Moral Hazard and Imperfect Competition in Financial Markets

Milena Wittwer*

March 25, 2026

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 compete with one another when trading, creating an interaction between incentives and market clearing. I develop a model to study how frictions arising from moral hazard and imperfect competition interact, and show that greater competition can reduce market efficiency by tightening intermediaries' incentive constraints. Using proprietary data from Canadian equity exchanges, I illustrate the empirical relevance of these frictions and their interaction, highlighting the need to coordinate competition policy and conduct regulation.

Keywords: Asset markets, moral hazard, imperfect competition, financial intermediation, price impact

JEL: D4, D44, D61, D8, D82, G1, G24, L1

*Correspondence to Milena Wittwer, Columbia University, NBER, and CEPR: 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 thank Pavel Zryumov for his valuable discussions; in addition to Jason Allen, Ana Babus, Eric Budish, Piotr Dworczak, William Fuchs, Simon Gervais, Sergei Glebkin, Zhiguo He, Joshua Mollner, Daniel Neuhann, Christian Opp, Cecilia Parlatore, Indira Puri, Uday Rajan, Bernard Salanie, Michael Sockin, Adi Sunderam, Pietro Tebaldi, and all participants of IO and finance seminars at Columbia University and the following conferences: 2025 Fall FTG meeting, 2025 Chicago Market Design Conference, 2025 Germans Abroad Christmas Meeting, and 2025 Workshop for Assistant Professors in Finance.

1 Introduction

Most theoretical work on moral hazard builds on the principal–agent tradition, following Ross (1973), Mirrlees (1976), Holmström (1979), and Holmström and Milgrom (1987), and studies incentives within bilateral relationships. In financial markets, intermediaries (agents) invest on behalf of clients (principals), giving rise to moral hazard concerns (He and Krishnamurthy, 2012). Yet intermediaries do not operate in isolation. When trading assets, they compete with one another in centralized markets such as exchanges (Kyle, 1989; Vayanos, 1999; Vives, 2011; Rostek and Weretka, 2012). While moral hazard and imperfect competition are each well understood in isolation, we lack a framework to understand their interaction in centrally cleared asset markets.

In this paper, I develop a model that unifies moral hazard and imperfect competition and assess its empirical relevance using data from equity exchanges. Because the two frictions interact, stronger competition can tighten incentive constraints and reduce welfare. This highlights the need for coordinated competition and conduct regulation to avoid unintended consequences.

In the model, N clients seek exposure to a risky asset with normally distributed returns but lack direct market access. Each client delegates trading to an intermediary (dealer). Clients and dealers have CARA preferences over wealth, potentially with different risk tolerances. The game unfolds over four periods. First, clients simultaneously offer contracts consisting of a lump-sum payment and a share of investment wealth. Second, dealers decide whether to accept the contract and, conditional on acceptance, whether to exert effort when acquiring information about the expected asset return relevant for their client. Effort yields a precise signal, whereas shirking yields a private benefit but a noisy signal. Third, dealers compete in a centralized market by submitting (net) demand schedules, and the market clears at a single price where aggregate demand equals supply. Fourth, returns and transactions realize.

I characterize a symmetric equilibrium in which clients offer identical contracts, dealers exert effort, and submit symmetric (linear) demand functions. Under imperfect competition, dealers internalize their price impact and strategically scale back demand to improve their terms of trade (Kyle, 1989; Ausubel et al., 2014). Moral hazard, instead, shapes contract choice: larger private benefits from shirking require allocating dealers a greater share of wealth to induce effort (Holmström and Tirole, 1997; He and Krishnamurthy, 2012). This feeds back into trading by effectively decreasing dealers' risk tolerance and the positions they are willing to take for a given expected return. When incen-

tive constraints are slack contracts coincide with the first best under perfect competition; when they bind, contracts depend on the number of dealers, and thus on competition, creating an interaction between frictions.

To clarify how these frictions interact, I study whether stronger competition, measured by the number of dealers, mitigates or exacerbates moral hazard, as reflected in the contract share dealers must receive to exert effort. The answer depends on how competition affects the relative profitability of effort versus shirking, and this depends non-trivially on the noise generated by shirking. When shirking noise is sufficiently high, contract shares decline in the number of dealers, indicating that competition mitigates moral hazard (favorable contract regime). When shirking noise is low, contract shares instead rise, implying that competition amplifies moral hazard (unfavorable contract regime).

Three forces determine which contract regime prevails: how competition affects price impact, the gap between dealers' individual return assessments and the market price, and the risk compensation they receive for absorbing the asset. The last force makes shirking relatively more attractive as competition intensifies. A shirking dealer, being less informed, trades more cautiously and holds a smaller position, so the shrinking risk compensation falls disproportionately on effort dealers who hold larger positions. To deter a deviation to shirking, clients must offer larger contract shares.

Whether increased competition through dealer entry raises or reduces welfare depends on the prevailing contract regime. When the incentive constraint is slack and contracts do not respond to entry, welfare rises for the standard reason in trading models following Kyle (1989); Vayanos (1999); Vives (2011): improved risk sharing. When contracts instead adjust with dealer entry, an additional indirect welfare channel emerges through changes in the equilibrium contract. When the contract increases with the number of dealers, this indirect effect (which reflects weaker risk sharing and less efficient use of information per unit of risk) can dominate the direct effect, leading welfare to decline as competition intensifies.

I next illustrate how the model can be used to assess the relative importance of moral hazard and imperfect competition using proprietary trade data from Canadian exchanges. The data cover roughly 60% of Canadian equity trading and come from two exchanges that list the same securities and are owned by the same company, which helps rule out many confounding factors that are otherwise difficult to control for. Moreover, the data identify whether trades are executed for intermediaries' own accounts or for their clients, which is an essential feature for detecting moral hazard.

Identifying moral hazard is challenging because incentives and contractual arrange-

ments are rarely observed directly and, when they are observable, reflect equilibrium outcomes (Einav et al., 2021). In financial markets, client–intermediary contracts are proprietary, and institution-specific, making systematic collection of contract-level data of all market participants largely infeasible.

My approach is to use trade data to infer the share of wealth that dealers must receive to exert effort. Because trading activity is observable to regulators even when contracts are private, this approach provides a feasible first pass, albeit one that relies on strong assumptions. To bring the model to the data, I introduce a small number of targeted adjustments to rationalize the observed empirical patterns at an aggregate level. For instance, I model prices as arising from frequently held batch auctions (Budish et al., 2015) for subsets of trades, such as client-account transactions on a given exchange.

I estimate two parameters that govern the severity of the two frictions in equilibrium: dealers' risk tolerance and the contract share. Average price differences for comparable proprietary trades across exchanges identify dealers' risk tolerance; while average price differences between client and proprietary trades for comparable trades identify the contract share. The estimate of dealers' risk tolerance is noisy but broadly consistent with existing evidence (Wittwer and Allen, 2026). The estimated contract share is around 60 percent, implying that intermediaries retain well over half of the wealth generated on behalf of their clients, qualitatively consistent with evidence from the literature on delegated investment management (Lan et al., 2013; Berk and van Binsbergen, 2015).

To analyze the model's central implication, that moral hazard interacts with imperfect competition, I estimate contract shares separately across days with varying levels of dealer participation. I find that the dealer's wealth share increases with the number of active dealers. While this pattern may partly reflect time-varying market conditions or other confounding factors that trade-level data alone cannot fully address, it is consistent with the model's unfavorable-contract regime, in which greater competition tightens moral-hazard frictions. In this regime, increased competition can reduce welfare. Empirically, when clients are less risk-tolerant than dealers, adding an additional dealer reduces welfare by more than eliminating moral hazard altogether would increase it.

Taken together, my findings imply that the efficiency of asset markets depends jointly on the severity of incentive frictions between intermediaries and their clients and on the degree of market competition. Once we account for market clearing, incentive contracts endogenously adjust to the competitive environment, generating an interaction between competition and moral hazard. This mechanism is not tied to a particular modeling choice and is likely to extend across alternative contract design, preference specifications,

and market-clearing frameworks.

For policymakers, the key implication is that pro-competition measures may have unintended effects in tightening incentive constraints and ultimately reducing welfare. Examples include policies, such as lowering entry barriers or expanding platform access, as well as capital and leverage rules that may affect dealer competition and incentives. Effective regulation therefore requires a coordinated approach that jointly considers market structure and incentive provision, for instance by combining competition policy with rules on compensation design, disclosure of intermediary incentives, or targeted oversight of client–dealer contracting practices.

Literature. My main 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 Holmström and Milgrom (1987). In my setting, the clients are the principals and the dealers the agents. In addition, because dealers 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.

I instead model strategic interactions among dealers in the spirit of Wilson (1979) and Kyle (1989); see Rostek and Yoon (2025) for a literature overview. This allows me to study how imperfect competition and moral hazard interact and jointly shape equilibrium contracts and welfare. This also distinguishes my work from Kyle et al. (2011), who model one informed agent trading against perfectly competitive market makers, and from Baldauf et al. (2024), who analyze contracting under moral hazard with exogenous price

impact.[1]

The paper also relates to the banking and corporate finance literature that studies the interaction between imperfect competition and moral hazard. In banking, greater competition is often argued to compress margins and weaken incentives, attracting riskier borrowers and thereby undermining financial stability (Keeley, 1990; Crawford et al., 2018).[2] In corporate finance, product-market competition shapes managerial effort and incentive design, with competition either mitigating or exacerbating moral hazard depending on its effect on rents and performance sensitivity (Schmidt, 1997; Scharfstein, 1988; Raith, 2003).

My central welfare result—that competition can either mitigate or exacerbate moral hazard in a non-monotonic way—parallels existing findings but arises through different mechanisms. Whereas prior models capture competition through Bertrand or Cournot interactions over prices or quantities, dealers in my model compete through demand (or supply) functions, a formulation better suited to trading environments with endogenous price formation. As a result, welfare losses arise from distortions in dealers' information–risk trade-off when acting on private information—a channel absent from the existing literature.

2 Model with imperfect competition and moral hazard

The goal is to understand how moral hazard and imperfect competition interact to shape welfare in intermediated asset markets. To this end, I borrow canonical features from separate strands of the literature: imperfect competition in asset markets (Kyle, 1989; Rostek and Yoon, 2025), and moral hazard (Ross, 1973; Holmström and Milgrom, 1987) in intermediary asset pricing (He and Krishnamurthy, 2012). Moral hazard here arises because dealers must exert costly effort to acquire information about asset returns, linking the model to the literature on costly information acquisition (Stigler, 1961; Grossman and Stiglitz, 1980).

[1] 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 (2026).

[2] Building on Stiglitz and Weiss (1981), subsequent work shows that this relationship is not monotonic: competition can both discipline and destabilize banks, depending on how it affects lending incentives (Boyd and Nicolo, 2005; Martinez-Miera and Rapullo, 2010).

2.1 Setup and equilibrium

In the baseline model, dealers compete to acquire assets on behalf of clients, consistent with the intermediary asset-pricing literature's focus on investment; extensions to trading are straightforward. I present the model first, then discuss the key tractability assumptions, and finally analyze the interaction between moral hazard and imperfect competition. Proofs are in Appendix A.6.

Setting. $N > 2$ clients seek to invest in one risky asset of fixed supply $A > 0$, but have no direct access to the asset market. Therefore, each hires a dealer to invest on their behalf.[3] Equivalently, I could assume a continuum of clients partitioned into N groups, each collectively hiring a dealer.

The game unfolds over four periods. In period 1, each client i offers a dealer i a two-part contract (Holmström and Milgrom, 1987; He and Krishnamurthy, 2012). It consists of a fixed amount $K_i \geq 0$, and a fraction $\phi_i \geq 0$ of wealth ω_i that will be generated from their investment at the end of the game. Thus, when accepting the job the dealer's total compensation is $K_i + \phi_i \omega_i$; when declining it, the dealer receives a reservation utility of $\overline{U}$. Wealth results from purchasing amount a_i at price p of the asset, which generates a per-unit gross return of R_i:

$$\omega_i = a_i(R_i - p). \tag{1}$$

Returns are normally distributed, $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, with known variance $\Sigma > 0$. The mean ν_i captures individual beliefs, such as market forecasts; it can also represent balance-sheet exposures, funding costs, or hedging needs that shape appetite for additional asset holdings. These means are drawn independently from a common prior, $\nu_i \sim \mathcal{N}(\nu, \Sigma_\nu)$, with $\nu \in \mathbb{R}$ and $\Sigma_\nu > 0$. Uncertainty about expected returns is thus driven by idiosyncratic factors, with any common component being absorbed by the known common prior—an assumption I relax in Appendix A.2

In period 2, dealers form beliefs about their (and their client's) expected return, ν_i. They can exert effort ($s_i = 0$), in which case they observe $\nu_i \in \mathbb{R}$, and know the conditional return distribution. Alternatively, they can shirk ($s_i = 1$) which implies that they only observe the return distribution with noise. Conditional on their effort-decision, they

[3] The assumption $N > 2$ is standard in the uniform-price auction literature, as linear equilibria do not exist for $N \leq 2$ (Rostek and Yoon, 2025).

see the following private return signal in period 3:

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

where $\sigma_\epsilon > 0$. Unlike in a standard principal-agent setting exerting effort does not create larger output, but rather changes the information available to the dealer in the subsequent trading stage (more similar to the information acquisition literature).

Shirking generates a private benefit to the dealer, $b > 0$, and a (potential) loss for the client $\Delta \geq 0$. In practice, the benefit from shirking may reflect the value of leisure from working less, while the client's loss reflects reduced attention to non-modeled responsibilities, such as reporting, communication, or compliance duties, which can impose administrative or regulatory costs on the client.

Given their private signal, dealers compete against one another to purchase shares of the asset. 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 dealer 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 signal (and their effort-decision). The submitted demand functions are aggregated, and the market clears at prices where aggregate demand equals supply.[4] From the perspective of dealer i who takes choices of all other dealers as given:

$$p^c(s_i) : \sum_{i=1}^{N} a_i(p^c(s_i), \mu_i(s_i)) = A. \tag{2}$$

For a fixed realization of the other dealers' signals, there are two possible clearing prices, depending on whether the dealer 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 dealers, not just dealer i.

Clients and dealers have CARA utility over end-of-game wealth, a standard assumption in models of market power in trade settings (Rostek and Yoon, 2025). Risk tolerances

[4] In case of excess demand—which never happens in equilibrium—dealers are rationed pro-rata on the margin.

may differ between clients and dealers ($\rho_I > 0$, $\rho_H > 0$), and the special case $\rho_H \to \infty$ nests the common assumption in the principal-agent literature that the principal is risk-neutral.

Specifically, a dealer who chooses their demand function expects to obtain the following utility when choosing demand $a^c(s_i) = a_i(p, \mu_i(s_i))$ at clearing price $p = p^c(s_i)$:

$$\mathbb{E}\left[1 - \exp\left(-\frac{1}{\rho_I}\omega_I^c(s_i;\phi_i)\right)\middle|\mu_i(s_i)\right], \tag{3}$$

$$\text{with } \omega_I^c(s_i;\phi_i) = \phi_i\Big\{a^c(s_i)(R_i - p^c(s_i))\Big\} + s_i b + K_i. \tag{4}$$

Given that all random variables are normally distributed, asset returns conditional on return signals are also normally distributed: $R_i|\mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i)], Var[R_i \mid \mu_i(s_i)])$. Under normal return distributions, maximizing (3) is equivalent to maximizing the certainty equivalent $\mathcal{U}(s_i;\phi_i)$ point-wise at every clearing price $p^c(s_i)$:

$$\mathcal{U}(s_i;\phi_i) + K_i\ , \tag{5}$$

$$\begin{aligned}\text{where } \mathcal{U}(s_i;\phi_i) = {}& \phi_i\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)a_i^c(s_i) \\ & - \phi_i^2\frac{1}{2\rho_I}a_i^c(s_i)Var[R_i \mid \mu_i(s_i)]a_i^c(s_i) + s_i b.\end{aligned} \tag{6}$$

When the dealer chooses their effort-level, s_i, they do not know their private signal, $\mu_i(s_i)$. I assume that they take an expectation, integrating out the uncertainty over their signal.

Similarly, when choosing the contract, each client maximizes their expected certainty equivalent from their end-game wealth,

$$\omega_H^c(s_i;\phi_i) = (1-\phi_i)\Big\{a^c(s_i)(R_i - p^c(s_i))\Big\} - s_i\Delta - K_i, \tag{7}$$

integrating out the uncertainty about the dealer's signal. The client's certainty equivalent (conditional on the unobserved signal) is defined analogously to the dealer's certainty equivalent with four differences: the client's risk tolerance is ρ_H rather than ρ_I, the client obtains fraction $1-\phi_i$ of wealth generated by the asset rather than ϕ_i, and suffers a loss, $\Delta \geq 0$, when the dealer shirks, rather than a benefit, $b > 0$:

$$\mathcal{W}(\phi_i;s_i) - K_i, \tag{8}$$

$$\begin{aligned}\text{where } \mathcal{W}(\phi_i;s_i) = {}& (1-\phi_i)\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)a_i^c(s_i) \\ & - (1-\phi_i)^2\frac{1}{2\rho_H}a_i^c(s_i)Var[R_i \mid \mu_i(s_i)]a_i^c(s_i) - \Delta s_i.\end{aligned} \tag{9}$$

At the end of the game, clients observe their realized wealth $\omega_H^c(s_i;\phi_i)$, and in case their dealer shirked, their loss from shirking Δ. They don't observe the market outcome, $\{a^c(s_i), p^c(s_i)\}$, or the asset's realized return, R_i. This implies that they cannot infer their dealer's signal, or effort-decision, and thus aren't able to write effort-contingent contracts.[5] This assumption is in the spirit of He and Krishnamurthy (2012), who argue that clients lacking the expertise to participate directly in financial markets are unlikely to understand or monitor their intermediary's trading decisions or success. In the case of risk-neutral clients who do not suffer a loss from shirking ($\rho_H \to \infty$, $\Delta = 0$), clients could in principle observe $\{a^c(s_i), p^c(s_i)\}$ and would still be unable to write an effort-contingent contract, since the clearing price and allocation aggregates the signals of all dealers (for example, as in Corollary 1 below) so that the individual dealer's effort choice remains unidentified.

Summarizing, the sequence of events is as follows:

(1) Clients choose their contract, $\{K_i, \phi_i\}$, simultaneously and independently. Each dealer accepts their contract or walks away, in which case the game ends.

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

(3) Each dealer 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 auction clears, the asset pays its return, dealers observe the market outcome, clients observe their end-of-game wealth, and all transactions are settled.

While I introduced the model in a buy-side setting, it is straightforward and common in the finance literature to modify the model to allow for selling or trading. One can simply allow the exogenous net supply to take any real value, $A \in \mathbb{R}$, and interpret dealers' submitted schedules as net-demand curves, where positive quantities correspond to purchases and negative ones to sales.

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

[5] Formally, the client knows $\{\phi_i, K_i\}$ and observes $(1-\phi_i)a^c(s_i)(R_i - p^c(s_i)) - s_i\Delta - K_i$ but does not know the components $a^c(s_i)$, $p^c(s_i)$, and R_i separately. Since R_i is continuously distributed, for any observed wealth level there exist values of the unobserved components consistent with either $s_i = 0$ or $s_i = 1$, so the effort choice is not identified.

To derive an equilibrium, consider one client-dealer pair, i, and fix the behavior of all clients and dealers at the symmetric equilibrium of Proposition 1 stated below. Start in the stage 3, the trading stage, where dealer i receives signal $\mu_i(s_i)$, depending on their effort choice and the given contract, $\{K_i, \phi_i\}$, and chooses their demand schedule. They do so by maximizing certainty equivalent (6) point-wise. The necessary (and sufficient) condition for quantity a_i at the fixed market clearing price p and both $s_i \in \{0,1\}$ is

$$\mathbb{E}[R_i \mid \mu_i(s_i)] - \phi_i \frac{1}{\rho_I} Var[R_i \mid \mu_i(s_i)] a_i = p + \Lambda_i a_i, \text{ where } \Lambda_i = \left(\frac{dp}{da_i}\right). \tag{10}$$

Λ_i measures the price impact of the dealer's own demand on the clearing price. It is the amount by which the market clearing price moves when dealer i demands marginally more than quantity a_i. Since all other dealers j but i submit the equilibrium guess (17), price impact is

$$\Lambda_i = \frac{\phi^*}{\rho_I (N-2)} \Sigma. \tag{11}$$

Solving equation (10) for a_i pins down the dealer's demand function, $a_i(p, \mu_i(s_i))$ for $s_i \in \{0,1\}$.

In stage 2, the dealer chooses their effort level anticipating their demand function in the trading stage, but without yet observing the signal. The dealer exerts effort if and only if

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

Although suppressed in the notation, both sides depend on the contracts and effort choices of all other dealers and clients (held fixed at the equilibrium candidate) as well as the dealer's own trading response in stage 3, through market clearing.

In stage 1, the client anticipates how their dealer will trade and when they will exert effort. Since the client aims at incentivizing effort (to avoid the loss from shirking, Δ), they choose contract $\{\phi_i, K_i\}$ to maximize:

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

anticipating that the dealer will accept the job offer when

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

and that the dealer will exert effort when incentive constraint (IC) is fulfilled. Since

the dealer's incentive constraint is independent of the fixed part of the contract, K_i, it is optimal to set K_i as low as possible, so as to still incentivize the dealer to participate. This is the case when

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

Note that $K(\phi_i; 0)$ depends on ϕ_i, because ϕ_i affects the dealer's demand function in the trading game. Inserting the binding participation constraint (IR') into the client's objective function (12), the optimization problem simplifies to:

$$\max_{\phi_i} \left\{ \mathbb{E}[\mathcal{W}(\phi_i; 0] - K(\phi_i; 0) \right\} = \max_{\phi_i} \left\{ \mathbb{E}[\mathcal{W}(\phi_i; 0] + \mathbb{E}[\mathcal{U}(0; \phi_i] - \bar{U} \right\} \tag{12}$$

subject to (IC). There are two solutions, depending on whether the constraint binds or is slack. When it is slack, $\phi^* = \frac{\rho_I}{\rho_I + \rho_H}$, the dealer perfectly internalizes that they invest for a client, and correctly chooses demand, depending on their differing degree of risk-tolerance, i.e., maximizes risk-sharing.

When the (IC) constraint binds, which is the case when offering $\frac{\rho_I}{\rho_I + \rho_H}$ is insufficient to incentivize effort, the binding constraint pins down the client's best response to other clients choosing ϕ^*. In a symmetric equilibrium, each client i chooses the same contract as all others, i.e.,

$$\mathbb{E}[\mathcal{U}(0; \phi^*)] = \mathbb{E}[\mathcal{U}(1; \phi^*)] \Leftrightarrow b = \frac{A^2 \phi^{*2} \alpha(N) - \beta(N)}{\gamma(N)}, \tag{13}$$

where $\alpha(N) > 0, \beta(N) < 0$, and $\gamma(N) > 0$ are characterized in Proposition 1. Solving for the fixed-point yields a unique positive solution, denoted by $\Phi(N)$.

Proposition 1 summarizes all equilibrium conditions.

Proposition 1. *In a symmetric equilibrium (i) all clients offer contract* $\{K^*, \phi^*\}$*:*

$$\phi^* = \begin{cases} \Phi(N) \equiv \frac{1}{|A|} \sqrt{\frac{\beta(N) + b\gamma(N)}{\alpha(N)}} & \text{when } \frac{\rho_I}{\rho_I + \rho_H} \leq \Phi(N) \\ \frac{\rho_I}{\rho_I + \rho_H} & \text{otherwise,} \end{cases} \tag{14}$$

and

$$K^* = \bar{U} - \rho_I \frac{N-2}{2(N-1)} \frac{\Sigma_v}{\Sigma} - \frac{A^2 \Sigma}{2N(N-2)\rho_I} (\phi^*)^2, \tag{15}$$

where $\alpha(N) = (-1+N)\Sigma^2 \left(\tilde{\Sigma} - \Sigma\right) > 0$,
$\beta(N) = -(((N-2)N\rho_I^2 \sigma_\epsilon \Sigma_v((N-1)\Sigma + (N-2)\Sigma_v))(\sigma_\epsilon + \Sigma_v)^{-1} < 0$,
and $\gamma(N) = 2(-1+N)N\rho_I \Sigma(2\Sigma + (-2+N)\tilde{\Sigma}) > 0$, *given* $\tilde{\Sigma} = Var(R_i | \mu_i(1))$.

(ii) All dealers exert effort, $s^* = 0$, and submit the following demand function:

$$a^*(p, \mu_i(s^*)) = \rho_I(\Lambda^* \rho_I + \phi^* Var[R_i \mid \mu_i(s^*)])^{-1}(\mathbb{E}[R_i \mid \mu_i(s^*)] - p), \tag{16}$$

$$\text{with } \Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\Sigma, \tag{17}$$

where $Var[R_i \mid \mu_i(0)] = \Sigma$ and $\mathbb{E}[R_i \mid \mu_i(0)] = \nu_i$.

In Appendix A.1 I provide sufficient conditions for equilibrium existence for two environments. In the first environment, clients are risk-neutral ($\rho_H \to \infty$) and do not suffer a direct loss from shirking ($\Delta = 0$), following the standard simplifying assumption in the principal-agent literature that the principal bears no risk. In the second environment, clients are risk-averse ($\rho_H < \infty$), possibly more so than dealers, and suffer a loss ($\Delta \geq 0$). This seems more natural when clients represent households or non-financial firms and dealers are well-diversified financial institutions with greater capacity to bear risk.

With risk-neutral clients the equilibrium of Proposition 1 exists when the dealer's private benefit lies in an interval $b \in [\underline{b}, \overline{b}]$. The lower bound $\underline{b} = -\beta(N)/\gamma(N)$ ensures that shirking is sufficiently attractive to make the incentive constraint bind; below it, the equilibrium contract collapses to the unconstrained share: $\phi_{UC} = \frac{\rho_I}{\rho_I + \rho_H} \to 0$. The upper bound ensures that compensating an effort-exerting dealer does not require a prohibitively large lump-sum payment, which would otherwise create a profitable deviation toward contracts that tolerate shirking. The interval $[\underline{b}, \overline{b}]$ is non-empty when signal noise σ_ϵ is sufficiently large relative to the variance of the dealer's signal under effort Σ_ν. This is intuitive: if the information gain from inducing effort is large, no client finds it profitable to deviate by tolerating shirking while all others incentivize effort.

Risk-averse clients complicate the existence proof relative to the risk-neutral case. This is because the unconstrained contract no longer assigns the dealer a zero share but instead optimizes risk-sharing between the client and their dealer at $\phi_{UC} = \frac{\rho_I}{\rho_I + \rho_H}$. Therefore, the best feasible deviation from the effort-inducing contract is itself interior and depends on the contract shares and demands chosen by all other clients and dealers, as well as all other model parameters in a way that does not admit a clean analytical characterization. To provide clean sufficient conditions, I follow He and Krishnamurthy (2012) in assuming that clients incur a loss $\Delta \geq 0$ when their dealer shirks. When Δ is sufficiently large, no client finds it profitable to deviate toward a shirking-inducing contract, since the loss from shirking outweighs the gain from offering weaker incentives.

I show that in any symmetric equilibrium with linear demand functions—a common refinement in the related literature (Rostek and Yoon, 2025)—dealers exert effort as long as $\Delta \geq b$. The only symmetric linear equilibria are therefore those characterized in Proposition 1: either the effort-inducing contract when $\Delta \geq b$, or its unconstrained counterpart $\phi^* = \phi_{UC}$ when the gains from shirking are too small to bind the incentive constraint. Intuitively, when the client's loss from shirking is at least as large as the dealer's private benefit, the client always finds it profitable to deviate away from any contract that induces shirking toward one that induces effort, since effort improves the quality of the dealer's information and leads to better trading outcomes for both parties.

Given the contracts, the dealer's shirking decisions, and their demand functions, the equilibrium price (18), and the equilibrium allocations are then pinned down via market clearing.

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

$$p^* = \frac{1}{N}\sum_i \nu_i - \frac{\phi^*}{\rho_I}\Sigma\left(\frac{N-1}{N-2}\right)\frac{A}{N}, \tag{18}$$

and dealer i purchases amount

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

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

Model discussion and extensions. The central insight of the model is that, once market clearing is taken into account, incentive contracts for dealers depend on the degree of competition in asset markets, creating an interaction between competition and moral hazard. While the specific predictions may vary with alternative modeling choices, this idea generalize across different approaches to contract design, preferences, and market clearing.

Before analyzing the interaction between moral hazard and imperfect competition in more detail, I discuss the key modeling choices that enable me to combine strategic trading, incentives to shirk, and costly information acquisition within a single tractable framework. I focus on assumptions that are central to this paper's contribution rather than on standard ingredients of the related literatures (such as CARA utility, ex-ante identical dealers, normally distributed returns, or 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, which is a common contract space restriction following Holmström and Milgrom (1987).[6] The contract resembles common institutional arrangements, combining a fixed base salary with a performance-based bonus ϕ_i that reflects average outcomes across repeated market clearings, and does not depend on one-time market outcomes. With outcome-contingent contracts dealers' demand would no longer be linear in prices—a feature that is known to render uniform-price auctions intractable.

Second, I assume that dealers 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 their decision.[7] This modeling choice preserves tractability and reflects the idea that dealers and clients do not make a single, one-shot decision based on a unique signal. Instead, trading occurs repeatedly across many rounds, and decisions are likely to reflect average outcomes across these rounds, which the model captures through expectations.

Third, the baseline model abstracts from common-value concerns: all dealers share the same prior ν about the asset's expected return, so the clearing price conveys no information about asset value. In Appendix A.2, I relax this assumption by decomposing the expected return ν_i into a common component, reflecting general market conditions, and an idiosyncratic component, reflecting institution-specific beliefs, tastes, or constraints. This extension generates two departures from the baseline. First, it induces correlation across dealers' expected returns ν_i. Second, it creates an asymmetry between effort and shirking with respect to price informativeness: under effort, the dealer conducts thorough market research and learns ν_i exactly, so the clearing price adds no information; under shirking, the dealer observes only a noisy signal of ν_i, and the clearing price, which aggregates other dealers' signals, becomes informative about the unknown common com-

[6] In the classic principal–agent framework, where the agent's effort directly affects output, Holmström and Milgrom (1987) show that linear contracts are optimal when the agent has CARA preferences and output noise is normally distributed. My environment is more complex, as it includes strategic interaction between agents, endogenous price formation, and an indirect mapping from effort to outcomes. This makes solving for the optimal contract difficult. Rather than pursuing optimality, I model contracts that mirror those used in practice.

[7] 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 (3), where expectations are taken jointly over wealth and signals.

ponent. The key insight that ϕ^* is changes in the degree of competition carries over to this richer setting.

Fourth, following He and Krishnamurthy (2012), I model the dealer's effort decision as binary and focus on equilibria in which clients incentivize effort. Both choices deliver tractability and a clean closed-form characterization of the equilibrium contract. However, relaxing either assumption preserves the paper's main message. Under continuous effort, sketched in Appendix A.2, the equilibrium contract depends on the number of dealers, N, even in the unconstrained case, making the interaction between imperfect competition and moral hazard even more pervasive. In a shirking-inducing equilibrium, which arises when the dealer's private benefit b is sufficiently large (characterized in Appendix A.6), the contract also varies with N, again reinforcing the core insight.[8]

Finally, the model features a single trading stage and abstracts from inter-temporal considerations. Even when focusing solely on trading—and setting aside interactions with contract choices—dynamic auction environments are complex (e.g., Vayanos, 1999; Rosket and Weretka, 2015; Du and Zhu, 2017). A promising, though challenging, extension would be to build on this literature by studying a dynamic common-value environment in which prices convey information about asset values and clients can revise their contracts after observing market outcomes. Another promising, though difficult, extension would be to incorporate systematic asymmetries across clients or dealers, allowing for asymmetric equilibria (building on Malamud and Rostek (2017), as sketched in Appendix A.2).

2.2 How imperfect competition and moral hazard affect choices

My goal is to analyze how imperfect competition and moral hazard interact in affecting asset allocations, prices, and ultimately welfare.

Imperfect competition. In an imperfectly competitive market, dealers have price impact (Kyle, 1989), as characterized (in my model) by expression (17) in equilibrium. Strategic dealers take this into account when choosing how to trade: they deliberately reduce their demand for additional units to lower the market-clearing price, thereby improving the terms of their winning bids. Strategic demand reduction reduces the market clearing price (compared to the price that would arise if dealers were price-takers) and distorts allocations. In particular, dealers with above-average return expectations win too little, and those with below-average expectations win too much compared to the efficient

[8] I characterize the symmetric shirking inducing equilibrium in the proof of Proposition 5.

outcome.[9]

Moral hazard. Moral hazard frictions 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 dealer, it does not affect allocative efficiency. In the remainder of the paper, I therefore focus on the flexible component, ϕ_i, which I call the 'contract'.

When the benefit of shirking is sufficiently low so that the dealer has no incentives to shirk, the contract, $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$, is unconstrained. This contract achieves the first best-solution, maximizing risk sharing across all dealers and clients, when the market is perfectly competitive ($N \to \infty$). When the incentive constraint binds, which is the case when $\frac{\rho_I}{\rho_I+\rho_H}$ is too low, the contract is set to ensure that the dealer exerts effort: $\phi^* = \Phi(N)$. While the expression is messy, the key property (discussed below) is that it depends non-trivially on the number of dealers, and therefore the degree of competition.

The contract, in turn, affects trading because it affects the dealer's risk-return trade-off. Increasing the contract raises the dealer's 'effective' risk-aversion $\frac{\phi^*}{\rho_I}$. Since dealers weigh the extra profit they expect from the purchase against the additional risk it brings, dealers are less comfortable taking on larger (risky) positions for a fixed expected return. As a result, the price (18) needs to decline to persuade them to absorb extra amounts of the asset, and price impact (17) increases. This ultimately affects asset allocations, and allocative efficiency. Dealers with above-average return expectations purchase too little, and dealers with below-average expectations too much when the constrained binds. The higher the contract, i.e., the worse the moral hazard friction, the larger this distortion.[10]

How the frictions interact. Intuitively, it is ambiguous whether stronger competition (measured by the number of dealers) alleviates or amplifies moral hazard (captured by the size of the contract).

To build intuition, I first consider the limiting case $N \to \infty$, where prices are fully informative and no dealer has price impact. In this limit, the only force determining whether a dealer wants to shirk is how much their signal deteriorates relative to exerting

[9] To see this formally, compare the market-clearing allocation of dealer i when everyone is a price-taker, $a^o(p^o, \nu_i)$, to the equilibrium allocation with strategic dealers for a fixed contract: $a^o(p^o, \nu_i) - a(p^*, \nu_i) = \frac{\rho_I}{\phi^*}\frac{1}{N-1}(\nu_i - \frac{1}{N}\sum_i \nu_i)\Sigma^{-1}$, where $\nu_i = \mathbb{E}[R_i \mid \mu_i(0)]$ and $\Sigma = Var(R_i \mid \mu_i(0))$. This expression is positive when ν_i exceeds the average and negative otherwise.

[10] To see this formally, compare the market-clearing allocation of dealer i when moral hazard is not binding, $a^*_{i,UC}$, to the allocation when the constrained is binding, $a^*_{i,C}$. This gives: $a^*_{i,C} - a^*_{i,UC} = [(\rho_I + \rho_H) - \frac{\rho_I}{\Phi(N)}]\frac{N-2}{N-1}(\frac{1}{N}\sum_i \nu_i - \nu_i)\Sigma^{-1}$.

effort.

Corollary 2. *Let* $\gamma_\infty(\sigma_\epsilon) = \frac{\rho_I \sigma_\epsilon \Sigma_\nu (\Sigma + \Sigma_\nu)}{2\Sigma(\sigma_\epsilon \Sigma_\nu + \Sigma(\sigma_\epsilon + \Sigma_\nu))}$*, and note that* $\gamma'_\infty(\sigma_\epsilon) > 0$*.*

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

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

Under perfect competition, there is a favorable contract regime and an unfavorable one. When the noise from shirking, σ_ϵ, is high relative to the private benefit from shirking, so that shirking is unattractive, the contract is unconstrained and achieves the first best; when the noise is low, making shirking tempting, no finite contract can induce effort. To see why this is the case, note that in the competitive limit the dealer's expected payoff takes a familiar mean–variance structure. It is proportional to the information–risk ratio:

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

A dealer 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 dealer exerts effort, they expect the following payoff:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(0;\phi_i)] = \frac{\rho_I}{2} \frac{Var(\mathbb{E}[R_i|\mu_i(0)])}{Var(R_i|\mu_i(0))} = \frac{\rho_I}{2} \frac{\Sigma_\nu}{\Sigma}. \tag{20}$$

When they shirk, they expect to get:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(1;\phi_i)] = \frac{\rho_I}{2} \frac{Var(\mathbb{E}[R_i|\mu_i(1)])}{Var(R_i|\mu_i(1))} + b = \frac{\rho_I}{2} \frac{\Sigma_\nu^2 (\Sigma_\nu + \sigma_\epsilon)^{-1}}{\Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}} + b. \tag{21}$$

The larger is the noise dealers observe when shirking (higher σ_ϵ), the lower is their expected payoff from shirking. Greater noise weakens the signal's ability to predict returns while increasing the residual uncertainty that dealers must bear. Both effects reduce the information–risk ratio under shirking and, in turn, lower the expected payoff.

To capture the relationship between competition and moral hazard with finitely many dealers, I analyze how the constraint contract, $\Phi(N)$, changes when we add an additional dealer. When the dealer needs to obtain a higher fraction of wealth to still exert effort, I say that moral hazard intensifies. When the opposite is true, moral hazard is relaxed.

Corollary 3. *Define* $\gamma(\sigma_\epsilon) \equiv \frac{\rho_I \sigma_\epsilon \Sigma_\nu (2(N-1)^3 \Sigma + (N-2)(2+N(2N-3))\Sigma_\nu}{4(N-1)^2 \Sigma (N\Sigma\sigma_\epsilon - \sigma_\epsilon \Sigma_\nu + N(\Sigma+\sigma_\epsilon)\Sigma_\nu)}$*; note that* $\gamma'(\sigma_\epsilon) > 0$*.*

(i) When σ_ϵ is sufficiently low 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.

Corollary 3 underscores the non-linear interaction between moral hazard and imperfect competition when the incentive constraint binds, foreshadowed in Corollary 2.[11] When noise from shirking is small, increasing the number of dealers worsens the moral hazard problem—the fraction of wealth the dealer must obtain to exert effort increases in N (at the margin); otherwise it decreases. Given that the cutoff $\gamma(\sigma_\epsilon)$ varies in N, however, the relationship is complex.

To build an intuition for Corollary 3, consider a dealer contemplating a unilateral deviation to shirking, taking the contract and all other dealers' behavior as given. Understanding how this (potential) deviation affects the dealer's demand and the clearing price allows me to pin down how attractive shirking is relative to effort, and therefore how the contract that deters shirking must respond as N changes.

When the dealer shirks, their trading demand (16) is affected, because their private information becomes noisier: instead of knowing their expected return ν_i precisely, their conditional expectation is shrunk toward the common prior ν, $\mathbb{E}[R_i \mid \mu_i(1)] = \nu + \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}((\nu_i + \epsilon_i) - \nu)$, reflecting that the dealer can no longer perfectly distinguish their expected return from the common prior. Moreover, because their signal is less precise, they trade less aggressively: the conditional variance increases from $Var[R_i|\mu_i(0) = \Sigma$ to $Var[R_i|\mu_i(1) = \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, causing the denominator of the demand function (16) to grow and scaling down the submitted order. In expectation, a shirking dealer wins less than when exerting effort.

Moreover, the clearing price becomes less accurate, as the shirking dealer's noisy signal enters the price instead of a precise one. In addition, because the shirking dealer trades less aggressively, their order is smaller, and the remaining dealers must absorb more of the fixed asset supply, pushing the price down to compensate them for bearing additional risk. In expectation, the price therefore declines when one dealer shirks, and the decline is larger the noisier shirking is.

The effects on demand and price together determine how much the dealer expects to

[11] Appendix A.4 details the intertwined forces behind this interaction, including how competition affects price impact and price predictability.

gain in terms of payoff, depending on their effort-choice, $s_i \in \{0, 1\}$:

$$\mathbb{E}[\mathcal{U}(s_i; \phi^*)] = \underbrace{\phi^* \frac{\rho_I}{2} \frac{2\Lambda^* \rho_I + \phi^* Var[R_i \mid \mu_i(s_i)]}{(\Lambda^* \rho_I + \phi^* Var[R_i \mid \mu_i(s_i)])^2}}_{\text{conversion rate}} \times \underbrace{\mathbb{E}\Big[\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)^2\Big]}_{\text{value of information}} + s_i b. \quad (22)$$

The expected payoff depends on the expected squared gap between the dealer's own expected return and the clearing price, which captures how the dealer's informational advantage over the market translates into expected payoff—akin in spirit to the 'value of information' in Grossman and Stiglitz (1980), though here it arises in an imperfectly competitive setting with strategic dealers.[12] The value of information is multiplied by a 'conversion rate', which captures how effectively the dealer can exploit their informational advantage given the costs of trading that arise from bearing risk and price impact, Λ^*. When $\Lambda^* = 0$, the conversion rate is independent of N and only depends on the dealer's risk-tolerance, ρ_I, and the residual risk, $Var[R_i \mid \mu_i(s_i)]$. When $\Lambda^* > 0$, the dealer's own order moves the price against them, and its extent depends on N. Price impact enters twice: once when the order is submitted and once through market clearing.[13] The larger price impact, the lower the conversion rate, reducing the share of their informational advantage that the dealer can convert into profits.

As N increases, three things happen. First, the value of information declines under both effort and shirking, because with more dealers sharing the burden of absorbing the fixed asset supply A, the price discount required to compensate each dealer for bearing risk shrinks. Importantly, this decline is faster under effort than under shirking (since the effort dealer trades more aggressively, holds a larger position, and is therefore more exposed to the shrinking risk discount), making shirking relatively more attractive as N grows.

Second, the value of information also changes because the price converges toward the common prior ν as dealers enter. For a dealer who exerts effort, this implies that their own expected return valuation differs more strongly from the clearing price (which

[12] The gap is squared because, as in standard mean-variance portfolio theory, the dealer's optimal demand is proportional to the signal-price gap, so that the resulting payoff (combining the price and quantity effect) is quadratic in the gap.

[13] Specifically, the denominator of the conversion rate $(\Lambda^* \rho_I + \phi^* \text{Var}[R_i \mid \mu_i(s_i)])^2$ is the square of the total cost of trading, combining the price impact cost $\Lambda^* \rho_I$ and the risk cost $\phi^* \text{Var}[R_i \mid \mu_i(s_i)]$—recall optimality condition (10). The numerator $2\Lambda^* \rho_I + \phi^* \text{Var}[R_i \mid \mu_i(s_i)]$ is a weighted combination of the two cost components, where the price impact cost receives double weight, reflecting that price impact affects profits twice; once when the dealer submits their order and once through the market clearing price.

averages over the valuations of all market participants). A shirking dealer experiences an analogous effect when noise from shirking is small, in which case their assessment is sufficiently close to the true individual expected return. In contrast, when σ_ϵ is large, the dealer's signal is so noisy that their assessment, $\mathbb{E}[R_i \mid \mu_i(1)]$, is heavily pulled toward ν (in the limit when σ_ϵ is infinite, $\mathbb{E}[R_i \mid \mu_i(1)] \to \nu$). The key is that their assessment is closer to the common prior ν than the clearing price. Therefore, as the price converges toward ν, it moves toward the dealer's assessment rather than away from it, narrowing the gap and reducing the value of information.[14] This implies that when shirking noise is high, this component of the value of information falls under shirking while it rises under effort, making shirking less attractive as N increases.

Third, the conversion rate increases with N under both effort and shirking, as the reduction in price impact Λ^* makes it cheaper to exploit a given informational advantage. The rate grows faster under effort than under shirking, because a shirking dealer perceives greater residual uncertainty, $Var[R_i \mid \mu_i(1)] > Var[R_i \mid \mu_i(0)]$. Since under effort the risk cost is lower, the same decline in Λ^* has a proportionally larger effect, raising the conversion rate by more and making shirking relatively less attractive as N grows, and increasingly so as σ_ϵ increases.

Taken together, when σ_ϵ is large, two forces make shirking less attractive as N grows: the value of information gap narrows under shirking while widening under effort, and the conversion rate rises faster under effort. Together these outweigh the shrinking price discount that hurts effort dealers disproportionately, so clients can offer a smaller ϕ^* and $\Phi(N)$ decreases in N. When σ_ϵ is small, the gap evolves similarly under both effort and shirking, and the conversion rate effect is outweighed by the shrinking price discount that hurts effort dealers disproportionately. Shirking becomes relatively more attractive as N grows, clients must raise ϕ^*, and $\Phi(N)$ increases in N.

Whether the incentive constraint binds depends on all model parameters; Corollaries 4 and 5 in Appendix A.3 characterize the conditions under which each regime arises when clients are risk-neutral, and when they are risk-averse. In the latter case, unfavorable contract regime arises when both shirking noise and (exogenous) supply are sufficiently small. When asset supply is sufficiently large, the constraint never binds, because a larger supply forces each dealer to absorb a larger position. The net benefit of doing so

[14] Formally, this component of the value of information decreases as N increases when $g(\sigma_\epsilon) = \gamma_0 + \gamma_1\sigma_\epsilon + \gamma_2\sigma_\epsilon^2 < 0$, where $\gamma_0 = (N-1)N\Sigma^2\Sigma_\nu^2 > 0$, $\gamma_1 = \Sigma\Sigma_\nu\big(2(N-1)\Sigma + (3N-2)\Sigma_\nu\big) > 0$, and $\gamma_2 = -(N-2)(N-1)\Sigma^2 + (7N-2N^2-4)\Sigma\Sigma_\nu - (N-2)^2\Sigma_\nu^2 < 0$. Since g is a downward-opening parabola that is increasing at $\sigma_\epsilon = 0$, this component declines with N only for sufficiently large σ_ϵ.

is greater under effort than under shirking, since shirking raises perceived variance and therefore the cost of bearing that risk.

Figure 1 illustrates the two regimes through numerical examples. In the first example, corresponding to the unfavorable regime, noise from shirking is low, $\sigma_\epsilon = 0.5$, and $\Phi(N)$ increases in N for large enough N. Given the other parameters ($\rho_I = \rho_H = 0.5, \Sigma = \Sigma_\nu = 1, b = 0.2$), the constraint always binds for some N, irregardless of the size of supply. In order to generate a region at which the contract is unconstrained, I set supply relatively high to 10. In the second example, corresponding to the favorable regime, noise from shirking is high, $\sigma_\epsilon = 100$. $\Phi(N)$ decreases in N for large enough N, and eventually becomes unconstrained. I scale down supply to $A = 2$ to generate a region at which the constraint binds, and also to illustrate the potential non-linearity of the contract. When A is larger, e.g., 10, as in the first example, the contract is always unconstrained.

Take away. The finding that contract shares vary with the number of dealers suggests that policies promoting competition in asset markets, such as lowering entry barriers for intermediaries, may inadvertently intensify moral hazard. This risk is most pronounced when asset returns can be predicted with reasonable precision even without exerting effort, as in transparent markets such as government bonds or large-cap equities. In opaque markets, where returns are hard to forecast, such as corporate bonds, structured credit, or OTC derivatives, competition is less likely to undermine incentives.

2.3 How imperfect competition and moral hazard affect welfare

Intuitively, because allocations depend on the equilibrium contract—and the contract itself varies with the number of dealers N—the welfare effects of increased competition should hinge on how the contract responds to N. If stronger competition tightens incentive constraints, welfare may decline even as markets become more competitive. To analyze whether this is the case, I compute the allocative efficiency of the equilibrium allocation:

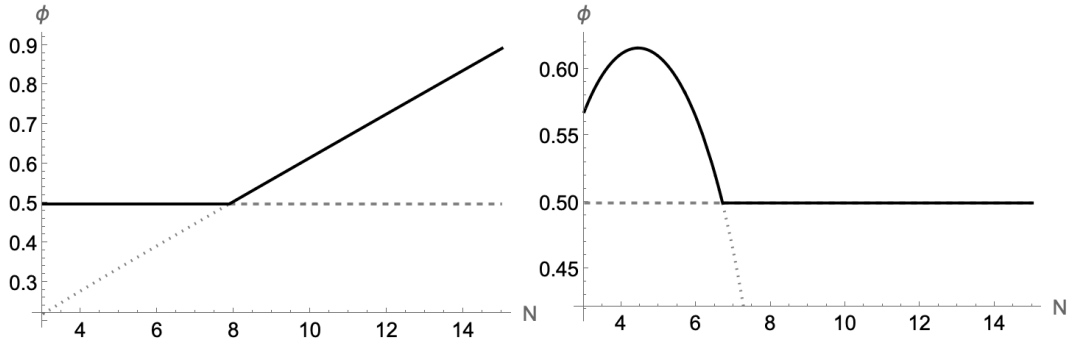
$$W(\phi^*, N) = \sum_i W_i(\phi^*, N), \tag{23}$$

$$W_i(\phi^*, N) = a^*(p^*, \nu_i)\nu_i - \sum_{k\in\{h,i\}} \frac{1}{2}\phi_k^{*2}\frac{1}{\rho_k} a^*(p^*, \nu_i)\Sigma\, a^*(p^*, \nu_i), \tag{24}$$

where $\phi_h = 1 - \phi_i$. Following the market-design literature, I adopt an ex-ante perspective and define 'welfare' as expected allocative efficiency, $\mathbb{E}[W(\phi^*, N)]$. This definition

Figure 1: Examples of how the flexible contract, ϕ^*, can change in N

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



Notes: Figure 1 plot the equilibrium contract ϕ^* (black line) as a function of the number of dealers N for two numeric examples. The chosen contract is a piecewise function (the thick black line) 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, $\rho_I = \rho_H = 0.5, \Sigma = \Sigma_v = 1, b = 0.2$. In Figure 1a shirking-noise is low, $\sigma_\epsilon = 0.5$ and supply is large, $A = 10$, in Figure 1b it is high, $\sigma_\epsilon = 100$ and supply is small $A = 2$. Appendix Figures A1 and A2 shows the corresponding fixed part of the contract, K^*, and the expected gain from exerting effort relative to shirking under the equilibrium contract.

ensures that welfare comparisons do not depend on a particular realization of return expectations.

Welfare is shaped by three forces: how wealth is split between dealers and clients, their relative risk tolerances, and how much useful information dealers 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 dealer ($N = 1$) purchases an asset with normally distributed returns, $\mathcal{N}(v, \Sigma)$, and keeps a share ϕ of the payoff. Welfare is

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

where $\beta(\phi, N)$ is given by expression (27) 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, $N > 2$ dealers have private information about their expected asset return, v_i, like in my model but they keep all assets ($\phi = 1$). Welfare is

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

with $\alpha(\phi, N)$ given by expression (26) 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 dealers are risk-tolerant. In the full model with moral hazard, welfare is driven by the same three forces. 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) both affect welfare but in different ways.

Proposition 2. *Welfare is given by*

$$\mathbb{E}[W(\phi^*, N)] = A\nu - \beta(\phi^*, N)\Sigma A^2 + \alpha(\phi^*, N)\beta(\phi^*, N)\frac{\Sigma_\nu}{\Sigma}, \tag{25}$$

with

$$\alpha(\phi^*, N) = \frac{(-2+N)N\rho_I^2((2+N(-2+\phi^*)-2\phi^*)\phi^*\rho_H + (-2+N)(-1+\phi)^{*2}\rho_I)}{-(-1+N)\phi^{*2}(\rho_I - 2\phi^*\rho_I + \phi^{*2}(\rho_H+\rho_I))}, \tag{26}$$

$$\beta(\phi^*, N) = \frac{1}{2N}\left(\frac{1}{\rho_I}\phi^{*2} + \frac{1}{\rho_H}(1-\phi^*)^2\right), \tag{27}$$

where $\alpha(\phi^*, N) \geq 0$ for $\phi^* < 1$, and $\beta(\phi^*, N) \geq 0$.

(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 dealers (N) increases when

$$\underbrace{\frac{\partial \mathbb{E}[W(\phi, N)]}{\partial N}}_{\text{Direct effect}} - \underbrace{\frac{\partial \beta(\phi, N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\Sigma A^2 + \frac{\partial \alpha(\phi, N)\beta(\phi, N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\frac{\Sigma_\nu}{\Sigma}}_{\text{Indirect effect through } \Phi(N)} < 0, \tag{28}$$

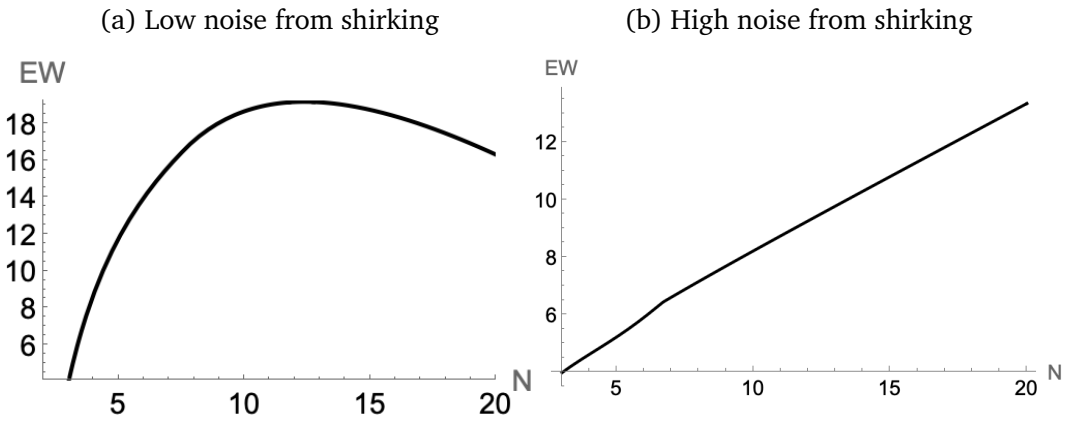
where $\phi = \Phi(N)$, and increases otherwise.

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

Reducing moral hazard is always welfare enhancing. As the contract share ϕ^* declines, the loss from risk aversion falls ($\frac{\partial \beta(\phi^*, N)}{\partial \phi} \geq 0$). This is because the dealer is effectively more risk-averse than the client ($\frac{1-\phi^*}{\rho_H} < \frac{\phi^*}{\rho_I}$) when the incentive constraint is binding ($\frac{\rho_I}{\rho_I+\rho_H} \leq \phi^*$). Therefore, shifting some fraction of the asset from the dealer to the client reduces the total risk bearing cost, and improves risk-sharing.[15] In addition, because the dealer absorbs a smaller share of the risky position, they react more aggressively to informative signals, which improves welfare. Formally, this is captured

[15] For instance, if dealer and client have equal risk tolerance ($\rho_H = \rho_I$), welfare rises when moving from a situation in which the dealer receives more than half the surplus toward a more balanced split.

Figure 2: Examples of how equilibrium welfare can change in N



Notes: Figure 2 plots equilibrium welfare, $\mathbb{E}[W(\phi^*, N)]$ from Proposition 2, as a function of the number of dealers N for the same two numerical examples as in Figure 1. In both examples, parameters are $\rho_I = \rho_H = 0.5, \Sigma = \Sigma_\nu = 1, b = 0.2$ and $\nu = 2$. Figure 1a considers a low-shirking-noise, high-supply environment ($\sigma_\epsilon = 0.5, A = 10$), while Figure 1b considers a high-shirking-noise, low-supply environment ($\sigma_\epsilon = 100, A = 2$).

by $\frac{\partial(\alpha(\phi^*,N)\beta(\phi^*,N))}{\partial\phi} \leq 0$, which shows that lowering ϕ^* strengthens the information-driven component of welfare.

Increasing competition by adding dealers can either raise or reduce welfare. When the contract does not adjust with N—that is, when incentive constraints are slack—more competition unambiguously increases welfare, for two reasons. First, as the number of dealers grows, each dealer (and thus each client) holds a smaller share of the risky asset, A/N, so the total loss from risk aversion declines. Formally, the term $-\frac{\partial\beta(\phi^*,N)}{\partial N}\Sigma A^2 \geq 0$ contributes positively to welfare. Second, the clearing-price aggregates information across a larger set of signals, which 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 can decline because changes in the number of dealers affect welfare indirectly through the equilibrium contract. This indirect effect is negative when the contract becomes more generous to dealers as N rises, that is, in the unfavorable contract regime. Welfare declines when this negative indirect effect—reflecting reduced risk sharing and suboptimal use of signal information per unit of risk (the second line of (28))—outweighs the positive direct effect of competition (the first line of (28)).

To illustrate this, I return to the two numerical examples. In the low–σ_ϵ example of Figure 1a, the contract is constrained at $N = 15$ (with $\Phi(15) = 0.89$) and is increasing in N. At this point, the indirect marginal effect of competition on welfare is negative.

Although the direct effect is positive, the contract's unfavorable response to competition leads to an overall decline in welfare. By contrast, in the high–σ_ϵ example of Figure 1b, the contract is decreasing at $N = 6$ (with $\Phi(6) = 0.56$). 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 dealers 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 dealers have price impact and distort allocations because of it). Appendix A.5 compares the welfare impact of moral hazard and imperfect competition and identifies when each dominates. The results inform whether policy should focus on promoting competition or mitigating moral hazard, recognizing their interaction.

3 Quantification

The goal of the remainder of the paper is to provide a rough empirical assessment of the relative importance of moral hazard and imperfect competition using trade data. The advantage of this approach is that such data are observable to regulators, in contrast to contractual terms or direct measures of effort, which are private. Given that I only observe trade data and not contracts or effort, my quantification exercise leverages modeling assumptions described in Section 2 with some additional modifications, which I describe below.

3.1 Institutional environment and data

I use proprietary trade data from the Canadian equity market, which provides a useful setting to empirically evaluate how moral hazard and imperfect competition interact in asset pricing.

Institutional environment. Canada (like other major markets such as the European Union and the United Kingdom) grants dealers greater discretion in trade execution than the United States because it does not enforce a strict National Best Bid and Offer (NBBO) regime. As a result, prices and allocations can vary more with local trading conditions,

creating greater scope for moral hazard and imperfect competition to affect outcomes in empirically distinguishable ways.[16] Furthermore, exchange trading in Canada is representative of a typical equity transaction, because Canadian regulation requires equity trades to be executed on a regulated marketplace.

I focus on the Toronto Stock Exchange (TMX), which is the primary exchange, and Alpha, its smaller counterpart trading the same securities. Combined, they represent about 60% of total market activity. Both venues share common microstructure features of modern stock markets, with most trading occurring on fairly competitive electronic limit order books. Only member firms, who are typically banks, large broker-dealers, and asset managers, are permitted to directly submit orders. All other investors, including large institutional investors, access the market through these members, which actively execute trades both for their own account and on behalf of clients.

In handling client orders, member firms retain considerable discretion. Although their execution decisions are subject to principles-based best-execution requirements under the Canadian Investment Regulatory Organization (CIRO) and the supervisory framework of the Universal Market Integrity Rules (UMIR), these rules largely rely on firms to design, implement, and monitor their own execution policies and procedures.[17]

Data. My data cover all regular trades on TSX and Alpha in 2021 during the continuous trading session (9:30 am to 4:00 pm). For each trade, I observe the time of execution, the traded security (the 'symbol'), trade size, price, and the national best bid and offer (prior trade execution). I also know whether the trade is active—meaning it immediately executes against resting orders in the limit order book and thus consumes liquidity—or passive; I focus on active trades throughout the analysis.

The data have two features that make them well suited to studying the interaction between imperfect competition and moral hazard. First, I observe the exchange member firm executing each trade, which I refer to as the 'dealer', recognizing that multiple individuals may operate under a single firm identifier. Second, I observe whether a trade is executed for the member's own account or on behalf of a client, a distinction that is central to analyzing moral hazard.

[16] While dealers in the United States also retain discretion over order routing, timing, venue choice, queue position, and execution speed (and face heterogeneous competitive conditions across venues) NBBO price protection limits price dispersion, making these effects harder to detect empirically.

[17] Despite this, many institutional investors choose not to become exchange members, because membership is relatively costly: firms must be affiliated with CIRO, have access to the clearing system, and pay a C$65,000 entry fee plus monthly charges.

Suggestive evidence. Before bringing the model to the data, I examine whether the raw data display patterns consistent with the two frictions emphasized in the model, which also serves to illustrate the variation that identifies the key model parameters. Appendix B.1 provides further evidence.

To provide suggestive evidence of imperfect competition, I exploit variation across TSX, which is the primary Canadian exchange, and Alpha, which is its smaller counterpart, that list the same set of stocks. Both exchanges are well established, have operated for over a decade, are owned by the same parent company (TMX), and operate under similar market rules and regulatory requirements. This shared institutional structure rules out many, though not all, confounding factors unrelated to competition that could otherwise explain price differences.

As the smaller of the two exchanges, Alpha operates with systematically higher market concentration than TSX (see Appendix Figure A3) and exhibits worse execution prices (see Appendix Table A1). Controlling for trade size as well as dealer–stock–day–hour fixed effects, dealers buying on Alpha pay prices that are on average about C$0.004 higher—roughly 2 basis points of the median stock price or 40% of the median bid-ask spread—and conversely receive lower prices when selling. The regression results further confirm early empirical studies on the NYSE and regional U.S. exchanges (such as Bernhardt et al. (2005)), according to which larger trades are executed at worse prices. This suggests that dealers have price impact—a key feature of the model.

Detecting evidence of moral hazard in financial markets is challenging because incentives and contractual arrangements are rarely observed. I provide suggestive evidence based on the idea that, when moral-hazard frictions exist between dealers and their clients, trades executed on behalf of clients should occur at systematically worse prices than trades for a dealer's own account.

Appendix Table A2 and Appendix Figure A4 show that, on average, a client-account trade executed on the same exchange, for the same stock, and within the same hour and day occurs at a price about C$0.005 worse—roughly 2 basis points of the median stock price or 50% of the median bid-ask spread—after controlling for trade size. To my knowledge, this type of evidence has not previously been documented, as trade accounts are usually unobserved.

In practice, several forces consistent with moral hazard could lead to worse execution for client accounts. Dealers may simply pass through client orders without optimizing execution, or exert less effort in trading on behalf of clients. In both cases, the dealer does not exert the same effort to improve the client's performance. My model captures

these forces in a stylized way: prices are worse when the dealer invests on the client's behalf because the dealer captures part of the surplus as compensation, pushing prices up when investors buy and down when they sell.

3.2 Bringing the model to the data

To link the theoretical framework to these empirical patterns, I introduce a few targeted adjustments to match aggregate rather than order-level outcomes. The model is designed to approximate the forces that drive average prices on exchanges (such as risk tolerance and return volatility) not to replicate limit-order-book microstructure or match order-level execution dynamics. The mapping therefore involves deliberate simplifications, which I describe below.

Model-data mapping. Since my data does not contain order-level information, I approximate the price formation through a sequence of frequently held batch auctions (Budish et al., 2015), discretizing the trading day into Γ periods. I focus on average prices and aggregate trade volume, abstracting from individual execution terms that depend on limit order book dynamics I do not observe. In each batch auction, active orders are treated as net demand and passive resting orders as fixed supply.

To reflect the observed price differences across accounts and exchanges, I model separate auctions for client-account and inventory-account trades on each exchange, as detailed in Appendix A.2. Client and inventory books are institutionally ring-fenced by mandate, client-priority rules, and distinct risk budgets, making joint optimization across accounts operationally costly and often infeasible in real time. Allowing for optimal bidding in a single auction with two asymmetric account types would also introduce substantial analytical complexity and would generate a single clearing price—implying no price differences across accounts by construction, at odds with the data. Similarly, I model separate auctions for the buy- and sell-side to avoid a mechanical bias from averaging transaction prices over a short interval, given that bid–ask spreads imply buys and sells occur at different prices in practice.

Finally, I treat each security separately so that prices and trade sizes are not averaged across securities with different price levels, market capitalizations, or liquidity. Each auction for a given security clears independently: demand and supply are aggregated only within that security and not across securities. This reflects how stock markets operate, and implies that dealers choose their demand for one security independently of all other securities.

Observables. I focus on liquid S&P/TSX 60 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 $e \in \{\text{TSX, Alpha}\}$, period τ in a day t, security s, account-type $k \in \{\text{CL, IN}\}$, and side $l \in \{\text{buy, sell}\}$. I average over prices with these average prices reflecting the (hypothetical) auction prices, $p^{el}_{tsk\tau}$; and I aggregate traded quantities to obtain the total amount traded in each auction, $A^{el}_{tsk\tau}$; see Appendix Figures A5a and A5b. For each day–security–account-type–side combination, I count the number of distinct dealer IDs executing a trade and use this count as the number of potential bidders in the corresponding auction, N^{el}_{tsk}; see Appendix Figure A5c.[18]

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 common 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 A5d.

3.3 Estimating risk-tolerance and contract shares

The goal is to learn about the risk tolerance of the dealer, ρ_I, and the client, ρ_H, as well as the contract size, ϕ^*. In practice, dealers and clients may differ in their risk tolerance and contractual arrangements. The estimated parameters should therefore be interpreted as market-level averages summarizing the aggregate behavior of the dealer population, rather than the characteristics of any individual institution. To estimate these

[18]This measure is an approximation: In the model, at least three active dealers each receive a signal and submit one bid, whereas in the data I observe executed trades (not all bids) within each auction window, and the associated dealer (firm) IDs. One alternative would count each executed trade within a period as a separate bidder, implicitly assuming that multiple trades by the same ID reflect independent private signals; this would overstate competition, particularly since dealers within the same firm may not compete independently. Another approach would aggregate trades by dealer ID within each auction window and count distinct IDs. I do not pursue this approach because it would likely overstate coordination within firms and understate competition, and because there are too often fewer than three dealer IDs, especially for the inventory account on Alpha.

parameters, I follow the GMM procedure detailed in Appendix B.2; here I only summarize the key moments that inform identification.

Moments. Price differences between own-account trades on TSX and Alpha identify the dealer's risk tolerance, ρ_I; while price differences between client- and own-account trades on a given exchange identify the contract size, ϕ^*.[19] To see why this is the case, 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:

$$p^{el}_{t\tau sk} = \frac{1}{N^{el}_{tsk}} \sum_i \nu_{ti\tau sk} - \frac{\phi^*_k}{\rho_I} \Sigma_{ts} \left(\frac{N^{el}_{tsk} - 1}{N^{el}_{tsk} - 2} \right) \frac{A^{el}_{t\tau sk}}{N^{el}_{tsk}}, \tag{18'}$$

with $\phi^*_{IN} = 1$ for own-accounts, and $\phi^*_{CL} = \phi^* \geq 0$ for client-accounts (see Appendix A.2). Averaging over intraday periods yields the following daily average price (per security, account, trade-side, and exchange):

$$p^{el}_{tsk} = \frac{1}{\Gamma} \sum_\tau p^{el}_{t\tau sk} \approx \nu_{ts} - \frac{\phi^*_k}{\rho_I} \Sigma_{ts} \left(\frac{N^{el}_{tsk} - 1}{N^{el}_{tsk} - 2} \right) \frac{1}{N^{el}_{tsk}} \frac{1}{\Gamma} \sum_\tau A^{el}_{t\tau sk}. \tag{18"}$$

Here I have relied on the assumption that $\nu_{ti\tau sk}$'s are drawn iid within a day-security from $\mathcal{N}(\nu_{ts}, \Sigma_{ts})$. For inventory trades the only unknown parameters in expression (18") are the dealer's risk tolerance ρ_I, and ν_{ts}, which drops out when taking differences across exchanges: $p^{TSX,l}_{ts,IN} - p^{Alpha,l}_{ts,IN}$ therefore inform us about ρ_I. Given ρ_I, the cross-account price differences, $p^{TSX,l}_{ts,CL} - p^{TSX,l}_{ts,IN}$, pin down the contract size, ϕ^*.

To learn about the client's risk tolerance, ρ_H, I next exploit the equilibrium condition from Proposition 1 that the contract must satisfy $\frac{\rho_I}{\rho_H + \rho_I} \leq \phi^*$. This inequality reflects the fact that clients must offer sufficiently strong incentives to deter dealers from shirking, and tighter incentives are required when clients are more risk-averse. Rearranging the inequality condition yields a lower bound on the client's risk-tolerance parameter, ρ^{lb}_H, consistent with the observed contract.

I use data from both buy- and sell-side auctions, and drop observations when a security is not bought (or sold) by at least three dealers during all periods in a day. This ensures that taking daily averages provides a reasonable approximation to the expected auction price for a given account type and trade side. 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

[19] 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.

empirical percentiles of the bootstrap draws, which provide more reliable coverage for median-based estimators with skewed sampling distributions.

Median risk-tolerance and contract. The median estimate of risk tolerance, ρ_I, is about 670 (see Table 1). Although imprecisely estimated, the estimate is broadly consistent with Wittwer and Allen (2026), who infer slightly lower risk tolerance ($\rho_I \approx 250$) for Canadian primary dealers, which account for the largest stock trading volume in Canada (Wittwer and Uthemann, 2024). While numerical values are not directly comparable across the two settings, the qualitative conclusion is the same.

The median estimated contract share, ϕ^*, is measured with high precision and is consistently around 0.6, that is, 60%, across specifications. This implies that intermediaries capture more than half of the surplus generated by a trade, indicating economically meaningful moral-hazard frictions. The finding is consistent with evidence that intermediaries capture the value they create in asset management (Fama and French, 2010; French, 2008; Berk and van Binsbergen, 2015), and that total manager compensation represents a large share of value created in hedge funds and private equity (Goetzmann et al., 2003; Metrick and Yasuda, 2010; Lan et al., 2013).[20]

Finally, the implied lower bound on clients' risk tolerance is around 385, which is about 60% of the dealer's risk-tolerance. Many of the active clients in my setting are large institutional investors (such as pension funds, endowments, foundations, or fully managed investment funds). These investors typically hold diversified portfolios, making them relatively risk-tolerant.

Discussion. Unobserved factors that systematically affect execution prices across exchanges for otherwise comparable inventory trades affect the dealer's risk-tolerance estimate, ρ_I. For example, if more difficult–to–execute or less informed inventory trades were disproportionately routed to Alpha while simpler orders are sent to the more liquid TSX, ρ_I could be biased upward. Similarly, unobserved factors that systematically differentiate execution prices (for comparable trades) between client and inventory trades on TSX could affect the estimated contract share ϕ^*. For instance, client orders may be more urgent or require faster completion, leading to worse execution prices relative to

[20]The 20% performance fee standard in hedge funds may seem lower than 60%, but the two are not directly comparable: the performance fee is levied on gross returns above a hurdle, not on the gain from trade. Once management fees are added and total compensation is expressed relative to the gain from trade, the manager's share is substantially higher. Lan et al. (2013) show this in a calibrated 2-and-20 fund: the manager captures approximately all of the surplus, with management fees accounting for the majority.

Table 1: Parameter estimates

Parameter	Estimate	SE (SD)	SE (Pct)	95% CI (lower)	95% CI (upper)
ρ_I	670.562	138.219	127.743	404.247	904.998
ρ_H^{lb}	385.241	82.914	80.283	221.820	536.528
ϕ^*	0.635	0.024	0.022	0.599	0.686

Notes: Table 1 reports parameter estimates from the main specification (2021, 15-minute intervals, with return variance measured daily at the security level). Appendix Table A6 reports robustness checks. 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.2 for details.

inventory trades for reasons outside the model. In that case, contract shares would be biased upward.

While transaction data alone cannot fully rule out these concerns, identification relies on comparisons of average execution prices, which helps smooth out trade-specific concerns. Moreover, the analysis focuses on trades executed within narrow time windows for the same security across two exchanges owned by the same company (TMX) and operating under similar market rules. This institutional alignment reduces the likelihood that unobserved factors generate quantitatively important biases. Reassuringly, the resulting estimates are comparable in magnitude to those reported in the existing literature.

3.4 Interaction of moral hazard and imperfect competition

My model highlights that moral hazard and imperfect competition interact, implying that the contract should vary with the number of dealers, unless incentive constraints do not bind (Corollary 3).

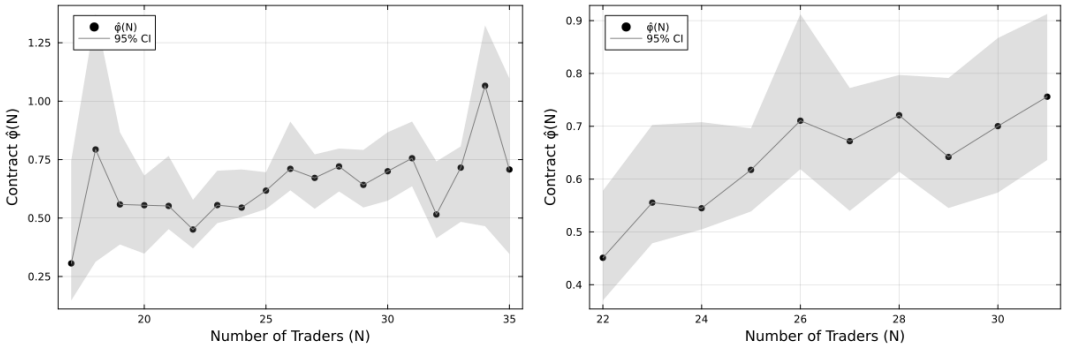
How the contract varies with the number of dealers. To test this, I estimate $\phi^*(N)$ separately for each number of active TSX client-account dealers per security and day, $N_{ts}^{TSX,l} = N$; see Appendix B.2 for details.

Figure 3a shows an upward trend but with substantial noise at the lowest and highest dealer counts, where data are sparse. Excluding these thinly populated dealer-count groups in Figure 3b reveals a pronounced pattern: the contract increases from about 0.45 to 0.76 (by 70%) as the number of active dealers rises from 22 to 31 (by 40%). This empirical pattern is consistent with the unfavorable contract regime, in which moral hazard frictions intensify in stronger competition.

With trade data alone, I cannot rule out that this empirical pattern is at least par-

Figure 3: How ϕ^* changes in N

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



Notes: Figure 3a reports the contract estimates and their 95% confidence intervals for each dealer-count group using all available observations. Figure 3b repeats the exercise but excludes groups with fewer than 50 observations. This restriction narrows the support of dealer counts from 17–35 to 22–31 and removes much of the noise in the extremes. Appendix B.2 provides further details on the estimation procedure.

tially driven by confounding factors, such as reflect time-varying market conditions.[21] Cleanly accounting for time-varying market conditions would require jointly estimating all model primitives and allowing key parameters—such as the noise and the private benefit from shirking, σ_ϵ and b—to vary in market conditions. This would require observing 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 effects. If we are willing to take 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 consider 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 25 dealers, which lie at the center of the empirical distribution (Figure 3b).

Because I do not know the exact values of the client's risk tolerance or the information–risk ratio, $\kappa = \overline{\Sigma}/\Sigma_v$, I compute welfare effects over a range of values. I vary the client's risk tolerance, ρ_H, from its lower bound (415) to values exceeding the dealer's

[21] 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 A6 shows that the opposite is true: $|A|$ tends to increase with N. I additionally examine whether return variance differs systematically across dealer-count groups and find no evidence of such differences.

Table 2: Welfare effects for median client auction

ρ		ϕ			Total		Decomp.	
$\hat{\rho}_I$	ρ_H	ϕ_{UC}	$\hat{\phi}^*$	$\hat{\phi}^*_+$	ΔW_ϕ	ΔW_N	ΔW_N^{d}	$\Delta W_N^{\mathrm{ind}}$
670	415	0.62	0.62	0.71	0.00	-0.15	0.52	-0.67
670	500	0.57	0.62	0.71	0.23	-0.55	0.55	-1.10
670	670	0.50	0.62	0.71	1.23	-1.04	0.59	-1.63
670	800	0.46	0.62	0.71	2.25	-1.27	0.60	-1.88
670	1000	0.40	0.62	0.71	4.03	-1.51	0.62	-2.13

Notes: Table 2 reports welfare effects (in percent) 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 $\Sigma_\nu = \overline{\Sigma} = 0.038$). Appendix Table A3 reports the corresponding welfare effects in dollar terms. Columns 1–2 list the estimated dealer risk tolerance $\hat{\rho}_I$ and the values of ρ_H. Columns 3–5 report the unconstrained contract $\phi_{UC} = \rho_I/(\rho_I + \rho_H)$, the estimated contract $\hat{\phi}^*(N)$ at $N = 25$, and the estimated contract $\hat{\phi}^*(N{+}1)$ at $N{+}1 = 26$. Columns 6–7 report total welfare changes: ΔW_ϕ is the gain from eliminating moral hazard at $N = 25$, while ΔW_N is the total effect of increasing competition from N to $N{+}1$ with endogenous contract adjustment. Columns 8–9 decompose ΔW_N into a direct effect, ΔW_N^{d}, holding the contract fixed at $\hat{\phi}^*(N)$, and an indirect effect, $\Delta W_N^{\mathrm{ind}}$, arising from the contract change to $\hat{\phi}^*(N{+}1)$.

risk tolerance;[22] and the information-risk ratio from 0 to 2, with 1 as the baseline. To express welfare changes in percentage terms, I set the dealer's expected signal ν equal to the median mid-quote (C$57.16); because this level is large relative to the return variance (0.038), percentage changes are small, so I focus on relative welfare differences across scenarios.

Table 2 displays the welfare effects of (i) removing moral hazard, while keeping competition constant; and (ii) of increasing competition by adding one more dealer for all parameter constellations. I focus on the theoretically ambiguous welfare effect of increasing competition. Adding dealers has a positive direct effect through improved risk sharing, but also an indirect effect operating through endogenous contract adjustments as N changes. While the direct effect is always positive, the indirect effect can be positive or negative depending on how contracts respond to greater competition.

In the data, the contract rises when moving from 25 to 26 dealers, generating a negative indirect effect that dominates the positive direct effect and leads to a net welfare loss from additional entry. When the contract instead falls the indirect effect becomes positive. Over wider ranges, such as from 22 to 31 dealers, where the contract predomi-

[22]This lower bound exceeds the median lower bound in Table 1 because the contract at $N = 25$ (0.617) is slightly smaller than the median estimate (0.635). For lower values of ρ_H, the unconstrained contract would exceed the observed one, which cannot occur in equilibrium and would mechanically reduce welfare.

nantly rises, the indirect effect is negative and grows in magnitude. The resulting welfare loss is amplified at lower information–risk ratios; Appendix Figure A7 for how welfare effects (i) and (ii) vary with this ratio.

Overall, the welfare loss generated by the interaction between moral hazard and imperfect competition is sizable relative to the gains from eliminating moral hazard in isolation and is robust across specifications. The effect is strongest when clients are less risk-tolerant than dealers, an empirically plausible configuration. This suggests that analyses that abstract from limited competition can substantially understate the severity of moral-hazard frictions.

Take away. Overall, the quantification suggests that inefficiency could be driven by the interaction between moral hazard and imperfect competition rather than by either friction alone. Understanding how this interaction varies with market conditions requires richer data, pointing to a natural direction for future research.

4 Conclusion

I develop a model in which imperfect competition and moral hazard jointly shape trading outcomes and welfare. Using proprietary trade data from Canadian equity exchanges, I illustrate that the model can help interpret observed price differences and provides a framework for thinking about the relative importance of each friction. The results imply that effective policy must address competition and incentives jointly.

References

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.

Du, S. and H. Zhu (2017). What is the optimal trading frequency in financial markets? *Review of Economic Studies 84*, 1606–1651.

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.

Goetzmann, W. N., J. E. Ingersoll, and S. A. Ross (2003). High-water marks and hedge fund management contracts. *Journal of Finance 58*(4), 1685–1718.

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.

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

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

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.

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.

Malamud, S. and M. Rostek (2017). Decentralized exchange. *American Economic Review 107*(11), 3320–3362.

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.

Rosket, M. and M. Weretka (2015). Dynamic thin markets. *Review of Financial Studies 28*, 2946–2992.

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

Rostek, M. and M. Weretka (2012). Price inference in small markets. *Econometrica 80*(2), 687–711.

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 (2026). Market power and capital constraints. American Economic Review.

Wittwer, M. and A. Uthemann (2024). Specialization in financial markets. Working paper.

APPENDIX

Moral Hazard and Imperfect Competition in Financial Markets

by Milena Wittwer

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

A Additional theoretical evidence

Section A.1 provides existence and uniqueness conditions.
Section A.2 provides model extensions.
Section A.3 provides sufficient conditions for the equilibrium contract regime.
Section A.4 provides an alternative decomposition of the forces driving $\Phi(N)$.
Section A.5 analyses when imperfect competition is worse than moral hazard.
Section A.6 presents mathematical auxiliary statements and all proofs.

A.1 Existence and uniqueness

In this appendix, I first focus on the simpler environment with risk-neutral clients, before considering risk-averse clients.

A.1.1 Risk-neutral clients

When clients are risk-neutral ($\rho_H \to \infty$) the contract of Proposition 1 collapses to $\Phi(N)$. The unconstraint contract would leave the dealer with a zero share ($\frac{\rho_I}{\rho_I+\rho_H} \to 0$), in which case the dealer would not want to exert effort.

This equilibrium exists under the sufficient conditions provided in 3. These conditions conditions do not rely on auxiliary parameter Δ, and are consistent with Corollary 2. When $b < \gamma_\infty(\sigma_\epsilon) = \lim_{N\to\infty} \underline{b}(N)$, the lower bound condition $b \geq \underline{b}$ is never satisfied as $N \to \infty$, meaning the incentive constraint never binds, consistent with part (ii) of the corollary. Instead, when $b > \gamma_\infty(\sigma_\epsilon) = \lim_{N\to\infty} \underline{b}(N)$, there exists no finite effort-inducing contract in a perfectly competitive market, because as $N \to \infty$, $\Phi(N) \to 0$ and $\overline{b}(\Phi(N)) \to -\infty$. Once $\overline{b} < \underline{b}$ no b satisfies both sufficient conditions simultaneously.

Proposition 3. *Set $\Delta = 0$ and $\rho_H \to \infty$. The effort-inducing equilibrium characterized in Proposition 1 exists if $b \in [\underline{b}, \overline{b}]$, where:*

(i) The lower bound $\underline{b}$ is the minimum benefit from shirking required for the incentive constraint to bind:

$$\underline{b} \equiv -\frac{\beta(N)}{\gamma(N)},$$

with $\beta(N)$ and $\gamma(N)$ defined in Proposition 1.

(ii) The upper bound $\overline{b}$ is the maximum benefit from shirking consistent with the client preferring the effort-inducing contract over offering a zero contract share:

$$\overline{b} \equiv f(\Phi(N), \Phi(N)) - g(0, \Phi(N)),$$

where $f(\phi_i, \phi)$ and $g(\phi_i, \phi)$ denote the client's expected payoff (net of $\bar{U}$) when the dealer exerts effort and shirks, respectively:

$$f(\phi_i, \phi) \equiv \mathbb{E}[\mathcal{W}(\phi_i; 0)] + \mathbb{E}[\mathcal{U}(0; \phi_i)], \tag{29}$$

$$g(\phi_i, \phi) + b \equiv \mathbb{E}[\mathcal{W}(\phi_i; 1)] + \mathbb{E}[\mathcal{U}(1; \phi_i)]. \tag{30}$$

Note that $\overline{b}$ is only implicitly defined as a function of b since $\Phi(N)$ depends on b.

Under condition $(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_\nu < 0$, there exists a value of b satisfying both conditions (i) and (ii).

The equilibrium exists when the dealer's benefit from shirking, b, lies within an interval: $[\underline{b}, \overline{b}]$. Condition (i), which determines $\underline{b}$, ensures that the dealer has incentives to shirk. If the private benefit from shirking, b, is too small, the dealer would instead exert effort even under a zero contract. To see this, note that when all clients choose ϕ^*, the dealer's expected payoff from exerting effort relative to shirking is

$$\mathbb{E}[\mathcal{U}(0; \phi^*)] - \mathbb{E}[\mathcal{U}(1; \phi^*)] = \frac{\alpha(N)}{\gamma(N)}(\phi^* A)^2 - \frac{\beta(N)}{\gamma(N)} - b.$$

If $b < -\beta(N)/\gamma(N)$, this difference is positive even at $\phi^* = 0$. In that case, exerting effort strictly dominates shirking for all feasible contracts, so the incentive constraint never binds and the equilibrium contract is always unconstrained.

Condition (ii), which determines $\overline{b}$ says that the dealer's benefit cannot be too high. The highest benefit, $\overline{b}$, is the one at which the client is indifferent between choosing the

equilibrium contract (in response to all other clients choosing it), and earning expected payoff $f(\Phi(N), \Phi(N)$, and deviating to the best shirking-inducing contract. The best deviation is to offer the dealer a zero share, which leaves the client with $g(0, \Phi(N)) + \overline{b}$.

The interval $[\underline{b}, \overline{b}]$ is non-empty when signal noise is sufficiently large relative to the variance of the dealer's signal under effort:

$$\sigma_\epsilon > \frac{(N-2)^2}{3N-4}\Sigma_\nu.$$

Intuitively, when signal noise is high, the information gain from inducing effort is large so that no client finds it profitable to deviate by tolerating shirking while all others incentivize effort.

A.1.2 Risk-averse clients

When clients are risk-averse, proving equilibrium existence is more challenging than in the risk-neutral case. The reason is that the unconstrained optimal contract no longer assigns the dealer a zero share; instead, it achieves optimal risk-sharing between client and dealer at $\phi_{UC} = \frac{\rho_I}{\rho_I + \rho_H}$. This makes it difficult to rule out profitable unilateral deviations from the effort-inducing contract, since the optimal deviation depends on the equilibrium contract share and all other model parameters in a way that does not simplify cleanly.

To restore tractability, I assume that clients suffer a loss $\Delta \geq 0$ when their dealer shirks. This follows He and Krishnamurthy (2012), who assume that when the specialist (in our model the dealer) shirks, the dollar return delivered to households (here clients) falls, and impose that the loss is sufficiently large to guarantee that incentivizing effort is always optimal for households. In practice, such a loss could reflect reputational costs, regulatory penalties, or administrative burdens arising from poor dealer performance.

Proposition 4 provide existence conditions for the symmetric effort-inducing equilibrium (Proposition 1). Proposition 5 specifies conditions under which there is no symmetric equilibrium with linear demand in which dealers shirk.

Proposition 4. *Define, $f(\phi_i, \phi)$ is the payoff the client expects (net of $\bar{U}$) when choosing some contract ϕ_i in response to everyone else choosing some contract, ϕ, and the dealer exerts effort; while the clients expects payoff $g(\phi_i, \phi) + b - \Delta$ when the dealer shirks instead. Formally:*

$$f(\phi_i, \phi) \equiv \mathbb{E}[\mathcal{W}(\phi_i; 0)] + \mathbb{E}[\mathcal{U}(0; \phi_i)], \tag{29'}$$

$$g(\phi_i, \phi) + b - \Delta \equiv \mathbb{E}[\mathcal{W}(\phi_i; 1)] + \mathbb{E}[\mathcal{U}(1; \phi_i)]. \tag{30'}$$

The effort-inducing equilibrium characterized in Proposition 1 exists when (i) the dealer's benefit from shirking is sufficiently high

$$b \geq -\frac{\beta(N)}{\gamma(N)},$$

with $\beta(N)$ and $\gamma(N)$ defined in Proposition 1; and (ii) and the client's loss from shirking is sufficiently high:

$$\Delta \geq g\left(\phi_{UC}, \Phi(N)\right) - f\left(\Phi(N), \Phi(N)\right) + b \qquad \text{for } \phi_{UC} < \Phi(N), \tag{31}$$

$$\text{and } \Delta \geq g\left(\tilde{\phi}_i^S(\phi_{UC}), \phi_{UC}\right) - f\left(\phi_{UC}, \phi_{UC}\right) + b \qquad \text{for } \tilde{\phi}_i^S(\phi_{UC}) < \phi_{UC}. \tag{32}$$

Here ϕ_{UC} is the unconstrained equilibrium contract, $\frac{\rho_I}{\rho_I+\rho_H}$, $\Phi(N)$ is the constrained equilibrium contract in Proposition 1, and $\tilde{\phi}_i^S(\phi)$ is the contract that makes the dealer indifferent between shirking and effort when all other clients choose contract ϕ.

Condition (i) is the same condition as with risk-neutral clients (Proposition 3): It ensures that the dealer has incentives to shirk for sufficiently large contract shares.

Condition (ii) compares the client's expected payoff from the equilibrium contract ϕ^*—under which the dealer exerts effort—to the highest payoff the client could obtain by deviating to a shirking-inducing contract when all other clients choose ϕ^*. The condition requires that the loss from shirking, Δ, is large enough to offset the maximal gain from such a deviation. When this holds, no client has a profitable deviation to inducing shirking given ϕ^*, and the effort-inducing equilibrium of Proposition 1 exists.

In principle, condition (ii) could be expressed solely in terms of the dealer's private benefit b: as with risk-neutral clients, b must be neither too small (otherwise shirking would never be attractive and the equilibrium contract would collapse to the unconstrained risk-sharing rule $\phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$) nor too large, as clients would otherwise find it profitable to deviate toward contracts that tolerate shirking. In practice, however, establishing that a non-empty interval of b values supports equilibrium existence is difficult: since the constrained contract $\Phi(N)$ itself depends on b, the existence conditions become a fixed-point problem that does not admit a clean analytical solution. Introducing the client's loss $\Delta \geq 0$ as an auxiliary parameter sidesteps this difficulty, since Δ does not enter the equilibrium contract and therefore provides a tractable sufficient condition for ruling out shirking deviations.

Proposition 5 specifies conditions under which there is no symmetric equilibrium with linear demand in which dealers shirk.

Proposition 5. *Let $\Delta \geq b$. In any symmetric equilibrium with linear demand dealers exert effort.*

When $\Delta \geq b$ a client always finds it profitable to deviate away from any contract that induces shirking. The reason is intuitive: when the client's loss from shirking Δ is at least as large as the dealer's private benefit b, the client bears at least as much from shirking as the dealer gains. Since effort improves the quality of the dealer's information and leads to better trading outcomes for both parties, the client can always do better by offering a contract that induces effort instead. This holds regardless of what contract the client is currently offering and regardless of what other dealers are doing, so no shirking equilibrium can survive as a symmetric equilibrium.

A.2 Model extensions

I provide three model extensions. Appendix A.2.1 allows for unknown common values. Appendix A.2.2 introduces a continuous effort choice. Appendix A.2.3 sketches out what happens when allowing for heterogeneous degrees of risk-aversion. Appendix A.2.4 presents a framework with client and dealer accounts.

A.2.1 Common value

The baseline model assumes that the institution-specific asset's expected return, v_i, is drawn iid from a distribution with commonly known prior, ν. This shuts off common-value concerns. In this appendix, I relax this assumption by introducing an unknown common component v_0 shared across all market participants.

This extension is natural for two reasons. First, dealers in the same market are exposed to common underlying factors, such as interest rates, exchange rates, or aggregate credit conditions, so that their expected returns naturally co-move. Second, and more importantly, treating v_0 as unknown gives the effort decision a richer interpretation: exerting effort means conducting thorough market research that reveals both the common component v_0 and the idiosyncratic component $\tilde{v}_i$, while shirking means forgoing this research and observing only a noisy signal that conflates the two. This creates an asymmetry between effort and shirking with respect to the informational role of the clearing price, which I exploit to characterize how common values affect the effort-inducing contract below.

Modified setting. Returns are normally distributed, $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, with known variance $\Sigma > 0$. The expected return now has two components:

$$\nu_i = \nu_0 + \tilde{\nu}_i,$$

where $\nu_0 \sim \mathcal{N}(\bar{\nu}, \Sigma_0)$ is an unknown common value shared by all dealers, and $\tilde{\nu}_i \sim \mathcal{N}(0, \Sigma_\nu)$ is an idiosyncratic component drawn independently across dealers. As in the baseline, a dealer who exerts effort observes $\mu_i(0) = \nu_i$ exactly, while a dealer who shirks observes only the noisy signal $\mu_i(1) = \nu_i + \epsilon_i$. Effort therefore represents thorough market research that reveals all payoff-relevant information about ν_i (the sum of the common and idiosyncratic component) while shirking leaves the dealer with an imprecise and undifferentiated signal. When $\Sigma_0 = 0$, the common value is known and uncertainty is purely idiosyncratic, recovering the baseline model.

Equilibrium. There are two important changes in this setting relative to the baseline. First, the ν_i's are no longer independent. Writing $\nu_i = \nu_0 + \tilde{\nu}_i$, any two dealers' expected returns share the common component ν_0, so:

$$Cov(\nu_i, \nu_j) = \Sigma_0 > 0, \quad i \neq j.$$

This correlation affects the moments used to compute expected utilities: for example, $\mathbb{E}[\nu_i^2] = \bar{\nu}^2 + \Sigma_0 + \Sigma_\nu$ and $\mathbb{E}[\nu_i \nu_j] = \bar{\nu}^2 + \Sigma_0$ for $i \neq j$, both of which exceed their baseline counterparts by Σ_0. The correlation among the ν_i's thus affects the equilibrium contract through its impact on dealers' expected trading profits.

Second, inference is asymmetric depending on whether the dealer exerts effort or shirks. Under effort ($s_i = 0$), dealer i observes $\mu_i(0) = \nu_i = \nu_0 + \tilde{\nu}_i$ exactly. Since $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$ by assumption, knowing ν_i is sufficient for the distribution of R_i, and the clearing price p^c is uninformative:

$$\mathbb{E}[R_i \mid \mu_i(0), p^c] = \nu_i, \qquad Var[R_i \mid \mu_i(0), p^c] = \Sigma.$$

The inference channel stemming from unknown common values is therefore dormant under effort: having already learned both ν_0 and $\tilde{\nu}_i$ through their research, the dealer has no residual uncertainty that the price could resolve.

Under shirking ($s_i = 1$), dealer i observes only the noisy signal $\mu_i(1) = \nu_0 + \tilde{\nu}_i + \epsilon_i$, which does not reveal ν_i perfectly. In a linear equilibrium, the clearing price p^c can be

inverted to recover:

$$\hat{\nu}_{-i} \equiv \nu_0 + \frac{1}{N-1}\sum_{j\neq i}\tilde{\nu}_j \sim \mathcal{N}\left(\nu_0, \frac{\Sigma_\nu}{N-1}\right),$$

which is a noisy signal of ν_0 and hence informative about $\nu_i = \nu_0 + \tilde{\nu}_i$. The shirking dealer therefore uses both $\mu_i(1)$ and $\hat{\nu}_{-i}(p^c)$ to form their posterior about ν_i. Projecting ν_i onto $(\mu_i(1), \hat{\nu}_{-i})$ via standard normal updating:

$$\begin{aligned}\mathbb{E}[R_i \mid \mu_i(1), p^c] &= \mathbb{E}[\nu_i \mid \mu_i(1), \hat{\nu}_{-i}(p^c)] = c_0\bar{\nu} + c_s\mu_i(1) + c_p\hat{\nu}_{-i}(p^c),\\ Var[R_i \mid \mu_i(1), p^c] &= \tilde{\Sigma},\end{aligned}$$

where c_0, c_s, c_p are the projection coefficients onto $(\mu_i(1), \hat{\nu}_{-i})$—not onto $(\mu_i(1), p^c)$ directly—and $\tilde{\Sigma}$ is the posterior variance. These are derived in Lemma 2 in Appendix A.6. Since p^c is an invertible affine function of $\hat{\nu}_{-i}$ in a linear equilibrium, conditioning on p^c is equivalent to conditioning on $\hat{\nu}_{-i}$, and the conditional mean is linear in $(\mu_i(1), p^c)$ with c_p understood as the projection coefficient on $\hat{\nu}_{-i}$ rather than on the raw price level. When $\Sigma_0 > 0$, the inference channel is active under shirking and only under shirking: the shirking dealer, having failed to learn ν_0 through research, can partially recover it from the price, so that $c_p > 0$. When $\Sigma_0 = 0$, $c_p \to 0$ and the baseline is recovered.

Proposition 6. *In a symmetric equilibrium with common values ($\Sigma_0 > 0$):*
(i) All clients offer contract $\{K^*, \phi^*\}$:

$$\phi^* = \begin{cases} \Phi_{CV}(N) \equiv \frac{1}{|A|}\sqrt{\frac{\beta_{CV}(N)+b}{\alpha_{CV}(N)}} & \text{when } \frac{\rho_I}{\rho_I+\rho_H} \leq \Phi_{CV}(N),\\ \frac{\rho_I}{\rho_I+\rho_H} & \text{otherwise,}\end{cases} \tag{33}$$

and

$$K^* = \bar{U} - \frac{(N-2)\rho_I\big(2(N-1)\Sigma_0 + N\Sigma_\nu\big)}{2(N-1)N\Sigma}, \tag{34}$$

where

$$\alpha_{CV}(N) \equiv \frac{1}{2N\rho_I}\left(\frac{N-2}{\Sigma} + \frac{N}{\sigma_\epsilon} + \frac{N-1}{\Sigma_\nu} + \frac{1}{N\Sigma_0+\Sigma_\nu}\right)^{-1} \tag{35}$$

$$\beta_{CV}(N) \equiv -\frac{(N-2)\rho_I\big(2(N-1)\Sigma_0 + N\Sigma_\nu\big)}{2(N-1)N\Sigma} + \frac{(N-2)\rho_I\,\Sigma_\nu^2 R}{2(N-1)\,\mathcal{D}Q}, \tag{36}$$

with

$$R \equiv (N-1)\Sigma_0\big(2N^2\Sigma_0^2 + 2N\Sigma_0\sigma_\epsilon + N^2\Sigma_0\sigma_\epsilon + \sigma_\epsilon^2\big) + \big(-4N\Sigma_0^2 + 4N^2\Sigma_0^2 + N^3\Sigma_0^2 - 2\Sigma_0\sigma_\epsilon + 2N\Sigma_0\sigma_\epsilon + 2N^2\Sigma_0\sigma_\epsilon + \sigma_\epsilon^2\big)\Sigma_\nu + \big(-2\Sigma_0 + 2N\Sigma_0 + 2N^2\Sigma_0 + \sigma_\epsilon + N\sigma_\epsilon\big)\Sigma_\nu^2 + N\Sigma_\nu^3, \tag{37}$$

$$Q \equiv (N-1)\Sigma_0\sigma_\epsilon + N\Sigma_0\Sigma_\nu + \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu), \tag{38}$$

$$\mathcal{D} \equiv (N-1)N\Sigma\Sigma_0\sigma_\epsilon + N\big((\Sigma - 2\Sigma_0)\sigma_\epsilon + N\Sigma_0(\Sigma + \sigma_\epsilon)\big)\Sigma_\nu + \big(N(\Sigma + \sigma_\epsilon) - 2\sigma_\epsilon\big)\Sigma_\nu^2. \tag{39}$$

When $\Sigma_0 = 0$*:* $Q \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$*,* $\mathcal{D} \to \Sigma_\nu\big(N\Sigma(\sigma_\epsilon + \Sigma_\nu) + (N-2)\sigma_\epsilon\Sigma_\nu\big)$*,* $R \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)(\sigma_\epsilon + N\Sigma_\nu)$*, so that* $\alpha_{CV}(N) \to \alpha(N)$*,* $\beta_{CV}(N) \to \beta(N)$*, and* $\Phi_{CV}(N) \to \Phi(N)$*, recovering Proposition 1.*

(ii) All dealers exert effort, $s^* = 0$*, and submit the following demand function:*

$$a^*(p, \mu_i(s^*)) = \rho_I\Big(\Lambda^*\rho_I + \phi^* Var[R_i \mid \mu_i(s^*)]\Big)^{-1}\Big(\mathbb{E}[R_i \mid \mu_i(s^*)] - p\Big), \tag{40}$$

$$\text{with } \Lambda^* = \frac{\phi^*\Sigma}{\rho_I(N-2)}, \tag{41}$$

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

The equilibrium structure mirrors the baseline in several respects, but the optimal contract differs in important ways. I discuss each in turn. Under effort, the equilibrium demand function and price impact are identical to the baseline: $\Lambda^* = \frac{\phi^*\Sigma}{\rho_I(N-2)}$ depends only on Σ, and dealers trade on their exact signal ν_i. This is a direct consequence of the dormant inference channel under effort.

Turning to the contract, the effort-inducing contract share $\Phi_{CV}(N)$ differs from the baseline $\Phi(N)$ through two channels, both operating via the common value parameter Σ_0. The first channel operates through K^*. The fixed payment now compensates the dealer for a larger information rent: compared to the baseline, K^* is reduced by an additional term $\frac{(N-2)\rho_I\Sigma_0}{N\Sigma}$, reflecting the fact that correlated ν_i's increase dealers' expected trading profits through the Σ_0 component of $\mathbb{E}[\nu_i^2]$ and $\mathbb{E}[\nu_i\nu_j]$.

The second channel operates through ϕ^*. The coefficients $\alpha_{CV}(N)$ and $\beta_{CV}(N)$ replace their baseline counterparts $\alpha(N)/\gamma(N)$ and $\beta(N)/\gamma(N)$, capturing the effect of common values on the incentive compatibility constraint. Two forces are at work. On one hand, the active inference channel under shirking allows the shirking dealer to partially recover ν_0 from the price, changing the profitability of shirking relative to effort. On the other hand, the correlation among ν_i's alters the variance of trading profits under both effort and shirking. The net effect on ϕ^* is in general ambiguous without further

parametric restrictions.

Take away. While less tractable than the baseline, the extension confirms the main message of the paper: the contract a client chooses in anticipation of market clearing depends on how many market participants interact in market clearing.

A.2.2 Continuous effort

The baseline model assumes binary effort, which implies that whenever the incentive constraint is slack, the contract is independent of the number of dealers. With continuous effort, this is no longer the case: the contract affects the effort level chosen by the dealer, and therefore always depends on N. I adopt the binary effort model as the baseline for tractability, though the continuous effort extension lends itself to numerical analysis.

Modified setting. The setting is identical to Section 2, with one modification: instead of choosing between effort and shirking, $s_i \in \{0,1\}$, each dealer chooses a continuous effort level $e_i \in [0,\bar{e}]$ for some $\bar{e} > 0$.

Effort determines the precision of the dealer's private signal: a dealer exerting effort e_i observes

$$\mu_i(e_i) = \nu_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon(e_i)),$$

where $\sigma_\epsilon(e_i) = 1/e_i$ is decreasing in effort. Higher effort thus yields a more precise signal. Exerting effort e_i costs the dealer $k(e_i) = \frac{c}{2}e_i^2$, with $c > 0$, which is subtracted from the dealer's certainty equivalent:

$$\mathcal{U}(e_i;\phi_i) = \phi_i\big(\mathbb{E}[R_i \mid \mu_i(e_i)] - p^c(e_i)\big)a_i^c(e_i) - \phi_i^2\frac{1}{2\rho_I}a_i^c(e_i)Var[R_i \mid \mu_i(e_i)]a_i^c(e_i) - k(e_i). \quad (42)$$

The rest of the model is unchanged. It suffices to replace s_i with e_i in all equations.

Equilibrium. The trading stage is unchanged: given effort level e_i and contract ϕ_i, the dealer's optimal demand function takes the same form as in Proposition 1 (ii), with

$$Var[R_i \mid \mu_i(e_i)] = \tilde{\Sigma}(e_i), \text{ where } \tilde{\Sigma}(e_i) = \Sigma + \left(\Sigma_\nu^{-1} + e_i\right)^{-1} \text{ and } \Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\tilde{\Sigma}(e^*).$$

In the effort stage, the dealer chooses e_i to maximize their expected certainty equivalent, taking the contract ϕ_i and all other dealers' strategies as given. The first-order condition for the dealer's optimal effort level e^* is:

$$\frac{\partial}{\partial e_i}\mathbb{E}[\mathcal{U}(e_i;\phi_i)]\bigg|_{e_i=e^*} = 0, \quad \text{(EOC)}$$

where

$$\mathbb{E}\left[\mathcal{U}(e_i;\phi_i)\right] = \phi_i \frac{A^2(-1+N)\phi^2\tilde{\Sigma}(e)^2(\sigma_\epsilon(e_i)+\Sigma_\nu)+(-2+N)^2\rho_I^2\Sigma_\nu(+N\Sigma_\nu))}{2(-2+N)(-1+N)\rho_I(+\Sigma_\nu)(2\phi\tilde{\Sigma}(e)+(-2+N)\phi_i\tilde{\Sigma}(e_i))} - k(e_i), \quad (43)$$

with $\tilde{\Sigma}(e_i) = \Sigma + \left(\Sigma_\nu^{-1} + e_i\right)^{-1}$ and $\sigma_\epsilon(e_i) = 1/e_i$. Note that $\mathbb{E}\left[\mathcal{U}(e_i;\phi_i)\right]$ is strictly concave. Therefore, the first-order condition pins down the dealer's effort-choice taken as given their own contract, ϕ_i, and the contracts of all other clients, ϕ, as well as the effort-choice of the other dealers, e, which I denote by $e_i(\phi_i;\phi,e)$.

The client anticipates $e_i(\phi_i;\phi,e)$ and chooses ϕ_i to maximize their payoff. Unlike the baseline model, there is no explicit incentive constraint: in the baseline, the client must ensure the dealer prefers effort over shirking, which requires ϕ_i to be large enough to make effort weakly dominant. Here, since effort is continuous and the dealer always chooses positive effort given any contract ϕ_i through (EOC), the client need not impose a lower bound on ϕ_i to induce effort. Instead, the contract affects the level of effort chosen by the dealer, and the client uses ϕ_i to induce the desired effort level. The client therefore faces only the participation constraint:

$$\mathbb{E}[\mathcal{U}(e^*(\phi_i;\phi);\phi_i)] + K_i \geq \bar{U}. \quad \text{(IR}')$$

Substituting the participation constraint (IR′) in the client's objective, the optimization problem becomes:

$$\max_{\phi_i} \left\{ \mathbb{E}[\mathcal{W}(\phi_i; e_i(\phi_i;\phi,e))] + \mathbb{E}[\mathcal{U}(e_i(\phi_i;\phi,e);\phi_i)] - \bar{U} \right\}. \quad \text{(M}')$$

Since we know how dealers trade in the third stage, we can compute the client's expected payoff for any e_i:

$$\begin{aligned}\mathbb{E}[\mathcal{W}(\phi_i;e_i)] + \mathbb{E}[\mathcal{U}(e_i;\phi_i)] = &\left(A^2(N-1)\phi^2\tilde{\Sigma}(e)^2(\sigma_\epsilon(e_i)+\Sigma_\nu) + (N-2)^2\rho_I^2\Sigma_\nu(\sigma_\epsilon(e_i)+N\Sigma_\nu)\right) \\ &\times \left(2\phi\rho_H\tilde{\Sigma}(e) - (N-2)\left(\rho_I + (\phi_i-2)\phi_i(\rho_H+\rho_I)\right)\tilde{\Sigma}(e_i)\right) \\ &\times \left(2(N-2)(N-1)\rho_H\rho_I(\sigma_\epsilon(e_i)+\Sigma_\nu)\left(2\phi\tilde{\Sigma}(e) + (N-2)\phi_i\tilde{\Sigma}(e_i)\right)^2\right)^{-1} \\ &- k(e_i). \end{aligned} \quad (44)$$

The first-order condition with respect to ϕ_i, where I abbreviate $e_i(\phi_i;\phi,e)$ by e_i, is

$$\frac{\partial}{\partial\phi_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;e_i)] + \mathbb{E}[\mathcal{U}(e_i;\phi_i)]\right\} + \frac{\partial}{\partial e_i}\left\{\mathbb{E}[\mathcal{W}(\phi_i;e_i)]\right\} \times \frac{\partial e_i}{\partial\phi_i} = 0. \quad (45)$$

The first line captures the direct effect of ϕ_i on risk-sharing, holding effort fixed, and is

maximized at $\phi_i = \frac{\rho_I}{\rho_I+\rho_H}$. The second line captures the effort inducement effect: the client distorts ϕ_i away from the risk-sharing optimum to induce their optimal effort. The optimal contract therefore balances risk-sharing against effort inducement, and does not generally equal $\frac{\rho_I}{\rho_I+\rho_H}$. Since e_i is the root of a cubic polynomial, solving analytically for the contract ϕ_i satisfying the first-order condition (45) is not tractable, but the solution is straightforward to obtain numerically, as is verifying that it constitutes a maximum.

At this point, the dealer's trading behavior and effort choice are characterized as functions of ϕ_i, ϕ, and e, and the client's optimal contract is characterized as a function of ϕ and e. Imposing symmetry, the equilibrium requires $e_i(\phi^*;\phi^*,e^*) = e^*$ and $\phi_i(\phi^*,e^*) = \phi^*$, and the equilibrium lump-sum payment K^* follows from the binding participation constraint. The equilibrium contract ϕ^* therefore depends on the equilibrium effort level e^*, even though effort is unobservable. This is not problematic since the contract is chosen ex ante, before any realization of signals or trading outcomes: the client anticipates the dealer's effort response when designing the contract, without being later able to verify it.

Take away. While significantly less tractable than the baseline (with no simple closed-form solution), this extension brings a key insight: the optimal contract is generally a function of N, since both the dealer's and client's expected payoffs depend on the number of dealers through market clearing.

A.2.3 Heterogeneous risk tolerance

The baseline model assumes that all dealers share the same risk tolerance ρ_I and all clients share the same risk tolerance ρ_H. This assumption delivers a tractable symmetric equilibrium but abstracts from an important dimension of heterogeneity: in practice, dealers and clients differ substantially in their capacity to bear risk, and these differences are likely to affect both the optimal contract design and the degree of market power. In this appendix, I illustrate how to extent the baseline model to allow for fully heterogeneous risk tolerances $\{\rho_{Ii}\}_i$ across dealers and $\{\rho_{Hi}\}_i$ across clients, so that each client-dealer pair i is characterized by a distinct pair (ρ_{Ii},ρ_{Hi}).

Equilibrium incentives. I build on Proposition 1 of Malamud and Rostek (2017). In the trading stage, each dealer i effectively has risk aversion $\frac{\phi_i}{\rho_{Ii}}$, since the contract share ϕ_i scales the dealer's exposure to trading profits. Applying Proposition 1 of Malamud

and Rostek (2017) with $\alpha^i = \frac{\phi_i}{\rho_{Ii}}$, the price impact of dealer i is:

$$\Lambda_i = \frac{\psi_i \phi_i}{\rho_{Ii}} \Sigma, \tag{46}$$

where ψ_i is a dealer-specific scalar:

$$\psi_i = \frac{2}{\frac{\kappa\phi_i}{\rho_{Ii}} - 2 + \sqrt{\left(\frac{\kappa\phi_i}{\rho_{Ii}}\right)^2 + 4}}, \tag{47}$$

and $\kappa \in \mathbb{R}_+$ is the unique positive solution to:

$$\sum_i \left(\frac{\kappa\phi_i}{\rho_{Ii}} + 2 + \sqrt{\left(\frac{\kappa\phi_i}{\rho_{Ii}}\right)^2 + 4} \right)^{-1} = \frac{1}{2}. \tag{48}$$

Note that κ now depends on all contracts $\{\phi_i\}_i$ through equation (48), so the price impacts $\{\Lambda_i\}_i$ and contracts $\{\phi_i^*\}_i$ are jointly determined. In the symmetric case $\phi_i = \phi^*$ and $\rho_{Ii} = \rho_I$ for all i, $\psi_i = \frac{1}{N-2}$ and $\Lambda_i = \frac{\phi^* \Sigma}{\rho_I (N-2)}$, recovering the baseline price impact.

Given $\{\Lambda_i\}_i$, the demand functions under effort and shirking follow from the first-order condition (10) as in the baseline, with Λ_i now dealer-specific. The clearing prices $p^c(0)$ and $p^c(1)$ follow from market clearing, and the dealer's expected certainty equivalents $\mathbb{E}[\mathcal{U}(0;\phi_i)]$ and $\mathbb{E}[\mathcal{U}(1;\phi_i)]$ are computed by taking expectations over signals as in the baseline proof.

The participation constraint binds at:

$$K_i^* = \bar{U}_i - \mathbb{E}[\mathcal{U}(0;\phi_i^*)], \tag{49}$$

and the unconstrained contract maximizes $\mathbb{E}[\mathcal{W}(\phi_i;0)] + \mathbb{E}[\mathcal{U}(0;\phi_i)] - \bar{U}_i$ over ϕ_i, taking other dealers' contracts $\{\phi_j\}_{j\neq i}$ as given. Unlike the baseline, this unconstrained optimum does not admit a simple closed form, as ϕ_i enters both the dealer's effective risk aversion $\frac{\phi_i}{\rho_{Ii}}$ and the price impact Λ_i through the scalar κ in (48). The constrained contract is pinned down by the binding incentive compatibility constraint and takes the form:

$$\phi_i^* = \Phi_i\left(N, \rho_{Ii}, \rho_{Hi}, \{\phi_j^*, \rho_{Ij}, \rho_{Hj}\}_{j\neq i}\right), \tag{50}$$

where $\Phi_i(\cdot)$ depends on all other dealers' contracts and risk tolerances through the demand slopes, price impacts, and aggregate market quantities.

Unlike the symmetric baseline, this system does not admit a closed-form solution in general. The coupling between contracts $\{\phi_i^*\}_i$ through the price impact scalar κ, and between price impacts through the contracts, makes the fixed point analytically intractable. The equilibrium could however be computed numerically.

Take away. The main message of this extension is that the interaction between moral hazard frictions and imperfect competition is robust to heterogeneity in risk tolerance. While the closed-form characterization of the baseline no longer obtains (because contracts, price impacts, and risk tolerances are jointly determined across all client-dealer pairs) the equilibrium contract ϕ_i^* of each pair continues to depend on the number of market participants N through the price impact scalar κ. The degree of competition in the asset market therefore remains a key determinant of the optimal incentive contract, even when dealers and clients are asymmetric.

A.2.4 Client and dealer accounts

I extend the baseline model to allow (some or all) dealers to trade on both a client account and an inventory account, each clearing in a separate auction with separate objective functions.

This approach is motivated by three considerations. First, client and inventory books are institutionally ring-fenced by mandate, client-priority rules, and distinct risk budgets, making joint optimization across accounts operationally costly and often infeasible. Second, separate auctions generate separate equilibrium prices p^{CL*} and p^{IN*}, consistent with the observed price differences across client and inventory trades in the data. Third, allowing dealers to jointly optimize across both accounts in a single auction would render the trading game asymmetric, since dealers would effectively act as two distinct types of bidders—trading for themselves and for their clients—with different objective functions. Asymmetric uniform-price auctions do not generally admit closed-form solutions, precluding an analytical characterization of the trading stage.

Modified setting. There are $N_{CL} > 2$ dealers participating in the client auction (given supply A_{CL}) and $N_{IN} > 2$ in the inventory auction (given supply A_{IN}). Since the two accounts are optimized independently, the model is silent on whether the same dealers participate in both auctions or not.

In the client auction, the setting is identical to the baseline model. Dealer i trades on behalf of client i, and maximizes certainty equivalent (6). The equilibrium of the client auction follows directly from Proposition 1 with supply A_{CL}. From Corollary 1 we know

that the market clears at

$$p^*_{CL} = \frac{1}{N_{CL}} \sum_i v_{i,CL} - \frac{\phi^*}{\rho_I} \Sigma \left(\frac{N_{CL}-1}{N_{CL}-2} \right) \frac{A_{CL}}{N_{CL}}, \tag{51}$$

where I have relabeled v_i to $v_{i,CL}$ to highlight that those signals are for the client auction.

In the inventory auction, dealer i trades for their own account, with private belief $R_i \sim \mathcal{N}(v_{i,IN}, \Sigma)$, where $v_{i,IN} = v + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma_\varepsilon)$ captures idiosyncratic reasons to trade, such as balance-sheet constraints or individual market views. These reasons are privately known to the dealer, while all other information is publically known.

The dealer chooses demand so as to maximizes their certainty equivalent:

$$\mathcal{U}_{IN} = \big(v_{i,IN} - p_{IN} \big) a_{i,IN} - \frac{\Sigma}{2\rho_I} \big(a_{i,IN} \big)^2, \tag{52}$$

where p_{IN} clears the inventory auction, and a_{IN} is the amount the dealer wins.

Since the dealer bears the full risk of their inventory position, dealer demand is like in the the baseline model but with $\phi^* = 1$ and $v_{i,IN}$ instead of v_i. Therefore, the auction clears at price

$$p^*_{IN} = \frac{1}{N_{IN}} \sum_i v_{i,IN} - \frac{1}{\rho_I} \Sigma \left(\frac{N_{IN}-1}{N_{IN}-2} \right) \frac{A_{IN}}{N_{IN}}. \tag{53}$$

When brining the model to the data I consider separate auctions for each exchange $e \in$ {TSX, Alpha}, period τ in a day t, security s, account-type $k \in$ {CL, IN}, and trade-side $l \in$ {buy, sell}. To distinguish them, I include super-and subscripts. For example, $v_{i,IN}$ in expression (53) is replaced by $v_{ti\tau sk}$ for $k = IN$. To account to systematic changes over time or across securities, I only impose $v_{ti\tau sk} \sim N(v_{ts}, \Sigma_{ts})$ to be iid across i, τ, k within a day/security.

Future work. In the paper, I present a model of dealers trading on behalf of clients and use the data to form aggregate moment conditions for the Canadian equities market. Future work could exploit several interesting aspects at the order-submission level (that I abstract from given my data). For example, a dealer's inventory position may affect their willingness to take the other side of a client trade, creating endogenous linkages between the two accounts that the model rules out by construction. Moreover, a dealer's client order flow may be informative about asset fundamentals, potentially influencing their inventory trading decisions. Finally, by treating the two auctions as independent, the model rules out strategic interactions whereby dealers use their inventory account to influence prices in the client auction, or vice versa.

A.3 Characterization of equilibrium regimes

In this appendix we characterize when the favorable versus unfavorable contract regime arises. I consider two cases in turn: risk-neutral clients who do not suffer a loss from shirking ($\rho_H \to \infty$, $\Delta = 0$) and risk-averse clients who suffer a loss ($\rho_H < \infty$, $\Delta > 0$).

When clients are risk-neutral, the contract regime is determined solely by the dealer's private benefit from shirking relative to the threshold $\gamma(\sigma_\epsilon)$ of Corollary 3. When $b > \gamma(\sigma_\epsilon)$, the incentive constraint binds and clients must offer a contract share that increases with competition, placing them in the unfavorable regime. When $b \leq \gamma(\sigma_\epsilon)$, either the incentive constraint does not bind or competition relaxes rather than tightens the incentive distortion, placing clients in the favorable regime.

Corollary 4 (Risk-neutral clients)**.** *Let $\rho_H \to \infty$, $\Delta = 0$. Define $\underline{b}, \overline{b}$ as characterized in Proposition 3, and $\gamma(\sigma_\epsilon)$ as in Corollary 3.*

(i) The unfavorable contract regime arises when $b \in [\underline{b}, \overline{b}]$ and $b > \gamma(\sigma_\epsilon)$.

(ii) The favorable contract regime arises when $b \in [\underline{b}, \overline{b}]$ and $b \leq \gamma(\sigma_\epsilon)$, or when $b < \underline{b}$.

When clients are risk-averse, whether the favorable or unfavorable contract regime arises depends on all model parameters, of which supply A plays a particularly important role for finite N. In the limit $N \to \infty$ (Corollary 2), supply becomes irrelevant: shared among infinitely many dealers, its impact on prices and allocations vanishes. For finite N, however, supply matters.

Corollary 5 (Risk-averse clients)**.** *Let $\rho_H < \infty$ and $\Delta > 0$ be high enough to guarantee equilibrium existence according to Proposition 4. Define $\bar{A}^2(\sigma_\epsilon) \equiv \max_{N>2} \frac{b\gamma(N)+\beta(N)}{\alpha(N)} \left(\frac{\rho_I+\rho_H}{\rho_I}\right)^2$ and $\gamma(\sigma_\epsilon)$ as in Corollary 3; note that $\frac{\partial \bar{A}^2(\sigma_\epsilon)}{\partial \sigma_\epsilon} < 0$.*

(i) The unfavorable contract regime arises when $A^2 < \bar{A}^2(\sigma_\epsilon)$ and $b > \gamma(\sigma_\epsilon)$.

(ii) The favorable contract regime arises when $A^2 \geq \bar{A}^2(\sigma_\epsilon)$ or $b \leq \gamma(\sigma_\epsilon)$.

A larger supply forces each dealer to absorb a larger position, and the net benefit of doing so is larger under effort than under shirking. Although a shirking dealer benefits more from the resulting price reduction, since their less aggressive demand requires a larger price adjustment to clear the market, this is outweighed by the higher risk-bearing cost of holding a larger position under shirking ($Var(R_i|\mu_i(0)) < Var(R_i|\mu_i(1))$).

Formally, given all clients choose the same contract ϕ, and all dealers other than i behave as in the equilibrium, expected payoff under effort is:

$$\mathbb{E}[\mathcal{U}(0;\phi)] = \Sigma(2N(N-2)\rho_I)^{-1}A^2\phi^2 + (N-2)\rho_I\Sigma_\nu(2(N-1)\Sigma)^{-1},$$

and expected payoff under shirking is

$$\begin{aligned}\mathbb{E}[\mathcal{U}(1;\phi)] = (\Sigma^2(\sigma_\epsilon+\Sigma_\nu))(2(-2+N)^2\rho_I\sigma_\epsilon\Sigma_\nu \\ + \Sigma(2(-2+N)N\rho_I\sigma_\epsilon + 2(-2+N)N\rho_I\Sigma_\nu))^{-1}A^2\phi^2 \\ + (-2+N)\rho_I\Sigma_\nu(\sigma_\epsilon+N\Sigma_\nu)(2(-2+N)(-1+N)\sigma_\epsilon\Sigma_\nu \\ + \Sigma(2(-1+N)N\sigma_\epsilon + 2(-1+N)N\Sigma_\nu))^{-1} + b.\end{aligned}$$

The positive coefficient on $A^2\phi^2$, which captures the marginal effect of supply on expected payoff, is larger under effort than under shirking.

The incentive constraint therefore binds only when supply is sufficiently small, with the threshold, $\bar{A}^2(\sigma_\epsilon)$, defined in Corollary 5, falling in σ_ϵ. This is because higher shirking noise makes effort relatively more attractive, so a smaller supply suffices to keep the constraint slack.

A.4 Alternative decomposition of how ϕ changes in N

Whether stronger competition (measured by the number of dealers) alleviates or amplifies moral hazard (captured by the size of the contract) depends on three intertwined forces.

(i) Price-impact effect: The first force is that greater competition reduces each dealer's individual price impact,

$$\Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\Sigma, \tag{17}$$

altering their incentive to strategically scale back demand. To be willing to exert effort, dealers need to obtain a larger fraction of wealth as soon as the incentive constraint binds.

The price-impact effect is related to, yet somewhat distinct from, the idea that increased competition compresses the rents dealers can extract from market power, thereby raising the compensation required to sustain effort. In my trade model, market-power rents depend not only on price impact but also on the information–risk ratio, which measures how informative a dealer's signal is relative to the residual uncertainty borne when

investing, Σ_ν/Σ. The influence of this ratio on market-power rents varies with the number of dealers in a way not captured by the price-impact effect.

To show this formally, I consider two alternative measures of market-power rents. The first is from the perspective of an individual dealer is to compare the expected payoff when acting strategically to the payoff when behaving as a price-taker while all other dealers remain strategic: $\Lambda^* A^2 \phi^* (2(N-1)^2 N)^{-1} + (N-2)\rho_I (2(N-1)^3)^{-1} \Sigma_\nu/\Sigma$. The second measure compares the expected payoff of a dealer (who exerts effort) when all dealers behave strategically with the payoff when all dealers behave as price-takers, holding contracts fixed: $\Lambda^* \phi^* (A/N)^2 - \rho_I (2N(N-1))^{-1} \Sigma_\nu/\Sigma$.

(ii) Price-information effect: Adding more dealers makes prices—and thus allocations—more predictable. In the limit with infinitely many dealers, market outcomes become perfectly predictable. Because contracts and effort are chosen prior to trading, changes in predictability affect both decisions.

To see this formally, consider a client i choosing a contract ϕ_i, taking all other clients' contracts as given at ϕ^*. As the number of dealers N increases, the client can better predict the clearing price—and thus its market-clearing allocation—since, ceteris paribus, the variance of the clearing price declines in N: $Var(p^c(0)) = (2\phi^* + (N-2)\phi_i)^{-2}(\phi^{*2} + (\phi^* + (N-2)\phi_i)^2 (N-1)^{-1})\Sigma_\nu$.

To highlight the role of the price-information effect, I shut down the price-impact effect by assuming that dealers behave as price-takers and derive the resulting equilibrium in Proposition 7 below. The constrained contract in that equilibrium, $\tilde{\Phi}(N)$, shares similarities to the contract with strategic demand reduction, $\Phi(N)$, emphasizing the central role of the price-information effect in the interaction between moral hazard and imperfect competition.

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

Proposition 7. *In a symmetric equilibrium:*

(i) All clients offer the following contract $\{K^*, \phi^*\}$*:*

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

where

$$\Phi_C(N) = \frac{1}{|A|}\sqrt{\frac{\tilde{\beta}(N) + b\tilde{\gamma}(N)}{\tilde{\alpha}(N)}} \tag{54}$$

$$\Phi_U(N) = \frac{1}{2}\left(\frac{\sqrt{(\rho_H+\rho_I)((N-2)^2\rho_H + N^2\rho_I)}}{\rho_H+\rho_I} - (N-2)\right) > 0, \tag{55}$$

and

$$K^* = \bar{U} - \rho_I \frac{N-1}{2N}\frac{\Sigma_\nu}{\Sigma} - \frac{A^2\Sigma}{2N^2\rho_I}(\phi^*)^2, \text{ where} \tag{56}$$

$\tilde{\alpha}(N) = \Sigma^2(\Sigma - \tilde{\Sigma})(\Sigma - (N-1)^2\tilde{\Sigma}) > 0,$
$\tilde{\beta}(N) = -(N-1)N\rho_I^2(\Sigma^2\Sigma_\nu + \Sigma((N-2)\Sigma_\nu - \bar{m}^2(N-1)N(\sigma_\epsilon + \Sigma_\nu))\tilde{\Sigma} + (N-1)^2\Sigma_\nu\tilde{\Sigma}^2) < 0,$
$\tilde{\gamma}(N) = 2N^2\rho_I\Sigma(\Sigma + (N-1)\tilde{\Sigma})^2 > 0$, *given* $\bar{m} = \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}$ *and* $\tilde{\Sigma} \equiv Var[R_i|\mu_i(1)]$.

(ii) All dealers exert effort, and submit the following demand function:

$$a^*(p, \mu_i(0)) = \rho_I(\phi^* Var[R_i|\mu_i(0)])^{-1}(\mathbb{E}[R_i \mid \mu_i(0)] - p). \tag{57}$$

The equilibrium with price-taking dealers is closely related to the equilibrium with strategic dealers characterized in Proposition 1, but differs in two ways. First, by assumption, dealers submit demand functions as in equation (17) while setting $\Lambda_i = 0$, effectively ignoring their own price impact. Relative to the strategic benchmark, dealers therefore overestimate how aggressively they can trade without affecting prices, leading them to behave as if competition were stronger than it truly is. This misperception alters the functional form of individual demands, and changes how those demands aggregate and, in equilibrium, how prices respond to order flow.

Second, and crucially, clients correctly anticipate this behavior and adjust the contracts they offer accordingly. Even though dealers act as if they were price takers, clients understand that this belief is incorrect and take into account how dealers' misperceived incentives translate into equilibrium prices and allocations. As a result, contract choice differs from the strategic-trading case both when the incentive constraint binds and when it does not. In particular, clients internalize that dealers' failure to account for price impact changes incentives, and they respond by adjusting the strength of the contract to

partially offset this mistake.

Importantly, clients' ability to correct dealers' misperception is limited. Contracts are set ex ante and are not auction-specific; in particular, they do not condition on realized clearing prices or on the dealer's contribution to price formation. As a consequence, clients cannot directly discipline dealers for ignoring their price impact, nor can they fully restore the strategic-trading benchmark through contract design alone. This limitation is reflected in the structure of the unconstrained contract $\Phi_U(N)$, which depends explicitly on the number of dealers. Unlike in the strategic equilibrium, $\Phi_U(N)$ does not coincide with the second-best benchmark $\rho_I/(\rho_I+\rho_H)$ for finite N. Instead, it converges to this benchmark only as $N \to \infty$, when dealers truly become price takers and the misperception underlying their behavior vanishes.

A.5 When is imperfect competition worse than moral hazard?

In this appendix I investigate 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 dealers are price-takers (i.e., do not reduce their demand strategically) to the corresponding welfare when dealers 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 dealers 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 dealers—and the welfare when moral hazard constraints bind for all dealers, $\phi^* = \Phi(N)$.

Proposition 8. *Consider the scenario where moral hazard constraints bind,* $\frac{\rho_I}{\rho_H+\rho_I} < \Phi(N)$.

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

$$f(\phi) \leq N \text{ , where } f(\phi) = 1 + \frac{\phi\sqrt{\rho_H(\rho_H+\rho_I)}}{-\rho_I + \phi(\rho_H+\rho_I)} \text{ at } \phi = \Phi(N). \tag{58}$$

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

$$\Phi(N) \geq \frac{\rho_I(N-1)}{(N-1)(\rho_H+\rho_I) - \sqrt{\rho_H(\rho_H+\rho_I)}}, \tag{59}$$

with this cutoff being larger than $\frac{\rho_I}{\rho_I+\rho_H}$.

The first part of the corollary (58) identifies the minimal number of dealers necessary for imperfect competition to distort asset allocations less than moral hazard for a given contract ϕ.[23] As moral hazard becomes less sever (b declines which decreases ϕ), $f(\phi)$ increases. This means the number of dealers 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 dealers such that that imperfect competition is less welfare harming than moral hazard.

The second condition (59) states that the welfare loss from imperfect competition is smaller than the loss from moral hazard when $\Phi(N)$ is sufficiently high. In the unfavorable contract regime, where $\Phi(N)$ increases in N and moral hazard frictions worsen as the number of dealers rises, the left-hand side of the condition increases while the right-hand side declines in N, making the condition more likely to hold. Moreover, Corollary 2 implies that, in the limit, when $b > \gamma(\sigma_\epsilon)$, the condition must be satisfied, so that the welfare loss from imperfect competition is dominated by the loss from moral hazard.

Proposition 8 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 dealers. Addressing moral hazard, by contrast, would require redesigning the contractual relationship between dealer institutions and dealers—such as by structuring compensation to more directly reward effort that improves returns.

A.6 Additional mathematical results and proofs

In this appendix I provide additional mathematical statements, and all proofs.

A.6.1 Auxiliary lemmas

I start with deriving the unconditional and conditions distribution of returns in the baseline model and in the model extension with common values (Lemma 1 and Lemma 2, respectively).

[23] The relationship is visualized in Appendix Figure A8 for a simplified environment in which dealers and clients have the same risk tolerance ($\rho_I = \rho_H$).

Lemma 1. *(i)* $R_i \sim \mathcal{N}(\nu, \Sigma+\Sigma_\nu)$. *(ii)* $R_i|\mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i)], Var[R_i \mid \mu_i(s_i)])$, *where*

$$\mathbb{E}[R_i \mid \mu_i(s_i)] = \begin{cases} \nu + \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}\big((\nu_i + \epsilon_i) - \nu\big) & when s_i = 1 \\ \nu_i & when\ s_i = 0. \end{cases} \tag{60}$$

$$Var[R_i \mid \mu_i(s_i)] = \begin{cases} \Sigma + \left(\Sigma_\nu^{-1} + \sigma_\epsilon^{-1}\right)^{-1} & when\ s_i = 1 \\ \Sigma & when\ s_i = 0. \end{cases} \tag{61}$$

Lemma 2. *(i)* $R_i \sim \mathcal{N}(\bar{\nu}, \Sigma + \Sigma_0 + \Sigma_\nu)$. *(ii)* $R_i \mid \mu_i(s_i), p^c \sim \mathcal{N}(\mathbb{E}[R_i \mid \mu_i(s_i), p^c], Var[R_i \mid \mu_i(s_i), p^c])$, *where:*

$$\mathbb{E}[R_i \mid \mu_i(s_i), p^c] = \begin{cases} c_0\bar{\nu} + c_s\mu_i(1) + c_p\hat{\nu}_{-i}(p^c) & when\ s_i = 1, \\ \nu_i & when\ s_i = 0, \end{cases} \tag{62}$$

$$Var[R_i \mid \mu_i(s_i), p^c] = \begin{cases} \tilde{\Sigma} & when\ s_i = 1, \\ \Sigma & when\ s_i = 0, \end{cases} \tag{63}$$

where $\hat{\nu}_{-i}(p^c)$ *is the signal about* ν_0 *recovered by inverting the linear equilibrium price function:*

$$\hat{\nu}_{-i}(p^c) = \frac{\big(C_1(N-1) + C_{21}\big)p^c + A - C_{21}\big(c_0\bar{\nu} + c_s\mu_i(1)\big)}{C_1(N-1)}, \tag{64}$$

with $C_1 = \rho_I\left(\frac{\phi}{N-2}\Sigma + \phi\Sigma\right)^{-1}$ *and* $C_{21} = \rho_I\left(\frac{\phi}{N-2}\Sigma + \phi_i\tilde{\Sigma}\right)^{-1}$, *and where:*

$$c_s = \frac{\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{Q}, \qquad c_p = \frac{(N-1)\Sigma_0\sigma_\epsilon}{Q}, \qquad c_0 = \frac{\sigma_\epsilon\Sigma_\nu}{Q}, \tag{65}$$

$$\tilde{\Sigma} = \Sigma + \frac{\sigma_\epsilon\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{Q}, \tag{66}$$

$$Q \equiv (N-1)\Sigma_0\sigma_\epsilon + N\Sigma_0\Sigma_\nu + \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu). \tag{67}$$

Here c_s *and* c_p *are the projection coefficients of* ν_i *onto* $\mu_i(1)$ *and* $\hat{\nu}_{-i} \equiv \nu_0 + \frac{1}{N-1}\sum_{j\neq i}\tilde{\nu}_j$ *respectively, where* $\hat{\nu}_{-i} \sim \mathcal{N}(\nu_0, \frac{\Sigma_\nu}{N-1})$. *Since* p^c *is an invertible affine function of* $\hat{\nu}_{-i}$ *in a linear equilibrium, conditioning on* p^c *is equivalent to conditioning on* $\hat{\nu}_{-i}$. *Note that* $c_0 + c_s + c_p = 1$. *When* $\Sigma_0 = 0$, $Q \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$, *so* $c_p \to 0$, $c_s \to \bar{m} = \frac{\Sigma_\nu}{\sigma_\epsilon + \Sigma_\nu}$, $c_0 \to 1 - \bar{m}$, *and* $\tilde{\Sigma} \to \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, *recovering Lemma 1.*

A.6.2 Proofs

I provide all proofs for the propositions, followed by proofs for all corollaries, and lemmas.

Proof of Proposition 1. To derive the equilibrium, I follow a guess-and verify strategy, and solve the game backwards. Starting in the last period, I consider the choice of dealer i and assume all other dealers play the equilibrium strategy. The dealer 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:

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

Fo this, the dealer 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:

$$\mathbb{E}\left[1 - \exp\left(-\frac{1}{\rho_I}\omega^c(s_i;\phi_i)\right)\middle|\mu_i(s_i)\right]. \tag{3}$$

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 (3) is equivalent to standard mean-variance optimization:

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

where $a_i = a_i(p,\mu_i(s_i))$. The necessary condition for quantity a_i at the fixed market clearing price p and both $s_i \in \{0,1\}$ is

$$\mathbb{E}[R_i \mid \mu_i(s_i)] - \phi_i\frac{1}{\rho_I}Var[R_i \mid \mu_i(s_i)]a_i = p + \Lambda_i a_i, \text{ where } \Lambda_i = \left(\frac{dp}{da_i}\right). \tag{10}$$

Λ_i measures the price impact of the dealer's own demand on the clearing price. Since all other dealers j but i submit the equilibrium guess (17), 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 (10) pins down the dealer's best reply $a_i(p,\mu_i(s_i))$ for $s_i \in \{0,1\}$. When dealer 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 (17).

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

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

where

$$\begin{aligned}\mathcal{U}(s_i;\phi_i) =& \phi_i(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i))a_i^c(s_i) \\ &- \frac{1}{2}\frac{\phi_i^2}{\rho_I} a_i^c(s_i) Var[R_i \mid \mu_i(s_i)] a_i^c(s_i) + bs_i.\end{aligned} \tag{6}$$

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

$$\begin{aligned}\mathcal{W}(\phi_i;s_i) =& (1-\phi_i)\big(\mathbb{E}[R_i \mid \mu_i(s_i)] - p^c(s_i)\big)a_i^c(s_i) \\ &- (1-\phi_i)^2 \frac{1}{2\rho_H} a_i^c(s_i) Var[R_i \mid \mu_i(s_i)] a_i^c(s_i) - \Delta s_i.\end{aligned} \tag{23}$$

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:

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

Note that $K(\phi_i;0)$ depends on ϕ_i, because ϕ_i affects the dealer's demand function in the trading game. Inserting the binding participation constraint (IR') into the client's objective function (12), the optimization problem simplifies to:

$$\max_{\phi_i} \Big\{\mathbb{E}[\mathcal{W}(\phi_i;0] - K_i(\phi_i;0)\Big\} = \max_{\phi_i} \Big\{\mathbb{E}[\mathcal{W}(\phi_i;0] + \mathbb{E}[\mathcal{U}(0;\phi_i] - \bar{U}\Big\} \tag{12}$$

subject to (IC), assuming the client wants to incentivize effort (which is verified in Proposition 4). There are two possible solutions for ϕ_i: an unconstrained contract, $\phi_i = \frac{\rho_I}{\rho_I+\rho_H}$, and a constrained contract, $\phi_i = \Phi(N)$. In either case, the equilibrium lump-sum payment, K^* then follows from (IR').

Consider first the constrained scenario, which arises when $\Phi(N) \geq \frac{\rho_I}{\rho_I+\rho_H}$. The constrained contract is given by the binding incentive constraint, defined below, in expressions (69) and (70), respectively. Since all clients must choose the same contract in the symmetric equilibrium, I can solve $\mathbb{E}[\mathcal{U}(0;\phi)] = \mathbb{E}[\mathcal{U}(1;\phi)]$ for ϕ. This gives a unique positive solution, which I call $\Phi(N)$.

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 $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$ when

$$\phi_i < \text{cutoff}_0(\phi) \equiv \frac{\phi}{N-2} + \frac{3\rho_I}{2(\rho_I+\rho_H)} \text{ and } \frac{\rho_I}{\rho_I+\rho_H} \leq \text{cutoff}_0(\phi). \tag{68}$$

This is the case for any contract ϕ that all other clients are choosing (including $\phi = \frac{\rho_I}{\rho_I+\rho_H}$). 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.

It remains to show that when the solution is unconstrained, the dealer (actually) exerts effort, i.e., the incentive constraint is satisfied. To check this, compute the dealer's expected certainty equivalents for an arbitrary ϕ_i for some given ϕ chosen by the other clients, by relying on market clearing and equilibrium trading behavior in period 3:

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] = \phi_i \frac{A^2(N-1)\phi^2\Sigma^2 + (N-2)^2 N \rho_I^2 \Sigma_\nu)}{2(N-2)(N-1)(2\phi + (N-2)\phi_i)\rho_I \Sigma}, \tag{69}$$

$$\mathbb{E}[\mathcal{U}(1;\phi_i)] = b + \phi_i \frac{A^2(-1+N)\phi^2\Sigma^2(\sigma_\epsilon+\Sigma_\nu) + (-2+N)^2\rho_I^2\Sigma_\nu(\sigma_\epsilon + N\Sigma_\nu))}{2(-2+N)(-1+N)\rho_I(\sigma_\epsilon+\Sigma_\nu)(2\phi\Sigma + (-2+N)\phi_i\tilde{\Sigma})}. \tag{70}$$

To ensure that the dealer has no incentive to deviate to shirking when the contract is unconstrained, $\phi_i = \frac{\rho_I}{\rho_H+\rho_I}$, 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 I 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. □

Proof of Proposition 2. I first compute realized welfare (23) in the equilibrium. Then I take expectations over return expectations, and rely on the simplifying assumptions that these expectations are independent across dealers i. Then to show that $\alpha(\phi^*, N)$ and $\beta(\phi^*, N)$ are non-negative for $\frac{\rho_I}{\rho_I+\rho_H} \leq \phi^* < 1$. □

To show statement (i) I take the partial derivative of welfare, characterize in Corollary 2, with respect to ϕ. Given $\frac{\rho_I}{\rho_I+\rho_H} \leq \phi^*$: $\frac{\partial \alpha(\phi^*,N)\beta(\phi^*,N)}{\partial \phi} \leq 0$, $\frac{\partial \beta(\phi^*,N)}{\partial \phi} \geq 0$. Therefore, welfare 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 dealers, N, shows statement (ii). □

Proof of Proposition 3. Let $\rho_H \to \infty$ and $\Delta = 0$. Condition (i) requires $b \geq -\beta(N)/\gamma(N)$, which ensures $\Phi(N)$ is real. Condition (ii) rules out a profitable deviation to the best shirking-inducing contract. When clients are risk-neutral, the unconstrained optimal contract gives the dealer nothing, so the relevant deviation is to offer $\phi_i = 0$. This deviation is unprofitable when

$$f(\Phi(N), \Phi(N)) \geq g(0, \Phi(N)) + b,$$

i.e. $b \leq \overline{b} \equiv f(\Phi(N), \Phi(N)) - g(0, \Phi(N))$. When $b = \overline{b}$ the client is indifferent between the equilibrium contract and deviating.

It remains to show $\overline{b} > \underline{b}$, i.e. that the interval $[\underline{b}, \overline{b}]$ is non-empty. The difficulty is that $\overline{b}$ depends on $\Phi(N)$, which itself depends on b. To highlight the dependence, I write $\Phi(N, b)$ in this proof. The key observation is that at $b = \underline{b}$, we have $\Phi(N, \underline{b}) = 0$ by definition of $\underline{b}$, since this is the boundary at which the incentive constraint just begins to bind. Since $\Phi(N, b)$ is continuous and strictly increasing in b, evaluating $\overline{b}(\Phi(N, b))$ as $b \to \underline{b}^+$ is equivalent to evaluating $\overline{b}(\phi)$ as $\phi \to 0^+$, where my notation highlights that the benefit upper bound depends on the contract.

From the definitions of f and g, and market clearing:

$$\overline{b}(\phi) \equiv f(\phi, \phi) - g(0, \phi) = -\frac{A^2\phi(N-2+2\phi)\Sigma}{4N^2\rho_I} - \frac{(N-2)\rho_I\Sigma_\nu\big((4-3N+2(N-2)\phi)\sigma_\epsilon + (N-2)(N-2+2\phi)\Sigma_\nu\big)}{4(N-1)N\phi\,\Sigma(\sigma_\epsilon + \Sigma_\nu)}.$$

As $\phi \to 0^+$, the first term vanishes since it is $O(\phi)$, while the second term behaves as

$$\overline{b}(\phi) \sim -\frac{(N-2)\rho_I \Sigma_\nu \big((4-3N)\sigma_\epsilon + (N-2)^2 \Sigma_\nu\big)}{4(N-1)N\phi\, \Sigma(\sigma_\epsilon + \Sigma_\nu)} \to +\infty \quad \text{as } \phi \to 0^+,$$

whenever $(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_\nu < 0$. Since $\Phi(N, \underline{b}) = 0$ by definition of $\underline{b}$, and $\Phi(N, b)$ is continuous and strictly increasing in b, we have $\Phi(N, b) \to 0^+$ as $b \to \underline{b}^+$. Therefore

$$\lim_{b \to \underline{b}^+} \overline{b}(\Phi(N, b)) = +\infty > \underline{b},$$

where $\underline{b}$ is finite. Hence there exists $\varepsilon > 0$ such that $\overline{b}(\Phi(N, b)) > b$ for all $b \in (\underline{b}, \underline{b} + \varepsilon)$, establishing that any such b satisfies both conditions (i) and (ii). □

Proof of Proposition 4. Let $b \geq -\beta(N)/\gamma(N)$ so that $\Phi(N)$ is a real number. To show that the symmetric effort-inducing equilibrium exists, I need to rule out that there are profitable deviations. There are two types of deviations.

First, the client could offer a different contract than the others, but one that still incentivizes the dealer to exert effort. When the incentive constraint is slack, the client has no profitable deviation, since ϕ^* maximizes expected payoff. When the incentive constraint binds, the only deviation that preserves effort is to offer a higher contract share. Such a deviation lowers the client's payoff and is therefore unprofitable.

Second, I need to rule out that clients do not have a unilateral incentive to deviate from the effort-inducing contract and allow their dealer to shirk. To find the optimal shirking-inducing reply, consider a client who seeks to induce shirking while all other clients choose the equilibrium contract ϕ^*. Such a client solves

$$\max_{\phi_i} \Big\{ \mathbb{E}[\mathcal{W}(\phi_i; 1)] - K_i(\phi_i; 1) \Big\} = \max_{\phi_i} \Big\{ \mathbb{E}[\mathcal{W}(\phi_i; 1)] + \mathbb{E}[\mathcal{U}(1; \phi_i)] - \bar{U} \Big\} \tag{71}$$

$$\text{subject to } \mathbb{E}[\mathcal{U}(0; \phi_i)] \leq \mathbb{E}[\mathcal{U}(1; \phi_i)]. \tag{72}$$

The best reply is:

$$\phi_i^S(\phi^*) = \begin{cases} \tilde{\phi}_i^S(\phi^*) & \text{if } \tilde{\phi}_i^S(\phi^*) < \dfrac{\rho_I}{\rho_I + \rho_H} \quad \text{(shirking-IC binds)}, \\ \dfrac{\rho_I}{\rho_I + \rho_H} & \text{otherwise} \quad \text{(shirking-IC slack)}, \end{cases} \tag{73}$$

where $\tilde{\phi}_i^S(\phi^*)$ makes the dealer indifferent between shirking and effort, i.e., $\mathbb{E}[\mathcal{U}(1; \tilde{\phi}_i^S(\phi^*))] =$

$\mathbb{E}[\mathcal{U}(0;\tilde{\phi}_i^S(\phi^*))]$. Using expressions (69) and (70):

$$\begin{aligned}\tilde{\phi}_i^S(\phi^*) = &\Big(8b(N-1)\phi^{*2}\rho_I\Sigma^2\Big)\Big[-(N-2)(N-1)\phi^*\rho_I\Sigma\Big(-\rho_I\Sigma_\nu+\tilde{m}^2\rho_I(\sigma_\epsilon+\Sigma_\nu)+2b(\Sigma+\tilde{\Sigma})\Big)\Big)\\ &\pm\Big\{\phi^{*2}\rho_I\Sigma^2\Big((2-3N+N^2)^2\rho_I(-\rho_I\Sigma_\nu+\tilde{m}^2\rho_I(\sigma_\epsilon+\Sigma_\nu)+2b(\Sigma+\tilde{\Sigma}))^2\\ &-8b(-1+N)\Big(A^2(-1+N)\phi^{*2}\Sigma^2(\Sigma-\tilde{\Sigma})+(-2+N)^2\rho_I\Big(\tilde{m}^2(-1+N)\rho_I\Sigma(\sigma_\epsilon+\Sigma_\nu)\\ &+2b(-1+N)\Sigma\tilde{\Sigma}+\rho_I\Sigma_\nu(\Sigma-N\tilde{\Sigma})\Big)\Big)\Big)\Big\}^{1/2}\Big]^{-1},\end{aligned} \tag{74}$$

where $\tilde{m}=\Sigma_\nu(\Sigma_\nu+\sigma_\epsilon)^{-1}$, $\tilde{\Sigma}=Var(R_i|\mu_i(1))$.

Consider the case where, in equilibrium, the effort-constraint is binding, i.e., $\phi^*=\Phi(N)$ and $\frac{\rho_I}{\rho_I+\rho_H}<\Phi(N)$. When the client offers less than $\Phi(N)$, their dealer shirks. The optimal shirking deviation is to offer $\phi_i^S(\Phi(N))=\frac{\rho_I}{\rho_I+\rho_H}$. This is not profitable when

$$\Delta \geq g\left(\phi_{UC},\Phi(N)\right)-f\left(\Phi(N),\Phi(N)\right)+b \text{ for } \phi_{UC}<\Phi(N), \tag{31}$$

where $\phi_{UC}=\frac{\rho_I}{\rho_I+\rho_H}$. Here $f(\phi_i,\phi)$ is the payoff the client expects (net of $\bar{U}$) when choosing some contract ϕ_i in response to everyone else choosing some contract, ϕ, and the dealer exerts effort; while the clients expects payoff $g(\phi_i,\phi)+b-\Delta$ when the dealer shirks instead.[24] Formally:

$$\begin{aligned}f(\phi_i,\phi)\equiv&\mathbb{E}[\mathcal{W}(\phi_i;0)]+\mathbb{E}[\mathcal{U}(0;\phi_i)],\\ =&\Big((2\phi\rho_H+(-2+N)(-\rho_I-(-2+\phi_i)\phi_i(\rho_H+\rho_I)))\\ &(A^2(-1+N)\phi^2\Sigma^2+(-2+N)^2N\rho_I^2\Sigma_\nu)\Big)\\ &\Big(2(-2+N)(-1+N)(2\phi+(-2+N)\phi_i)^2\rho_H\rho_I\Sigma\Big)^{-1}.\end{aligned} \tag{75}$$

$$\begin{aligned}g(\phi_i,\phi)+b-\Delta\equiv&\,\mathbb{E}[\mathcal{W}(\phi_i;1)]+\mathbb{E}[\mathcal{U}(1;\phi_i)]\\ =&\Big((A^2(-1+N)\phi^2\Sigma^2+(-2+N)^2\rho_I^2(\Sigma_\nu+\tilde{m}^2(-1+N)\\ &(\sigma_\epsilon+\Sigma_\nu)))(2\phi\rho_H\Sigma-(-2+N)(\rho_I+(-2+\phi_i)\phi_i(\rho_H+\rho_I))\tilde{\Sigma})\Big)\\ &\Big(2(-2+N)(-1+N)\rho_H\rho_I(2\phi\Sigma+(-2+N)\phi_i\tilde{\Sigma})^2\Big)^{-1}.\end{aligned} \tag{76}$$

Now consider the case where, in equilibrium, the dealer chooses to exert effort even under the unconstrained contract, $\phi^*=\phi_{UC}$. In that case the client would have to offer a lower contract to make the dealer shirk, i.e., $\tilde{\phi}_i^S(\phi_{UC})<\phi_{UC}$. This is not profitable

[24] Note that both $f(\phi_i,\phi)$ and $g(\phi_i,\phi)$ are for any fixed ϕ strictly concave in ϕ_i as long as $\phi_i<\text{cutoff}(\phi)\equiv\frac{\phi}{N-2}\frac{\Sigma}{Var(R_i|\mu_i(s_i))}+\frac{3\rho_I}{2(\rho_H+\rho_I)}$, where $\text{cutoff}(\phi)>\frac{\rho_I}{\rho_I+\rho_H}$.

when

$$\Delta \geq g\left(\tilde{\phi}_i^S(\phi_{UC}), \phi_{UC}\right) - f\left(\phi_{UC}, \phi_{UC}\right) + b \text{ for } \tilde{\phi}_i^S(\phi_{UC}) < \phi_{UC}, \tag{32}$$

where $\phi_{UC} = \frac{\rho_I}{\rho_I + \rho_H}$. □

Proof of Proposition 5. To rule out that there is no symmetric shirking inducing equilibrium, I first follow the same steps as in deriving necessary conditions for the effort-inducing equilibrium to obtain the analogous conditions for the shirking-inducing symmetric equilibrium candidate. In this candidate (i) all clients offer:

$$\phi^{S*} = \begin{cases} \frac{\rho_I}{\rho_I + \rho_H} & \text{when } \frac{\rho_I}{\rho_I + \rho_H} \leq \Phi^S(N) \\ \Phi^S(N) \equiv \frac{1}{|A|}\sqrt{\frac{b + \beta^S(N)}{\alpha^S(N)}} & \text{otherwise,} \end{cases} \tag{77}$$

$$\mathbb{E}\left[\mathcal{U}(0;\phi)\right] \leq \mathbb{E}\left[\mathcal{U}(1;\phi)\right] \Leftrightarrow b \geq C^S(\phi) \equiv A^2\phi^2\alpha^S(N) - \beta^S(N) \tag{78}$$

where:

$$\alpha^S(N) = \frac{(\tilde{\Sigma} - \Sigma)\tilde{\Sigma}}{2N\rho_I\left((N-2)\Sigma + 2\tilde{\Sigma}\right)} > 0,$$
$$\beta^S(N) = -\frac{(N-2)\rho_I\Sigma_\nu}{2(N-1)}\left(\frac{N}{(N-2)\Sigma + 2\tilde{\Sigma}} - \frac{\sigma_\epsilon + N\Sigma_\nu}{N(\sigma_\epsilon + \Sigma_\nu)\tilde{\Sigma}}\right),$$

given $\tilde{\Sigma} = Var[R_i \mid \mu_i(1)]$. Note that $\alpha^S(N) > 0$ since $\tilde{\Sigma} > \Sigma$, and that $C^S(\phi)$ is increasing in ϕ^2, so that higher contract shares make shirking less attractive for the dealer. (ii) All dealers shirk, $s^* = 1$, and submit the demand function of Proposition 1 but now with $s^* = 1$ with with ϕ^{S*} instead of ϕ^*.

This equilibrium is a valid candidate only when $b \geq -\frac{\beta^S(N)}{\alpha^S(N)}$. When $b < -\frac{\beta^S(N)}{\alpha^S(N)}$, the incentive compatibility condition $b \geq C^S(\phi)$ fails for all $\phi > 0$, meaning that for any contract a client could offer, the dealer strictly prefers to exert effort.

To rule out that this candidate is an equilibrium for $b \geq -\frac{\beta^S(N)}{\alpha^S(N)}$, I have to argue that there is a unilateral profitable deviation. Consider the constrained case, $\Phi^S(N) < \frac{\rho_I}{\rho_H + \rho_I} \equiv \phi_{UC}$. The best deviation would be to ϕ_{UC}, where the client's dealer would exert effort. This is profitable when

$$g^S(\Phi^S(N), \Phi^S(N)) + b - \Delta < f^S(\phi_{UC}, \Phi^S(N)) \text{ for } \Phi^S(N) < \phi_{UC}, \tag{??}$$

where g^S, f^S are defined analogously to Proposition 4 with the difference being that in Proposition 4 all dealers $j \neq i$ choose the demand schedule following effort, while now they choose the demand schedule following shirking. Formally:

$$f^S(\phi_i, \phi) = \frac{(N-2)^2 N \phi_i \rho_I^2 \Sigma_\nu + A^2 (N-1) \phi^2 \phi_i \tilde{\Sigma}^2}{2(N-2)(N-1)\rho_I \big((N-2)\phi_i \Sigma + 2\phi \tilde{\Sigma}\big)}. \tag{79}$$

$$g^S(\phi_i, \phi) = \frac{(N-2)^2 \phi_i \rho_I^2 \left(\Sigma_\nu + \frac{(N-1)\Sigma_\nu^2}{\sigma_\epsilon + \Sigma_\nu}\right) + A^2 (N-1) \phi^2 \phi_i \tilde{\Sigma}^2}{2(N-2)(N-1)(2\phi + (N-2)\phi_i)\rho_I \tilde{\Sigma}}. \tag{80}$$

Since $f^S(\phi_{UC}, \phi) - g^S(\phi, \phi) > 0$ for any $\phi > 0$ the client would find it profitable to deviate to offer the unconstraint contract when $\Delta \geq b$.

Next consider the unconstrained case, i.e., $\phi_{UC} \leq \Phi^S(N)$. The client could deviate to a contract at which the dealer is indifferent between shirking and exerting effort, characterized in expression (74). This is profitable when

$$g^S(\phi_{UC}, \phi_{UC}) + b - \Delta < f^S(\tilde{\phi}_i^E(\phi_{UC}), \phi_{UC}) \text{ for } \phi_{UC} \leq \tilde{\phi}_i^E(\phi_{UC}), \tag{81}$$

where $\tilde{\phi}^E(\phi)$ is the contract that makes dealer i indifferent between shirking and exerting effort given that all others play the shirking equilibrium candidate. Since $f^S(\phi_i, \phi_{UC}) - g^S(\phi_{UC}, \phi_{UC}) > 0$ for any $\phi_i > 0$ this a profitable deviation as long as $\Delta \geq b$. □

Proof of Proposition 6. The proof follows the same guess-and-verify strategy as the baseline. I highlight only the changes relative to Proposition 1.

In the trading stage, the dealer's optimization problem is identical to the baseline, with the only difference that the conditional moments $\mathbb{E}[R_i \mid \mu_i(s_i), p^c(s_i)]$ and $Var[R_i \mid \mu_i(s_i), p^c(s_i)]$ now depend on Σ_0 as characterized in Lemma 2. Under effort, the conditional moments are unchanged relative to the baseline, so the equilibrium demand and price impact are identical. Under shirking, the conditional mean depends on $\hat{\nu}_{-i}(p^c(1))$ through the coefficient $c_p > 0$. Market clearing (with the shirking dealer's demand $C_{21}(c_0 \bar{\nu} + c_s \mu_i(1) + c_p \hat{\nu}_{-i} - p^c(1))$ and aggregate effort demand $C_1(N-1)(\hat{\nu}_{-i} - p^c(1))$) gives:

$$p^c(0) = \frac{C_1 \sum_{j \neq i} \nu_j + C_{20} \nu_i - A}{C_1(N-1) + C_{20}},$$

$$p^c(1) = \frac{C_1(N-1)\hat{\nu}_{-i} + C_{21}(c_0 \bar{\nu} + c_s \mu_i(1)) - A}{C_1(N-1) + C_{21}},$$

where $C_1 = \rho_I \left(\frac{\phi^*}{N-2}\Sigma + \phi^* \Sigma\right)^{-1}$, $C_{20} = \rho_I \left(\frac{\phi^*}{N-2}\Sigma + \phi_i \Sigma\right)^{-1}$, and $C_{21} = \rho_I \left(\frac{\phi^*}{N-2}\Sigma + \phi_i \tilde{\Sigma}\right)^{-1}$.

Anticipating market clearing, the dealer exerts effort in stage 2 if and only if $\mathbb{E}[\mathcal{U}(0;\phi_i)] \geq \mathbb{E}[\mathcal{U}(1;\phi_i)]$. The expected certainty equivalents under effort and shirking are computed analogously to the baseline, using the following moment rules that account for $Cov(\nu_i,\nu_j)=\Sigma_0>0$:

$$\begin{aligned} \mathbb{E}[\nu_i^2] &= \bar{\nu}^2+\Sigma_0+\Sigma_\nu, & \mathbb{E}[\nu_i\nu_j] &= \bar{\nu}^2+\Sigma_0, \quad i\neq j,\\ \mathbb{E}[\nu_i\hat{\nu}_{-i}] &= \bar{\nu}^2+\Sigma_0, & \mathbb{E}[\hat{\nu}_{-i}^2] &= \bar{\nu}^2+\Sigma_0+\tfrac{\Sigma_\nu}{N-1},\\ \mathbb{E}[\epsilon_i^2] &= \sigma_\epsilon, & \mathbb{E}[\epsilon_i\cdot] &= 0. \end{aligned}$$

The closed-form expressions for $\mathbb{E}[\mathcal{U}(0;\phi_i)]$ and $\mathbb{E}[\mathcal{U}(1;\phi_i)]$ are omitted as they are unwieldy but can be verified symbolically in Mathematica. When $\Sigma_0=0$ they reduce to the baseline expressions (69)–(70).

In stage 1, there are again two solutions to the client's optimization problem. First, imposing $\phi_i=\phi^*$ and solving the binding incentive constraint $\mathbb{E}[\mathcal{U}(0;\phi^*)]=\mathbb{E}[\mathcal{U}(1;\phi^*)]$

$$b=\alpha_{CV}(N)A^2\phi^{*2}-\beta_{CV}(N) \tag{13'}$$

gives solution $\phi^*=\Phi_{CV}(N)$. Second, when the constraint is slack, the unconstraint contract is $\phi^*=\frac{\rho_I}{\rho_I+\rho_H}$. Finally, the participation constraint pins down the lum-sump fee:

$$K^*=\bar{U}-\mathbb{E}\left[\mathcal{U}(0;\phi^*)\right]=\bar{U}-\frac{(N-2)\rho_I\big(2(N-1)\Sigma_0+N\Sigma_\nu\big)}{2(N-1)N\Sigma}-\frac{A^2\Sigma}{2N(N-2)\rho_I}(\phi^*)^2,$$

where the additional term $\frac{(N-2)\rho_I\Sigma_0}{N\Sigma}$ relative to the baseline reflects the higher expected trading profits due to correlated ν_i's. □

Proof of Proposition 7. The proof is analogous to the proof of Proposition 1. The only difference is that we now set $\Lambda_i=0$. □

Proof of Proposition 8. 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 welfare when dealers are price-takers to the corresponding welfare when dealers act strategically and reduce their demand (as in Proposition 1).

To analyze the effect of moral hazard, I assume dealers 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 dealers—and the welfare when moral hazard constraints bind for all dealers, $\phi^*=\Phi(N)$.

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

$$LIC = +\underline{\gamma}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \underline{\delta}A\Sigma A, \text{ with} \tag{82}$$

$$\underline{\gamma} = \frac{\rho_H + \rho_I}{2(N-1)} \geq 0 \tag{83}$$

$$\underline{\delta} = 0. \tag{84}$$

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

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

$$LMH(N,\phi) = \tilde{\gamma}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \tilde{\delta}A\Sigma A, \text{ with} \tag{85}$$

$$\tilde{\gamma} = \frac{(\rho_I - \phi(\rho_H+\rho_I))^2(N-1)N\rho_I(\rho_H+\rho_I)}{2N\phi^2\rho_H\rho_I(\rho_H+\rho_I)} \geq 0 \tag{86}$$

$$\tilde{\delta} = -\frac{(\rho_I - \phi(\rho_H+\rho_I))^2\phi^2}{2N\phi^2\rho_H\rho_I(\rho_H+\rho_I)} \leq 0 \tag{87}$$

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 8.

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 $N \geq 2$:

$$1 + \frac{\Phi(N)\sqrt{\rho_H(\rho_H+\rho_I)}}{-\rho_I + \Phi(N)(\rho_H+\rho_I)} \leq N \iff \Phi(N)\left(\sqrt{\rho_H(\rho_H+\rho_I)} - (N-1)(\rho_H+\rho_I)\right) \leq -(N-1)\rho_I$$

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

Proof of Corollary 1. To derive the market clearing price in equilibrium it suffices to insert the equilibrium demand functions of all dealers 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. □

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

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(0;\phi_i)] = \frac{\rho_I}{2}\frac{\Sigma_\nu}{\Sigma} \tag{20}$$

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(1;\phi_i)] = \frac{\rho_I}{2}\frac{\Sigma_\nu^2(\Sigma_\nu+\sigma_\epsilon)^{-1}}{\Sigma+(\Sigma_\nu^{-1}+\sigma_\epsilon^{-1})^{-1}} + b. \tag{21}$$

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

$$b > \gamma_\infty(\sigma_\epsilon) \equiv \frac{\rho_I\sigma_\epsilon\Sigma_\nu(\Sigma+\Sigma_\nu)}{2\Sigma(\sigma_\epsilon\Sigma_\nu+\Sigma(\sigma_\epsilon+\Sigma_\nu))},$$

and is always fulfilled otherwise. □

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. □

Proof of Corollary 4. Consider statement (i). We need to show there exists $b \in [\underline{b}, \overline{b}]$ with $b > \gamma(\sigma_\epsilon)$, i.e. that $\overline{b} > \gamma(\sigma_\epsilon) > \underline{b}$. When $\rho_H \to \infty$, $\phi_{UC} = 0$ so the incentive constraint always binds. Therefore $\Phi(N, \gamma(\sigma_\epsilon)) > 0$, which is equivalent to $\gamma(\sigma_\epsilon)\gamma(N) + \beta(N) > 0$, i.e. $\gamma(\sigma_\epsilon) > \underline{b}$. This establishes the lower bound. For the upper bound, from Proposition 3, under condition $\sigma_\epsilon(3N-4) > (N-2)^2\Sigma_\nu$, $\overline{b}(\Phi(N,b)) \to +\infty$ as $b \to \underline{b}^+$. Since $\gamma(\sigma_\epsilon)$ is finite, there exist values of $b \in (\gamma(\sigma_\epsilon), \overline{b})$, establishing that the unfavorable contract regime arises for b in a non-empty interval above $\gamma(\sigma_\epsilon)$.

Now consider statement (ii). There are two cases. The first case is $b \in [\underline{b}, \overline{b}]$ and $b \leq \gamma(\sigma_\epsilon)$. From the proof of statement (i), $\gamma(\sigma_\epsilon) > \underline{b}$ when $\rho_H \to \infty$, so the interval $[\underline{b}, \gamma(\sigma_\epsilon)]$ is non-empty. From Corollary 3, $b \leq \gamma(\sigma_\epsilon)$ implies $\Phi'(N) \leq 0$: the constrained contract share is decreasing in N, so competition relaxes rather than tightens the incentive distortion. The second case is $b < \underline{b}$. Then the incentive constraint is satisfied at $\phi^* = \phi_{UC} = 0$ since $-\beta(N)/\gamma(N) = \underline{b} > b$, so the dealer exerts effort under the unconstrained contract. □

Proof of Corollary 5. Consider statement (i). The incentive constraint binds, i.e. $\Phi(N) > \phi_{UC}$, when offering the unconstrained contract would lead the dealer to shirk:

$$b > \frac{A^2\phi_{UC}^2\alpha(N) - \beta(N)}{\gamma(N)}.$$

Substituting $\phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$ and rearranging, this is equivalent to:

$$A^2 < \frac{b\gamma(N)+\beta(N)}{\alpha(N)}\left(\frac{\rho_I+\rho_H}{\rho_I}\right)^2.$$

For this to hold for some $N > 2$, it suffices that $A^2 < \bar{A}^2(\sigma_\epsilon) \equiv \max_{N>2} \frac{b\gamma(N)+\beta(N)}{\alpha(N)}\left(\frac{\rho_I+\rho_H}{\rho_I}\right)^2$. To incentivize effort clients choose $\Phi(N)$. By Corollary 3 this contract increases in N given $b > \gamma(\sigma_\epsilon)$.

The proof for statement (ii) is analogous. Note that $\bar{A}^2(\sigma_\epsilon)$ decreases in σ_ϵ because

$$\frac{\partial}{\partial\sigma_\epsilon}\frac{b\gamma(N)+\beta(N)}{\alpha(N)} = -\frac{2bN^2\rho_I}{\sigma_\epsilon^2} < 0. \quad \square$$

Proof of Lemma 1. (i) 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 $Var(R_i) = \mathbb{E}[Var(R_i|\nu_i)] + Var(\mathbb{E}[R_i|\nu_i]) = \Sigma + \Sigma_\nu$.

(ii) To derive the conditional expectation, note:

$$\mathbb{E}[R_i \mid \mu_i(0)] = \nu_i,$$
$$\mathbb{E}[R_i \mid \mu_i(1)] = \mathbb{E}[\nu_i \mid \nu_i + \epsilon_i] = \nu + Cov(\nu_i, \nu_i+\epsilon_i)Var(\nu_i+\epsilon_i)^{-1}\big((\nu_i+\epsilon_i) - \nu\big),$$

where

$$Cov(\nu_i, \nu_i+\epsilon_i) = \Sigma_\nu, \qquad Var(\nu_i+\epsilon_i) = \Sigma_\nu + \sigma_\epsilon.$$

Therefore,

$$\mathbb{E}[\nu_i \mid \nu_i+\epsilon_i] = \nu + \Sigma_\nu(\Sigma_\nu+\sigma_\epsilon)^{-1}\big((\nu_i+\epsilon_i) - \nu\big).$$

To derive the conditional variance, I use the law of total variance: $Var(R_i \mid \mu_i(s_i)) = \mathbb{E}[Var(R_i \mid \nu_i) \mid \mu_i(s_i)] + Var(\mathbb{E}[R_i \mid \nu_i] \mid \mu_i(s_i))$. Since $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, we have:

$$Var(R_i \mid \mu_i(s_i)) = \Sigma + Var(\nu_i \mid \mu_i(s_i)).$$

To compute $Var(\nu_i \mid \mu_i(s_i))$, 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: $\mu_i(s_i) = \nu_i + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon(s_i))$ with $\sigma_\epsilon(0) = 0$ and $\sigma_\epsilon(1) = \sigma_\epsilon > 0$. Therefore:

$$Var(\nu_i \mid \mu_i(1)) = \left(\Sigma_\nu^{-1} + \sigma_\epsilon^{-1}\right)^{-1} \text{ and } Var(\nu_i \mid \mu_i(0)) = 0. \quad \square$$

Proof of Lemma 2. Part (i) follows from the law of total variance: since $R_i = \nu_0 + \tilde{\nu}_i + \varepsilon_i$ with $\nu_0 \sim \mathcal{N}(\bar{\nu}, \Sigma_0)$, $\tilde{\nu}_i \sim \mathcal{N}(0, \Sigma_\nu)$, $\varepsilon_i \sim \mathcal{N}(0, \Sigma)$, all mutually independent, it follows that $R_i \sim \mathcal{N}(\bar{\nu}, \Sigma + \Sigma_0 + \Sigma_\nu)$.

For part (ii), the effort case $s_i = 0$ is immediate since $\mu_i(0) = \nu_i$ and $R_i \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$ by assumption, with $Var[R_i \mid \nu_i] = \Sigma$ independent of ν_i. The price p^c is uninformative conditional on ν_i since the dealer already knows both components ν_0 and $\tilde{\nu}_i$ exactly.

For the shirking case $s_i = 1$, we apply the multivariate normal projection formula. The dealer observes two signals about $\nu_i = \nu_0 + \tilde{\nu}_i$. Signal 1: $\mu_i(1) = \nu_i + \epsilon_i$, with $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon)$ independent of ν_i. Signal 2: $\hat{\nu}_{-i}(p^c) = \nu_0 + \frac{1}{N-1}\sum_{j \neq i} \tilde{\nu}_j$, recovered by inverting (64), with $\hat{\nu}_{-i} \sim \mathcal{N}(\nu_0, \frac{\Sigma_\nu}{N-1})$.

The prior on ν_i is $\mathcal{N}(\bar{\nu}, \Sigma_0 + \Sigma_\nu)$. Stacking $y \equiv (\mu_i(1), \hat{\nu}_{-i})'$, the joint covariance matrix of y and the cross-covariance of y with ν_i are:

$$\Sigma_{yy} = \begin{pmatrix} \Sigma_0 + \Sigma_\nu + \sigma_\epsilon & \Sigma_0 \\ \Sigma_0 & \Sigma_0 + \frac{\Sigma_\nu}{N-1} \end{pmatrix}, \qquad \Sigma_{\nu_i y} = \begin{pmatrix} \Sigma_0 + \Sigma_\nu & \Sigma_0 \end{pmatrix}. \tag{88}$$

By the normal projection formula:

$$\mathbb{E}[\nu_i \mid \mu_i(1), \hat{\nu}_{-i}] = \bar{\nu} + \Sigma_{\nu_i y}\Sigma_{yy}^{-1}(y - \bar{\nu}\mathbf{1}) = c_0\bar{\nu} + c_s\mu_i(1) + c_p\hat{\nu}_{-i}, \tag{89}$$

where $(c_s, c_p) = \Sigma_{\nu_i y}\Sigma_{yy}^{-1}$ and $c_0 = 1 - c_s - c_p$. Computing Σ_{yy}^{-1} with $|\Sigma_{yy}| = (\Sigma_0 + \Sigma_\nu + \sigma_\epsilon)(\Sigma_0 + \frac{\Sigma_\nu}{N-1}) - \Sigma_0^2$, one can verify that:

$$c_s = \frac{\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{Q}, \qquad c_p = \frac{(N-1)\Sigma_0\sigma_\epsilon}{Q}, \qquad c_0 = \frac{\sigma_\epsilon\Sigma_\nu}{Q}, \tag{90}$$

by checking $\Sigma_{\nu_i y} = (c_s, c_p)\Sigma_{yy}$, where $Q = (N-1)\Sigma_0\sigma_\epsilon + N\Sigma_0\Sigma_\nu + \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$. The posterior variance is:

$$Var[\nu_i \mid \mu_i(1), \hat{\nu}_{-i}] = (\Sigma_0 + \Sigma_\nu) - \Sigma_{\nu_i y}\Sigma_{yy}^{-1}\Sigma_{y\nu_i} = \frac{\sigma_\epsilon\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{Q}. \tag{91}$$

Since $Var[R_i \mid \nu_i] = \Sigma$ is independent of ν_i, the law of total variance gives:

$$Var[R_i \mid \mu_i(1), \hat{\nu}_{-i}] = \Sigma + \frac{\sigma_\epsilon \Sigma_\nu (N\Sigma_0 + \Sigma_\nu)}{Q} = \tilde{\Sigma}. \tag{92}$$

Since p^c is an invertible affine function of $\hat{\nu}_{-i}$ via (64), conditioning on p^c is equivalent to conditioning on $\hat{\nu}_{-i}$, so $\mathbb{E}[R_i \mid \mu_i(1), p^c] = c_0 \bar{\nu} + c_s \mu_i(1) + c_p \hat{\nu}_{-i}(p^c)$ and $Var[R_i \mid \mu_i(1), p^c] = \tilde{\Sigma}$.

The baseline limits follow by setting $\Sigma_0 = 0$: $Q \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$, $c_p \to 0$, $c_s \to \frac{\Sigma_\nu}{\sigma_\epsilon + \Sigma_\nu} = \bar{m}$, $c_0 \to 1 - \bar{m}$, and $\tilde{\Sigma} \to \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, recovering Lemma 1. □

B Additional empirical evidence

Appendix B.1 provides additional suggestive evidence.
Appendix B.2 describes the estimation procedure used to take the model to the data.
Appendix Tables and Figures are presented at the end.

B.1 Additional suggestive evidence and discussion

This appendix provides additional discussion and robustness evidence related to the price differences between TSX and Alpha documented in Appendix Table A1, and related to the price differences between client-account and own-account trades documented in Appendix Table A2.

Imperfect competition. One potential concern is that the observed price gaps reflect execution-related liquidity frictions. A smaller venue may exhibit thinner order-book depth or slower resiliency, increasing execution costs. Moreover, orders routed to Alpha—whether client or proprietary—may differ in urgency or aggressiveness, for example by walking further into the book or demanding faster execution, which would mechanically worsen prices.

Importantly, not all such liquidity differences constitute confounding factors. Many of them translate into higher price impact, and are therefore conceptually similar to a reduction in competition. What cannot be fully ruled out with transaction data alone are execution-related frictions orthogonal to competition-related changes in price impact.

Several features of the evidence suggest that these (unobservable) execution-related frictions are unlikely to be the sole drivers of the observed price difference. TSX and Alpha are owned by the same parent company and operate under similar rule books

and execution protocols, limiting the scope for institutional differences. The Alpha price gap is symmetric across trade direction—higher prices for purchases and lower prices for sales—consistent with a general price-impact mechanism rather than one-sided urgency or execution failures. Moreover, as shown in Appendix Table A4, Panel D, the Alpha effect persists—and if anything becomes larger—when attention is restricted to small trades, which are least likely to mechanically walk up a thinner the limit order book.

Another concern is strategic venue selection. Dealers may route orders across exchanges based on order characteristics that are imperfectly observed by the econometrician—for example, sending more difficult or information-sensitive orders to Alpha while routing simpler orders to TSX. To address this, the regressions condition on dealer × symbol × trade date × hour fixed effects, so that identification comes from comparing the same dealer executing the same stock, on the same side of the market, within the same narrow time window, across the two venues. Any remaining selection would therefore need to operate at this level to explain the results.

Appendix Table A4 further shows that the estimated Alpha coefficient is stable across increasingly restrictive fixed-effects specifications. The similarity of the estimates when moving from symbol–day fixed effects (Panel A), to hour-level controls (Panel B), and to dealer-level fixed effects (Panel C) indicates that the price gap persists within narrowly defined dealer–asset–time cells and is therefore unlikely to be driven by coarse liquidity differences, intraday market conditions, or persistent dealer-specific routing strategies.

Taken together, the evidence is consistent with imperfect competition affecting prices, though—as is typical with transaction-level data—it does not establish a causal relationship.

Moral hazard. One concern is that client orders differ from own-account trades in ways that are imperfectly observed by the econometrician—for example, they may be more urgent, less strategically timed, or subject to constraints imposed by clients or internal execution protocols. If such differences were pervasive, worse prices for client-account trades could reflect higher execution difficulty rather than distorted incentives. Importantly, to the extent that dealers exert less effort, monitor less closely, or optimize less aggressively when trading on behalf of clients—because they do not internalize the full surplus—these frictions are part of the moral-hazard mechanism itself.

To mitigate concerns about unobservable differences across account-types that are orthogonal to moral hazard frictions, regressions condition on symbol × trade date × hour × exchange fixed effects, so that client and own-account trades are compared within the

same security, on the same venue, and within the same narrow time window, absorbing systematic differences in asset-level liquidity and intraday market conditions.[25] The estimated price gap is symmetric across trade direction: client purchases occur at higher prices and client sales at lower prices than own-account trades. This pattern is consistent with a general deterioration in execution quality rather than one-sided urgency or directional information effects. Moreover, Appendix Table A5, Panel C shows that the price gap persists when attention is restricted to small trades, which may be more comparable across account types in terms of trade motive and informational content and less likely to differ mechanically in how they interact with the limit order book.

Another concern is selection across account types. Dealers may choose whether to execute a given order on behalf of a client or for their own account based on private information about expected execution quality or short-term price dynamics. If dealers systematically internalized easier or more profitable opportunities while allocating more difficult trades to client accounts, the observed price gap could reflect strategic selection rather than incentive distortions. According to the rules of the exchange, however, such selection is not permitted.

Taken together, while the evidence does not establish moral hazard in a causal sense, it documents a robust—and previously unobserved—empirical pattern that is consistent with incentive distortions in delegated trading (Appendix Table A5). In the model, moral hazard manifests as reduced effort or suboptimal execution when dealers act on behalf of clients, leading to systematically worse prices for client trades. The empirical findings align closely with this mechanism.

B.2 Estimation procedure

I estimate ρ_I and ϕ^* parameters using median-based moment conditions in a two-stage GMM procedure, and then compute the implied lower bound on client risk tolerance. 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 and asymptotically equivalent to the joint estimator under

[25] Including dealer fixed effects is not feasible in this setting because relatively few dealers operate on both account types within the same security and hour, leaving insufficient within-dealer variation to support reliable estimation. Appendix Table A5, Panel D shows that when dealer fixed effects are included without hour interactions, the estimated effects remain in the same direction as in the baseline specifications and are statistically significant on the sell side.

regularity conditions (Newey and McFadden, 1994).[26] I rely on medians rather than means to limit the influence of outliers.

Estimating ρ_I and ϕ^*. Recall the average daily price per auction type (18"):

$$p^{el}_{tsk} = \nu_{ts} - \frac{\phi^*_k}{\rho_I}\Sigma_{ts}\left(\frac{N^{el}_{tsk}-1}{N^{el}_{tsk}-2}\right)\frac{A^{el}_{tsk}}{N^{el}_{tsk}}, \qquad (e,l)\in\{\text{TSX},\text{Alpha}\}\times\{b,q\}, \tag{18"}$$

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

$$m^{el}_{tsk} = \Sigma_{ts}\left(\frac{N^{el}_{tsk}-1}{N^{el}_{tsk}-2}\right)\frac{A^{el}_{tsk}}{N^{el}_{tsk}}, \tag{93}$$

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 (18") and rearrange for ρ_I:

$$R^{(1)}_\iota \equiv \frac{m^{\text{TSX},l}_{ts,IN} - m^{\text{Alpha},l}_{ts,IN}}{p^{\text{Alpha},l}_{ts,IN} - p^{\text{TSX},l}_{ts,IN}}. \tag{94}$$

Similarly, take the difference between TSX inventory and TSX client prices using expression (18") and rearrange for ϕ^* to obtain the second moment:

$$R^{(2)}_\iota(\rho) \equiv \frac{m^{\text{TSX},l}_{ts,IN} - \rho\left(p^{\text{TSX},l}_{ts,CL} - p^{\text{TSX},l}_{ts,IN}\right)}{m^{\text{TSX},l}_{ts,CL}}. \tag{95}$$

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

$$\bar{g}^{(1)}_n(\rho) = \frac{1}{n}\sum_{\iota=1}^{n}\psi\big(R^{(1)}_\iota - \rho\big), \qquad \bar{g}^{(2)}_n(\phi^*;\rho) = \frac{1}{n}\sum_{\iota=1}^{n}\psi\big(R^{(2)}_\iota(\rho) - \phi^*\big). \tag{96}$$

[26] 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 parameters (ρ_I, ϕ^*) are estimated in two stages as just-identified quantile (Z-) estimators. The first-stage estimator $\hat{\rho}_I$ solves

$$\bar{g}_n^{(1)}(\hat{\rho}_I) = 0 \quad \Longleftrightarrow \quad \hat{\rho}_I = \text{median}\big\{R_\iota^{(1)}\big\}, \tag{97}$$

i.e., the sample median of $R_\iota^{(1)}$ across all valid observations. Conditional on $\hat{\rho}_I$, the second-stage estimator $\hat{\phi}^*$ solves

$$\bar{g}_n^{(2)}(\hat{\phi}^*; \hat{\rho}_I) = 0 \quad \Longleftrightarrow \quad \hat{\phi}^* = \text{median}\big\{R_\iota^{(2)}(\hat{\rho}_I)\big\}, \tag{98}$$

the median of the adjusted ratios computed using the plug-in estimate $\hat{\rho}_I$.

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.

Bootstrap standard errors are computed as the empirical standard deviations of the bootstrap estimates,

$$\widehat{\text{SE}}(\hat{\theta}) = \text{sd}\big(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}\big), \qquad \hat{\theta} \in \{\hat{\rho}_I, \hat{\phi}^*\}. \tag{99}$$

Normal-approximation confidence intervals are given by

$$\text{CI}_{95}^{\text{norm}}(\hat{\theta}) = \big[\hat{\theta} \pm 1.96\, \widehat{\text{SE}}(\hat{\theta})\big]. \tag{100}$$

I also report percentile-based bootstrap confidence intervals,

$$\text{CI}_{95}^{\text{pct}}(\hat{\theta}) = \big[Q_{0.025}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}),\ Q_{0.975}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)})\big], \tag{101}$$

and the corresponding implied standard errors,

$$\widehat{\text{SE}}^{\text{pct}}(\hat{\theta}) = \frac{Q_{0.975}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}) - Q_{0.025}(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)})}{2 \times 1.96}. \tag{102}$$

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,

$$\rho_H^{\text{lb}} = \rho_I\Big(\frac{1}{\phi^*} - 1\Big). \tag{103}$$

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

For inference, I construct the bootstrap distribution of the bound jointly with $(\hat{\rho}_I, \hat{\phi}^*)$: in each bootstrap replication b, I compute the corresponding $\hat{\rho}_H^{\text{lb},(b)}$, 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 computed analogously to those for $\hat{\rho}_I$ and $\hat{\phi}^*$. I also report the corresponding percentile–implied standard error, constructed using the same mapping from percentile bounds as above. By re–estimating all underlying objects in each bootstrap replication, this procedure ensures that uncertainty in the bound $\hat{\rho}_H^{\text{lb}}$ 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 the number of dealers N, I apply the same two–stage procedure as in the baseline estimation, restricting attention to subsamples with a fixed dealer count. Specifically, I first estimate ρ_I using the full sample, as described above. Conditional on $\hat{\rho}_I$, I estimate $\phi^*(N)$ as the median of the ratio in (95), which combines information from TSX inventory and client auctions, restricting attention to observations for which the corresponding TSX client auction involves exactly $N_{ts,CL}^{\text{TSX},,l} = N$ active dealers. Inference follows the same clustered bootstrap by trading day as in the baseline case.

Appendix Table A1: Prices are worse on the more concentrated exchange (Alpha)

	(Buy)		(Sell)	
Alpha	0.004***	(0.000)	-0.004***	(0.000)
quantity	0.154***	(0.011)	-0.115***	(0.018)
N	136,715,587		131,862,280	

Notes: Appendix Table A1 reports estimates from regressions of transaction prices on an indicator for the Alpha exchange (Alpha) and trade size (quantity), controlling for symbol × trade date × hour × dealer fixed effects. Appendix Table A4 reports robustness checks. Column (Buy) uses asset purchases, while column (Sell) uses sales. Prices are measured in Canadian dollars, and quantities in millions of shares. In both specifications, the R^2 is close to one due to the rich set of fixed effects, and standard errors are clustered at the daily level. For comparison, the median trade quantity is 0.0001 million C$, the median price is about 22.5C$, the median bid-ask spread using the national best bid and offer is 0.01C$.

Appendix Table A2: Clients trade at worse prices than dealers for their own account

	(Buy)		(Sell)	
client-account	0.005***	(0.000)	-0.005***	(0.000)
quantity	0.156***	(0.014)	-0.114***	(0.018)
N	133,615,215		128,407,743	

Notes: Appendix Table A2 reports estimates from regressions of transaction prices on an indicator for client-account trades, traded quantity, and symbol × trade date × hour × exchange fixed effects. Appendix Table A5 reports robustness checks. I do not further interact these fixed effects with dealer identifiers because there are insufficient observations of dealers operating across both account types and exchanges for the same symbol within the same hour. Own-account trades are the baseline. Column (Buy) uses asset purchases, while column (Sell) uses sales. Prices are measured in Canadian dollars, and quantities in millions of of shares. In both specifications, the R^2 is close to one due to the rich set of fixed effects, and standard errors are clustered at the daily level. For comparison, the median trade quantity is 0.0001 million C$, the median price is about 22.5C$, the median bid-ask spread using the national best bid and offer is 0.01C$.

Appendix Table A3: Welfare effects for median client auction in dollar values

ρ		ϕ			Total		Decomp.	
$\hat{\rho}_I$	ρ_H	ϕ_{UC}	$\hat{\phi}^*$	$\hat{\phi}^*_+$	ΔW_ϕ	ΔW_N	ΔW_N^{d}	$\Delta W_N^{\mathrm{ind}}$
670	415	0.62	0.62	0.71	0.00	-0.15	0.54	-0.70
670	450	0.60	0.62	0.71	0.06	-0.35	0.56	-0.91
670	500	0.57	0.62	0.71	0.24	-0.58	0.58	-1.16
670	670	0.50	0.62	0.71	1.31	-1.11	0.62	-1.73
670	800	0.46	0.62	0.71	2.40	-1.36	0.64	-2.01
670	1000	0.40	0.62	0.71	4.33	-1.62	0.67	-2.29

Notes: Appendix Table A3 is analogous to Table 2 but displays welfare effects in dollar values, in thousand C$. Percentage and level welfare changes are similar because baseline welfare is large (approximately C$105,000), dominated by the level term Av, while risk-adjustment components are small. The table reports welfare effects in levels (thousand C$) for different values of the client's risk tolerance ρ_H in a representative TSX client-account auction ($N = 25$, $A =$ C$1,604, $v =$ C$57.16, and $\Sigma_v = \overline{\Sigma} = 0.038$). Columns 1–2 list the estimated dealer risk tolerance $\hat{\rho}_I$ and the values of ρ_H. Columns 3–5 report the unconstrained contract $\phi_{UC} = \rho_I/(\rho_I + \rho_H)$, the estimated contract $\hat{\phi}^*(N)$ at $N = 25$, and the estimated contract $\hat{\phi}^*(N{+}1)$ at $N{+}1 = 26$. Columns 6–7 report total welfare changes: ΔW_ϕ is the welfare gain from eliminating moral hazard at $N = 25$, while ΔW_N is the total effect of increasing competition from N to $N{+}1$ with endogenous contract adjustment. Columns 8–9 decompose ΔW_N into a direct effect, ΔW_N^{d}, holding the contract fixed at $\hat{\phi}^*(N)$, and an indirect effect, $\Delta W_N^{\mathrm{ind}}$, arising from the contract change to $\hat{\phi}^*(N{+}1)$.

Appendix Table A4: Robustness—Prices are worse on the more concentrated exchange (Alpha)

	(Buy)		(Sell)	
Panel A: Symbol × Trade Date Fixed Effects				
Alpha	0.005***	(0.001)	-0.003*	(0.001)
quantity	0.219***	(0.028)	-0.203***	(0.032)
N	139,626,867		134,387,016	
Panel B: Symbol × Trade Date × Hour Fixed Effects				
Alpha	0.006***	(0.000)	-0.006***	(0.000)
quantity	0.129***	(0.013)	-0.104***	(0.016)
N	139,302,602		134,088,687	
Panel C: Symbol × Trade Date × Hour × Dealer Fixed Effects				
Alpha	0.004***	(0.000)	-0.004***	(0.000)
quantity	0.154***	(0.011)	-0.115***	(0.018)
N	136,715,587		131,862,280	
Panel D: Symbol × Trade Date × Hour × Dealer Fixed Effects (Small Trades)				
Alpha	0.006***	(0.001)	-0.005***	(0.000)
quantity	67.345	(37.727)	-68.919	(48.785)
N	96,753,660		91,755,344	

Notes: Appendix Table A4 reports robustness regressions of transaction prices on an indicator for the Alpha exchange and trade size. Panels differ by the granularity of fixed effects. Panel A controls for symbol × trade date fixed effects; Panel B adds symbol × trade date × hour fixed effects; Panel C further includes dealer fixed effects, conditioning on execution prices within narrowly defined symbol–time–dealer cells. Panel D replicates the Panel C specification but restricts the sample to small trades, defined as trades below the 25th percentile of the trade-size distribution. Column (Buy) uses asset purchases, while column (Sell) uses sales. Prices are measured in Canadian dollars, and quantities in millions of shares. Standard errors are clustered at the daily level. Across all specifications, trades executed on Alpha occur at worse prices than those on TSX.

Appendix Table A5: Robustness—Client-Account Price Differences Across Specifications

	(Buy)		(Sell)	
Panel A: Symbol × Trade Date × Exchange Fixed Effects				
Client-account	0.010***	(0.001)	-0.013***	(0.001)
quantity	0.253***	(0.030)	-0.213***	(0.034)
N	134,116,675		128,870,237	
Panel B: Symbol × Trade Date × Exchange × Hour Fixed Effects				
Client-account	0.005***	(0.000)	-0.005***	(0.000)
quantity	0.156***	(0.014)	-0.114***	(0.018)
N	133,615,215		128,407,743	
Panel C: Symbol × Trade Date × Exchange × Hour Fixed Effects (Small Trades)				
Client-account	0.005***	(0.000)	-0.005***	(0.000)
quantity	85.586*	(41.785)	-98.354*	(45.592)
N	95,083,035		89,801,803	
Panel D: Symbol × Trade Date × Exchange × Dealer Fixed Effects				
Client-account	0.001	(0.002)	-0.008***	(0.002)
quantity	0.187***	(0.026)	-0.198***	(0.032)
N	132,993,000		127,838,651	

Notes: Appendix Table A5 reports robustness regressions of transaction prices on an indicator for client-account trades and traded quantity. Panels differ by the granularity of fixed effects and sample restrictions. Panel A includes symbol × trade date × exchange fixed effects. Panel B further adds hour-of-day fixed effects, comparing client and own-account trades executed in the same security, on the same exchange, and within the same hour. Panel C restricts the sample to small trades, defined as trades below the 25th percentile of the trade-size distribution. Panel D includes dealer fixed effects interacted with symbol, trade date, and exchange, conditioning on price differences within narrowly defined dealer–asset–date cells. Own-account trades serve as the baseline category. Column (Buy) uses asset purchases, while column (Sell) uses sales. Prices are measured in Canadian dollars, and quantities in millions of shares. Standard errors are clustered at the daily level.

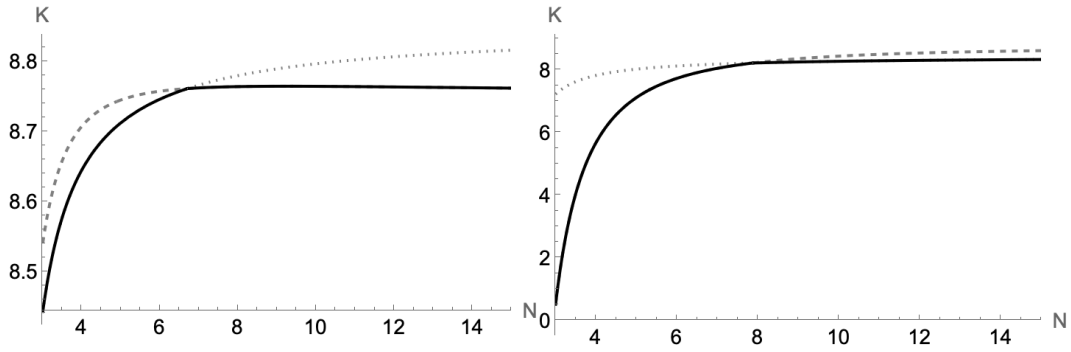
Appendix Table A6: Robustness—Parameter estimates across specifications

Parameter	Estimate	SE (SD)	SE (Pct)	95% CI Low	95% CI High
10 minutes					
ρ_I	805.528	461.094	429.936	216.588	1901.94
ϕ^*	0.666	0.042	0.040	0.602	0.761
ρ_H^{lb}	403.689	220.855	218.851	112.004	969.902
20 minutes					
ρ_I	506.907	65.4669	64.7832	390.245	644.195
ϕ^*	0.608	0.014	0.014	0.581	0.637
ρ_H^{lb}	326.405	47.5308	48.0952	234.604	423.137
30 minutes					
ρ_I	455.981	44.642	42.6444	366.303	533.469
ϕ^*	0.576	0.009	0.009	0.558	0.594
ρ_H^{lb}	334.799	34.9132	33.2562	262.411	392.775
Weekly return variance					
ρ_I	3610.320	875.799	930.790	1999.730	5648.430
ϕ^*	0.616	0.019	0.018	0.582	0.657
ρ_H^{lb}	2243.720	566.188	563.248	1212.500	3420.430
Monthly return variance					
ρ_I	18689.300	4838.380	4862.650	11276.200	30337.700
ϕ^*	0.634	0.0222	0.021	0.591	0.674
ρ_H^{lb}	10754.500	2820.610	2737.740	6539.260	17271.200

Notes: Appendix Table A6 reports robustness results for the estimated dealer risk tolerance ρ_I, the contract share ϕ^*, and the implied lower bound on client risk tolerance ρ_H^{lb}. In the three upper sections, I vary the time interval used to aggregate 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 aggregation window becomes coarser, the estimated dealer risk tolerance declines, but confidence intervals overlap across specifications. In the two lowest sections, I use alternative measures of return variance. The baseline uses within-day mid-quote variation at the symbol level. Using weekly or monthly return variances mechanically shifts the level of estimated risk tolerance, since risk tolerance is defined relative to the unit of risk, which changes in the time-horizon. Formally, prices depend on the ratio Σ/ρ_I, implying that an increase in Σ must be offset by a proportional increase in ρ_I. Across all specifications, the estimated contract share remains close to 0.6.

Appendix Figure A1: Numeric example—K^*

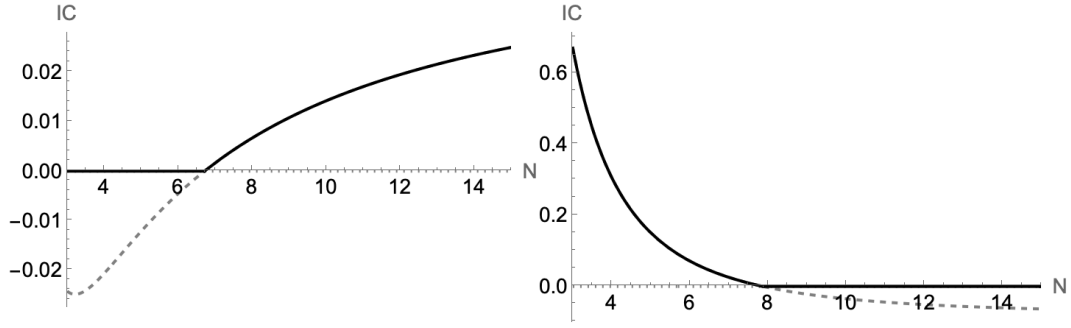
(a) K^*: Low noise from shirking (b) K^*: High noise from shirking



Notes: Appendix Figure A1 plot the equilibrium contract K^* (black line) as a function of the number of dealers N for two numeric examples. The contract is a piecewise function (thick black line) implied by the unconstrained contract (dashed gray line) and the constrained contract (dotted gray line). In both examples, $\rho_I = \rho_H = 0.5, \Sigma = \Sigma_\nu = 1, b = 0.2, \bar{U} = 9$. In Figure A1b shirking-noise is low, $\sigma_\epsilon = 0.5$ and supply is large, $A = 10$, in Figure A1a it is high, $\sigma_\epsilon = 100$ and supply is small $A = 2$. Note that the level of K^* is uninformative as it changes depending on my choice of $\bar{U}$.

Appendix Figure A2: Numeric example—Expected gain from effort relative to shirking

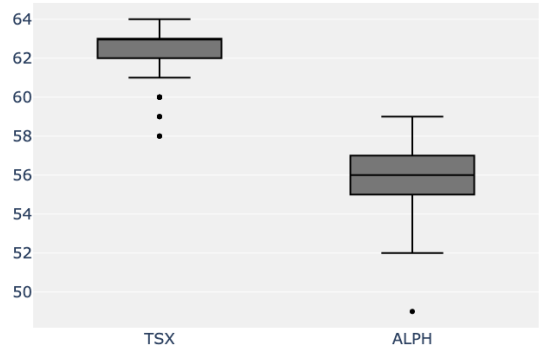
(a) IC: Low noise from shirking (b) IC: High noise from shirking



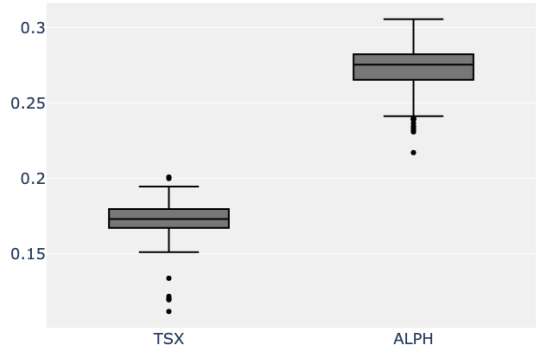
Notes: Appendix Figure A2 plot the dealer's expected payoff when exerting effort net of the expected payoff under shirking, evaluated at the equilibrium contract, for two numeric examples. When this difference is positive, the incentive constraint is slack and the contract unconstrained. The expected gain from effort is a piecewise function (thick black line) implied by the unconstrained contract (dashed gray line) and the constrained contract (dotted gray line). In both examples, $\rho_I = \rho_H = 0.5, \Sigma = \Sigma_\nu = 1, b = 0.2$, $\bar{U} = 9$. In Figure A2a shirking-noise is low, $\sigma_\epsilon = 0.5$ and supply is large, $A = 10$, in Figure A2b it is high, $\sigma_\epsilon = 100$ and supply is small $A = 2$.

Appendix Figure A3: Market concentration is higher on Alpha

(a) Number of active dealers



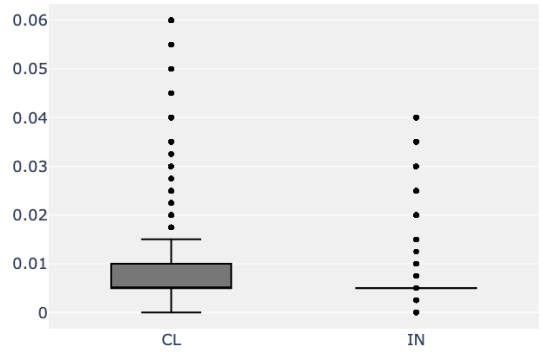
(b) Daily HHI



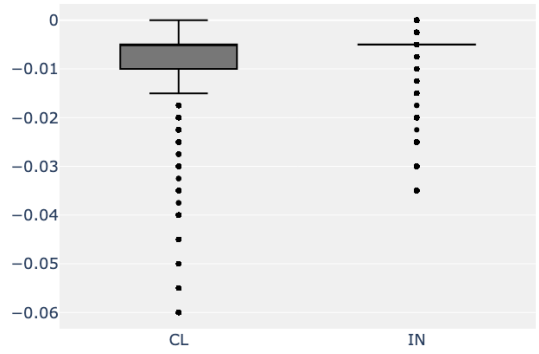
Notes: Appendix Figure A3a shows the distribution of the number of active dealers on a day for the TSX and Alpha exchange. Figure A3b shows the distribution of daily HHI indices (which is between 0 and 1) on TSX and Alpha. The $HHI_t = \sum_i (s_{ti})^2$ is computed using the fraction of total daily trade value executed by a dealer i on day t as $s_{ti} \in [0, 1]$.

Appendix Figure A4: Clients trade at worse prices than dealers for their own account—raw data

(a) Purchases



(b) Sales

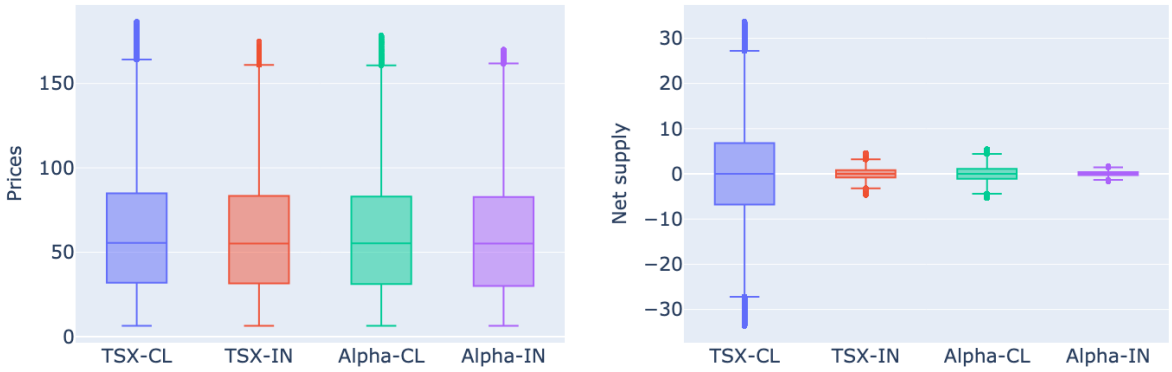


Notes: These boxplots show the distribution of the difference between the trade-price and the mid-price before the trade executes, excluding upper and lower 2.5% quantiles, using trades on TSX from July 2021 for client (CL) and own inventory (IN) accounts. The mid-price is the average of the national best bid and ask before the trade.

Appendix Figure A5: Distributions of prices, net supply, dealer counts, and return variance

(a) Prices by auction type

(b) Net supply by auction type



(c) Number of active dealers

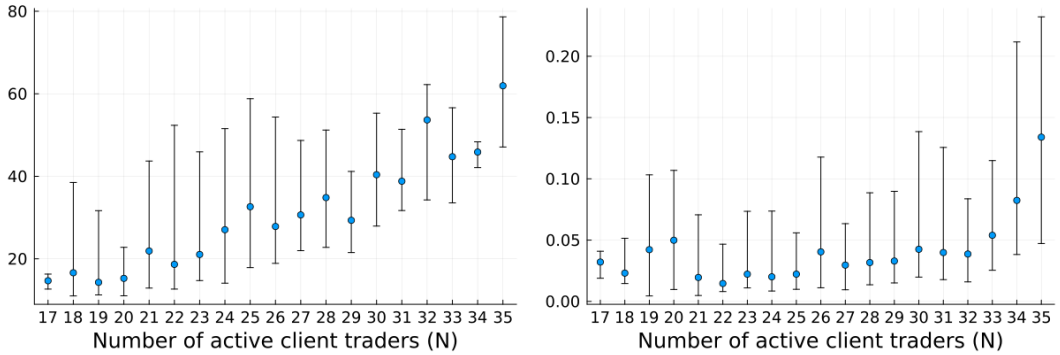
(d) Return variance, Σ_{ts}



Notes: Appendix Figure A5 shows the distribution of transaction prices (in C$), net supply (in thousand C$), the number of active dealers, and my proxy for the variance of daily returns (Σ_{ts}) across all discretized "auctions". 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 A6: Amount traded and return variance by number of dealers

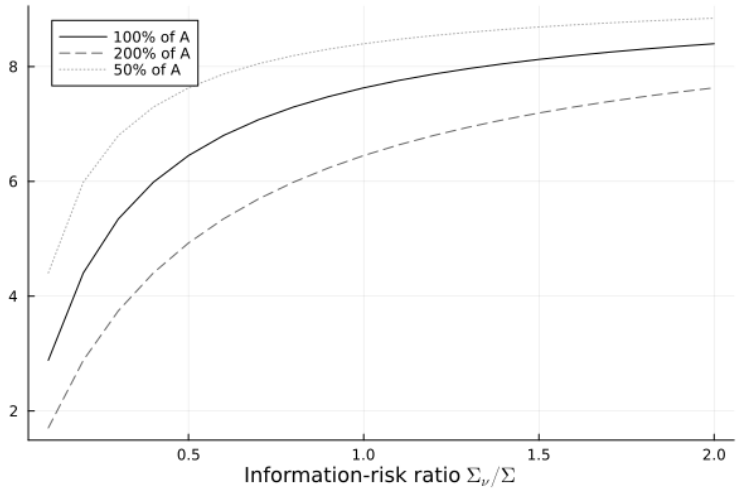
(a) Total amount traded $|A|$ (b) Return variance Σ



Notes: Appendix Figure A6a 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 dealers, N, active in that auction. The distribution is over days, securities, and sides of the auction. Figure A6b shows the distribution of return variances, Σ_{ts}, per N over days and securities. Supply is expressed in 1,000 shares (unites), and return variance in Canadian dollars.

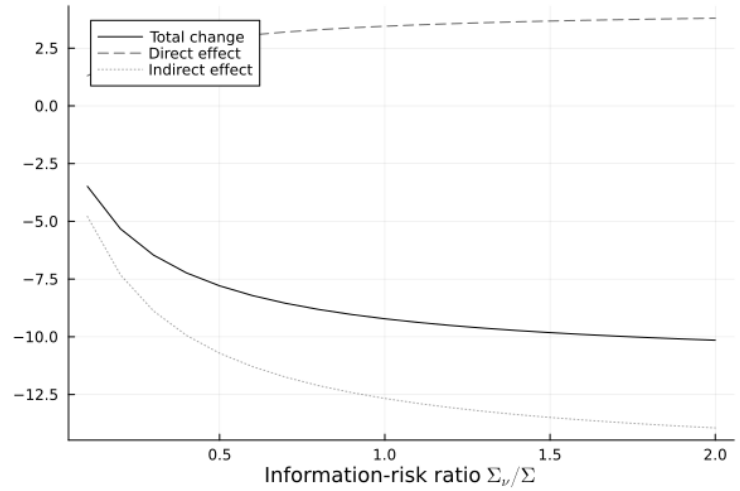
Appendix Figure A7: Welfare effects across information risk and supply

(a) Eliminating moral hazard



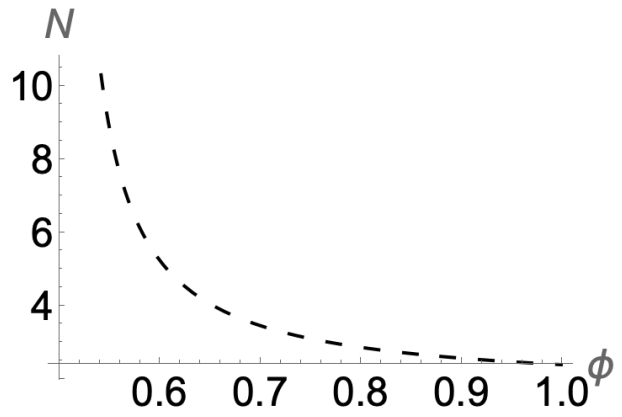
Notes: Appendix Figure A7a plots welfare effects—measured in percentage points relative to total welfare before relaxing the respective friction—for a representative TSX client-account auction, where clients and dealers share the same risk tolerance (670), as a function of the information–risk ratio, $\kappa = \Sigma_\nu/\Sigma$. The curves correspond to different auction sizes, scaled relative to the median net supply: $A/2$, A, and $2A$. The percentage gain declines in supply, because the improvement from stronger incentives grows much more slowly with supply, A, than baseline welfare, which rises roughly linearly in $A\nu$.

(b) Adding one dealer



Notes: Appendix Figure A7b shows the percentage welfare loss from increasing the number of active dealers from N to $N+1$, decomposed into the direct effect of adding a dealer (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 A8: When imperfect competition is worse than moral hazard ($\rho_I = \rho_H$)



Notes: Appendix Figure A8 shows cutoff function, $f(\phi)$, of Corollary 8 when $\rho_I = \rho_H$ in the dashed gray line. The x-axis shows ϕ from $\frac{1}{2}$, when moral hazard frictions are not binding, to 1, when all assets go to the dealer to incentivize effort.