrium demand functions of all traders into the market clearing condition, $\sum_i a^*(p^*, \mu_i(0)) = A$, and solve for the price. The equilibrium allocations can then be found by inserting the clearing price into the equilibrium demand. $\square$

Proof of Corollary 2. To see this, I analyze the incentive constraint as the market becomes perfectly competitive:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(\mu_i(0); 0)] = \frac{\rho_I}{2}\frac{\Sigma_\nu}{\Sigma} \tag{58}$$
$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(\mu_i(1); 1)] = \frac{\rho_I(\sigma_\epsilon + \Sigma_\nu)}{2\tilde{\Sigma}_i(1)} + b = \frac{\rho_I(\sigma_\epsilon + \Sigma_\nu)^2}{2(\sigma_\epsilon \Sigma_\nu + \Sigma(\sigma_\epsilon + \Sigma_\nu))} + b, \tag{59}$$

From here we see that the incentive constrained cannot be fulfilled for any ϕ_i when

$$b > \frac{\rho_I}{2}\left(\frac{\Sigma_\nu}{\Sigma} - \frac{(\sigma_\epsilon + \Sigma_\nu)}{\tilde{\Sigma}_i(1)}\right) = \gamma_\infty(\sigma_\epsilon).$$

Note that the cutoff decreases in σ_ϵ when $\sigma_\epsilon > \Sigma_\nu \frac{\Sigma_\nu - \Sigma}{\Sigma + \Sigma_\nu}$. Thus as σ_ϵ grows, it eventually hits this cutoff, and then decreases the RHS of the constraint on b. $\square$

Proof of Corollary 3. $\Phi'(N) < 0$ when $(\beta(N) + b\gamma(N))\alpha'(N) > (\beta'(N) + b\gamma'(N))\alpha(N)$; else $\Phi'(N) \geq 0$. Inserting for $\alpha(N), \beta(N)$, and $\gamma(N)$ and rearranging gives the cutoff of the corollary. Further taking the partial derivative of the cutoff with respect to σ_ϵ and rearranging gives the σ_ϵ cutoff above which $\gamma'(\sigma_\epsilon) < 0$. $\square$

B Additional empirical evidence

Appendix B.1 describes the estimation procedure used to take the model to the data.
Appendix B.2 explains the robustness analysis.
Appendix Tables and Figures are presented at the end.

B.1 Estimation details

Estimating ρ_I and ϕ^*. Recall the average daily price per auction type (19"):

$$p^{el}_{tsk} = \nu_{ts} - \frac{\phi^*_k}{\rho_I}\Sigma_{ts}\left(\frac{N^{el}_{tsk}-1}{N^{el}_{tsk}-2}\right)\frac{A^{el}_{tsk}}{N^{el}_{tsk}}, \qquad (e,l)\in\{\text{TSX},\text{Alpha}\}\times\{b,q\}, \tag{19"}$$

with $\phi^*_k = 1$ for $k = IN$ and $\phi^*_k > 0$ for $k = CL$, and where p^{el}_{tsk}, N^{el}_{tsk}, $A^{el}_{tsk} = \frac{1}{\Gamma}\sum_\tau A_{t\tau sk}$, and Σ_{ts} are observed (or estimated), while (ρ_I, ϕ^*) are unknown and ν_{ts} cancels in price differences. Define

$$m^{el}_{tsk} = \Sigma_{ts}\left(\frac{N^{el}_{tsk}-1}{N^{el}_{tsk}-2}\right)\frac{A^{el}_{tsk}}{N^{el}_{tsk}}, \tag{60}$$

and denote an observation, i.e., a (t,s,e,l,k) cell by ι.

To obtain the first moment for estimation, take the difference between Alpha inventory trades and TSX inventory prices using expression (19") and rearrange for ρ_I:

$$R^{(1)}_\iota \equiv \frac{m^{\text{TSX},l}_{ts,IN} - m^{\text{Alpha},l}_{ts,IN}}{p^{\text{Alpha},l}_{ts,IN} - p^{\text{TSX},l}_{ts,IN}}. \tag{61}$$

Similarly, take the difference between TSX inventory and TSX client prices using expression (19") and rearrange for ϕ^* to obtain the second moment:

$$R^{(2)}_\iota(\rho) \equiv \frac{m^{\text{TSX},l}_{ts,IN} - \rho\left(p^{\text{TSX},l}_{ts,CL} - p^{\text{TSX},l}_{ts,IN}\right)}{m^{\text{TSX},l}_{ts,CL}}. \tag{62}$$

Let $\psi(u) = \frac{1}{2} - \mathbf{1}\{u < 0\}$ denote the median (quantile $\tau = \frac{1}{2}$) score function. The corresponding sample moment conditions are

$$\bar{g}^{(1)}_n(\rho) = \frac{1}{n}\sum_{\iota=1}^{n}\psi\big(R^{(1)}_\iota - \rho\big), \qquad \bar{g}^{(2)}_n(\phi^*;\rho) = \frac{1}{n}\sum_{\iota=1}^{n}\psi\big(R^{(2)}_\iota(\rho) - \phi^*\big). \tag{63}$$

The two parameters (ρ_I, ϕ^*) are estimated in two stages as just–identified quantile (Z–)estimators

based on these moment conditions. The first-stage estimator $\hat{\rho}_I$ solves the just–identified median moment

$$\bar{g}_n^{(1)}(\hat{\rho}_I) = 0 \Leftrightarrow \hat{\rho}_I = \text{median}\{R_\iota^{(1)}\}, \tag{64}$$

that is, the sample median of the ratios $R_\iota^{(1)}$ across all valid observations. Conditional on $\hat{\rho}_I$, the second-stage estimator $\hat{\phi}^*$ is obtained by solving

$$\bar{g}_n^{(2)}(\hat{\phi}^*; \hat{\rho}_I) = 0 \Leftrightarrow \hat{\phi}^* = \text{median}\{R_\iota^{(2)}(\hat{\rho}_I)\}, \tag{65}$$

that is, the median of the adjusted ratios $R_\iota^{(2)}(\hat{\rho}_I)$ computed using the plug–in estimate $\hat{\rho}_I$ from Step 1. Both steps correspond to minimizing an absolute–deviation objective:

$$\hat{\rho}_I = \arg\min_\rho \sum_\iota \left|R_\iota^{(1)} - \rho\right|, \qquad \hat{\phi}^* = \arg\min_\phi \sum_\iota \left|R_\iota^{(2)}(\hat{\rho}_I) - \phi\right|. \tag{66}$$

I estimate the two parameters, the trader's risk tolerance ρ_I and the contract size ϕ^*, in two stages rather than jointly. The first parameter, ρ_I, is identified from price differences between inventory–account trades on TSX and Alpha, which provide only a noisy signal of risk tolerance because the data sample is relatively small, and the differences in execution prices across exchanges are small and affected by venue-specific microstructure variation. In contrast, ϕ^* is identified from the price difference between client and inventory trades on the same exchange, which yields a much sharper signal. Estimating both parameters jointly would therefore force the contract–size estimate to absorb noise from the weakly identified first moment, reducing efficiency and stability. The two–stage approach isolates identification of ρ_I, treating it as a nuisance parameter, and then conditions on its plug–in estimate in the second stage. Under standard regularity conditions for two–step Z–estimators, this procedure remains consistent and asymptotically equivalent to the joint estimator (see Newey and McFadden, 1994).

To obtain standard errors and confidence intervals, I implement a nonparametric cluster bootstrap at the daily level. Let days $t \in \mathcal{T}$ denote the clustering units, allowing arbitrary dependence within a day but independence across days. For each bootstrap replication $b = 1, \ldots, B$:

1. Resample $|\mathcal{T}|$ days from $\mathcal{T}$ with replacement.

2. Reconstruct the bootstrap sample containing all observations belonging to the resampled days.

3. Recompute $(\hat{\rho}_I^{(b)}, \hat{\phi}^{*(b)})$ using the same two–step procedure.

The bootstrap standard errors are computed as the empirical standard deviations of the bootstrap estimates:

$$\widehat{\mathrm{SE}}(\hat{\rho}_I) = \mathrm{sd}\big(\hat{\rho}_I^{(1)}, \ldots, \hat{\rho}_I^{(B)}\big), \qquad \widehat{\mathrm{SE}}(\hat{\phi}^*) = \mathrm{sd}\big(\hat{\phi}^{*(1)}, \ldots, \hat{\phi}^{*(B)}\big), \tag{67}$$

where $\hat{\rho}_I^{(b)}$ and $\hat{\phi}^{*(b)}$ denote the estimates obtained from the b-th bootstrap replication. The corresponding normal-approximation confidence intervals reported in the tables are

$$\mathrm{CI}_{95}^{\mathrm{norm}}(\hat{\theta}) = \Big[\, \hat{\theta} - 1.96\, \widehat{\mathrm{SE}}(\hat{\theta}),\ \hat{\theta} + 1.96\, \widehat{\mathrm{SE}}(\hat{\theta}) \,\Big], \qquad \hat{\theta} \in \{\hat{\rho}_I, \hat{\phi}^*\}. \tag{68}$$

In addition, I also compute percentile-based intervals directly from the empirical distribution of the bootstrap draws:

$$\mathrm{CI}_{95}^{\mathrm{pct}}(\hat{\theta}) = \Big[\, Q_{0.025}\big(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}\big),\ Q_{0.975}\big(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}\big) \,\Big], \tag{69}$$

where $Q_\tau(\cdot)$ denotes the empirical τ–quantile. For ease of comparison, I also report the standard errors implied by these percentile intervals,

$$\widehat{\mathrm{SE}}^{\mathrm{pct}}(\hat{\theta}) = \frac{Q_{0.975}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}) - Q_{0.025}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)})}{2 \times 1.96}. \tag{70}$$

Percentile intervals are my preferred specification, as they make no symmetry or normality assumption about the sampling distribution of the estimator and are therefore more reliable when the bootstrap distribution is skewed or heavy–tailed, as is typical for ratio– and median–based estimators. By re–estimating ρ_I in each bootstrap replication, the inference procedure correctly accounts for first–stage estimation uncertainty. In practice, the percentile and normal–approximation intervals yield very similar results.

Implied lower bound on client risk tolerance. In equilibrium $\frac{\rho_I}{\rho_I + \rho_H}$ is weakly below the contract, ϕ^*. Thus the model implies a lower bound on client risk tolerance,

$$\rho_H^{\mathrm{lb}} = \rho_I\Big(\frac{1}{\phi^*} - 1\Big). \tag{71}$$

Plugging in my estimates $(\hat{\rho}_I, \hat{\phi}^*)$, I can compute the implied $\hat{\rho}_H^{\mathrm{lb}}$.

For inference, I construct the bootstrap distribution of the bound jointly with $(\hat{\rho}_I, \hat{\phi}^*)$:

in each bootstrap replication b,

$$\hat{\rho}_H^{\mathrm{lb},(b)} = \hat{\rho}_I^{(b)}\Big(\frac{1}{\hat{\phi}^{*(b)}} - 1\Big), \tag{72}$$

whenever the ratio is well defined (which is always the case in the data). The bootstrap standard error and percentile 95% confidence interval are then

$$\widehat{\mathrm{SE}}(\hat{\rho}_H^{\mathrm{lb}}) = \mathrm{sd}\big(\hat{\rho}_H^{\mathrm{lb},(1)}, \ldots, \hat{\rho}_H^{\mathrm{lb},(B)}\big), \tag{73}$$

$$\mathrm{CI}_{95}^{\mathrm{pct}}(\hat{\rho}_H^{\mathrm{lb}}) = \Big[\, Q_{0.025}\big(\hat{\rho}_H^{\mathrm{lb},(1)}, \ldots, \hat{\rho}_H^{\mathrm{lb},(B)}\big),\; Q_{0.975}\big(\hat{\rho}_H^{\mathrm{lb},(1)}, \ldots, \hat{\rho}_H^{\mathrm{lb},(B)}\big) \,\Big], \tag{74}$$

and I report the corresponding percentile–implied standard error using the same formula as for the other parameters. This procedure ensures that uncertainty about the bound reflects sampling variation in both $\hat{\rho}_I$ and $\hat{\phi}^*$.

Estimating the contract as a function of N. To estimate ϕ^* as a function of N, I again proceed in two steps. First, I estimate the trader's risk tolerance by the median ratio implied by the TSX–Alpha inventory price difference, as before. Second, for a given trader count N, define the per–observation ratio

$$r_\iota(\rho) \;=\; \frac{m_{ts,IN}^{\mathrm{TSX},l} - \rho\big(p_{ts,CL}^{\mathrm{TSX},l} - p_{ts,IN}^{\mathrm{TSX},l}\big)}{m_{ts,CL}^{\mathrm{TSX},l}}, \tag{75}$$

where $N_{ts,CL}^{\mathrm{TSX},l}$ denotes the number of active traders in the TSX client auction for (t,s,l). I estimate the contract share at that N by the within–N median

$$\hat{\phi}^*(N) \;=\; \mathrm{median}\Big\{\, r_\iota(\hat{\rho}_I) \;:\; N_{ts,CL}^{\mathrm{TSX},l} = N,\; r_\iota(\hat{\rho}_I) \text{ finite} \,\Big\}. \tag{76}$$

I report $\hat{\phi}^*(N)$ for those N with at least κ usable observations (in the code, `min_obs`).

As before, I form clusters by trading day t. For $b = 1, \ldots, B$, I resample days with replacement, rebuild the bootstrap sample, re–estimate $\hat{\rho}_I^{(b)}$ and, using it, $\hat{\phi}^{*(b)}(N)$. For each parameter, I report (i) the bootstrap standard deviation and (ii) the percentile 95% confidence interval given by the empirical 2.5% and 97.5% quantiles of the bootstrap draws. Because some N may be absent in a bootstrap resample, the number of effective draws used for $\hat{\phi}^*(N)$ can be smaller than B; the code records this count.

B.2 Robustness analysis and discussion

Robustness. I conduct several robustness checks to assess the stability of the median estimates, summarized in Appendix Table A4.

First, I vary the time interval used to collapse trades into hypothetical auctions. Because Alpha is relatively illiquid, intervals shorter than 10 minutes yield too few observations to cluster standard errors reliably. As the discretization becomes coarser, the estimated risk tolerance declines slightly, but confidence intervals overlap throughout. The estimated contract share also remains stable, ranging between 0.57 and 0.66.

Second, I consider alternative measures of return variance. The baseline uses within-day mid-quote variation for each symbol. Re-estimating parameters using weekly or monthly variances changes the level of estimated risk tolerance in a mechanical way: the inverse of dealer risk tolerance is multiplied by a larger variance term, while prices, trader counts, and net supply remain unchanged. This simply reflects that risk tolerance is measured relative to the unit of risk, which increases when variance is computed over longer horizons. Importantly, the estimated contract share remains close to 0.6 across all specifications.

Finally, I replicate the estimation for the additional years in my sample for which national best mid-price data are available. For 2022, I estimate higher risk-tolerance, because the return variance is higher compared to 2021. However, the point-estimate is noisy and the confidence intervals overlap. The contract is similar to the baseline specification.

Interpretation of ρ_I. The key source of identifying variation for traders' risk tolerance is the price difference between inventory-account trades on Alpha and TSX, where market concentration—and thus the effective number of traders and implied net supply—differs (recall Figure 1 and moment condition 61). Any unobservable factors that cause prices to differ across venues for the same number of traders N and the same aggregated trade volume A (the same side of the trade and symbol) over a short interval can therefore affect the estimated risk-tolerance parameter. A natural concern is strategic venue selection: traders may route orders across exchanges based on liquidity conditions or venue-specific microstructure rules. If inventory-account traders send harder-to-execute orders to Alpha—leading to systematically worse execution prices there—and simpler orders to TSX, then the model would attribute these price differences to higher risk tolerance and thus overstate ρ_I. The reverse pattern would bias the estimate in the opposite direction.

To illustrate this limitation, I calibrate the model using 2020 data, in which the mar-

ket was under severe distress. In that year, the model produces a negative estimate of risk tolerance—implying risk-loving dealers—because Alpha prices were not systematically worse than TSX prices for inventory accounts at comparable levels of net supply (Appendix Table A3). This breaks the empirical pattern required for the model to imply a positive ρ_I. One explanation is the COVID-19 period's exceptional market stress: traders appear to have routed a disproportionate share of order flow to TSX, the more liquid venue, and may have been willing to pay a liquidity premium for faster or more certain execution. Such shifts in venue choice can push TSX prices above Alpha prices, reversing the usual price ranking and thereby flipping the identifying moment in the model.

Interpretation of ϕ^*. For a given ρ_I the contract size is identified from the price difference between client-accounts and inventory accounts that is implied by observed auction supply and number of bidders in the client-account.

Even after conditioning very narrowly on symbol, 15-minute interval, trade side, exchange, and restricting to active trades, price differences between client and inventory accounts may still arise from execution features unrelated to the number of active traders or net-supply traded (within that cell). For example, within the same cell, client orders may require more urgent or more aggressive execution—because they are larger, have tighter completion constraints, or routinely walk further into the book—generating worse CL prices than IN prices for reasons the model interprets as a higher ϕ^*. A downward bias can arise in the opposite situation: even within a narrow cell, inventory-account trades may be executed more aggressively—for instance, if dealers are adjusting short-term positions or acting on high-frequency signals—making IN prices worse than CL prices for reasons possibly unrelated to moral hazard frictions. In that case the estimated ϕ^* would be biased downward.

To fully understand what drives the contract estimate, we would need much richer information—such as detailed limit-order-book states, the execution logic behind each trade, and ideally the contractual terms governing each dealer–client relationship—to assess the extent to which traders acted in line with or against clients' interests. In light of these data limitations, I view the calibration as a first, rough pass that can reveal potential symptoms of moral-hazard frictions, but not a substitute for a more comprehensive analysis with richer data.

Appendix Table A1: Robustness—Prices are worse on the more concentrated exchange (Alpha)—Subset of uncollapsed trade-data

	price					
	(Buy 1)	(Buy 2)	(Buy 3)	(Sell 1)	(Sell 2)	(Sell 3)
alpha	0.005***	0.005***	0.004***	-0.006***	-0.006***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
quantity	0.148***	0.124***	0.155***	-0.111***	-0.102***	-0.115***
	(0.014)	(0.011)	(0.011)	(0.017)	(0.016)	(0.019)
Symbol-date-hour-accout-type fes	Yes	Yes		Yes	Yes	
Trader fes		Yes			Yes	
Symbol-date-hour-account-type-trader fes			Yes			Yes
N	135,349,397	135,349,397	132,672,020	130,486,020	130,486,020	128,198,570
R^2	1.000	1.000	1.000	1.000	1.000	1.000

Notes: Appendix Table A1 tests whether prices are higher on Alpha than on TSX, using uncollapsed trade-level data. I restrict to active 2021 trades during continuous trading hours for inventory, client, or market maker accounts in common stocks and ETFs. Columns (1 Buy)–(3 Buy) regress prices on an Alpha indicator, trade size,and increasingly detailed fixed effects. Prices are in C$, trade size in million C$. Standard errors are clustered at the daily level.

Appendix Table A2: Robustness—Clients trade at worse prices than dealers

	price					
	(Buy 1)	(Buy 2)	(Buy 3)	(Sell 1)	(Sell 2)	(Sell 3)
account_type: CL	0.009***	0.005***	0.001	-0.013***	-0.005***	-0.008***
	(0.001)	(0.000)	(0.002)	(0.001)	(0.000)	(0.002)
quantity	0.074***	0.051***	0.057***	-0.082***	-0.041***	-0.079***
	(0.010)	(0.005)	(0.009)	(0.007)	(0.003)	(0.007)
Symbol-date-exchange fes	Yes			Yes		
Symbol-date-hour-exchange fes		Yes			Yes	
Symbol-date-trader-exchange fes			Yes			Yes
N	145,472,740	144,773,194	143,950,353	138,943,933	138,279,152	137,534,111
R^2	1.000	1.000	1.000	1.000	1.000	1.000

Notes: Appendix Table A2 is analogous to Table 2 but relies on a subset of the uncollapsed trade-level data (active trades with common stocks or ETFS during the continuous trading phase of 2021 for client and inventory accounts). In all cases, the market inventory account is the baseline. Prices are in C$, trade size in million C$. Standard errors are clustered at the daily level.

Appendix Table A3: During COVID (2020): Prices are worse on the less concentrated (and less liquid) exchange

	price			
	(Buy 1)	(Buy 2)	(Sell 1)	(Sell 2)
trading_venue_code: ALPH	-0.004*	-0.014***	0.004***	0.016***
	(0.002)	(0.002)	(0.001)	(0.002)
normalized_volume	0.000***	-0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Symbol-date-account-type fes	Yes	Yes	Yes	Yes
N	147,743,952	16,676,804	144,919,494	17,486,938
R^2	1.000	1.000	1.000	1.000

Notes: Appendix Table A3 tests whether prices are higher on Alpha than on TSX, using uncollapsed trade-level data. I restrict to active 2020 trades during continuous trading hours for inventory, client, or market maker accounts in common stocks and ETFs. Columns (Buy 1), and (Sell 1) regress prices on an Alpha indicator, trade size, date-symbol-account-type fixed effects for the buy-side, and sell-side respectively. In columns (Buy 2) and (Sell 2) I restrict attention to inventory accounts. Prices are in C$, trade size in million C$. Standard errors are clustered at the daily level.

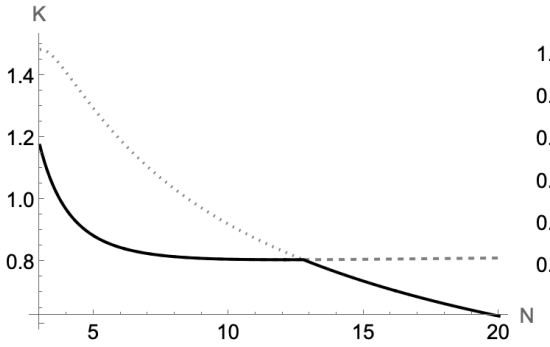
Appendix Table A4: Robustness—Parameter estimates across specifications

Parameter	Estimate	SE (SD)	SE (Pct)	95% CI Low	95% CI High
10 minutes					
ρ_I	805.528	455.642	424.044	203.300	1865.550
ϕ^*	0.666	0.0400	0.0403	0.604212	0.762141
ρ_H^{lb}	403.689	223.960	222.776	89.246	962.527
20 minutes					
ρ_I	506.907	70.6011	67.5948	389.702	654.673
ϕ^*	0.608	0.0141	0.0147	0.581047	0.638478
ρ_H^{lb}	326.405	50.0869	51.9017	230.691	434.146
30 minutes					
ρ_I	455.981	46.9629	45.9809	359.206	539.451
ϕ^*	0.577	0.0096	0.0090	0.559488	0.594898
ρ_H^{lb}	334.799	36.8342	36.3307	254.612	397.028
Weekly return variance					
ρ_I	3610.320	838.470	824.624	2064.040	5296.570
ϕ^*	0.617	0.0199	0.0196	0.582120	0.658867
ρ_H^{lb}	2243.720	538.908	523.603	1265.140	3317.670
Monthly return variance					
ρ_I	18689.300	4590.790	4488.520	11381.800	28976.800
ϕ^*	0.635	0.0218	0.0217	0.589530	0.674456
ρ_H^{lb}	10754.500	2660.210	2630.330	6340.440	16651.300
2020					
ρ_I	-347.586	113.022	115.208	-581.792	-130.178
ϕ^*	0.549	0.0140	0.0133	0.523004	0.575002
ρ_H^{lb}	-285.313	96.3273	97.9068	-492.165	-108.371
2022					
ρ_I	1223.780	306.382	272.567	670.670	1739.130
ϕ^*	0.612	0.0188	0.0187	0.578618	0.652114
ρ_H^{lb}	777.063	215.197	204.692	387.816	1190.210

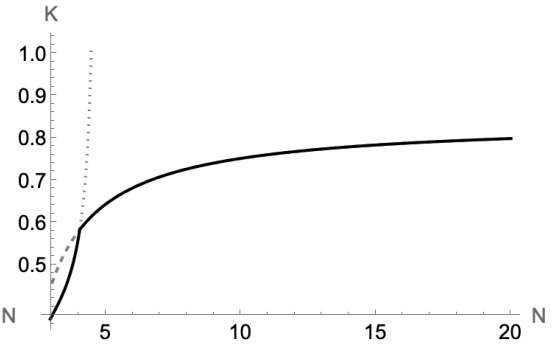
Notes: Appendix Table A4 reports robustness results for the trader risk tolerance ρ_I, the estimated contract share ϕ^*, and the implied lower bound on client risk tolerance ρ_H^{lb}. Further details on the estimation procedure are provided in Appendix B.1.

Appendix Figure A1: Examples of how the fixed part of the contract, K^*, can change in N

(a) K^*: High noise from shirking



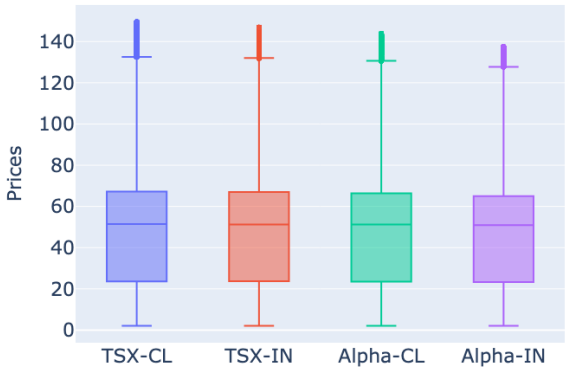
(b) K^*: Low noise from shirking



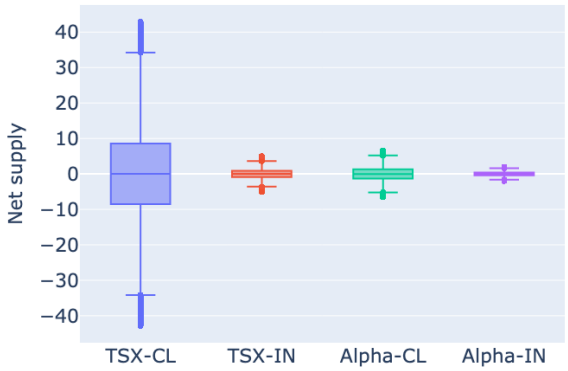
Notes: Appendix Figure 3 plot the equilibrium contract K^* (black line) as a function of the number of traders N for two numeric examples. The chosen contract is a piecewise function consisting of the unconstrained contract, $\frac{\rho_I}{\rho_I+\rho_H}$, the dashed gray line, and the constrained contract, $\Phi(N)$, the dotted gray line. In both examples, $\rho_I = 0.5, \rho_H = 0.5, \Sigma = .9, \Sigma_\nu = 1.5, b = 0.002, \bar{U} = 0.01$. In Figure A1a shirking-noise is high, $\sigma_\epsilon = 1.5$ and supply is large, $A = 5$, in Figure A1b it is low, $\sigma_\epsilon = 0.001$ and supply is small $A = 1$.

Appendix Figure A2: Distributions of prices, net supply, trader counts, and return variance

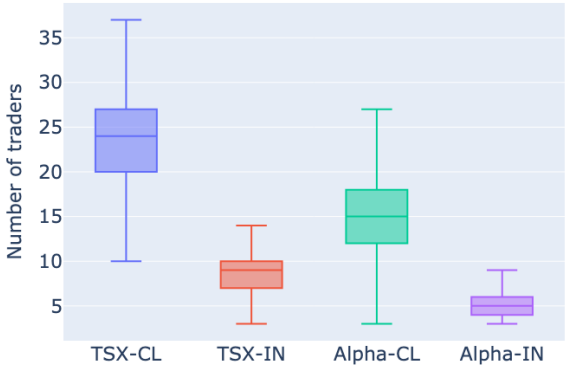
(a) Prices by auction type



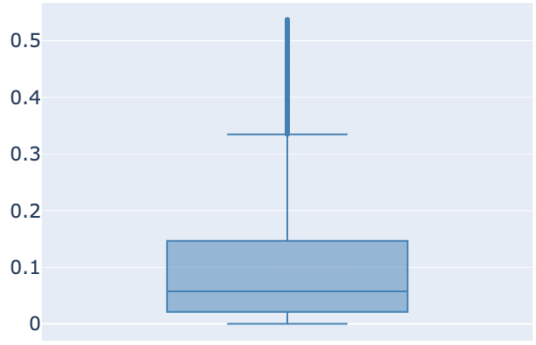
(b) Net supply by auction type



(c) Number of active traders



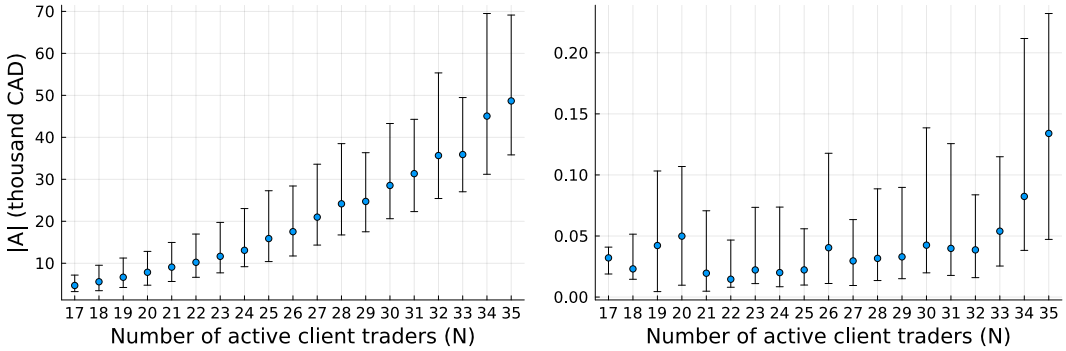
(d) Return variance, Σ_{ts}



Notes: Appendix Figure A2 shows the distribution of transaction prices (in C$), net supply (in thousand C$), the number of active traders, and our proxy for the variance of daily returns (Σ_{ts}) across all discretized "auctions" in the calibration sample. Each box plot reports TSX inventory (TSX–IN), TSX client (TSX–CL), Alpha inventory (Alpha–IN), and Alpha client (Alpha–CL) auctions separately. Boxes report the interquartile range, whiskers exclude outliers, and individual points indicate outliers.

Appendix Figure A3: Amount traded and return variance by number of traders

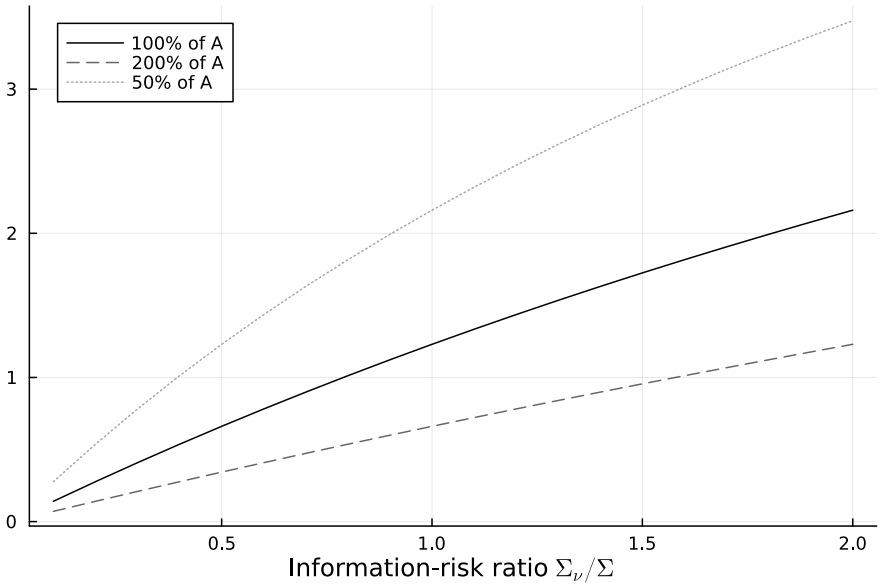
(a) Total amount traded $|A|$ (b) Return variance Σ



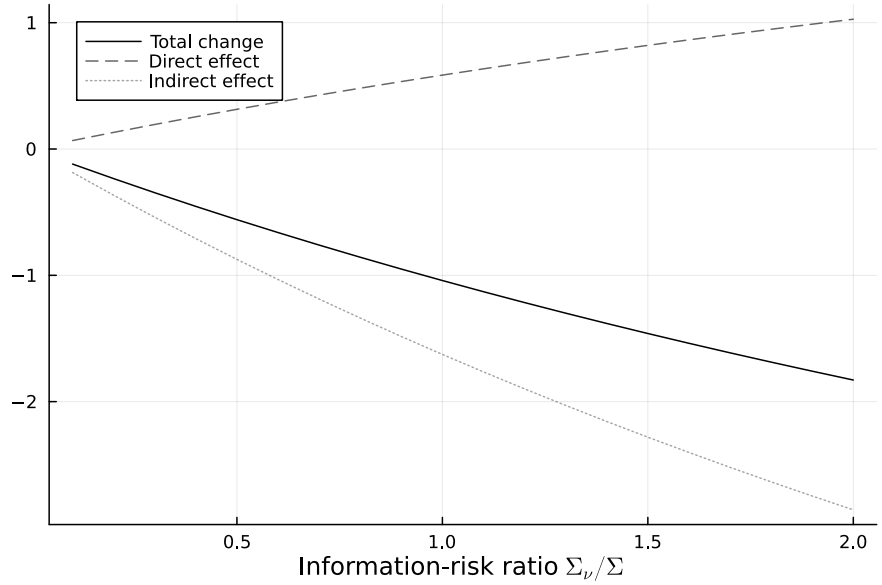
Notes: Appendix Figure A3a shows the distribution of the average daily amount traded in a client auction (either on the buy- or sell-side) on TSX per number of traders, N, active in that auction. The distribution is over days, securities, and sides of the auction. Figure A3b shows the distribution of return variances, Σ_{ts}, per N over days and securities. Supply is expressed in thousands of Canadian dollars, and return variance in Canadian dollars.

Appendix Figure A4: Welfare effects across information risk and supply

(a) Eliminating moral hazard



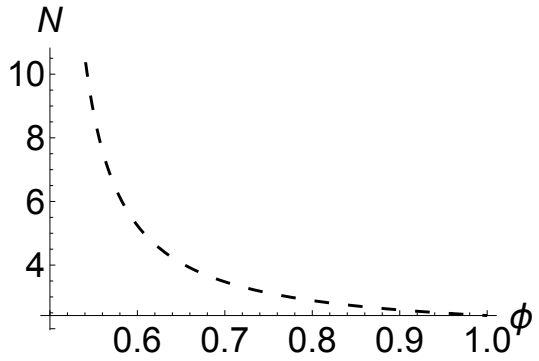
(b) Adding one trader



Notes: Appendix Figure A4a plots welfare effects—measured in percentage points relative to total welfare before relaxing the respective friction—for a representative TSX client-account auction, where clients and traders share the same risk tolerance (670), as a function of the information–risk ratio, $\kappa = \Sigma_\nu/\Sigma$. The curves correspond to different auction sizes, scaled relative to the median net supply: $A/2$, A, and $2A$. Figure A4a shows the percentage welfare gain from eliminating moral hazard. Figure A4b shows the percentage welfare loss from increasing the number of active traders from N to $N+1$, decomposed into the direct effect of adding a trader (holding the contract fixed) and the indirect effect operating through the change in the contract, $\hat{\phi}^*(N+1) - \hat{\phi}^*(N)$. Because welfare in the model is affine in the information–risk ratio, both experiments produce similarly shaped curves in κ, and the informative differences lie in their levels, signs, and how they scale with auction size and contract changes.

Appendix Figure A5: When imperfect competition is worse than moral hazard ($\rho_I = \rho_H$) assuming ϕ does not depend on N

(a) Welfare



Notes: Appendix Figure A5 shows the two cutoff function, $f(\phi)$, of Corollary 4 when $\rho_I = \rho_H$ in the dashed gray line, and the black line, respectively. The x-axis shows ϕ from $\frac{1}{2}$, when moral hazard frictions are not binding, to 1, when all assets go to the trader to incentivize effort. The loss from imperfect competition is lower than the loss from moral hazard when the number of traders is in between the gray and dashed back line.