Moral Hazard and Imperfect Competition in Financial Markets

Milena Wittwer*

July 16, 2026

Abstract

When financial intermediaries invest on behalf of clients, they exert effort that clients cannot contract on and compete when trading assets. I develop a model in which the resulting moral hazard and degree of competition interact through market clearing, so that the incentive contracts clients offer intermediaries depend on how many intermediaries compete. I show that greater competition can tighten incentive constraints and reduce welfare. Pricing patterns in proprietary Canadian equity data are consistent with both frictions, and, read through the lens of the model, suggest a welfare loss from greater competition. This calls for coordinating competition 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 discussion. For comments and suggestions I also thank Jason Allen, Ana Babus, Eric Budish, Piotr Dworczak, William Fuchs, Simon Gervais, Sergei Glebkin, Lawrence Glosten, Zhiguo He, Jian Li, Joshua Mollner, Daniel Neuhann, Christian Opp, Cecilia Parlatore, Indira Puri, Uday Rajan, Bernard Salanie, Michael Sockin, Adi Sunderam, Pietro Tebaldi, Laura Veldkamp, and all participants of seminars at Columbia University (Finance and IO), and London School of Economics, as well as the following conferences: 2025 Fall FTG meeting, 2025 Chicago Market Design Conference, 2025 Germans Abroad Christmas Meeting, 2025 Workshop for Assistant Professors in Finance, and 2026 Conference on Industrial Organization of Financial Markets.

1 Introduction

Much of the 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 and contracting within bilateral relationships. In financial markets, intermediaries invest on behalf of clients, giving rise to moral hazard concerns (He and Krishnamurthy, 2012).[1] 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). This competition is imperfect: recent empirical studies document that intermediaries hold market power across many asset markets (see Corhay et al., 2025).

Both frictions—moral hazard and imperfect competition—have been studied extensively in asset markets, but largely in separate literatures. Most work on contracting takes asset prices as given, while work on trading, where asset prices are endogenous, abstracts from principal-agent frictions. However, the same market clearing that determines the price of an asset also connects competing intermediaries who participate in it. This makes the design of incentive contracts inseparable from competition among intermediaries in the asset market and raises the question: How does competition affect investment incentives, contracts, and ultimately market efficiency?

To answer this question, I develop a model that unifies moral hazard and competition in asset markets, and relate it to proprietary data from equity exchanges. Because the two frictions interact, stronger competition can tighten incentive constraints and reduce market efficiency. This highlights the need for coordinated competition policy and conduct regulation to avoid unintended consequences.

In the model, 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 Constant Absolute Risk Aversion (CARA) preferences over wealth, potentially with different risk tolerances. The game unfolds over four periods. First, clients separately but simultaneously offer contracts consisting of a lump-sum payment and a share of investment wealth (hereafter referred to as contract share). Second, each dealer decides whether to accept the contract and, conditional on acceptance, whether to exert effort when forming individual beliefs about the expected asset return. Effort yields a correct belief, whereas shirking generates a noisy signal but a private

[1] Broadly understood, such concerns arise whenever the intermediary's handling of the client's money is not fully contractible, in delegated asset management as much as in trade execution.

benefit. Third, dealers compete in a centralized market by submitting (net-) demand schedules, specifying how much they are willing to pay for different amounts of the asset. Fourth, the market clears at the price at which aggregate demand equals supply, and 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 changing how much risk dealers are willing to bear for a given expected return. When incentive 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 both 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 much shirking degrades the quality of the dealer's signal, and therefore on how easy it is to form their return expectation. 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).

Two forces determine which contract regime prevails. The first operates through the risk-discount in the price that compensates (risk-averse) dealers for absorbing the asset. An effort dealer, being better informed, trades more aggressively and expects to hold a larger position, making them more exposed to the risk-discount. As competition intensifies and more dealers share this burden, the discount shrinks. This makes shirking relatively more attractive, so that the contract share must rise to deter it. The second force operates through information rents. As more dealers enter, the market price converges toward the market average. This widens the informational advantage of exerting effort and correctly forming individual return expectations, making shirking relatively less attractive. The contract share can fall. When the shirking noise is small, the risk-discount force dominates and competition amplifies moral hazard; when it is large, the information force dominates and competition mitigates it.

Whether increased competition raises or reduces welfare depends on the prevailing contract regime. When the incentive constraint is slack, contracts do not respond to dealer entry and welfare rises as the price-impact distortion declines, improving risk sharing (Kyle, 1989; Vayanos, 1999; Vives, 2011). When the incentive constraint binds, contracts endogenously adjust with dealer entry and an additional indirect welfare channel emerges. In the unfavorable regime, the contract share increases with competition. This distorts the dealer's information–risk trade-off by lowering their effective risk tolerance (further) below the level that would achieve efficient risk-sharing with their client, since the dealer does not correctly internalize the client's stake in the investment. This leads to inefficient risk-sharing and distorts asset allocations. The indirect channel can dominate the direct price-impact effect, causing welfare to decline as competition intensifies.

I next use the model to shed light on the relative importance of moral hazard and imperfect competition in practice. I draw on proprietary trade data covering roughly 60% of Canadian equity trading, which record whether a trade is executed for a dealer's own (proprietary) account or on behalf of a client. This distinction is essential for studying moral hazard, and motivates an extension of the model in which dealers trade both for their own accounts and for clients.

Identifying moral hazard cleanly, even for a single agent, is challenging: incentives and contracts are rarely observed directly and, when they are, reflect equilibrium outcomes (Einav et al., 2021). In financial markets the difficulty compounds, since client-intermediary contracts are private and institution-specific. Assembling contract-level data for all participants who jointly set prices (what testing the interaction of moral hazard and competition would require) is therefore largely infeasible.

Rather than attempt a clean test, I ask what trade data alone—which are observable by regulators—can reveal: patterns indicative of these frictions. Consistent with the prediction of the extended model, I show that dealers obtain systematically better prices when they trade for their own accounts than when they trade on behalf of clients. Read through the lens of the model, this price gap implies a contract share. Because it varies with the number of active dealers, I can illustrate the central prediction numerically. For liquid assets, where return expectations are relatively easy to form, welfare falls as competition rises, because the contract share dealers must be paid to exert effort increases with their number.

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 asset market competition. Once we account for market clearing, the incentive

contract between any pair of intermediary and client endogenously adjusts to the competitive environment, generating an interaction between competition and moral hazard. This mechanism is not tied to a particular modeling choice and likely extends across alternative contract design, preference specifications, and market-clearing frameworks. At a high level, my findings motivate more work that incorporates general-equilibrium interactions into contract design.

For policymakers, the key implication is that the welfare effects of pro-competitive measures depend on how easily intermediaries can form beliefs about an asset's expected return, and thus market opacity. Competition can mitigate moral hazard in opaque markets (structured credit, some derivatives, corporate bonds) but amplify it in transparent ones (government bonds, equities).[2] As a result, policies that promote competition, including lowering entry barriers or expanding platform access, may have unintended and context-dependent consequences. Effective regulation therefore requires a coordinated approach that jointly accounts for market structure and incentive provision, for instance by combining competition policy with rules on compensation, disclosure of intermediary incentives, or targeted oversight of client–dealer contracting practices.

Literature. My main contribution is to provide a theoretical framework that combines moral hazard between clients and intermediaries with strategic competition in an asset market, and to show how market clearing connects these two frictions to shape welfare. Doing so, I connect to several strands of literature.

At a high level, my paper builds on the principal–agent literature (Ross, 1973; Mirrlees, 1976; Holmström, 1979; Holmström and Milgrom, 1987), and its applications to delegated asset management (Bhattacharya and Pfleiderer, 1985; Starks, 1987) and intermediary asset pricing (He and Krishnamurthy, 2012). A common feature of these large and active literatures is their focus on a single principal–agent relationship.

I zoom out of the bilateral contracting problem, similar to a strand of the delegated asset management literature that embeds the principal–agent problem in equilibrium models of asset markets. Contributions in this literature study a range of frictions under perfectly competitive asset prices, including information acquisition (Garcia and Vanden, 2009; Sockin and Xiaolan, 2023), benchmarking and investment constraints (Cuoco and Kaniel, 2011; Basak and Pavlova, 2013; Buffa et al., 2022), search (Gârleanu and Peder-

[2]This distinction has direct bearing on recent regulatory initiatives: MiFID II, for instance, introduced pre- and post-trade transparency requirements and best-execution mandates across asset classes that differ markedly in opacity, applying a largely uniform framework to markets where welfare effects may in fact diverge.

sen, 2018), and moral hazard in benchmarking contracts (Kashyap et al., 2023). I instead study how imperfect competition in the asset market shapes asset prices and ultimately incentive contracts. This focus is motivated by recent empirical studies documenting the importance of market power across many asset markets (see Corhay et al., 2025).

To do this, I embed moral hazard and contracting between clients and intermediaries in a model where intermediaries compete strategically in a centrally cleared asset market in the spirit of Wilson (1979) and Kyle (1989). Existing contributions in this tradition abstract from moral hazard and contracting (Vayanos, 1999; Vives, 2011; Rostek and Weretka, 2012; Rostek and Yoon, 2021; Dávila and Parlatore, 2021; Kacpercyk et al., 2025); see Rostek and Yoon (2025) for an overview.

Introducing moral hazard into this setting allows me to study how strategic competition and incentive contracting interact through market clearing. This also distinguishes my work from Kyle et al. (2011), who consider a single client-intermediary pair facing perfectly competitive market makers, and from Baldauf et al. (2024), who study bilateral contracting with exogenous price impact. My framework instead features multiple client-intermediary pairs, with intermediaries competing strategically under endogenous price impact.

Because effort in my model is costly information acquisition, the paper speaks to the classic question of what sustains incentives to become informed (Stigler, 1961; Grossman and Stiglitz, 1980): here the reward for information is not the trading profit itself but a contracted share of it, which market clearing ties to the number of competitors.[3]

Finally, the paper relates to literatures in banking, corporate finance, and insurance markets that analyze the interaction between moral hazard and imperfect competition in different economic environments. In banking, greater competition is often argued to compress margins and weaken incentives, attracting riskier borrowers and undermining financial stability (Keeley, 1990; Crawford et al., 2018).[4] 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 perfor-

[3]More remotely, the paper connects to work on competitive information provision in search markets. Closest in spirit, Mekonnen and Pakzad-Hurson (2026) study intermediaries who sell information to a searching agent and show that competition between intermediaries reduces efficiency by distorting bargaining power. By contrast, I study financial intermediaries that acquire information in order to trade on behalf of clients in a centrally cleared asset market.

[4]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).

mance sensitivity (Schmidt, 1997; Scharfstein, 1988; Raith, 2003). In insurance markets, a parallel result holds for asymmetric information more broadly: Rothschild and Stiglitz (1976) show that competition can unravel efficient risk-sharing under adverse selection, an insight extended to moral hazard in competing financial exchanges by Santos and Scheinkman (2001).

My paper shares the finding that competition and moral hazard interact nonmonotonically, but differs from these literatures in that I study competition within an asset market, where prices form endogenously through central market clearing and dealers compete by submitting demand functions under incomplete information. This formulation is canonical in the trading literature but has, to the best of my knowledge, not previously been combined with moral hazard. The combination gives rise to a novel mechanism: market clearing connects incentive contracts, which distorts dealers' information–risk trade-off and generates welfare losses.

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 arises because dealers must exert costly effort to acquire information about asset returns, linking the model also to the literature on costly information acquisition (Stigler, 1961; Grossman and Stiglitz, 1980).

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 model extensions, and finally analyze the interaction between moral hazard and imperfect competition. Proofs are in Appendix A.4.

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.[5] Therefore, each hires a dealer to invest on their

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

behalf. The N clients can be interpreted as a subset drawn from a continuum of investors, with each dealer matched to one client. Under this interpretation, a market with more dealers simply serves more investors from the same underlying population.[6]

The game unfolds over four periods. In period 1, each client i offers dealer i a two-part contract (Holmström and Milgrom, 1987; He and Krishnamurthy, 2012). It consists of a fixed amount K_i, 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 $\overline{U}$ (which is low enough that each client prefers hiring their dealer over not investing). Wealth results from purchasing amount a_i at price p of the asset, which generates an unknown per-unit gross return of R at the end of the game:

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

From dealer i's perspective, the asset's return is normally distributed: $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, with known variance $\Sigma > 0$. Here ν_i is dealer i's fully informed belief about the expected return, which they can pin down precisely only through conducting research (as described below). Posterior means are drawn independently from a common prior: $\nu_i \sim \mathcal{N}(\nu, \Sigma_\nu)$, with $\nu \in \mathbb{R}$ and $\Sigma_\nu > 0$. In the baseline, disagreement about expected returns is thus driven by idiosyncratic factors, and any common component is absorbed into the known prior mean ν. In a model extension presented in Appendix A.1.1, I allow for an unknown common component: there, the expected return is $\nu_i = \nu_0 + \tilde{\nu}_i$, where $\nu_0 \sim \mathcal{N}(\bar{\nu}, \Sigma_0)$ is a common value shared by all dealers and $\tilde{\nu}_i \sim \mathcal{N}(0, \Sigma_\nu)$ is a dealer-specific idiosyncratic component.

In both cases, I maintain private dealer-specific beliefs to avoid two known problems at opposite extremes. With symmetric information, uniform-price auctions (described below) are subject to a known indeterminacy result dating back to Wilson (1979) and Grossman (1981): they admit a continuum of equilibria, so that anything goes. With a pure common value that all dealers research, linear equilibria in which demand depends on private signals break down without random asset supply, because price impacts diverge (Kyle, 1989; Rostek and Yoon, 2025), and introducing random supply would preclude a closed-form characterization of the equilibrium contract. Dealer-specific beliefs

[6] An alternative specification keeps the investor population fixed across all values of N. Specifically, a unit mass of investors, each with risk tolerance ρ_H, is partitioned among the N dealers, so that each dealer serves a mass $1/N$ with aggregate risk tolerance ρ_H/N. The analysis can be carried out analogously, replacing ρ_H with ρ_H/N.

preserve tractability while keeping the essential content: effort yields precise information, while shirking does not.

In period 2, dealers decide how much research to conduct to form their belief. A dealer who exerts effort ($s_i = 0$) observes v_i precisely. A dealer who shirks ($s_i = 1$) obtains a private benefit, $b > 0$, which in practice may reflect the value of leisure from reduced effort, and observes only a noisy signal of v_i. Thus, conditional on their effort decision, dealers observe the following private signal in period 3:

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

where $\sigma_\epsilon > 0$ denotes the variance of the noise. The client observes neither the effort decision nor the signal, so contracts cannot condition on them. Unlike in a standard principal-agent setting, exerting effort does not increase output directly; instead, it improves the quality of the dealer's information in the subsequent trading stage.

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 continuously differentiable 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 the price at which aggregate demand equals supply.[7] From the perspective of dealer i who takes choices of all other dealers as given:

$$p^c(s_i) : \sum_{j=1}^{N} a_j(p^c(s_i), \mu_j(s_j)) = A. \tag{2}$$

For a fixed realization of signals, there are two possible clearing prices, depending on whether dealer i exerts effort or shirks. I write this as $p^c(s_i)$ for brevity, but the clearing price also depends on the contracts and effort choices of all other dealers, not just dealer i's.

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

Clients and dealers have CARA utility over end-of-game wealth, a common assumption in models of market power in trade settings.[8] Risk tolerances 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 submits 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 $p = p^c(s_i)$:

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

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

Given that all random variables are normally distributed, asset returns conditional on return signals are also normally distributed: $R|\mu_i(s_i) \sim \mathcal{N}(\mathbb{E}[R \mid \mu_i(s_i)], Var[R \mid \mu_i(s_i)])$. Therefore, maximizing expected CARA utility (3), conditional on the signal $\mu_i(s_i)$, is equivalent to maximizing the certainty equivalent (CE), the wealth level that, received with certainty, gives the same expected utility as their random wealth:

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

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

The client's CE from their end-game wealth,

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

is defined analogously to the dealer's, with risk tolerance ρ_H instead of ρ_I and wealth share $1-\phi_i$ instead of ϕ_i:

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

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

[8] The CARA-normal framework, which combines CARA preferences with normally distributed returns, is standard in the related literature because it admits closed-form equilibria in uniform-price auctions. A well-known feature is the absence of wealth effects: the optimal investment does not depend on initial wealth. Wealth effects could be introduced by letting risk tolerance depend on initial wealth, but I abstract from them since they are not the focus of this paper.

Neither the dealer, when choosing effort, nor the client, when choosing the contract, knows the signal $\mu_i(s_i)$. They integrate out the uncertainty and maximize their expected CE, taken over the dealer's future signal. The expected CE is a common objective for valuing costly information acquisition (Van Nieuwerburgh and Veldkamp (2010); Veldkamp (2011)).[9] Conditional on a signal it coincides with expected CARA utility; the two differ in how they aggregate over signals, with the expected CE rewarding earlier resolution of uncertainty.[10]

What renders the model tractable is that the expected CE is quadratic in ϕ_i given demand, which yields a closed-form equilibrium contract. I could thus equivalently impose the mean–variance preferences (6) and (9) directly, rather than derive them from CARA utility, as is common in the auction literature to preserve linearity in marginal utility and the existence of linear equilibria (e.g., Vives, 2011; Rostek and Weretka, 2012; Chen and Duffie, 2021; Rostek and Yoon, 2021; Wittwer and Allen, 2026). In that case, dealers and clients would simply maximize their expected utilities in periods 2 and 3 of the game.

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 observes their private expected return signal, depending on their chosen effort-level, $\mu_i(s_i)$, and competes to buy shares of the asset in a uniform-price auction.

(4) The market clears, the asset pays its return, 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

[9] Dealers and clients adopt the same return expectations which eliminates potential frictions arising from disagreement between clients and dealers over asset returns. This allows me to isolate the distortions generated by the two core frictions in my model—moral hazard and imperfect competition—rather than those caused by heterogeneous return expectations across dealers and their clients.

[10] Maximizing the expected CE, 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. Conditional on a signal the two coincide, since the CE is a monotone (logarithmic) transformation of expected utility; choosing effort adds an expectation over signals, and the expected CE takes the logarithm inside it rather than outside. As the logarithm is concave the two differ, with the former rewarding earlier resolution of uncertainty (Van Nieuwerburgh and Veldkamp, 2010).

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 clients and dealers choose identical strategies when observing the same signal. The equilibrium objects are the contracts $\{K_i, \phi_i\}$, the dealers' effort decisions, $s_i \in \{0,1\}$, and their demand functions, $a_i(\cdot, \mu_i(s_i))$ for $s_i \in \{0,1\}$ for all i.

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 stage 3, the trading stage, where dealer i receives signal $\mu_i(s_i)$ 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 is

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

Optimal demand a_i at price p equates marginal benefit (LHS) with marginal cost (RHS). The marginal benefit is the expected asset return net of a risk adjustment that increases with the dealer's position. This adjustment depends on the return variance, the dealer's risk tolerance, and the contract share: the more the dealer keeps of the asset, the lower their effective risk tolerance. The marginal cost is the price plus the dealer's own price impact Λ_i, which measures how much the clearing price moves with an additional unit of demand, multiplied by a_i. When all other dealers submit the equilibrium demand (18),

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

Solving (10) for a_i pins down the demand function, $a_i(p, \mu_i(s_i))$.

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

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

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 ex-

ert effort. Since the client aims at incentivizing effort, 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

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

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 (M), the optimization problem simplifies to:

$$\max_{\phi_i} \left\{ \mathbb{E}[\mathcal{W}(\phi_i;0)] + \mathbb{E}[\mathcal{U}(0;\phi_i)] - \bar{U} \right\} \text{ subject to (IC)}. \tag{12}$$

There are two solutions, depending on whether the constraint binds or is slack. When it is slack, $\phi^* = \phi_{UC} = \frac{\rho_I}{\rho_I + \rho_H}$, the dealer perfectly internalizes that they also invest for their 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 ϕ_{UC} 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.,

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

where $\alpha(N) > 0$ and $\beta(N) < 0$ 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 3 in the appendix specifies the analogous equilibrium with a common value component.

Proposition 1. *Let the dealer's benefit from shirking be neither too low nor too high, $b \in [\underline{b}, \overline{b}]$, with bounds characterized in the proof of the Proposition. The interval is non-empty*

whenever $\rho_H < \infty$*, and in the risk-neutral limit* ($\rho_H \to \infty$) *whenever* $(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_\nu < 0$.

There is an equilibrium in which (i) all clients offer contract $\{K^*, \phi^*\}$*:*

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

and

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

where

$$\alpha(N) \equiv \frac{1}{2N\rho_I}\left(\frac{N-2}{\Sigma} + \frac{N}{\sigma_\epsilon} + \frac{N}{\Sigma_\nu}\right)^{-1} > 0, \tag{16}$$

$$\beta(N) \equiv -\frac{(N-2)\,\rho_I\left(\tilde{\Sigma}-\Sigma\right)\left((N-1)\Sigma + (N-2)\Sigma_\nu\right)}{2(N-1)\,\Sigma\left(2\Sigma + (N-2)\tilde{\Sigma}\right)} < 0, \tag{17}$$

given $\tilde{\Sigma} = Var(R|\mu_i(1))$.

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

$$a^*(p, \mu_i(s^*)) = \rho_I(\Lambda^*\rho_I + \phi^* Var[R \mid \mu_i(s^*)])^{-1}(\mathbb{E}[R \mid \mu_i(s^*)] - p), \tag{18}$$

$$\text{with } \Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\Sigma, \tag{19}$$

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

The symmetric effort-inducing equilibrium exists as long as the dealer's benefit from shirking is neither too low nor too high: $b \in [\underline{b}, \overline{b}]$. The lower bound, $\underline{b} = -\beta(N)$, guarantees that shirking is tempting enough that a dealer with a sufficiently small contract share would shirk; below it, the dealer would exert effort, and the contract would always collapse to the unconstrained risk-sharing rule $\phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$. The upper bound $\overline{b}$ rules out deviations to contracts that tolerate shirking: because the lump-sum fee adjusts to the dealer's participation constraint, a client who tolerates shirking pockets the dealer's private benefit, which becomes attractive when b is large.

Ruling out deviations toward tolerating shirking is what makes proving existence hard. When a client deviates to a different contract share, the dealer responds on two margins: it trades differently, and it may stop exerting effort. Because the dealer's payoffs under effort and shirking form through market clearing with all other dealers, solving for the contract shares at which the dealer would shirk is not tractable. One therefore cannot

simply split deviations into those that preserve effort and those that induce shirking, and check each group separately.

The trick that nevertheless rules out all such deviations is to realize that the optimal division of risk between client and dealer does not depend on the dealer's action. The share that maximizes their joint surplus depends only on the two risk tolerances, $\phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$, and not on the quality of the dealer's information. In other words, shirking degrades the surplus, but not how best to divide it. A client who tolerates shirking can therefore never earn more than the shirking payoff at ϕ_{UC}, whether or not that share actually induces shirking. This bounds all such deviations by a single benchmark, without characterizing which contracts induce shirking: they are unprofitable as long as this benchmark payoff, plus the pocketed benefit b, stays below the equilibrium payoff, which is the condition that defines $\overline{b}$.

When clients are risk averse ($\rho_H < \infty$), the interval $[\underline{b}, \overline{b}]$ is non-empty for all admissible parameters, and it contains values of b for which the incentive constraint is slack as well as values for which it binds. The constrained contract $\Phi(N)$ increases in b, starting from zero at $b = \underline{b}$, and the constraint binds once $\Phi(N)$ exceeds the unconstrained risk-sharing rule $\phi_{UC} > 0$, so the interval opens with slack values of b. Where the constraint starts to bind, tolerating shirking still does not pay. A client who tolerates shirking loses twice: the dealer's expected payoff from trading falls, a loss the client absorbs through the fee, and the client's own expected payoff from trading deteriorates as well. At that point, the pocketed benefit covers exactly the first loss but not the second. The upper bound $\overline{b}$ therefore lies strictly above the point where the constraint starts to bind, and with it a range of binding values of b.

Only in the risk-neutral limit ($\rho_H \to \infty$) does non-emptiness require the additional condition stated in the proposition: the shirking noise must be large relative to the signal variance under effort, because the unconstrained contract then leaves the dealer with nothing: $\phi_{UC} \to 0$. Thus, when the information gain from inducing effort is large, no risk-neutral client finds it profitable to tolerate shirking.

Given contracts, effort decisions, and demand functions, market clearing pins down the equilibrium price (20) and allocations.

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

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

and dealer i purchases amount

$$a^*(p^*, \nu_i) = \frac{A}{N} + \frac{\rho_I}{\phi^*}\frac{N-2}{N-1}\left(\nu_i - \frac{1}{N}\sum_i \nu_i\right)\Sigma^{-1}, \tag{21}$$

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

Summarizing, 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 the asset market, creating an interaction between competition and moral hazard. Although each friction is well understood in isolation, their interaction through market clearing has not been pointed out to the best of my knowledge. While the specific predictions may vary with alternative modeling choices, this idea should generalize across different approaches to contract design, preferences, and market clearing.

Discussion and extensions. Before analyzing the interaction between competition and moral hazard in more detail, I discuss the key modeling choices that let me combine strategic trading, contracting, and costly information acquisition in a single 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/mean-variance preferences, normally distributed returns, or linear equilibria).

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.1.1 I relax this assumption by decomposing the expected return ν_i into an unknown common component, reflecting general market conditions, and an idiosyncratic component, reflecting institution-specific beliefs. 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 component. The key insight that ϕ^* changes in the degree of competition carries over to this richer setting; so does the main welfare finding that more competition can reduce welfare, because equilibrium contracts endogenously adjust.

Clients offer simple, linear compensation contracts, K_i, ϕ_i, consisting of a fixed fee and a constant share of the future wealth generated by the asset. Restricting attention to linear contracts is common in this literature and has a foundation in Holmström and Mil-

grom (1987), where linearity arises endogenously as the optimal contract.[11] It also mirrors common arrangements in asset management, hedge funds, and investment banking, which pair a fixed component (a salary or management fee) with a performance-based component (a bonus, performance fee, or carried interest) reflecting cumulative success over a review period rather than the outcome of any single market event.[12] Such compensation rewards absolute performance; in Appendix A.1.2 I show that benchmark-relative compensation (as in Admati and Pfleiderer, 1997; Kashyap et al., 2023, among others) raises the contract share required to induce effort, and that competition still shapes equilibrium contracts through market clearing.

Following He and Krishnamurthy (2012), I focus on equilibria in which clients incentivize effort. While no client gains from deviating away from this equilibrium toward tolerating shirking, a symmetric equilibrium in which all clients tolerate shirking may coexist. In it, the contract would again depend on N, since the client caps the dealer's share at the level that keeps shirking incentive-compatible—the analogue of (13)—and this cap varies with N through market clearing. The paper's core insight therefore does not depend on this focus. Shirking equilibria can be ruled out altogether by allowing clients to suffer a direct loss $\Delta \geq 0$ when their dealer shirks, as in He and Krishnamurthy (2012): when $\Delta \geq b$, the equilibrium of Proposition 1 is the only symmetric equilibrium with linear demand, which is a common refinement in the related literature (Rostek and Yoon, 2025).

I also follow He and Krishnamurthy (2012) in modeling the dealer's effort decision as binary. This delivers tractability and a clean closed-form characterization of the equilibrium contract. Under continuous effort, sketched in Appendix A.1.3, 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.

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;

[11] 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.

[12] Market outcome-contingent contracts would break tractability in the model, since dealers' demands would no longer be linear in prices, a feature known to render uniform-price auctions intractable.

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)).

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, as characterized (in my model) by expression (19) 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.[13] 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 outcome.[14]

Moral hazard. Moral hazard affects 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^*, which I often refer to as the contract.

When the benefit of shirking is sufficiently low so that the dealer has no incentives to shirk, the contract is unconstrained: $\phi^* = \phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$. This contract achieves the first-best solution, maximizing risk sharing across all dealers and clients, when the market

[13] Proposition 5 in Appendix A.1.4 shows how demands and contracts change when dealers behave as price-takers. The contract still depends on the number of dealers: the central message does not rely on strategic demand reduction, only on finitely many dealers whose orders clear centrally.

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

is perfectly competitive ($N \to \infty$). When the incentive constraint binds, which is the case when ϕ_{UC} 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 lowers the dealer's effective risk-tolerance ρ_I/ϕ^*, thus makes the dealer more risk-averse. 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 (20) needs to decline to persuade them to absorb extra amounts of the asset, and price impact (19) 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 constraint binds. The higher the contract, i.e., the worse the moral hazard friction, the larger this distortion.[15]

How the frictions interact. It is ambiguous whether stronger competition (measured by the number of dealers) alleviates or amplifies moral hazard (captured by the size of the contract necessary to incentivize effort). Throughout, I treat N as a continuous variable when taking derivatives, which is common in the literature studying entry and exit, and focus on parameter constellations under which entering dealers play the effort-inducing equilibrium of Proposition 1.

To build intuition for the effects of stronger competition, I first consider the limiting case $N \to \infty$, where all dealers are price takers. In this limit, the only force determining whether a dealer wants to shirk is how much their return signal deteriorates relative to exerting effort, i.e., the shirking signal noise, σ_ϵ.

Corollary 2. *Define* $\gamma_\infty(\sigma_\epsilon) \equiv \frac{\rho_I \sigma_\epsilon \Sigma_\nu (\Sigma+\Sigma_\nu)}{2\Sigma(\sigma_\epsilon \Sigma_\nu + \Sigma(\sigma_\epsilon+\Sigma_\nu))}$, *and 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 $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$.

[15] 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 constraint is binding, $a^*_{i,C}$. This gives: $a^*_{i,C} - a^*_{i,UC} = [(\rho_I+\rho_H) - \frac{\rho_I}{\Phi(N)}]\frac{N-2}{N-1}(\frac{1}{N}\sum_i \nu_i - \nu_i)\Sigma^{-1}$.

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:

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

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:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(0;\phi_i)] = \frac{\rho_I}{2}\frac{Var[\mathbb{E}[R|\mu_i(0)]]}{Var[R|\mu_i(0)]} = \frac{\rho_I}{2}\frac{\Sigma_\nu}{\Sigma}. \tag{22}$$

When they shirk, they expect to get:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(1;\phi_i)] = \frac{\rho_I}{2}\frac{Var[\mathbb{E}[R|\mu_i(1)]]}{Var[R|\mu_i(1)]} + b = \frac{\rho_I}{2}\frac{\Sigma_\nu^2(\Sigma_\nu+\sigma_\epsilon)^{-1}}{\Sigma+(\Sigma_\nu^{-1}+\sigma_\epsilon^{-1})^{-1}} + b. \tag{23}$$

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

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

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

Since the cutoff $\gamma_N(\sigma_\epsilon)$ *itself varies with* N, $\Phi(N)$ *need not be monotone in* N. *Both regimes are consistent with equilibrium existence: for any* $N > 2$, *there are parameters under which* b *lies in the valid range* $[\underline{b}, \overline{b}]$ *of Proposition 1, the incentive constraint binds, and* $b > \gamma_N(\sigma_\epsilon)$; *and likewise with* $b < \gamma_N(\sigma_\epsilon)$.

Corollary 3 underscores the non-linear interaction between moral hazard and imperfect competition when the incentive constraint binds, foreshadowed in Corollary 2. Whether increasing the number of dealers worsens or relaxes the moral hazard problem depends on the noise from shirking: when σ_ϵ is small, the fraction of wealth the dealer must obtain to exert effort increases in N; when σ_ϵ is large, it decreases.

For intuition, recall that the contract $\Phi(N)$ must leave the dealer indifferent between effort and shirking. When the dealer shirks while all others exert effort, their signal becomes noisier. Their conditional return expectation shrinks toward the common prior, $\mathbb{E}[R \mid \mu_i(1)] = \nu + \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}((\nu_i + \epsilon_i) - \nu)$, and they trade less aggressively. This is because the conditional variance increases from $Var[R \mid \mu_i(0)] = \Sigma$ to $Var[R \mid \mu_i(1)] = \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, which scales down the submitted order (characterized by (10)). The clearing price becomes less accurate as the shirking dealer's noisy signal enters the price instead of a precise one.

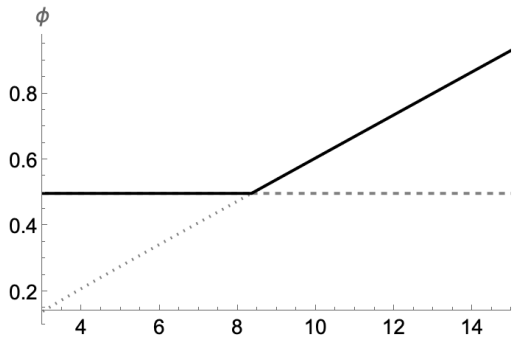
As N changes, two forces—formally characterized in Appendix A.2—shift the dealer's indifference point between effort and shirking in opposite directions, with their relative strength determined by the size of the shirking noise, σ_ϵ. The first operates through the price risk discount (equivalently, the risk premium on the asset's return). A dealer who exerts effort holds a larger position than one who shirks, and therefore benefits more from the price discount that compensates risk-averse dealers for absorbing the asset. As N grows and more dealers share the burden of absorbing the asset, this discount shrinks, reducing the risk-discount advantage of effort and making shirking relatively more attractive. The second is an information-driven force. As N grows, the clearing price converges toward the common prior ν, pulling away from the effort dealer's precise individual belief ν_i while moving toward the shirking dealer's assessment, which is already shrunk toward ν. This widens the information-driven advantage of effort, making effort relatively more attractive.

Which force dominates depends on the degree by which information degenerates when shirking, i.e., the shirking noise, σ_ϵ. When σ_ϵ is small the risk-discount force dominates. Shirking becomes relatively more attractive as N grows, and $\Phi(N)$ must increase in N. When σ_ϵ is large, the information-driven force strengthens and dominates, making shirking relatively less attractive as N grows, and $\Phi(N)$ decreases in N.

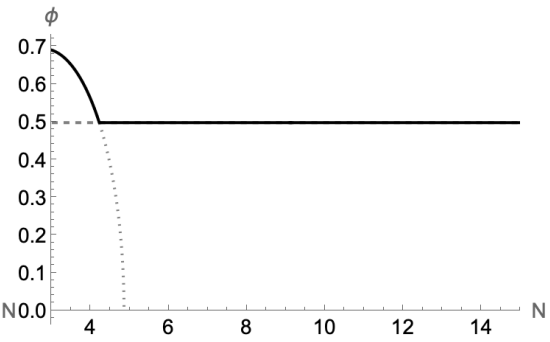
This result generalizes when expected returns share a common component (Corollary 4 in Appendix A.1.1). A third force then appears: the clearing price now aggregates information about the common component, so a shirking dealer can recover part of it from the price. Since information aggregation becomes more precise the more deal-

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

(a) Unfavorable regime



(b) Favorable regime



Notes: Figure 1 plots 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, $\rho_I = \rho_H = 0.5, \Sigma = 1$. In Figure 1a shirking-noise is low relative to the signal variance ($\sigma_\epsilon = 12$, $\Sigma_\nu = 16$), the shirking benefit is large ($b = 11.5$), and supply is high enough that the constraint is slack when dealers are few ($A = 46$); in Figure 1b shirking-noise is high relative to the signal variance ($\sigma_\epsilon = 100$, $\Sigma_\nu = 1$), $b = 0.16$, and supply is low enough that the constraint binds when dealers are few ($A = 1.3$). At every N, $b \in [\underline{b}, \overline{b}]$, except for $N \geq 5$ in Figure 1b, where $b < \underline{b}$, so that shirking is never tempting and the contract is unconstrained.

ers trade, this force makes shirking more attractive as N grows. However, its effect is bounded, because the price never reveals a dealer's own return as precisely as effort does. Therefore, this third force does not overturn the fact that there are two contract regimes, analogous to Corollary 3.

Whether the incentive constraint binds, so that $\Phi(N)$ becomes relevant, depends on all model parameters, in particular, the shirking benefit and signal noise, as well as asset supply. When clients are risk-averse, the incentive constraint binds when the shirking noise and asset supply are sufficiently small; when asset supply is large, the incentive constraint never binds, because a large supply forces each dealer to absorb a large position, which is too costly under shirking given higher perceived risk.[16]

Figure 1 illustrates the two contract regimes through numerical examples: in the unfavorable-regime example the constraint starts to bind as N grows and the contract share rises with competition; in the favorable-regime example the constraint binds only when dealers are few, and the contract share falls with competition until it becomes unconstrained.

[16]When clients are risk-averse, the constraint binds at N if and only if $A^2 < \frac{b+\beta(N)}{\alpha(N)}\left(\frac{\rho_I+\rho_H}{\rho_I}\right)^2$; otherwise the contract is unconstrained. When clients are risk-neutral the size of supply does not matter.

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 some 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:

$$W(\phi^*, N) = \sum_i W_i(\phi^*, N), \tag{24}$$

$$W_i(\phi^*, N) = a^*(p^*, \nu_i)\nu_i - \sum_{k\in\{h,i\}} \frac{1}{2}\phi_k^{*2}\frac{1}{\rho_k} a^*(p^*, \nu_i)\Sigma\, a^*(p^*, \nu_i), \tag{25}$$

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

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

where $\zeta(\phi, N)$ is given by expression (28) 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 ex-

pected asset return, v_i, like in my model but they keep all assets ($\phi = 1$). Welfare is

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

with $\eta(\phi,N)$ given by expression (27) in Proposition 2. Relative to the first benchmark, an additional force appears: the information–risk ratio Σ_v/Σ, which raises welfare when information is precise.

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*

$$\mathbb{E}[W(\phi^*,N)] = Av - \zeta(\phi^*,N)\Sigma A^2 + \eta(\phi^*,N)\zeta(\phi^*,N)\frac{\Sigma_v}{\Sigma}, \tag{26}$$

with

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

$$\zeta(\phi^*,N) = \frac{1}{2N}\left(\frac{1}{\rho_I}\phi^{*2} + \frac{1}{\rho_H}(1-\phi^*)^2\right), \tag{28}$$

where $\eta(\phi^*,N) \geq 0$ *for* $\frac{\rho_I}{\rho_I+\rho_H} \leq \phi^* < 1$, *and* $\zeta(\phi^*,N) \geq 0$.

(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

$$\underbrace{\frac{\partial \mathbb{E}[W(\phi,N)]}{\partial N}}_{\text{Direct effect}} \underbrace{- \frac{\partial \zeta(\phi,N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\Sigma A^2 + \frac{\partial \eta(\phi,N)\zeta(\phi,N)}{\partial \phi}\frac{\partial \Phi(N)}{\partial N}\frac{\Sigma_v}{\Sigma}}_{\text{Indirect effect through } \Phi(N)} < 0, \tag{29}$$

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

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

Reducing moral hazard is always welfare enhancing: statement (i). As the contract share ϕ^* declines, the loss from risk aversion falls ($\frac{\partial \zeta(\phi^*,N)}{\partial \phi} \geq 0$). This is because the dealer is effectively more risk-averse than the client, $(1-\phi^*)/\rho_H < \phi^*/\rho_I$, when the incentive constraint binds, i.e., $\frac{\rho_I}{\rho_I+\rho_H} \leq \phi^*$. Therefore, shifting some fraction of the

asset from the dealer to the client reduces the total risk bearing cost, and improves risk-sharing.[17] 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 by $\frac{\partial(\eta(\phi^*,N)\zeta(\phi^*,N))}{\partial\phi} \leq 0$, which shows that lowering ϕ^* strengthens the information-driven component of welfare.

Increasing competition by adding dealers can either raise or reduce welfare—statement (ii). 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 $-\frac{\partial\zeta(\phi^*,N)}{\partial N}\Sigma A^2 \geq 0$ 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 $\frac{\partial(\eta(\phi^*,N)\zeta(\phi^*,N))}{\partial N}\frac{\Sigma_v}{\Sigma} \geq 0$.

When the contract is constrained, welfare can decline, because the number of dealers then also acts on it indirectly, through the equilibrium contract. This finding generalizes to the setting with common values (Appendix A.1.1). The 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, which reflects reduced risk sharing and suboptimal use of signal information per unit of risk, outweighs the positive direct effect of competition.

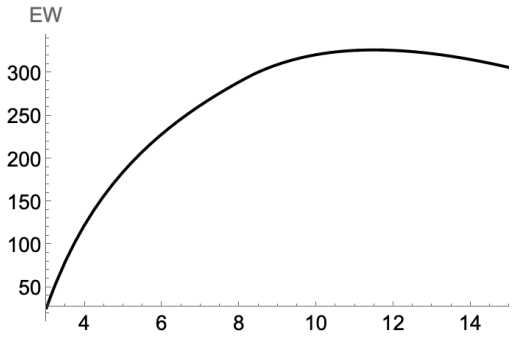
To illustrate the welfare result, I return to the two numerical examples of Figure 1. In the unfavorable-regime example, the constrained contract rises with N, and this negative indirect effect eventually outweighs the positive direct effect: welfare peaks at $N \approx 11$ and falls thereafter (Figure 2a). In the favorable-regime example, the contract falls with N where the constraint binds, so the indirect effect reinforces the direct effect, and welfare rises at every N (Figure 2b).

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 increased competition promotes efficiency (absent other frictions). Appendix A.3 compares the welfare impact of

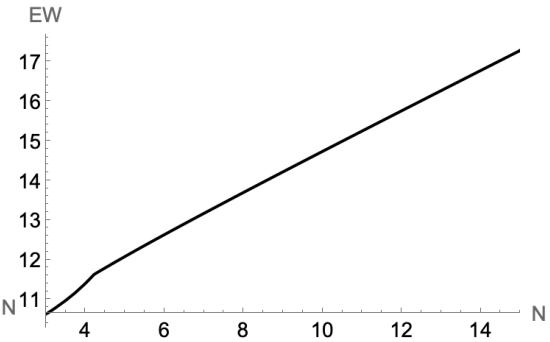
[17] For instance, if dealer and client have equal risk tolerance ($\rho_H = \rho_I$), welfare rises when moving from a situation in which the dealer receives more than half the surplus toward a more balanced split.

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

(a) Unfavorable regime



(b) Favorable regime



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, with $\nu = 8$. In Figure 2a (unfavorable regime) welfare peaks at $N \approx 11$ and declines thereafter; in Figure 2b (favorable regime) welfare increases at every N.

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 Suggestive empirical evidence

In the remainder of the paper, I turn to proprietary trade data from Canadian equity exchanges, with two goals. The first is to show that prices display patterns consistent with the two frictions; the second is to interpret these empirical patterns through the lens of the model to illustrate its central welfare prediction numerically.

Cleanly identifying moral hazard is notoriously hard even for a single principal-agent pair, because incentives, contracts, and effort are rarely observed; even if they are, they are equilibrium outcomes. My question is harder still: the interaction I study works through market clearing, so a clean test would require observing these for all participants whose orders jointly set the price, not just one, so I do not attempt it.

I ask instead what trade data alone can reveal using data regulators see even when contracts and effort are private. A regulator, who observes these data, could run this as a first screen, then request contract and effort data from the markets it flags.

3.1 Institutional environment

I use proprietary data from the Canadian equity market, which provides a useful setting.

Canadian equity market. Canadian regulation requires equity trades to be executed on a regulated marketplace, and grants dealers greater discretion in trade execution than the United States by not enforcing a strict National Best Bid and Offer (NBBO) regime—as do other major markets such as the European Union and the United Kingdom. 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.[18]

I focus on the Toronto Stock Exchange (TSX), 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 are permitted to directly submit orders. All other investors access the market through these members, which actively execute trades both for their own account and on behalf of clients. My model can accommodate this two-account structure: in Appendix A.1.5, each dealer trades for clients and on its own account, with everything else unchanged.

Member firms are heterogeneous; some are asset managers, who fit the model most closely, while others are banks or broker-dealers. I call them dealers like in the model. I expect clients to range from retail investors to large institutions; the data, however, do not allow me to identify them. For more institutional details, see Wittwer and Uthemann (2024).

Moral hazard and imperfect competition on equity exchanges. In handling client orders, dealers 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.

Several forces consistent with moral hazard could worsen trade execution for clients. A dealer might be less attentive to order flow and the limit order book, which can reveal

[18]Dealers in the United States also retain discretion in trade execution and face heterogeneous competitive conditions across venues, so the moral-hazard concerns discussed below apply there as well. NBBO price protection does not remove them as it does not dictate how orders are prepared or handled. It does, however, limit price dispersion, making these effects harder to detect empirically.

information about the asset's resale value, at least in the short run. It might also route client orders across venues and time their execution with more or less care, work less hard to limit price impact when trading for a client rather than for itself, or reserve its fastest algorithms for its own trading. None of these forces requires clients to be less informed about the (long-run) fundamental asset values than their dealers: what matters is that the care with which the dealer prepares and executes their orders is not contractible.

My (extended) model captures these forces in a stylized way: a client's trade executes at a worse price because the dealer keeps part of the surplus, pushing the price up when the client buys and down when it sells. Holding all else equal, the price (20) at which a client buys is higher the less the dealer retains, ϕ^*—that is, the smaller its own stake in the trade. The contract enters this price only by scaling the dealer's effective risk tolerance, ρ_I/ϕ^*: the smaller the stake, the less compensation the dealer demands for absorbing a position, with trading on own account corresponding to $\phi^* = 1$. The prediction therefore does not hinge on the model's specific source of moral hazard (costly information acquisition).[19] The contract share can discipline whatever behavior the client cannot contract on, and a price gap between client and own-account trades is the imprint such incentive provision would leave in prices.[20]

Similar to various sources of moral hazard, several mechanisms consistent with imperfect competition could shape prices. Trading on the exchange is continuous. Dealers submit, cancel, and update orders throughout the day. My model does not capture this continuous submission process in detail—nor does my data contain it—but it captures its first-order consequence for prices: dealers have price impact, so prices lie below fundamental value when they buy and above it when they sell, and this wedge shrinks as competition increases. How large the wedge is depends on how many dealers are active and how much they trade in a given interval.

[19] Even taking the information-acquisition channel literally, the model can be interpreted more broadly as an abstract order-submission game. Demand schedules correspond to the (net) orders dealers submit, R is the security's resale value over the relevant trading horizon, and v_i is dealer i's assessment of that value based on the information available when executing the order, such as proprietary data feeds and its own client order flow. Effort then corresponds to paying attention while working the order (monitoring incoming order flow and the state of the limit order book). Because the dealer internalizes only a fraction of the resulting gains or losses, increasing its contractual stake strengthens its incentive to exert attention.

[20] In practice, dealers might not be paid an explicit share ϕ^* of the trading surplus, but through arrangements such as commissions and relationship business. Such compensation does not remove the misalignment: the dealer is paid regardless of the price it obtains for the client. In the model, ϕ^* is exactly this stake, i.e., how much the dealer internalizes the price it obtains for the client.

Table 1: 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: Table 1 reports estimates from regressions of transaction prices on an indicator for client-account trades, trade-size, and symbol × trade date × hour × exchange fixed effects. 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 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 shares, the median price is about C$22.5, the median bid-ask spread using the national best bid and offer is C$0.01.

3.2 Data and pricing patterns

Data. The 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 (i.e., its symbol), trade size, price, and the national best bid and offer prior to 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, which are rarely observable to researchers yet known to regulators, and are crucial for studying 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 dealer's own account or on behalf of a client, a distinction that is central to analyzing moral hazard.

Pricing patterns. Two pricing patterns are consistent with the model's frictions, though suggestive rather than causal. I document the first pattern in more detail, since it is new to the literature, and treat the second briefly because it confirms existing evidence; Appendix B.1 provides robustness evidence and discussion.

Clients trade at systematically worse prices than dealers do on their own account. Table 1 (visualized in Appendix Figure A2) shows 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. This is what moral hazard predicts—a dealer has weaker incentives executing for a client than for itself—

and, to my knowledge, the pattern has not been documented in equity markets, since the account behind a trade is rarely observed.

Prices are worse on the less competitive exchange (Alpha). On Alpha, fewer dealers are active per day, and daily concentration indices (HHI) are systematically higher than on TSX (Appendix Figure A3). If competition disciplines pricing, Alpha should exhibit worse execution prices.[21] A regression controlling for trade size and account-type-stock-day-hour fixed effects shows that dealers buying on Alpha pay prices that are on average approximately C$0.006 higher (roughly 3 basis points of the median stock price or 60% of the median bid–ask spread) with the reverse holding for sales (Appendix Table A2).[22] The direction of the effect is consistent with existing evidence that competition shapes pricing on exchanges. The regression further confirms findings from early empirical studies on the NYSE and regional U.S. exchanges (Bernhardt et al., 2005) that larger trades are executed at worse prices, consistent with price impact—a key feature of the model.

3.3 From pricing patterns to contract shares and welfare

Read through the lens of the (extended) model, these pricing patterns imply a contract share. Because it varies with the number of active dealers, I can use it to illustrate the central prediction: for liquid assets, welfare falls as competition rises, since the contract share dealers must be paid to exert effort increases with their number. The exercise is a numerical illustration of the mechanism, not a causal or fully structural estimation of the model. I only sketch it here; Appendix B.2 develops it in full.

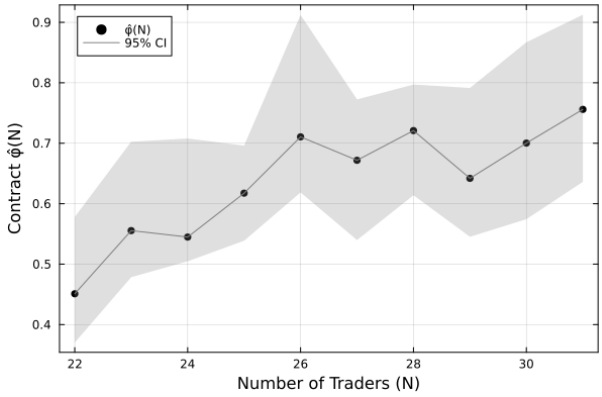
Calibration. To map the model's one-shot auction to continuous trading, I restrict the sample to liquid stocks and discretize the trading day into 15-minute intervals per exchange, security, trade side, and account type. The average price in each interval is interpreted *as if* it were the clearing price of a batch auction (Budish et al., 2015), with active orders as net demand and passive resting orders as supply.

The average price gap between client and own-account trades (within such an interval) then depends on the number of active dealers (N), the total amount traded, the stock's return volatility, the dealer's risk tolerance (ρ_I), and the contract share ($\phi^*(N)$).

[21] Recall that 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; a shared institutional structure that rules out many, though not all, confounding factors unrelated to competition.

[22] Note that this finding does not imply that the trade executed on Alpha could have obtained a better price on TSX, as this counterfactual is unobserved.

Figure 3: How ϕ^* changes in N



Notes: The figure reports $\hat{\phi}^*(N)$ and its 95% confidence interval for each dealer-count group with at least 50 observations. Estimates from thinner groups are too noisy to be informative, so the figure restricts the support of dealer counts to 22–31; the full, unrestricted range (17–35) is shown in Appendix Figure A4.

I observe all of these but the last two, and recover them in two steps. First, the cross-exchange price gap for own-account trades does not involve the contract share. When a dealer trades for its own book it keeps the entire position, so the relevant share is one rather than $\phi^*(N)$. Therefore, this gap is governed only by ρ_I and observables, and pins down ρ_I. Second, given ρ_I, the client–own-account price gap on a single exchange depends, among the unknowns, only on $\phi^*(N)$, and so identifies the contract share at a given N.

Figure 3 shows that the implied contract $\phi^*(N)$ increases from about 0.45 to 0.76 (by 70%) as the number of active dealers rises from N =22 to 31 (by 40%). The figure restricts attention to dealer-count groups with at least 50 observations, since estimates from thinly populated groups are noisy; the full, unrestricted range is shown in Appendix Figure A4. The level of these contract shares is relatively large, stemming from the sizable price gap between client and own-account trades (50% of the median bid–ask spread). This result is qualitatively consistent with existing evidence that intermediaries capture a substantial part of the surplus they create; Appendix B.2 discusses this evidence and potential estimation biases.

This empirical pattern of Figure 3 is consistent with the unfavorable contract regime, in which moral hazard frictions intensify in competition. Of course, the implied contract shares should not be taken literally: contracts do not adjust to the number of active dealers on a given day. A security's competitive environment, however, can be persistent, and incentive arrangements can be set with it in mind. Read through the lens of the model, the estimates suggest that such arrangements (whether explicit contract terms or

informal compensation norms) reflect the competition dealers face, across securities and over time.[23]

With trade data alone, I cannot rule out that this empirical pattern is at least partially driven by confounding factors, such as time-varying market conditions.[24] 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, to vary in market conditions. This in turn would require observing actual contract terms and effort choices, neither of which is available.

Model-implied welfare. If we are nevertheless willing to take the model at face-value, Proposition 2 offers a way to gauge the relative importance of moral hazard and imperfect competition. I illustrate this for a representative TSX client-account trade interval (with 25 active dealers) and compute the welfare effects of (i) removing moral hazard, holding competition constant, and (ii) adding one more dealer, across parameter constellations in Appendix Table A5.

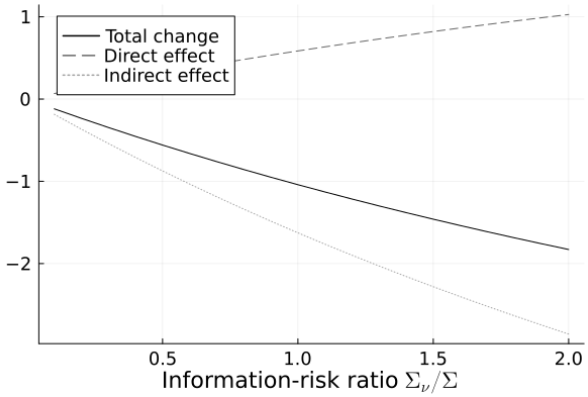
Here, I focus on the theoretically ambiguous welfare effect of increasing competition. Figure 4 shows the welfare effect of adding one dealer, as a function of the information–risk ratio, decomposed into a direct effect (improved risk sharing, always positive) and an indirect effect operating through the endogenous contract when clients are as risk-averse as dealers. This indirect effect is negative and dominates the positive direct effect: additional entry reduces welfare. The loss is amplified at higher information–risk ratios.

Take away. This exercise 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.

[23] An alternative interpretation of ϕ^* is that it captures how much surplus the dealer extracts from the client, unrelated to how hard the dealer works. The pattern that the estimated share rises in N is at odds with this interpretation: when more dealers are active, it is plausible that clients have better outside options and can more easily switch dealers, which should reduce the surplus the dealer can extract.

[24] 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 A5 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.

Figure 4: Welfare effects of adding one dealer



Notes: Figure 4 shows the percentage welfare change from increasing the number of active dealers from N to $N+1$ for a representative TSX client-account trade interval with $N = 25$ active dealers, when clients and dealers share the same risk tolerance ($\rho_I = \rho_H = 670$), as a function of the information–risk ratio, $\kappa = \Sigma_\nu/\Sigma$, given median return variance (0.038) and median net supply within the trade interval (C$1,604). The total effect is decomposed into the direct effect of adding a dealer (holding the contract fixed) and the indirect effect operating through the change in the contract, $\hat{\phi}^*(N+1) - \hat{\phi}^*(N)$. 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.

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

Admati, A. R. and P. Pfleiderer (1997). Does it all add up? Benchmarks and the compensation of active portfolio managers. *Journal of Business 70*(3), 323–350.

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.

Basak, S. and A. Pavlova (2013). Asset prices and institutional investors. *American Economic Review 103*(5), 1728–1758.

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.

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

Corhay, A., H. Kung, and L. Schmid (2025). Market power in finance. *Annual Review of Financial Economics 17*, 431–453.

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

Cuoco, D. and R. Kaniel (2011). Equilibrium prices in the presence of delegated portfolio management. *Journal of Financial Economics 101*(2), 264–296.

Dávila, E. and C. Parlatore (2021). Trading costs and information efficiency. *Journal of Finance 76*(3), 1471–1539.

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.

Gârleanu, N. and L. H. Pedersen (2018). Efficiently inefficient markets for assets and asset management. *Journal of Finance 73*(4), 1663–1712.

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. (1981). Nash equilibrium and the industrial organization of markets with large fixed costs. *Econometrica 49*, 1149–1172.

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.

Kacpercyk, M., J. Nosal, and S. Sundaresan (2025). Market power and price informativeness. *The Review of Economic Studies 92*(3), 1955–1986.

Kashyap, A. K., N. Kovrijnykh, J. Li, and A. Pavlova (2023). Is there too much benchmarking in asset management? *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.

Mekonnen, T. and B. Pakzad-Hurson (2026). Competition, persuasion, and search. *Theoretical Economics*.

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 63*(4), 1191–1243.

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

Rothschild, M. and J. E. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics 90*(4), 629–649.

Santos, T. and J. A. Scheinkman (2001). Competition among exchanges. *Quarterly Journal of Economics 116*(3), 1027–1061.

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.

Sockin, M. and M. Z. Xiaolan (2023). Delegated learning and contract commonality in asset management. *Review of Finance 27*(6), 1931–1975.

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.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and underdiversification. *Review of Economic Studies 77*(2), 779–805.

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

Veldkamp, L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton, NJ: Princeton University Press.

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

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

Wittwer, M. and J. Allen (2026). Market power and capital constraints. *American Economic Review 116*(4), 1309–1339.

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

INTERNET 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 empirical evidence.
Section C displays additional tables and figures.

A Additional theoretical evidence

Section A.1 provides model extensions.
Section A.2 provides mathematical details for how the contract changes in N.
Section A.3 analyses when imperfect competition is worse than moral hazard.
Section A.4 presents mathematical auxiliary statements and all proofs.

A.1 Model extensions

I provide five model extensions. Appendix A.1.1 allows for unknown common values. Appendix A.1.2 considers benchmarked compensation. Appendix A.1.3 introduces a continuous effort choice. Appendix A.1.4 derives an equilibrium assuming that dealers are price-takers, which provides a useful benchmark. Appendix A.1.5 presents a framework with client and dealer accounts.

A.1.1 Common value

The baseline model assumes that the institution-specific asset's expected return, ν_i, is drawn iid from a distribution with commonly known prior, ν. This shuts off concerns stemming from common-values and cross-dealer correlations. In this appendix, I relax this assumption by introducing an unknown common component ν_0 shared across all market participants.

Modified setting. From dealer i's perspective, returns are normally distributed, $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, with known variance $\Sigma > 0$. Their expected return 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 $\mu_i(1) = \nu_i + \epsilon_i$. Neither action reveals the two

components separately. When $\Sigma_0 = 0$, the common value is known and uncertainty is purely idiosyncratic, recovering the baseline model.

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, treating v_0 as unknown gives the effort decision a richer interpretation: exerting effort means conducting thorough market research that reveals v_i (the sum of the common and idiosyncratic components), while shirking means forgoing this research and observing only a noisier signal of the same sum. This creates an asymmetry between effort and shirking with respect to the informational role of the clearing price: the price is informative only to a shirking dealer, who can recover part of v_0 from it, whereas an effort dealer already knows v_i.

I maintain asymmetric information to avoid a well-known issue: in pure common-value settings with deterministic supply, linear equilibria in which demand depends on private signals break down because price impacts diverge. Any pure common-value formulation therefore requires introducing randomness in supply to restore equilibrium existence, which complicates the characterization of the equilibrium contract.[25]

Equilibrium. There are two important changes in this setting relative to the baseline. First, the v_i's are no longer independent. Writing $v_i = v_0 + \tilde{v}_i$, any two dealers' expected returns share the common component v_0, so: $Cov(v_i, v_j) = \Sigma_0 > 0, \quad i \neq j$. This correlation affects the equilibrium contract through its impact on dealers' expected payoff from shirking; for example, $\mathbb{E}[v_i^2] = \bar{v}^2 + \Sigma_0 + \Sigma_v$ and $\mathbb{E}[v_i v_j] = \bar{v}^2 + \Sigma_0$ for $i \neq j$.

Second, inference is asymmetric depending on whether the dealer exerts effort or shirks. Under effort ($s_i = 0$), dealer i observes $\mu_i(0) = v_i = v_0 + \tilde{v}_i$ exactly. Since $R \mid v_i \sim \mathcal{N}(v_i, \Sigma)$ by assumption, knowing v_i is sufficient for the distribution of R, and the clearing price p^c is uninformative: $\mathbb{E}[R \mid \mu_i(0), p^c] = v_i$, and $Var[R \mid \mu_i(0), p^c] = \Sigma$. The inference channel stemming from unknown common values is therefore dormant

[25] To illustrate, consider dealers acquiring imperfect signals about a common asset value through effort. With random supply, the clearing price becomes a noisy aggregator of private signals. This generates a feedback loop: the informativeness of the price depends on how aggressively dealers trade on their signals for a given contract, which in turn depends on the weight they place on the price when updating their beliefs. Pinning down the equilibrium therefore requires solving a joint fixed point in both the demand schedules and the contracts. A symmetric equilibrium contract could still be characterized numerically for any specific parameter values, but has no closed-form solution, making it difficult to analytically assess how it varies with the degree of competition.

under effort: the return depends on the two components only through their sum v_i, which research reveals, so the price carries no information about the return that the dealer does not already have.

Under shirking ($s_i = 1$), dealer i observes only the noisy signal $\mu_i(1) = v_0 + \tilde{v}_i + \epsilon_i$, which does not reveal v_i perfectly. In a linear equilibrium, the clearing price p^c can be inverted to recover:

$$\hat{v}_{-i} \equiv v_0 + \frac{1}{N-1}\sum_{j\neq i}\tilde{v}_j \sim \mathcal{N}\left(v_0, \frac{\Sigma_v}{N-1}\right),$$

which is a noisy signal of v_0 and hence informative about $v_i = v_0 + \tilde{v}_i$. The shirking dealer therefore uses both $\mu_i(1)$ and $\hat{v}_{-i}$ to form their posterior about v_i. Projecting v_i onto $(\mu_i(1), \hat{v}_{-i})$ via standard normal updating:

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

where c_0, c_s, c_p are the projection coefficients onto $(\mu_i(1), \hat{v}_{-i})$, that is, not onto $(\mu_i(1), p^c)$ directly, and $\tilde{\Sigma}_{CV}$ is the posterior variance. These are derived in Lemma 2 in Appendix A.4. When $\Sigma_0 > 0$, the inference channel is active under shirking: the shirking dealer's signal leaves residual uncertainty about v_i, part of which the price resolves because it is informative about the common component; so $c_p > 0$. When $\Sigma_0 = 0$, $c_p \to 0$ and the baseline is recovered.

Proposition 3. *Let the dealer's benefit from shirking be neither too low nor too high, $b \in [\underline{b}_{CV}, \overline{b}_{CV}]$, with bounds characterized in the proof of the Proposition; they reduce to the bounds of Proposition 1 when $\Sigma_0 = 0$. The interval is non-empty whenever $\rho_H < \infty$, and in the risk-neutral limit ($\rho_H \to \infty$) whenever $\Sigma_v\big[(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_v\big] + \Sigma_0\big[N(N-2)^2\Sigma_v - 4(N-1)^2\sigma_\epsilon\big] < 0$.*

There is an equilibrium with common values ($\Sigma_0 > 0$) in which: (i) All clients offer contract $\{K^*, \phi^*\}$:

$$\phi^* = \begin{cases} \Phi_{CV}(N) \equiv \frac{1}{|A|}\sqrt{\frac{\beta_{CV}(N)+b}{\alpha_{CV}(N)}} & \text{when } \frac{\rho_I}{\rho_I+\rho_H} \leq \Phi_{CV}(N),\\ \frac{\rho_I}{\rho_I+\rho_H} & \text{otherwise,}\end{cases} \tag{30}$$

and

$$K^* = \bar{U} - \rho_I\frac{N-2}{2(N-1)}\frac{\Sigma_v}{\Sigma} - \frac{A^2\Sigma}{2N(N-2)\rho_I}(\phi^*)^2, \tag{31}$$

where

$$\alpha_{CV}(N) \equiv \frac{1}{2N\rho_I}\left(\frac{N-2}{\Sigma}+\frac{N}{\sigma_\epsilon}+\frac{N-1}{\Sigma_\nu}+\frac{1}{N\Sigma_0+\Sigma_\nu}\right)^{-1}, \tag{32}$$

$$\beta_{CV}(N) \equiv -\frac{(N-2)\rho_I\left(\tilde{\Sigma}_{CV}-\Sigma\right)\left((N-1)\Sigma+(N-2)\Sigma_\nu\right)}{2(N-1)\,\Sigma\left(2\Sigma+(N-2)\tilde{\Sigma}_{CV}\right)}, \tag{33}$$

where $\tilde{\Sigma}_{CV} = Var[R \mid \mu_i(1), p^c]$ *is the shirking dealer's posterior variance, derived in Lemma 2. When* $\Sigma_0 = 0$, $\alpha_{CV}(N) \to \alpha(N)$ *and* $\beta_{CV}(N) \to \beta(N)$, *so that* $\Phi_{CV}(N) \to \Phi(N)$, *recovering Proposition 1.*

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

$$a^*(p, \mu_i(s^*)) = \rho_I(\Lambda^*\rho_I + \phi^* Var[R \mid \mu_i(s^*)])^{-1}(\mathbb{E}[R \mid \mu_i(s^*)] - p), \tag{34}$$

$$\text{with } \Lambda^* = \frac{\phi^*\Sigma}{\rho_I(N-2)}, \tag{35}$$

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

The equilibrium structure mirrors Proposition 1. 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, because under effort the common value inference channel is dormant. The effort-inducing contract differs from the baseline only through the contract share, $\Phi_{CV}(N)$. The fee K^* is unchanged: under effort, the dealer's expected payoff depends on expected returns only through $\nu_i - p^c(0)$, from which the common component cancels. In $\Phi_{CV}(N)$ the coefficients $\alpha_{CV}(N)$ and $\beta_{CV}(N)$ replace their baseline counterparts $\alpha(N)$ and $\beta(N)$, capturing the effect of common values on the incentive compatibility constraint.

Two forces are at work. 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. In addition, the correlation among ν_i's alters the expected payoff from shirking. The net effect on the level of the contract ϕ^* is in general ambiguous without further parametric restrictions.

Corollary 4. *Define* $\gamma_N^{CV}(\sigma_\epsilon, \Sigma_0) \equiv \frac{\beta'_{CV}(N)\,\alpha_{CV}(N) - \beta_{CV}(N)\,\alpha'_{CV}(N)}{\alpha'_{CV}(N)}$, *which is positive for all* $\Sigma_0 \geq 0$ *and satisfies* $\gamma_N^{CV}(\sigma_\epsilon, 0) = \gamma_N(\sigma_\epsilon)$ *of Corollary 3; note that* $\gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$ *increases in* σ_ϵ.

(i) When $b > \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$, $\Phi_{CV}(N)$ *increases in* N.

(ii) When $b < \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$, $\Phi_{CV}(N)$ *decreases in* N.

Since the cutoff $\gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$ itself varies with N, $\Phi_{CV}(N)$ need not be monotone in N. Both regimes are consistent with equilibrium existence: for any $N > 2$ and $\Sigma_0 > 0$, there are parameters under which b lies in the valid range $[\underline{b}_{CV}, \overline{b}_{CV}]$ of Proposition 3, the incentive constraint binds, and $b > \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$; and likewise with $b < \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$.

The corollary shows that the regime characterization is not an artifact of purely private values: for any degree of common-value uncertainty (Σ_0), the contract share rises with competition when the private benefit from shirking is high and falls when it is low. Moreover, a noisier signal still makes shirking less attractive.

Common-value uncertainty strengthens effort incentives: effort reveals the common component v_0 while shirking does not, so a larger Σ_0 makes effort more valuable relative to shirking. Fixing all other parameters, this has two implications. First, the regime cutoff rises (γ_N^{CV} increases in Σ_0) so the contract share falls with competition for a wider range of shirking benefits b; if b is small enough, a large enough Σ_0 places the market in this regime. Second, the incentive constraint is slack for a wider range of b: it cannot bind when $b < -\beta_{CV}(N)$, and $-\beta_{CV}(N)$ also increases in Σ_0.

However, both effects are bounded. Independent of how uncertain the common value is, a shirking dealer still observes the private signal $\mu_i(1) = v_i + \epsilon_i$, a noisy observation of v_i whose noise σ_ϵ does not grow with Σ_0. As Σ_0 increases, the prior on v_i becomes diffuse, but this signal stays equally informative and pins down what the shirking dealer knows; so their residual uncertainty about v_i—and with it their disadvantage relative to the effort dealer, who learns v_i exactly—stops growing. As $\Sigma_0 \to \infty$, the cutoff γ_N^{CV} and the slackness bound $-\beta_{CV}(N)$ therefore each rise only to a finite ceiling. In particular, once b exceeds $\lim_{\Sigma_0 \to \infty} \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$, the contract share rises with N for every degree of common-value uncertainty.

Welfare. The welfare analysis also carries over. To see this, fix any symmetric contract ϕ. When all dealers exert effort, each dealer's allocation depends on their expected return only through its distance to the average across dealers, $v_i - \frac{1}{N}\sum_j v_j$, from which the common component cancels. Allocations therefore have the same joint distribution as in the baseline and are independent of v_0, so every moment entering welfare (24) coincides with its baseline counterpart: expected welfare at the contract ϕ is given by the baseline expression (26).

In equilibrium, welfare still differs from the baseline, because it is evaluated at a different contract; the common component affects welfare only through $\Phi_{CV}(N)$. Condition (29) of Proposition 2 therefore applies with $\Phi_{CV}(N)$ in place of $\Phi(N)$: competition can

reduce welfare under common values when the market is in the unfavorable regime, $b > \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)$. Since the cutoff increases in Σ_0, this requires that the common-value uncertainty is not too strong.

What if effort is imperfect? In my common value extension, a dealer who exerts effort learns v_i perfectly. A natural next step would be to let effort improve information without perfecting it: effort would yield a signal $v_i + \epsilon_i$ with noise σ_e, shirking a noisier one with $\sigma_s > \sigma_e$; $\sigma_e \to 0$ recovers my current extension.

Deriving necessary conditions for an effort-inducing symmetric linear equilibrium is feasible. One can follow the same steps as in Lemma 2 and the proofs of Propositions 1 and 3. The equilibrium candidate has the same structure as before: the dealer's expected gain from exerting effort rather than shirking is again quadratic in the contract share, so the binding incentive constraint takes the form $b = \alpha A^2 \phi^{*2} - \beta$ with new coefficients α and β, and the constrained contract $\frac{1}{|A|}\sqrt{(\beta + b)/\alpha}$ depends on N through market clearing.

Proving that the candidate is an equilibrium is hard, for two reasons. First, the price is now informative for all dealers, not only for shirkers. When values are too correlated relative to the precision of private signals, dealers read a high price as good news rather than as a higher cost, demand schedules tilt upward, and no linear equilibrium exists. This complication is well known in demand-schedule competition with interdependent values (Vives, 2011; Rostek and Weretka, 2012). A linear equilibrium thus requires an effort signal that is precise enough relative to the common-value uncertainty. Second, the existence part of the proof of Proposition 3 rests on the properties that a deviating client's payoff is single-peaked at the risk-sharing contract ϕ_{UC}, whether their dealer works or shirks, and that cutting the dealer's share below the equilibrium level triggers shirking. Under perfect effort these hold for all parameters, in part because effort-path payoffs are unaffected by the common component. Under imperfect effort, beliefs anchor to the prior and load on the price, so deviation payoffs depend on (Σ_0, σ_e) and the properties could hold at most on the restricted region.

In sum, assuming that effort is perfect provides the tractability needed to prove existence analytically, but is unlikely to drive the model's main mechanism: the conjectured equilibrium preserves the dependence of contracts on competition through market clearing. What separates the conjecture from a theorem is an existence argument of a kind known to be demanding in the auction literature, even when we abstract from contracting.

A.1.2 Benchmarking

Contracts in my baseline model consist of a lump-sum fee and a share of absolute performance. In practice, compensation in asset management is often indexed to a benchmark—a contract form with a long tradition, dating back to Starks (1987) and Admati and Pfleiderer (1997). In this appendix I sketch how the analysis changes under such a contract.

Modified setting. A common benchmarking contract adds one instrument to the two-part contract: on top of the fixed fee K_i and a share $\hat{\phi}_i$ of the wealth the dealer generates, the dealer is paid a share ψ_i of performance relative to a benchmark. Following Kashyap et al. (2023), who study benchmarking in a single-asset environment (but with a continuum of price-taking dealers), the benchmark is one share of the risky asset, so that its return is $R - p^c(s_i)$ and the dealer's compensation is

$$K_i \;+\; \hat{\phi}_i\,\omega_i \;+\; \psi_i\Big(\omega_i - \big(R - p^c(s_i)\big)\Big) \;=\; K_i + \big(\phi_i\,a_i^c(s_i) - \psi_i\big)\big(R - p^c(s_i)\big),$$

where $\omega_i = a_i^c(s_i)\big(R - p^c(s_i)\big)$ is the wealth generated from the investment, as in the baseline, and $\phi_i \equiv \hat{\phi}_i + \psi_i$ is the dealer's total contract share.[26] The dealer's compensation is thus proportional to the active position, $\phi_i\,a_i^c(s_i) - \psi_i$, per unit of $R - p^c(s_i)$.

Equilibrium. I focus on equilibria in which the incentive constraint binds, because this is where benchmarking matters. When the constraint is slack, the baseline contract without benchmarking, $(\phi, \psi) = (\phi_{UC}, 0)$, delivers optimal risk sharing and satisfies the first- and second-order conditions of the joint problem in both instruments, so that benchmarking adds nothing. When clients are risk neutral, the slack case moreover never arises: the unconstrained contract leaves the dealer without a stake ($\phi_{UC} \to 0$), and a dealer without a stake strictly prefers to shirk.

Proposition 4. *Define $\Phi(N)$ as in Proposition 1. In a symmetric equilibrium:*

[26] Kashyap et al. (2023), equation (2), specify the manager's compensation as $w = \hat{a}\,r_x + b\,(r_x - r_b) + c = a\,r_x - b\,r_b + c$ with $a \equiv \hat{a} + b$, where $r_x = x(\Delta + \tilde{D} - p) + \varepsilon$ is the end-of-game payoff of the manager's position of x shares, net of the purchase price p, and $r_b = \tilde{D} - p$ is the payoff of the benchmark, normalized to one share of the risky asset. Mapping $(c, a, b, x, \tilde{D} - p)$ into $(K_i, \phi_i, \psi_i, a_i^c, R - p^c)$ and setting $\Delta = \varepsilon = 0$ yields the contract above. The abnormal return Δ and the compensation noise ε are ingredients of their moral-hazard problem—an unobservable portfolio choice with a private management cost—and play no role here, where moral hazard concerns information acquisition. The same absolute-plus-relative structure appears in Admati and Pfleiderer (1997), equation (2).

(i) Given a common benchmark intensity $\psi_i = \psi \geq 0$, when the incentive constraint binds, all clients offer the contract share

$$\phi^* \;=\; \Phi(N) + \frac{N\psi}{A}. \tag{36}$$

(ii) With risk-neutral clients ($\rho_H \to \infty$) choosing both instruments, assuming an interior solution exists where the incentive constraint binds, all clients offer the contract share and benchmark intensity

$$\phi^* \;=\; \Phi(N) + \frac{N\psi^*}{A}, \qquad \psi^* \;=\; u^*\phi^* \tag{36'}$$

$$\text{where} \quad u^* \;=\; \frac{A(N-2)}{N}\,\frac{A^2(N-1)\,\Phi(N)^2\,\Sigma^2\,\mathcal{P}_1 \;+\; (N-2)\,N\rho_I^2\,\Sigma_\nu\,\mathcal{P}_2}{A^2(N-1)^2\,\Phi(N)^2\,\Sigma^2\,\mathcal{Q}_1 \;+\; (N-2)\,N\rho_I^2\,\Sigma_\nu\,\mathcal{Q}_2},$$

and $\mathcal{P}_1, \mathcal{P}_2, \mathcal{Q}_1, \mathcal{Q}_2$ are polynomials in N, $\Phi(N)$, Σ, $\tilde{\Sigma} = Var[R \mid \mu_i(1)]$, and $\tilde{m} = \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}$, given in the proof.

(iii) In either case, the market clears at the baseline equilibrium price of Corollary 1.

To see where part (i) comes from, fix the benchmark intensity at a common value across clients, $\psi_i = \psi$. The dealer chooses the demand a_i taking the price impact $\partial p/\partial a_i = \Lambda$ as given. The first-order-condition—the analogue to (10)—rearranges to

$$a_i = \frac{\rho_I\big(\mathbb{E}[R \mid \mu_i(s_i)] - p\big)}{\Lambda\rho_I + \phi_i Var[R \mid \mu_i(s_i)]} + \frac{\psi}{\phi_i} = a^*\big(p, \mu_i(s_i)\big) + \frac{\psi}{\phi_i}.$$

This is the baseline demand (18) evaluated at contracts (ϕ_i, ϕ) and shifted by the constant (ψ/ϕ_i). Three things follow from this correspondence, which holds for all dealers. First, the slope is unchanged, so at a given contract the benchmark leaves the price impact $\Lambda = \phi\,\Sigma/(\rho_I(N-2))$ unchanged. Second, multiplying by ϕ_i, the net exposure is also unchanged at every price, $\phi_i\,a_i - \psi = \phi_i\,a^*\big(p, \mu_i(s_i)\big)$, implying that the benchmark drops out of the net exposure that governs the effort decision. Third, market clearing $\sum_j a_j = A$ clears the baseline demands at the reduced net supply

$$A_i^{\text{eff}} = A - \Big(\frac{\psi}{\phi_i} + (N-1)\frac{\psi}{\phi}\Big),$$

so the clearing price, the allocations, and hence dealer i's expected payoffs $\mathbb{E}[\mathcal{U}(s_i; \phi_i, \psi)]$ are those of the baseline model—at contracts (ϕ_i, ϕ), as in the proof of Proposition 1—evaluated at A_i^{eff} instead of A. Imposing symmetry ($\phi_i = \phi$), the binding incentive con-

straint depends on the contract and the supply only through $A\phi - N\psi$ and pins it down at $A\phi - N\psi = A\Phi(N)$, yielding (36). Benchmarking thus raises the contract share required to sustain effort by $N\psi/A$; linear in the benchmark intensity ψ and, for a positive intensity and positive net supply, increasing in the number of dealers, so imperfect competition amplifies the effect of benchmarking on the contract. Part (iii) follows because the symmetric clearing price also depends on the contract and the supply only through $\phi(A - N\psi/\phi) = A\phi - N\psi$, which the binding constraint pins at $A\Phi(N)$.

For part (ii) I let each client choose both instruments jointly. I solve the problem for risk-neutral clients ($\rho_H \to \infty$); with risk-averse clients, the slack case is covered above—$(\phi, \psi) = (\phi_{UC}, 0)$ satisfies the first- and second-order conditions—while the constrained case reduces, by the same steps as in the proof, to a quadratic equation whose closed-form roots are too long to be informative. With risk-neutral clients the baseline's unconstrained contract would leave the dealer a zero share, which cannot induce effort; conditional on a binding constraint, the joint problem has a closed-form solution.

At the optimum the constraint binds, but with two instruments the client can still move along it, trading ϕ_i against ψ_i while keeping the dealer indifferent between effort and shirking; in equilibrium, no such move raises the client's objective. Since the binding constraint pins the contract only through $A\phi - N\psi$, it traces out the menu of part (i)—$\Phi(N) + \frac{N\psi}{A}$ as a function of ψ—and the optimality condition selects a unique point on this menu, $u^* = \psi^*/\phi^*$, yielding part (ii). Equivalently, $\phi^* = \Phi(N) + N\psi^*/A$.

In summary, the extension confirms the main message of the paper: the contract a client chooses in anticipation of market clearing depends on how many dealers compete in market clearing. Benchmarking adds a new margin to this dependence. Every client's benchmark intensity enters the effective supply, so each dealer's incentive constraint depends on the intensity only through the aggregate $N\psi$—the number of dealers directly scales how benchmarking enters the incentive problem. Holding the intensity fixed, the wedge above the baseline share, $N\psi/A$, increases with the number of dealers; the optimally chosen intensity itself varies with N.

A.1.3 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. Here I sketch out why.

Modified setting. Instead of choosing between effort and shirking, each dealer chooses a continuous effort level $e_i \in [0, \bar{e}]$, $\bar{e} > 0$, which determines the precision of their private signal: a dealer exerting effort e_i observes $\mu_i(e_i) = \nu_i + \epsilon_i$ with $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1/e_i)$, so that higher effort yields a more precise signal. Effort costs $k(e_i) = \frac{c}{2}e_i^2$, with $c > 0$, replace the shirking benefit, and are subtracted from the dealer's certainty equivalent. 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: the dealer's demand takes the form of Proposition 1(ii), with $Var[R \mid \mu_i(e_i)] = \tilde{\Sigma}(e_i) = \Sigma + (\Sigma_\nu^{-1} + e_i)^{-1}$ and $\Lambda^* = \frac{\phi^*}{\rho_I(N-2)}\tilde{\Sigma}(e^*)$. In the effort stage, the dealer's expected certainty equivalent is strictly concave in e_i, and the first-order condition characterizes the optimal effort given the dealer's own contract and the contracts and effort choices of all others: $e_i(\phi_i; \phi^*, e^*)$.

Unlike the baseline, any positive contract induces some effort, so the contract steers the level of effort instead. Substituting the binding participation constraint into the client's objective, the client's first-order condition balances two motives: risk sharing alone would set $\phi_i = \frac{\rho_I}{\rho_I + \rho_H}$, while effort inducement distorts the contract away from this benchmark. Because the dealer's effort is the root of a cubic, the equilibrium—defined by $e_i(\phi^*; \phi^*, e^*) = e^*$ and $\phi_i(\phi^*, e^*) = \phi^*$, with K^* from the binding participation constraint—has no closed form, but it is straightforward to compute numerically.

A.1.4 Price-takers

In this appendix, I derive the equilibrium under price-taking behavior, shutting down price impact by assuming dealers treat prices as given, to show that the model's central insight does not arise due to strategic price impact but from central market clearing with finitely many dealers.

Proposition 5. *Let $b \geq -\tilde{\beta}(N)$, defined below. 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

$$\Phi_C(N) = \frac{1}{|A|}\sqrt{\frac{\tilde{\beta}(N) + b}{\tilde{\alpha}(N)}} \tag{37}$$

$$\Phi_U(N) = \frac{1}{2}\left(\frac{\sqrt{(\rho_H + \rho_I)((N-2)^2\rho_H + N^2\rho_I)}}{\rho_H + \rho_I} - (N-2)\right) > 0, \tag{38}$$

and

$$K^* = \bar{U} - \rho_I \frac{N-1}{2N}\frac{\Sigma_\nu}{\Sigma} - \frac{A^2\Sigma}{2N^2\rho_I}(\phi^*)^2, \tag{39}$$

where

$$\tilde{\alpha}(N) \equiv \frac{\Sigma(\Sigma-\tilde{\Sigma})(\Sigma-(N-1)^2\tilde{\Sigma})}{2N^2\rho_I\big(\Sigma+(N-1)\tilde{\Sigma}\big)^2} > 0,$$

$$\tilde{\beta}(N) \equiv -\frac{(N-1)\,\rho_I\Big(\Sigma^2\Sigma_\nu + \Sigma\big((N-2)\Sigma_\nu - \tilde{m}^2(N-1)N(\sigma_\epsilon+\Sigma_\nu)\big)\tilde{\Sigma} + (N-1)^2\Sigma_\nu\tilde{\Sigma}^2\Big)}{2N\Sigma\big(\Sigma+(N-1)\tilde{\Sigma}\big)^2} < 0,$$

given $\tilde{m} = \Sigma_\nu(\Sigma_\nu+\sigma_\epsilon)^{-1}$ *and* $\tilde{\Sigma} \equiv Var[R|\mu_i(1)]$.

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

$$a^*(p,\mu_i(0)) = \rho_I(\phi^* Var[R|\mu_i(0)])^{-1}(\mathbb{E}[R \mid \mu_i(0)] - p). \tag{40}$$

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 (18) 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, clients anticipate this behavior and adjust their contracts, but their ability to correct the misperception is limited: contracts are not auction-specific and cannot condition on realized prices. This is reflected in the unconstrained contract $\Phi_U(N)$, which depends on N and converges to the second-best benchmark $\rho_I/(\rho_I+\rho_H)$ only as $N \to \infty$, when the misperception vanishes.

A.1.5 Client and dealer accounts

To link the model to the data, where dealer firms trade for their own accounts and on behalf of their clients, and to rationalize the empirical patterns documented in Section 3.1, I extend the baseline model. I introduce proprietary trades alongside trades for clients, treating the two accounts as separate with prices determined through separate auctions. Appendix B.2 motivates this separation and discusses whether it is an adequate approximation for average prices.

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}). While the model is silent on whether these sets of participants overlap, it is more natural to think of them as separate. Large dealer firms typically organize own-account and client trading on distinct desks, with different traders and separate risk budgets.

For 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

$$p^*_{CL} = \frac{1}{N_{CL}} \sum_i v_{i,CL} - \frac{\phi^*}{\rho_I} \Sigma \left(\frac{N_{CL} - 1}{N_{CL} - 2} \right) \frac{A_{CL}}{N_{CL}}, \tag{41}$$

where I have relabeled v_i (which may be decomposed into a known common component, v, and an idiosyncratic rest with zero mean) 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 $R \sim \mathcal{N}(v_{i,IN}, \Sigma)$, where $v_{i,IN} = v + \varepsilon_i$ and $\varepsilon_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$ captures individual reasons to trade, such as balance-sheet constraints or individual market views. Conditional on all publicly available information, these are independent across dealers, but may be correlated with $v_{i,CL}$. The dealer chooses demand so as to maximize their certainty equivalent:

$$\mathcal{U}_{IN} = \big(v_{i,IN} - p_{IN} \big) a_{i,IN} - \frac{\Sigma}{2\rho_I} \big(a_{i,IN} \big)^2, \tag{42}$$

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 baseline model but with $\phi^* = 1$ and $v_{i,IN}$ instead of v_i. Therefore, the auction clears at price

$$p^*_{IN} = \frac{1}{N_{IN}} \sum_i v_{i,IN} - \frac{1}{\rho_I} \Sigma \left(\frac{N_{IN} - 1}{N_{IN} - 2} \right) \frac{A_{IN}}{N_{IN}}. \tag{43}$$

Calibration. When bringing the model to the data I consider separate auctions for each exchange $e \in \{\text{TSX, Alpha}\}$, period τ in a day t, security s, account-type $k \in \{\text{CL, IN}\}$, and trade-side $l \in \{\text{buy, sell}\}$. To distinguish them, I include super-and subscripts. For example, $v_{i,IN}$ in expression (43) is replaced by $v_{ti\tau sk}$ for $k = IN$. To account for systematic changes over time or across securities, I only impose $R_{ts} | v_{ti\tau sk} \sim \mathcal{N}(v_{ts}, \Sigma_{ts})$ with $v_{ti\tau sk} \sim \mathcal{N}(v_{ts}, \Sigma_{vts})$ being iid across i and τ for each account type k, with the same

distribution across account types, but not across day–securities.

A.2 Technical intuition for how the contract changes in N

In this appendix, I provide mathematical details to underpin the intuition for Corollary 3. To understand how the contract share changes in the number of dealers, consider a dealer contemplating a unilateral deviation to shirking, taking the contract share and all other dealers' behavior as given.

When the dealer exerts effort, the market clears at price (20). This price, $p^* = p^c(0)$, has two components. The first aggregates the private information of all dealers, equal to $\frac{1}{N}(\nu_i + \sum_{j\neq i}\nu_j)$, and reflects the market consensus about the asset's expected return. The second is a price discount that compensates risk-averse dealers for absorbing the fixed asset supply A, equal to $\frac{A(N-1)\phi^*\Sigma}{(N-2)N\rho_I}$, and drives a wedge between the price and the market's assessment of the asset's value. A similar decomposition holds for the clearing price under shirking, $p^c(1)$.

These clearing prices and implied allocations determine how much the dealer expects to profit given their effort choice $s_i \in \{0,1\}$:

$$\mathbb{E}\left[\mathcal{U}(s_i;\phi^*)\right] = \lambda(s_i) \times \mathbb{E}\Big[\Big(\mathbb{E}[R \mid \mu_i(s_i)] - p^c(s_i)\Big)^2\Big] + s_i b, \tag{44}$$

where

$$\lambda(s_i) \equiv \phi^* \frac{\rho_I}{2} \frac{2\Lambda^*\rho_I + \phi^* Var[R \mid \mu_i(s_i)]}{(\Lambda^*\rho_I + \phi^* Var[R \mid \mu_i(s_i)])^2}. \tag{45}$$

Here the expected return is $\mathbb{E}[R \mid \mu_i(0)] = \nu_i$ and $\mathbb{E}[R \mid \mu_i(1)] = \nu + \tilde{m}((\nu_i + \epsilon_i) - \nu)$ with $\tilde{m} = \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}$, the price $p^c(s_i)$ is the market-clearing price when dealer i chooses s_i, and price impact (19) is determined by the contract and demand of the other dealers and clients—held fixed at the equilibrium.

The expected payoff depends on the expected squared gap between the dealer's own expected return and the clearing price: $\mathbb{E}\Big[\Big(\mathbb{E}[R \mid \mu_i(s_i)] - p^c(s_i)\Big)^2\Big]$.[27] This gap is multiplied by a conversion rate, $\lambda(s_i)$, 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, $Var[R \mid \mu_i(s_i)]$. When $\Lambda^* > 0$, the dealer's own order moves the price against them, and its extent depends on

[27] 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.

N. Price impact enters twice: once when the order is submitted and once through market clearing. The larger price impact, the lower the conversion rate, reducing the share of their informational advantage that the dealer can convert into profits.

Like the price, the price gap decomposes into two parts, a supply effect, $S(s_i)$, and an information effect, $I(s_i)$. When the dealer exerts effort:

$$\begin{aligned}\mathbb{E}\left[\left(\mathbb{E}[R \mid \mu_i(0)] - p^c(0)\right)^2\right] &= I(0) + S(0) \\ &= \frac{(N-1)\Sigma_\nu}{N} + \frac{A^2\phi^{*2}(N-1)^2\Sigma^2}{(N-2)^2\rho_I^2 N^2}.\end{aligned}$$

Under shirking, the dealer's noisy signal garbles both components, introducing additional dependence on σ_ϵ and $\tilde{\Sigma} = Var[R \mid \mu_i(1)]$:

$$\begin{aligned}\mathbb{E}\left[\left(\mathbb{E}[R \mid \mu_i(1)] - p^c(1)\right)^2\right] &= I(1) + S(1) \\ &= \frac{\Sigma_\nu(\sigma_\epsilon + N\Sigma_\nu)\left(\Sigma + (N-2)\tilde{\Sigma}\right)^2}{(N-1)(\sigma_\epsilon + \Sigma_\nu)\left(2\Sigma + (N-2)\tilde{\Sigma}\right)^2} + \frac{A^2\phi^{*2}\Sigma^2\left(\Sigma + (N-2)\tilde{\Sigma}\right)^2}{(N-2)^2\rho_I^2\left(2\Sigma + (N-2)\tilde{\Sigma}\right)^2}.\end{aligned}$$

Since effort gives the dealer a more precise signal, the price gap is always higher under effort than under shirking.

A dealer exerts effort if

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

with $\alpha(N)$ and $\beta(N)$ as defined in Proposition 1; at $\Sigma_0 = 0$ they coincide with $\alpha_{CV}(N)$ and $\beta_{CV}(N)$ of Proposition 3. Note that

$$\begin{aligned}A^2\phi^{*2}\alpha(N) &= \lambda(0)S(0) - \lambda(1)S(1) \\ -\beta(N) &= \lambda(0)I(0) - \lambda(1)I(1).\end{aligned}$$

The supply-driven term, $A^2\phi^{*2}\alpha(N)$, captures how much more the effort dealer profits from the price discount relative to the shirking dealer, after accounting for the difference in conversion rates $\lambda(s_i)$. As N increases, this term declines, $\alpha'(N) < 0$: with more dealers sharing the burden of absorbing the fixed supply A, the price discount required to compensate each dealer for bearing risk shrinks, reducing the supply-driven advantage of effort over shirking.

The information-driven term, $-\beta(N)$, captures how much the effort dealer's belief differs from the price relative to the shirking dealer, after accounting for the difference

in conversion rates. As N increases, $-\beta(N)$ increases, $-\beta'(N) > 0$: as the price converges toward the common prior v with more dealers entering, it pulls away from the effort dealer's precise belief v_i, while moving toward the shirking dealer's belief, which is shrunk toward v. This widens the information-driven advantage of effort.

Whether the contract rises or falls in N is then the race between the two forces: $\Phi'(N) < 0$ exactly when $-\beta'(N) > -A^2\Phi(N)^2\alpha'(N)$, which is the case when σ_ϵ is large but not when it is small (Corollary 3).

A.3 When is imperfect competition worse than moral hazard?

In this appendix I investigate which friction has a larger impact on welfare and identify the conditions under which one dominates the other. 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 isolate its impact by shutting down strategic demand reduction mechanically, outside of equilibrium. To do this, I assume dealers submit price-taking demands, but I evaluate them at the baseline strategic contracts. The welfare loss from moral hazard is then the difference between the expected welfare evaluated at the optimal risk-sharing contract, $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$, and the expected welfare evaluated at the baseline constrained contract, $\phi^* = \Phi(N)$.

Proposition 6. *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

$$r(\phi) \leq N \text{ , where } r(\phi) = 1 + \frac{\phi\sqrt{\rho_H(\rho_H+\rho_I)}}{-\rho_I + \phi(\rho_H+\rho_I)} \text{ at } \phi = \Phi(N). \tag{46}$$

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

$$\Phi(N) \geq \frac{\rho_I(N-1)}{(N-1)(\rho_H+\rho_I) - \sqrt{\rho_H(\rho_H+\rho_I)}}, \tag{47}$$

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

The first part of the proposition (46) identifies the minimal number of dealers necessary for imperfect competition to distort asset allocations less than moral hazard for

a given contract ϕ. As moral hazard becomes less severe (b declines which decreases ϕ), $r(\phi)$ 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 imperfect competition is less welfare harming than moral hazard.

The second condition (47) 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 $b > \gamma_\infty(\sigma_\epsilon)$, the condition must be satisfied, so that the welfare loss from imperfect competition is dominated by the loss from moral hazard.

Proposition 6 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.4 Additional mathematical results and proofs

In this appendix I provide additional mathematical statements, and all proofs. For most of the proofs, I rely on Mathematica to solve and simplify complex algebraic expressions. To keep the proofs concise, I do not display all intermediate mathematical expressions in the manuscript. I invite the interested reader to consult the Mathematica notebooks for full algebraic details.

A.4.1 Auxiliary lemmas

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

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

$$\mathbb{E}[R \mid \mu_i(s_i)] = \begin{cases} \nu + \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}\big((\nu_i + \epsilon_i) - \nu\big) & \text{when } s_i = 1 \\ \nu_i & \text{when } s_i = 0. \end{cases} \tag{48}$$

$$Var[R \mid \mu_i(s_i)] = \begin{cases} \Sigma + \left(\Sigma_\nu^{-1} + \sigma_\epsilon^{-1}\right)^{-1} & when\ s_i = 1 \\ \Sigma & when\ s_i = 0. \end{cases} \tag{49}$$

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

$$\mathbb{E}[R \mid \mu_i(s_i), p^c] = \begin{cases} c_0 \bar{\nu} + c_s \mu_i(1) + c_p \hat{\nu}_{-i} & when\ s_i = 1, \\ \nu_i & when\ s_i = 0, \end{cases} \tag{50}$$

$$Var[R \mid \mu_i(s_i), p^c] = \begin{cases} \tilde{\Sigma}_{CV} & when\ s_i = 1, \\ \Sigma & when\ s_i = 0, \end{cases} \tag{51}$$

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

$$\begin{aligned} \hat{\nu}_{-i} &= \nu_0 + \frac{1}{N-1} \sum_{j \neq i} \tilde{\nu}_j \\ &= \frac{A - C_{21}\left(c_0 \bar{\nu} + c_s \mu_i(1)\right)}{C_1(N-1) + C_{21} c_p} + \left(\frac{C_1(N-1) + C_{21}}{C_1(N-1) + C_{21} c_p}\right) p^c, \end{aligned} \tag{52}$$

with $C_1 = \rho_I\left(\frac{\phi}{N-2}\Sigma + \phi\Sigma\right)^{-1}$ *and* $C_{21} = \rho_I\left(\frac{\phi}{N-2}\Sigma + \phi_i \tilde{\Sigma}_{CV}\right)^{-1}$, *where* ϕ *is the contract share of the other clients and* ϕ_i *dealer i's own share; these are the equilibrium demand slopes derived in the proof of Proposition 3. Moreover:*

$$c_s = \frac{\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{\mathcal{Q}}, \qquad c_p = \frac{(N-1)\Sigma_0 \sigma_\epsilon}{\mathcal{Q}}, \qquad c_0 = \frac{\sigma_\epsilon \Sigma_\nu}{\mathcal{Q}}, \tag{53}$$

$$\tilde{\Sigma}_{CV} = \Sigma + \frac{\sigma_\epsilon \Sigma_\nu (N\Sigma_0 + \Sigma_\nu)}{\mathcal{Q}}, \tag{54}$$

$$\mathcal{Q} \equiv (N-1)\Sigma_0 \sigma_\epsilon + N\Sigma_0 \Sigma_\nu + \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu). \tag{55}$$

Here c_s *and* c_p *are the projection coefficients of* ν_i *onto* $\mu_i(1)$ *and* $\hat{\nu}_{-i}$ *(which is a function of* p^c*), respectively. 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* $\Sigma_0 = 0$, $\mathcal{Q} \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$, *so* $c_p \to 0$, $c_s \to \tilde{m} = \frac{\Sigma_\nu}{\sigma_\epsilon + \Sigma_\nu}$, $c_0 \to 1 - \tilde{m}$, *and* $\tilde{\Sigma}_{CV} \to \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, *recovering Lemma 1.*

A.4.2 Proofs

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

Proof of Proposition 1. The proof has two parts. Part (a) derives the equilibrium objects as necessary conditions. Part (b) shows that no client–dealer pair has a profitable unilateral deviation when $b \in [\underline{b}, \overline{b}]$, and defines the bounds.

(a) Necessary conditions. The proof proceeds by guess and verify. I fix all client–dealer pairs other than pair i at the symmetric equilibrium candidate—the contract $\{K^*, \phi^*\}$, effort, and the demand schedule (18)—and show, solving the game backwards, that pair i's best response is to conform. The text preceding Proposition 1 derives the equilibrium conditions; I verify them here.

In the trading stage, dealer i observes $\mu_i(s_i)$ and maximizes (3) point-wise for each clearing price p, which yields the necessary condition (10). Since all other dealers j submit the equilibrium guess (18), price impact is $\Lambda_i = \frac{\phi^*}{\rho_I(N-2)}\Sigma$. The necessary conditions are sufficient, since $Var[R \mid \mu_i(s_i)]$ is positive semi-definite, $N > 2$ and $\phi_i > 0$. Equation (10) thus 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 (18).

Before trading and observing their private signal, the dealer chooses their effort level in period 2. They exert effort when the incentive constraint (IC) holds, where $\mathcal{U}(s_i; \phi_i)$ is the certainty equivalent (6), evaluated at the clearing quantity $a_i^c(s_i)$ and price $p^c(s_i)$.

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 $\mathcal{W}(\phi_i; s_i)$ given in (9), subject to the incentive constraint (IC) and the participation constraint (IR). Since the incentive constraint is independent of K_i, we can choose K_i to ensure participation: $K_i \equiv K(\phi_i; 0) = \bar{U} - \mathbb{E}[\mathcal{U}(0; \phi_i)]$.[28] Inserting the binding participation constraint (IR') into the client's objective function (M), the optimization problem simplifies to (12), subject to (IC), assuming the client wants to incentivize effort (which is verified in part (b)).

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^* follows from (IR'). The unconstrained contract arises when $\Phi(N) \leq \frac{\rho_I}{\rho_I + \rho_H}$: by property (i) in part (b), $\mathbb{E}[\mathcal{W}(\phi_i; 0)] + \mathbb{E}[\mathcal{U}(0; \phi_i)] - \bar{U}$ is maximized at $\phi_i = \frac{\rho_I}{\rho_I + \rho_H}$ for any contract ϕ chosen by the other clients, and since clients are ex-ante identical, all choose it. The constrained contract arises when $\Phi(N) \geq \frac{\rho_I}{\rho_I + \rho_H}$: the incentive constraint binds, and imposing symmetry and solving $\mathbb{E}[\mathcal{U}(0; \phi)] = \mathbb{E}[\mathcal{U}(1; \phi)]$ for ϕ—with both sides given

[28] Note that $K(\phi_i; 0)$ depends on ϕ_i, because ϕ_i affects the dealer's demand function in the trading game.

below in (56) and (57)—yields a unique positive solution, which I call $\Phi(N)$.

It remains to verify that the dealer exerts effort under the unconstrained contract. Using market clearing and equilibrium trading in period 3, the dealer's expected certainty equivalents for an arbitrary ϕ_i, given the contract ϕ of the other clients, are:

$$\mathbb{E}[\mathcal{U}(0;\phi_i)] = \phi_i \frac{A^2(N-1)\phi^2\Sigma^2 + (N-2)^2 N\rho_I^2\Sigma_\nu}{2(N-2)(N-1)(2\phi+(N-2)\phi_i)\rho_I\Sigma}, \tag{56}$$

$$\mathbb{E}[\mathcal{U}(1;\phi_i)] = b + \phi_i \frac{A^2(-1+N)\phi^2\Sigma^2(\sigma_\epsilon+\Sigma_\nu) + (-2+N)^2\rho_I^2\Sigma_\nu(\sigma_\epsilon+N\Sigma_\nu)}{2(-2+N)(-1+N)\rho_I(\sigma_\epsilon+\Sigma_\nu)(2\phi\Sigma+(-2+N)\phi_i\tilde{\Sigma})}. \tag{57}$$

Imposing symmetry, $\phi_i = \phi$, the incentive constraint simplifies to

$$\mathbb{E}[\mathcal{U}(0;\phi)] \geq \mathbb{E}[\mathcal{U}(1;\phi)] \Leftrightarrow b \leq \mathcal{C}(\phi) \equiv A^2\phi^2\alpha(N) - \beta(N), \tag{58}$$

with $\alpha(N)$ and $\beta(N)$ as defined in the proposition. Since $\mathcal{C}(\Phi(N)) = b$ by construction, $\mathcal{C}'(\phi) \geq 0$, and $\Phi(N) \leq \frac{\rho_I}{\rho_I+\rho_H}$ in the unconstrained case, $b \leq \mathcal{C}\left(\frac{\rho_I}{\rho_I+\rho_H}\right)$: the incentive constraint is satisfied.

(b) Existence. Let $b \geq \underline{b} \equiv -\beta(N)$ so that $\Phi(N)$ is a real number. To show that the symmetric effort-inducing equilibrium exists, I need to rule out profitable unilateral deviations. A deviating client chooses a share ϕ_i and a fee. For any ϕ_i, the fee optimally binds the dealer's participation constraint, as in (IR') but evaluated at the action the dealer chooses under ϕ_i: the fee affects neither the dealer's effort choice nor trading, and it enters the client's payoff linearly, so a higher fee strictly lowers the client's payoff and with a lower fee the dealer declines. A deviation is thus a choice of ϕ_i alone.

With the binding fee, the client's payoff, net of $\bar{U}$, is $f(\phi_i,\phi)$ when their dealer exerts effort and $g(\phi_i,\phi)+b$ when the dealer shirks—the fee passes the dealer's surplus, including the shirking benefit b, to the client—where ϕ is the contract of all other clients; the equilibrium payoff is $f(\phi^*,\phi^*)$. Formally:

$$\begin{aligned} f(\phi_i,\phi) \equiv& \mathbb{E}[\mathcal{W}(\phi_i;0)] + \mathbb{E}[\mathcal{U}(0;\phi_i)] \\ =& \Big((2\phi\rho_H + (-2+N)(-\rho_I - (-2+\phi_i)\phi_i(\rho_H+\rho_I))) \\ & (A^2(-1+N)\phi^2\Sigma^2 + (-2+N)^2 N\rho_I^2\Sigma_\nu)\Big) \\ & \Big(2(-2+N)(-1+N)(2\phi+(-2+N)\phi_i)^2\rho_H\rho_I\Sigma\Big)^{-1}, \end{aligned} \tag{59}$$

$$\begin{aligned} g(\phi_i,\phi) \equiv& \mathbb{E}[\mathcal{W}(\phi_i;1)] + \mathbb{E}[\mathcal{U}(1;\phi_i)] - b \\ =& \Big((A^2(-1+N)\phi^2\Sigma^2 + (-2+N)^2\rho_I^2(\Sigma_\nu + \tilde{m}^2(-1+N) \end{aligned} \tag{60}$$

$$(\sigma_\epsilon + \Sigma_\nu)))(2\phi\rho_H\Sigma - (-2+N)(\rho_I + (-2+\phi_i)\phi_i(\rho_H+\rho_I))\tilde{\Sigma})\Big)$$
$$\Big(2(-2+N)(-1+N)\rho_H\rho_I(2\phi\Sigma + (-2+N)\phi_i\tilde{\Sigma})^2\Big)^{-1},$$

with $\tilde{m} = \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}$, $\tilde{\Sigma} = Var(R \mid \mu_i(1))$, and $\phi_{UC} = \frac{\rho_I}{\rho_I+\rho_H}$. The dealer exerts effort when the gain from effort, $\mathbb{E}[\mathcal{U}(0;\phi_i)] - \mathbb{E}[\mathcal{U}(1;\phi_i)]$, is non-negative, breaking ties toward the action the contract intends.

The proof relies on four properties, verifiable with Mathematica, which hold for any fixed $\phi > 0$ and all admissible parameters:

(i) $f(\phi_{UC},\phi) - f(\phi_i,\phi)$ equals $(\phi_i - \phi_{UC})^2$ times a positive fraction, so ϕ_{UC} maximizes $f(\cdot,\phi)$;

(ii) $f(\phi_i,\phi)$ decreases in ϕ_i for $\phi_i \geq \phi_{UC}$;

(iii) $g(\phi_{UC},\phi) - g(\phi_i,\phi)$ equals $(\phi_i - \phi_{UC})^2$ times a positive fraction, so ϕ_{UC} maximizes $g(\cdot,\phi)$;

(iv) when $b = \mathcal{C}(\phi)$ given by (58), so that the symmetric incentive constraint binds at ϕ, the dealer's gain from effort, $\mathbb{E}[\mathcal{U}(0;\phi_i)] - \mathbb{E}[\mathcal{U}(1;\phi_i)]$, is negative for every $\phi_i \in (0,\phi)$: written as a single fraction, this gain has a positive denominator, and its numerator factors as $(\phi_i - \phi)$ times a linear function of ϕ_i that is positive on $[0,\phi]$.

First, consider deviations that preserve effort. When the incentive constraint is slack, $\phi^* = \phi_{UC}$, and (i) implies $f(\phi_i,\phi_{UC}) \leq f(\phi_{UC},\phi_{UC})$. When the constraint binds, $\phi^* = \Phi(N) \geq \phi_{UC}$ and $b = \mathcal{C}(\Phi(N))$, so by (iv) the dealer shirks under any contract below $\Phi(N)$; a deviation that preserves effort must therefore offer $\phi_i \geq \Phi(N)$, and (ii) implies $f(\phi_i,\Phi(N)) \leq f(\Phi(N),\Phi(N))$.

Second, consider deviations that induce shirking. By (iii), any such deviation earns at most $g(\phi_{UC},\phi^*) + b$, whether the incentive constraint binds or is slack, and without characterizing which contracts induce shirking. No shirking deviation is profitable when

$$b \;\leq\; \overline{b} \equiv f(\phi^*,\phi^*) - g(\phi_{UC},\phi^*). \tag{61}$$

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

It remains to show that $[\underline{b},\overline{b}]$ is non-empty. The incentive constraint is slack, $\phi^* = \phi_{UC}$, precisely when $b \leq \mathcal{C}(\phi_{UC}) = A^2\phi_{UC}^2\alpha(N) - \beta(N)$. For these b, the upper bound

equals $\overline{b} = f(\phi_{UC}, \phi_{UC}) - g(\phi_{UC}, \phi_{UC})$, and every slack b satisfies it automatically, $b \leq \mathcal{C}(\phi_{UC}) < \overline{b}$, since every factor of

$$f(\phi_{UC}, \phi_{UC}) - g(\phi_{UC}, \phi_{UC}) - \mathcal{C}(\phi_{UC})$$
$$= \frac{\rho_H \sigma_\epsilon \Sigma_\nu \Big[(N-2)N(\rho_H + \rho_I)^2 \big((N-1)\Sigma + (N-2)\Sigma_\nu\big) + (N-1)A^2\Sigma^2 \Big]}{2(N-1)N(\rho_H + \rho_I)^2 \Sigma (\sigma_\epsilon + \Sigma_\nu)\big(2\Sigma + (N-2)\tilde{\Sigma}\big)} > 0 \qquad (62)$$

is positive for $N > 2$. Hence every $b \in \big[\underline{b}, \mathcal{C}(\phi_{UC})\big]$ satisfies both bounds, and this range is non-degenerate whenever clients are risk-averse: $\mathcal{C}(\phi_{UC}) - \underline{b} = A^2 \phi_{UC}^2 \alpha(N) > 0$ for $\rho_H < \infty$. Note that the full interval $[\underline{b}, \overline{b}]$ extends beyond $\mathcal{C}(\phi_{UC})$, because for $b > \mathcal{C}(\phi_{UC})$ the constraint binds and the equilibrium exists as long as $b \leq \overline{b}$.

In the risk-neutral limit ($\rho_H \to \infty$), $\phi_{UC} \to 0$: the range $\big[\underline{b}, \mathcal{C}(\phi_{UC})\big]$ degenerates to the point $\underline{b}$, and the incentive constraint binds for all $b \geq \underline{b}$. The interval $[\underline{b}, \overline{b}]$ is nevertheless non-empty when $(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_\nu < 0$. To highlight the dependence of the constrained contract on b, write $\Phi(N, b)$: at $b = \underline{b}$ the incentive constraint just begins to bind, so $\Phi(N, \underline{b}) = 0$, and $\Phi(N, b)$ is continuous and strictly increasing in b since $\mathcal{C}$ is strictly increasing. Taking $\rho_H \to \infty$ in (59)–(60), the upper bound evaluated at contract ϕ becomes

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

As $\phi \to 0^+$, the first term vanishes since it is $O(\phi)$, while the second diverges to $+\infty$ whenever $(4-3N)\sigma_\epsilon + (N-2)^2\Sigma_\nu < 0$. Hence $\lim_{b \to \underline{b}^+} \overline{b}(\Phi(N, b)) = +\infty > \underline{b}$, and there exists $\varepsilon > 0$ such that every $b \in (\underline{b}, \underline{b} + \varepsilon)$ satisfies both bounds.

Allowing clients to suffer a direct loss $\Delta \geq 0$ when their dealer shirks, as in He and Krishnamurthy (2012), where the return delivered to households falls when the specialist shirks, would relax the upper bound in (61) to $\overline{b} + \Delta$; the loss is not needed for existence. □

Proof of Proposition 2. I first compute realized welfare (24) in the equilibrium and then take expectations over return means. To show statement (i) I take the partial derivative of welfare with respect to ϕ. Given $\frac{\rho_I}{\rho_I + \rho_H} \leq \phi^* < 1$, it is straightforward to verify that $\frac{\partial \eta(\phi^*, N)\zeta(\phi^*, N)}{\partial \phi} \leq 0$ and $\frac{\partial \zeta(\phi^*, N)}{\partial \phi} \geq 0$. 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,

proves statement (ii). □

Proof of Proposition 3. The proof follows the same guess-and-verify strategy as the baseline. I highlight only the changes relative to Proposition 1.

(a) Necessary conditions. In the trading stage, the dealer's optimization problem is identical to the baseline, with the only difference that the conditional moments $\mathbb{E}[R \mid \mu_i(s_i), p^c(s_i)]$ and $Var[R \mid \mu_i(s_i), p^c(s_i)]$ 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 $\hat{\nu}_{-i}$ through the coefficient $c_p > 0$. Market clearing gives:

$$p^c(0) = \frac{C_1 \sum_{j \neq i} \nu_j + C_{20}\nu_i - A}{C_1(N-1) + C_{20}},$$
$$p^c(1) = \frac{(C_1(N-1) + C_{21}c_p)\hat{\nu}_{-i} + C_{21}(c_0\bar{\nu} + c_s\mu_i(1)) - A}{C_1(N-1) + C_{21}},$$

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 $C_{21} = \rho_I\left(\frac{\phi^*}{N-2}\Sigma + \phi_i\tilde{\Sigma}_{CV}\right)^{-1}$.

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 $Cov(\nu_i, \nu_j) = \Sigma_0 > 0$:

$$\mathbb{E}[\nu_i^2] = \bar{\nu}^2 + \Sigma_0 + \Sigma_\nu, \qquad \mathbb{E}[\nu_i\nu_j] = \bar{\nu}^2 + \Sigma_0, \quad i \neq j,$$
$$\mathbb{E}[\nu_i\hat{\nu}_{-i}] = \bar{\nu}^2 + \Sigma_0, \qquad \mathbb{E}[\hat{\nu}_{-i}^2] = \mathbb{E}[\hat{\nu}_{-i}\nu_j] = \bar{\nu}^2 + \Sigma_0 + \tfrac{\Sigma_\nu}{N-1}, \quad j \neq i,$$
$$\mathbb{E}[\epsilon_i^2] = \sigma_\epsilon, \qquad \mathbb{E}[\epsilon_i x] = 0 \text{ for any other variable } x.$$

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; it is straightforward to verify (for example with the help of Mathematica) that they reduce to the baseline expressions (56)–(57) when $\Sigma_0 = 0$.

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 unconstrained contract is $\phi^* = \frac{\rho_I}{\rho_I + \rho_H}$. As in the baseline, the dealer exerts effort under the unconstrained

contract whenever $\Phi_{CV}(N) \leq \frac{\rho_I}{\rho_I+\rho_H}$, since $\alpha_{CV}(N) > 0$ makes the effort premium increasing in ϕ. The participation constraint pins down the lump-sum fee K^*. Computing the expectations and matching coefficients in (13") gives the closed forms of $\alpha_{CV}(N)$ and $\beta_{CV}(N)$ displayed in the proposition. Under effort, the dealer's payoff depends on expected returns only through $\nu_i - p^c(0)$, from which ν_0 cancels, so $\mathbb{E}[\mathcal{U}(0;\phi_i)]$ is free of Σ_0 and K^* coincides with the baseline. The same cancellation applies to the client's stage-1 objective, whose dependence on ϕ_i under effort coincides with the baseline's; the slack-case optimum $\phi^* = \frac{\rho_I}{\rho_I+\rho_H}$ therefore follows from property (i) in part (b) of the proof of Proposition 1. When $\Sigma_0 = 0$, $\alpha_{CV}(N) \to \alpha(N)$ and $\beta_{CV}(N) \to \beta(N)$, so that $\Phi_{CV}(N) \to \Phi(N)$, recovering Proposition 1.

(b) Existence. The argument of part (b) of the proof of Proposition 1 applies with three adjustments. First, the deviating client's payoff under effort, $f(\phi_i, \phi)$ in (59), is unchanged although the ν_i are now correlated: under effort all dealers, including the deviant, trade on exact beliefs, so the common component ν_0 (the only carrier of Σ_0) shifts every belief and the clearing price one-for-one and cancels from $\nu_i - p^c(0)$; both the dealer's and the client's payoffs are therefore free of Σ_0, as shown above. Second, the payoff under shirking becomes

$$\begin{aligned} g_{CV}(\phi_i,\phi) &\equiv \mathbb{E}[\mathcal{W}(\phi_i;1)] + \mathbb{E}[\mathcal{U}(1;\phi_i)] - b \\ &= \Big(\big(A^2(-1+N)\phi^2\Sigma^2 + (-2+N)^2\rho_I^2\,\mathcal{V}_{CV}\big) \\ &\quad \big(2\phi\rho_H\Sigma - (-2+N)(\rho_I + (-2+\phi_i)\phi_i(\rho_H+\rho_I))\,\tilde{\Sigma}_{CV}\big)\Big) \\ &\quad \Big(2(-2+N)(-1+N)\rho_H\rho_I\big(2\phi\Sigma + (-2+N)\phi_i\tilde{\Sigma}_{CV}\big)^2\Big)^{-1}, \end{aligned} \tag{63}$$

with $\mathcal{V}_{CV} \equiv (N-1)\,\mathbb{E}\Big[\big(\mathbb{E}[R \mid \mu_i(1), p^c] - \hat{\nu}_{-i}\big)^2\Big] = \Sigma_\nu^2(N^2\Sigma_0 + \sigma_\epsilon + N\Sigma_\nu)\,\mathcal{Q}^{-1}$ and $\mathcal{Q}$ from Lemma 2. Relative to (60), $\tilde{\Sigma}_{CV}$ replaces $\tilde{\Sigma}$, and $\mathcal{V}_{CV}$ replaces $\Sigma_\nu + \tilde{m}^2(N-1)(\sigma_\epsilon + \Sigma_\nu)$—the same object, $(N-1)$ times the variance of the gap between the deviant dealer's posterior mean and the average belief of the others—to which it reduces when $\Sigma_0 = 0$; hence g_{CV} reduces to g. Third, the binding-constraint function becomes $\mathcal{C}_{CV}(\phi) \equiv A^2\phi^2\alpha_{CV}(N) - \beta_{CV}(N)$ from (13").

Properties (i) and (ii) are then the baseline properties. Properties (iii) and (iv) continue to hold, which is straightforward to verify (e.g. with Mathematica): (iii) $g_{CV}(\phi_{UC},\phi) - g_{CV}(\phi_i,\phi)$ equals $(\phi_i - \phi_{UC})^2$ times a positive fraction, so ϕ_{UC} maximizes $g_{CV}(\cdot,\phi)$; and (iv) when $b = \mathcal{C}_{CV}(\phi)$, so that the symmetric incentive constraint binds at ϕ, the dealer's gain from effort, $\mathbb{E}[\mathcal{U}(0;\phi_i)] - \mathbb{E}[\mathcal{U}(1;\phi_i)]$, is negative for every $\phi_i \in (0,\phi)$: written as a single fraction, this gain has a positive denominator, and its numerator factors as

$(\phi_i - \phi)$ times a linear function of ϕ_i that is positive on $[0, \phi]$. The bounds are $\underline{b}_{CV} \equiv -\beta_{CV}(N) > 0$, so that $\Phi_{CV}(N)$ is a real number, and $\overline{b}_{CV} \equiv f(\phi^*, \phi^*) - g_{CV}(\phi_{UC}, \phi^*)$, implicitly defined when the constraint binds.

Non-emptiness follows as in the baseline. The constraint is slack precisely when $b \leq \mathcal{C}_{CV}(\phi_{UC})$. Every slack b satisfies the upper bound automatically: the difference between $f(\phi_{UC}, \phi_{UC}) - g_{CV}(\phi_{UC}, \phi_{UC})$ and $\mathcal{C}_{CV}(\phi_{UC})$ equals the client's own gain when the dealer works, $\mathbb{E}[\mathcal{W}(\phi_{UC}; 0)] - \mathbb{E}[\mathcal{W}(\phi_{UC}; 1)]$, which is positive for all $\Sigma_0 \geq 0$ and reduces to (62) at $\Sigma_0 = 0$ (again straightforward to verify). Moreover, $\mathcal{C}_{CV}(\phi_{UC}) - \underline{b}_{CV} = A^2 \phi_{UC}^2 \, \alpha_{CV}(N) > 0$ for $\rho_H < \infty$. In the risk-neutral limit, the argument of Proposition 1 applies with $\overline{b}_{CV}(\phi) \equiv f(\phi, \phi) - g_{CV}(0, \phi)$: as $\phi \to 0^+$, $\phi \, \overline{b}_{CV}(\phi)$ converges to

$$-\frac{(N-2)\rho_I \Sigma_\nu \Big(\Sigma_\nu \big[(4-3N)\sigma_\epsilon + (N-2)^2 \Sigma_\nu \big] + \Sigma_0 \big[N(N-2)^2 \Sigma_\nu - 4(N-1)^2 \sigma_\epsilon \big] \Big)}{4(N-1)N \, \Sigma \, \mathcal{Q}},$$

with $\mathcal{Q} > 0$ from Lemma 2, so $\overline{b}_{CV}(\phi)$ diverges to $+\infty$ precisely under the condition stated in the Proposition, which reduces to the baseline condition at $\Sigma_0 = 0$. Since this coefficient increases in Σ_0—its derivative in Σ_0 is $(N-2)\rho_I \sigma_\epsilon^2 \Sigma_\nu^2$ over a positive square—common values weaken the risk-neutral-limit condition. □

Proof of Proposition 4. I proceed as in the proof of Proposition 1: I fix all other clients at the symmetric contract (ϕ, ψ) and all other dealers at the behavior this contract induces—they exert effort and submit their equilibrium demand schedules. Dealer i accepts their contract—the fixed fee is pinned down by the binding participation constraint—exerts effort, and trades optimally given their contract; client i anticipates this response through market clearing, and in a symmetric equilibrium their optimal choice coincides with the contract of all others. Conditional on the signal, the certainty equivalent of the dealer's benchmarked wealth is

$$\mathcal{U}(s_i; \phi_i, \psi_i) = \big(\phi_i a_i - \psi_i \big) \big(\mathbb{E}[R \mid \mu_i(s_i)] - p \big) - \frac{1}{2\rho_I} \big(\phi_i a_i - \psi_i \big)^2 Var[R \mid \mu_i(s_i)].$$

Part (i). Let $\psi_i = \psi$ for all i. The dealer's first-order condition for the demand at price p—the analogue of (10)—rearranges to $a_i = a^*\big(p, \mu_i(s_i)\big) + \psi/\phi_i$: the baseline demand at contracts (ϕ_i, ϕ) shifted by a constant. Market clearing $\sum_j a_j = A$ therefore clears the baseline demands at the reduced net supply $A_i^{\text{eff}} = A - \big(\psi/\phi_i + (N-1)\psi/\phi \big)$, so the clearing price, and dealer i's expected payoffs are those of the baseline model—at contracts (ϕ_i, ϕ), as in the proof of Proposition 1—evaluated at A_i^{eff} instead of A.

Imposing symmetry ($\phi_i = \phi$), the effective supply is $A - N\psi/\phi$, and the incentive constraint $\mathbb{E}[\mathcal{U}(0;\phi,\psi)] \geq \mathbb{E}[\mathcal{U}(1;\phi,\psi)]$ is the baseline condition (13) evaluated at this supply: the dealer exerts effort when $b \leq (A\phi - N\psi)^2\alpha(N) - \beta(N)$, using $(A - N\psi/\phi)^2\phi^2 = (A\phi - N\psi)^2$. Substituting $b = A^2\Phi(N)^2\alpha(N) - \beta(N)$, which holds by definition of $\Phi(N)$, the constraint binds when $(A\phi - N\psi)^2 = A^2\Phi(N)^2$. Taking the positive root, which nests the baseline at $\psi = 0$, $A\phi - N\psi = A\Phi(N) \quad \Leftrightarrow \quad \phi = \Phi(N) + \frac{N\psi}{A}$, which is (36).

Part (ii). Inserting the binding participation constraint (IR') into the client's objective function (M), client i chooses (ϕ_i, ψ_i) to solve the two-instrument analogue of (12):

$$\max_{\phi_i,\psi_i} \left\{\mathbb{E}[\mathcal{W}(\phi_i,\psi_i;0)] + \mathbb{E}[\mathcal{U}(0;\phi_i,\psi_i)] - \bar{U}\right\} \text{ s.t. } \mathbb{E}[\mathcal{U}(0;\phi_i,\psi_i)] \geq \mathbb{E}[\mathcal{U}(1;\phi_i,\psi_i)].$$

With risk-neutral clients, the client's certainty equivalent $\mathbb{E}[\mathcal{W}(\phi_i,\psi_i;0)]$ is simply their expected payoff, $\mathbb{E}\big[\big((1-\phi_i)\,a_i^c(0) + \psi_i\big)(\nu_i - p^c(0))\big]$. The dealer's certainty equivalents are those derived above, with exposure $\phi_i\,a_i^c(s_i) - \psi_i$. Demands, price impact, clearing prices, and allocations are also computed as in the fixed-intensity derivation.

At the optimum the constraint binds, but with two instruments the client can still move along it, trading ϕ_i against ψ_i while keeping the dealer indifferent between effort and shirking. Optimality requires that no such move raise the client's objective, which holds when the gradients of the objective and of the constraint are proportional. Writing $\mathcal{S}_i$ for the objective and $\text{IC}_i \equiv \mathbb{E}[\mathcal{U}(0;\phi_i,\psi_i)] - \mathbb{E}[\mathcal{U}(1;\phi_i,\psi_i)]$, eliminating the Lagrange multiplier gives

$$\frac{\partial \mathcal{S}_i}{\partial \psi_i}\frac{\partial \text{IC}_i}{\partial \phi_i} = \frac{\partial \mathcal{S}_i}{\partial \phi_i}\frac{\partial \text{IC}_i}{\partial \psi_i},$$

with symmetry, $\phi_i = \phi$ and $\psi_i = \psi$, imposed after differentiating.

In the symmetric equilibrium, the binding incentive constraint depends on the contract only through $A\phi - N\psi$, as shown above, and pins it down at $A\phi - N\psi = A\Phi(N)$. I parametrize this locus by the demand shift $u = \psi/\phi$, so that $\phi = A\Phi(N)/(A - Nu)$, and substitute it into the optimality condition. The resulting equation is linear in u (under risk-neutrality) and the coefficient on u never vanishes. Expressed in terms of $\Phi(N)$, $\tilde{\Sigma} = Var[R \mid \mu_i(1)]$, and $\tilde{m} = \Sigma_\nu(\Sigma_\nu + \sigma_\epsilon)^{-1}$, the unique solution is

$$u^* = \frac{A(N-2)}{N}\,\frac{A^2(N-1)\,\Phi(N)^2\,\Sigma^2\,\mathcal{P}_1 \;+\; (N-2)\,N\rho_I^2\,\Sigma_\nu\,\mathcal{P}_2}{A^2(N-1)^2\,\Phi(N)^2\,\Sigma^2\,\mathcal{Q}_1 \;+\; (N-2)\,N\rho_I^2\,\Sigma_\nu\,\mathcal{Q}_2},$$

where, abbreviating $\Phi = \Phi(N)$,

$$\begin{aligned}
\mathcal{P}_1 &= \big(2-2\Phi+N(N-1+\Phi)\big)\Sigma^2 - \big(4-4\Phi+N(N-2+2\Phi)\big)\Sigma\tilde{\Sigma} + (N-2)(\Phi-1)\tilde{\Sigma}^2,\\
\mathcal{P}_2 &= \big(\tilde{m}(N-1)N(N-1+\Phi) - 4 + 4\Phi - N(N-3+3\Phi)\big)\Sigma^2\\
&\quad - (N-2)\big(4-4\Phi+N(N-2+2\Phi)\big)\Sigma\tilde{\Sigma} + (N-2)^2(\Phi-1)\tilde{\Sigma}^2,\\
\mathcal{Q}_1 &= \big(2+(N-1)N\big)\Sigma^2 - \big(4+(N-2)N\big)\Sigma\tilde{\Sigma} - (N-2)\tilde{\Sigma}^2,\\
\mathcal{Q}_2 &= \big(\tilde{m}(N-1)^3N + 12 - N(15+(N-6)N)\big)\Sigma^2\\
&\quad - (N-2)\big(N(12+(N-4)N) - 12\big)\Sigma\tilde{\Sigma} - (N-2)^2(2N-3)\tilde{\Sigma}^2.
\end{aligned}$$

The jointly optimal contract then follows from the locus: $\phi^* = A\Phi(N)/(A - Nu^*)$ and $\psi^* = u^*\phi^*$. Equivalently, $\phi^* = \Phi(N) + N\psi^*/A$: the fixed-intensity analysis traces out the menu of effort-inducing contracts, (36) as a function of ψ, and the joint problem selects the optimal point on this menu.

Part (iii). Market clearing at the shifted demands gives

$$p^* = \frac{1}{N}\sum_j v_j - \big(A\phi - N\psi\big)\frac{\Sigma}{\rho_I}\frac{N-1}{N(N-2)}.$$

This is the price of Corollary 1, whose discount term rearranges to $A\phi^*\Sigma\frac{N-1}{\rho_I N(N-2)}$, with $A\phi^*$ replaced by $A\phi - N\psi$. The binding incentive constraint pins $A\phi - N\psi$ at $A\Phi(N)$ in both part (i) and part (ii); since the baseline constrained contract is $\phi^* = \Phi(N)$, the two prices coincide. □

Proof of Proposition 5. The proof is analogous to the proof of Proposition 1. The only difference is that we now set $\Lambda_i = 0$. □

Proof of Proposition 6. 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 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

$$LIC = \underline{\xi}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \underline{\delta}A\Sigma A, \text{ with } \underline{\xi} = \frac{\rho_H + \rho_I}{2(N-1)} \geq 0, \tag{64}$$

and $\underline{\delta} = 0$. 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

$$\begin{aligned} LMH(N,\phi) &= \tilde{\xi}\left(\mathbb{E}[\nu_i\Sigma^{-1}\nu_i] - \mathbb{E}[\nu_i]\Sigma^{-1}\mathbb{E}[\nu_i]\right) - \tilde{\delta}A\Sigma A, \\ \text{with } \tilde{\xi} &= \frac{(\rho_I - \phi(\rho_H+\rho_I))^2(N-1)}{2\phi^2\rho_H} \geq 0 \text{ and } \tilde{\delta} = -\frac{(\rho_I - \phi(\rho_H+\rho_I))^2}{2N\rho_H\rho_I(\rho_H+\rho_I)} \leq 0, \end{aligned} \tag{65}$$

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 $\underline{\xi} \leq \tilde{\xi}$ 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 $r(\phi) \leq N$ with $r(\phi)$ given in Proposition 6.

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 $N \geq 2$:

$$\begin{aligned} 1 + \frac{\Phi(N)\sqrt{\rho_H(\rho_H+\rho_I)}}{-\rho_I + \Phi(N)(\rho_H+\rho_I)} \leq N &\iff \Phi(N)\left(\sqrt{\rho_H(\rho_H+\rho_I)} - (N-1)(\rho_H+\rho_I)\right) \leq -(N-1)\rho_I \\ &\iff \Phi(N) \geq \frac{(N-1)\rho_I}{(N-1)(\rho_H+\rho_I) - \sqrt{(\rho_H+\rho_I)\rho_H}}. \quad \square \end{aligned}$$

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. $\square$

Proof of Corollary 2. To derive the corollary, I compute the expected payoffs under effort and shirking as the market becomes perfectly competitive:

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(0;\phi_i)] = \frac{\rho_I}{2}\frac{\Sigma_\nu}{\Sigma} \tag{22}$$

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{U}(1;\phi_i)] = \frac{\rho_I}{2}\frac{\Sigma_\nu^2(\Sigma_\nu + \sigma_\epsilon)^{-1}}{\Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}} + b. \tag{23}$$

From here we see that the incentive constraint 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 $(\beta(N) + b)\alpha'(N) > \beta'(N)\alpha(N)$; else $\Phi'(N) \geq 0$. Inserting for $\alpha(N)$ and $\beta(N)$ and rearranging gives the cutoff of the corollary.

It remains to show that both regimes are consistent with equilibrium existence. Fix any $N > 2$ and any $\rho_I, \rho_H, \Sigma, \Sigma_\nu > 0$, and recall from the proof of Proposition 1 that the incentive constraint begins to bind at $b = \mathcal{C}(\phi_{UC})$, which lies strictly below $f(\phi_{UC}, \phi_{UC}) - g(\phi_{UC}, \phi_{UC})$ by (62). Both $\mathcal{C}(\phi)$ and $f(\phi, \phi) - g(\phi_{UC}, \phi)$ are continuous in ϕ (rational functions with positive denominators for $\phi > 0$), so the strict inequality (62) at $\phi = \phi_{UC}$ extends to an interval: there is $\varepsilon > 0$ such that every $\phi \in (\phi_{UC}, \phi_{UC} + \varepsilon)$ satisfies $\mathcal{C}(\phi) \leq f(\phi, \phi) - g(\phi_{UC}, \phi)$. Setting $b = \mathcal{C}(\phi)$ for such a ϕ, the incentive constraint binds at $\phi^* = \Phi(N) = \phi$, the upper bound (61) holds, and the equilibrium exists with a binding constraint.

For the unfavorable regime, note that $\mathcal{C}(\phi_{UC})$ is continuous and strictly increasing in A, and unbounded as $A \to \infty$, since $\alpha(N) > 0$, while the cutoff $\gamma_N(\sigma_\epsilon)$ does not depend on A. Hence there is $\bar{A}$ such that $\mathcal{C}(\phi_{UC}) > \gamma_N$ for all $A > \bar{A}$. For any such A, the construction above delivers $b = \mathcal{C}(\phi) > \mathcal{C}(\phi_{UC}) > \gamma_N$, since $\mathcal{C}$ increases in ϕ.

For the favorable regime, it is straightforward to verify that $\gamma_N(\sigma_\epsilon) > \underline{b}$ for all admissible parameters. As $A \to 0$, $\mathcal{C}(\phi_{UC}) \to \underline{b} < \gamma_N$, so there is $\underline{A} > 0$ such that $\mathcal{C}(\phi_{UC}) < \gamma_N$ for all $A < \underline{A}$. Shrinking ε if necessary, continuity of $\mathcal{C}$ keeps $b = \mathcal{C}(\phi) < \gamma_N$. □

Proof of Corollary 4. Treat N as continuous, as in Corollary 3. From the binding incentive constraint in Proposition 3, $b = A^2 \Phi_{CV}(N)^2 \alpha_{CV}(N) - \beta_{CV}(N)$, so

$$A^2 \Phi_{CV}(N)^2 = \frac{b + \beta_{CV}(N)}{\alpha_{CV}(N)} \equiv G(N).$$

$$G'(N) = \frac{\beta'_{CV}(N)\alpha_{CV}(N) - \big(b + \beta_{CV}(N)\big)\alpha'_{CV}(N)}{\alpha_{CV}(N)^2} = -\frac{\alpha'_{CV}(N)}{\alpha_{CV}(N)^2}\Big(b - \gamma_N^{CV}(\sigma_\epsilon, \Sigma_0)\Big),$$

where the second equality collects the numerator as $-\alpha'_{CV}(N)\, b + \big(\beta'_{CV}(N)\alpha_{CV}(N) - \beta_{CV}(N)\alpha'_{CV}(N)\big) = -\alpha'_{CV}(N)\big(b - \gamma_N^{CV}\big)$, using the definition of γ_N^{CV}.

By its definition in Proposition 3, $\alpha_{CV}(N) = \big(2N\rho_I J(N)\big)^{-1} > 0$, where $J(N) \equiv \frac{N-2}{\Sigma} +$

$\frac{N}{\sigma_\epsilon} + \frac{N-1}{\Sigma_\nu} + \frac{1}{N\Sigma_0+\Sigma_\nu} > 0$ collects the (positive) terms in its inverse. For $\alpha'_{CV}(N) < 0$, note

$$J'(N) = \frac{1}{\Sigma} + \frac{1}{\sigma_\epsilon} + \frac{1}{\Sigma_\nu} - \frac{\Sigma_0}{(N\Sigma_0+\Sigma_\nu)^2} > 0,$$

because $(N\Sigma_0+\Sigma_\nu)^2 \geq \Sigma_\nu^2 + 2N\Sigma_0\Sigma_\nu > \Sigma_0\Sigma_\nu$ implies $\Sigma_0(N\Sigma_0+\Sigma_\nu)^{-2} < 1/\Sigma_\nu$. Hence $\frac{d}{dN}\big(2N\rho_I J(N)\big) = 2\rho_I\big(J(N)+NJ'(N)\big) > 0$, and $\alpha'_{CV}(N) < 0$.

Since $\Phi_{CV}(N) > 0$ and $2A^2\Phi_{CV}(N)\Phi'_{CV}(N) = G'(N)$, the sign of $\Phi'_{CV}(N)$ equals the sign of $G'(N)$. Because $-\alpha'_{CV}(N)/\alpha_{CV}(N)^2 > 0$, the displayed decomposition gives $\Phi'_{CV}(N) > 0$ if and only if $b > \gamma_N^{CV}$, and $\Phi'_{CV}(N) < 0$ if and only if $b < \gamma_N^{CV}$.

Substituting the closed forms of $\alpha_{CV}(N)$ and $\beta_{CV}(N)$ into the definition of γ_N^{CV}, one can verify (e.g., with Mathematica) that $\gamma_N^{CV} > 0$ and $\partial\gamma_N^{CV}/\partial\Sigma_0 > 0$ for all admissible parameters, and that setting $\Sigma_0 = 0$ yields $\gamma_N(\sigma_\epsilon)$ of Corollary 3. Moreover, the cutoff is bounded in Σ_0: it increases toward the finite limit

$$\lim_{\Sigma_0\to\infty}\gamma_N^{CV}(\sigma_\epsilon,\Sigma_0) = \frac{\rho_I\,\sigma_\epsilon\Sigma_\nu\big(2(N-1)^3\Sigma + (N-2)(2N^2-3N+2)\Sigma_\nu\big)}{2(N-1)^2\Sigma\big((2N-1)\Sigma\sigma_\epsilon + 2\big(N\Sigma+(N-1)\sigma_\epsilon\big)\Sigma_\nu\big)},$$

which exceeds $\gamma_N(\sigma_\epsilon) > 0$. $-\beta_{CV}(N)$ increases in Σ_0 toward a finite limit as well: it is increasing in $\tilde{\Sigma}_{CV}$, since the derivative of $(\tilde{\Sigma}_{CV}-\Sigma)/(2\Sigma+(N-2)\tilde{\Sigma}_{CV})$ has numerator $N\Sigma > 0$, and from the closed form in Lemma 2, $\tilde{\Sigma}_{CV}$ increases in Σ_0 toward the finite limit $\Sigma + N\sigma_\epsilon\Sigma_\nu\big((N-1)\sigma_\epsilon + N\Sigma_\nu\big)^{-1}$. It is likewise straightforward to verify (e.g. with Mathematica) that $\partial\gamma_N^{CV}/\partial\sigma_\epsilon > 0$ for all admissible parameters, as in Corollary 3.

It remains to show that both regimes are consistent with equilibrium existence. The argument of the proof of Corollary 3 applies with $\mathcal{C}_{CV}$, g_{CV}, and γ_N^{CV} in place of $\mathcal{C}$, g, and γ_N: the binding construction near ϕ_{UC} uses the existence part of the proof of Proposition 3; the unfavorable regime uses that $\mathcal{C}_{CV}(\phi_{UC})$ is unbounded in A, since $\alpha_{CV}(N) > 0$, while γ_N^{CV} does not depend on A; and the favorable regime uses that $\gamma_N^{CV} > -\beta_{CV}(N) = \underline{b}_{CV}$ for all admissible parameters (straightforward to verify, e.g. with Mathematica), toward which $\mathcal{C}_{CV}(\phi_{UC})$ falls as $A \to 0$. □

Proof of Lemma 1. Part (i). Because $\nu_i \sim \mathcal{N}(\nu, \Sigma_\nu)$ and $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, the marginal distribution of R is given by: $R \sim \mathcal{N}(\nu, \Sigma_\nu + \Sigma)$, because $\mathbb{E}[R] = \mathbb{E}[\mathbb{E}[R|\nu_i]] = \mathbb{E}[\nu_i] = \nu$ and $Var(R) = \mathbb{E}[Var(R|\nu_i)] + Var(\mathbb{E}[R|\nu_i]) = \Sigma + \Sigma_\nu$.

Part (ii). For $s_i = 0$, $\mathbb{E}[R \mid \mu_i(0)] = \nu_i$. For $s_i = 1$, the normal projection formula gives $\mathbb{E}[R \mid \mu_i(1)] = \mathbb{E}[\nu_i \mid \nu_i + \epsilon_i] = \nu + Cov(\nu_i, \nu_i+\epsilon_i)\,Var(\nu_i+\epsilon_i)^{-1}\big((\nu_i+\epsilon_i) - \nu\big)$, with $Cov(\nu_i, \nu_i+\epsilon_i) = \Sigma_\nu$ and $Var(\nu_i+\epsilon_i) = \Sigma_\nu + \sigma_\epsilon$, which is the expression in the lemma.

To derive the conditional variance, I use the law of total variance: $Var(R \mid \mu_i(s_i)) = \mathbb{E}[Var(R \mid \nu_i) \mid \mu_i(s_i)] + Var(\mathbb{E}[R \mid \nu_i] \mid \mu_i(s_i))$. Since $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, we have: $Var(R \mid \mu_i(s_i)) = \Sigma + Var(\nu_i \mid \mu_i(s_i))$, with $Var(\nu_i \mid \mu_i(1)) = \left(\Sigma_\nu^{-1} + \sigma_\epsilon^{-1}\right)^{-1}$ and $Var(\nu_i \mid \mu_i(0)) = 0$. □

Proof of Lemma 2. Part (i) follows from the law of total variance: since $\nu_i = \nu_0 + \tilde{\nu}_i \sim \mathcal{N}(\bar{\nu}, \Sigma_0 + \Sigma_\nu)$ and $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$, we have $\mathbb{E}[R] = \mathbb{E}[\nu_i] = \bar{\nu}$ and $Var(R) = \mathbb{E}[Var(R \mid \nu_i)] + Var(\mathbb{E}[R \mid \nu_i]) = \Sigma + \Sigma_0 + \Sigma_\nu$, so that $R \sim \mathcal{N}(\bar{\nu}, \Sigma + \Sigma_0 + \Sigma_\nu)$.

For part (ii), the effort case $s_i = 0$ is immediate since $\mu_i(0) = \nu_i$ and $R \mid \nu_i \sim \mathcal{N}(\nu_i, \Sigma)$ by assumption, with $Var[R \mid \nu_i] = \Sigma$ independent of ν_i. The price p^c is uninformative about the return conditional on ν_i: since $R \mid \nu_i$ is independent of ν_0, i.e., $Cov(R, \nu_0 \mid \nu_i) = 0$, whatever the price conveys about ν_0 is irrelevant for R; the dealer need not observe the two components separately.

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: $\hat{\nu}_{-i}$, recovered by inverting (52), with $\hat{\nu}_{-i} \sim \mathcal{N}(\nu_0, \frac{\Sigma_\nu}{N-1})$. Note that my simplified notation hides the fact that $\hat{\nu}_{-i}$ is a function of p^c.

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:

$$\Sigma_{yy} = \begin{pmatrix} \Sigma_0 + \Sigma_\nu + \sigma_\epsilon & \Sigma_0 \\ \Sigma_0 & \Sigma_0 + \frac{\Sigma_\nu}{N-1} \end{pmatrix}, \qquad \Sigma_{\nu_i y} = \begin{pmatrix} \Sigma_0 + \Sigma_\nu & \Sigma_0 \end{pmatrix}. \tag{66}$$

By the normal projection formula:

$$\mathbb{E}[\nu_i \mid \mu_i(1), \hat{\nu}_{-i}] = \bar{\nu} + \Sigma_{\nu_i y}\Sigma_{yy}^{-1}(y - \bar{\nu}\mathbf{1}) = c_0\bar{\nu} + c_s\mu_i(1) + c_p\hat{\nu}_{-i}, \tag{67}$$

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:

$$c_s = \frac{\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{\mathcal{Q}}, \qquad c_p = \frac{(N-1)\Sigma_0\sigma_\epsilon}{\mathcal{Q}}, \qquad c_0 = \frac{\sigma_\epsilon\Sigma_\nu}{\mathcal{Q}}, \tag{68}$$

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

$$Var[\nu_i \mid \mu_i(1), \hat{\nu}_{-i}] = (\Sigma_0 + \Sigma_\nu) - \Sigma_{\nu_i y}\Sigma_{yy}^{-1}\Sigma_{y\nu_i} = \frac{\sigma_\epsilon\Sigma_\nu(N\Sigma_0 + \Sigma_\nu)}{\mathcal{Q}}. \tag{69}$$

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

$$Var[R \mid \mu_i(1), \hat{\nu}_{-i}] = \Sigma + \frac{\sigma_\epsilon \Sigma_\nu (N\Sigma_0 + \Sigma_\nu)}{\mathcal{Q}} = \tilde{\Sigma}_{CV}. \tag{70}$$

Since p^c is an invertible affine function of $\hat{\nu}_{-i}$ via (52), conditioning on p^c is equivalent to conditioning on $\hat{\nu}_{-i}$, so $\mathbb{E}[R \mid \mu_i(1), p^c] = c_0 \bar{\nu} + c_s \mu_i(1) + c_p \hat{\nu}_{-i}$ and $Var[R \mid \mu_i(1), p^c] = \tilde{\Sigma}_{CV}$. When $\Sigma_0 = 0$: $\mathcal{Q} \to \Sigma_\nu(\sigma_\epsilon + \Sigma_\nu)$, $c_p \to 0$, $c_s \to \frac{\Sigma_\nu}{\sigma_\epsilon + \Sigma_\nu} = \tilde{m}$, $c_0 \to 1 - \tilde{m}$, and $\tilde{\Sigma}_{CV} \to \Sigma + (\Sigma_\nu^{-1} + \sigma_\epsilon^{-1})^{-1}$, recovering Lemma 1. □

B Additional empirical evidence

Appendix B.1 provides additional suggestive evidence.
Appendix B.2 details the calibration: it maps the model to the data and estimates the dealer's risk tolerance and the contract share.

B.1 Additional suggestive evidence and discussion

This appendix examines alternative explanations for the two pricing patterns: the gap between client and own-account trades (Table 1), which I relate to moral hazard. And the gap between Alpha and TSX (Appendix Table A2), which I relate to imperfect competition, since Alpha is more concentrated than TSX (Appendix Figure A3). For each, I discuss possible confounding factors and present some robustness analysis.

Moral hazard. Client orders may differ from own-account trades in ways I do not observe. Not every such difference is a confounder. When a dealer times, routes, or works a client order differently because it keeps only part of the surplus (implicitly or explicitly), the resulting price difference represents the moral-hazard concern itself—understood broadly as non-contractible slack in how the dealer prepares and handles client orders (Section 3.1). One example is selection into account types, where booking the better execution opportunities as own-account captures surplus at the client's expense.

What would confound the comparison are differences that would remain even if the dealer prepared and handled client orders exactly like its own: properties of the drivers of the order, such as the information behind it or the urgency of the underlying trading need. I organize the discussion around this distinction, after conditioning on rich fixed effects. My preferred specification compares client and own-account trades within the same security, exchange, and narrow time window, absorbing systematic cross-exchange

Appendix Table A1: Robustness—Clients trade at worse prices than dealers for their own account

	(Buy)		(Sell)	
(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	
(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	
(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	
(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 A1 reports robustness regressions of transaction prices on an indicator for client-account trades and traded quantity, with own-account trades as the baseline category. (A) includes symbol × trade date × exchange fixed effects; (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; (C) restricts the sample to small trades, defined as trades below the 25th percentile of the trade-size distribution; (D) includes dealer fixed effects interacted with symbol, trade date, and exchange. In (C), traded quantity varies little by construction, so its coefficient is imprecisely estimated. 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.

and intraday patterns, which are known to play a role from previous studies (symbol × trade date × hour × exchange fixed effects; specification (B) of Appendix Table A1, with specification (A) omitting the hour controls).

The first group of remaining candidates concerns the drivers of the orders. Client trades could carry directional information, or serve urgent trading needs, so that the price concession buys speed rather than reflecting reduced care. Such differences confound only when the client initiates the order, which the data do not reveal; for example, some client-account orders are placed not by the client but by an asset manager trading on the client's behalf, so their timing and urgency are part of the intermediary's handling rather than the client's trading need.

Two features help reduce the concern that these differences are the sole drivers of the price gap. It is symmetric across trade direction (client purchases execute at higher prices and client sales at lower prices), consistent with a general deterioration in execution quality rather than one-sided information or one-sided urgency. Moreover, it persists

among small trades (Appendix Table A1, specification (C)), which are least likely to interact mechanically with the limit order book and may be more comparable across account types in motive and informational content. What I cannot examine directly is heterogeneity across clients, since the data flag whether a trade is executed for a client but not which client; for instance, differences in sophistication, or bundles of trades where a dealer sacrifices execution quality on one leg to complete a rebalancing quickly.

The second group concerns who executes. If different dealers handle the two account types, the gap could reflect differences across dealers rather than in how a given dealer treats each account. Part of these differences could reflect the moral-hazard effect itself: dealers handle client orders differently because of hidden incentives. To isolate the within-dealer component, I add dealer fixed effects. This is not feasible on top of the full controls, because few dealers trade both account types on both exchanges within the same security and hour, but it is in broader cells that retain sufficient variation, interacting dealer fixed effects with symbol, trade date, and exchange (specification (D)). The signs of the estimates persist, but the buy side loses statistical significance. This could indicate that clients may sort into dealers differently when buying than when selling, and it motivates analyzing purchases and sales separately.

To the extent that order- or client-driven differences unrelated to moral hazard remain, they inflate the measured price gap between client and own-account trades; I return to this when calibrating the model in Appendix B.2.

Imperfect competition. Execution conditions may differ between Alpha and TSX in ways I do not observe. Not every such difference is a confounder: thinner order-book depth, slower resiliency, or fewer active participants on the smaller venue translate into higher price impact and lower liquidity, and so represent the competition concern itself—understood broadly as the price effects of concentrated trading (Section 3.1), rather than literally as the model's dealer count.

What would confound the comparison are differences that would remain even if the two venues were equally liquid and competitive: properties of the orders routed to each venue, such as their urgency or information content. As before, I organize the discussion around this distinction, after conditioning on rich fixed effects: my preferred specification compares trades of the same account type, in the same security, on the same side of the market, and within the same narrow time window across the two venues (symbol × trade date × account-type × hour fixed effects; specification (B) of Appendix Table A2, with specification (A) omitting the hour controls). All specifications condition on account

Appendix Table A2: Robustness—Prices are worse on the more concentrated exchange (Alpha)

	(Buy)		(Sell)	
(A) Symbol × Trade Date × Account-Type Fixed Effects				
Alpha	0.005***	(0.001)	-0.002*	(0.001)
quantity	0.237***	(0.029)	-0.212***	(0.034)
N	139,516,860		134,287,187	
(B) Symbol × Trade Date × Account-Type × Hour Fixed Effects				
Alpha	0.006***	(0.000)	-0.006***	(0.000)
quantity	0.148***	(0.014)	-0.111***	(0.017)
N	138,958,482		133,779,455	
(C) Symbol × Trade Date × Account-Type × Hour Fixed Effects (Small Trades)				
Alpha	0.008***	(0.000)	-0.007***	(0.001)
quantity	93.453*	(42.009)	-117.988*	(46.228)
N	98,699,882		93,421,799	
(D) Symbol × Trade Date × Account-Type × Hour × Dealer Fixed Effects				
Alpha	0.004***	(0.000)	-0.004***	(0.000)
quantity	0.155***	(0.011)	-0.115***	(0.018)
N	136,164,192		131,418,991	

Notes: Appendix Table A2 reports robustness regressions of transaction prices on an indicator for trades executed on Alpha and traded quantity, with TSX trades as the baseline category. Specifications (A)–(C) mirror those of Appendix Table A1, with the roles of exchange and account type swapped: (A) includes symbol × trade date × account-type fixed effects; (B) further adds hour-of-day fixed effects, comparing trades of the same account type executed in the same security and within the same hour; (C) restricts the sample to small trades, defined as trades below the 25th percentile of the trade-size distribution. (D) includes dealer fixed effects interacted with symbol, trade date, account type, and hour—feasible here, unlike in Appendix Table A1, because dealers trade the same security on both venues within the same hour. In (C), traded quantity varies little by construction, so its coefficient is imprecisely estimated. 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.

type, so the venue gap does not reflect differences in the mix of client and own-account flow across exchanges, separating it from the account-type gap of Table 1.

The first group of remaining candidates concerns the drivers of the orders routed to each venue. Dealers may send more urgent, aggressive, or information-sensitive orders to Alpha. Two features suggest that these differences are unlikely to be the sole drivers. The gap is symmetric across trade direction (purchases execute at higher prices and sales at lower prices on Alpha) consistent with a general price-impact mechanism rather than one-sided urgency or one-sided execution failures. And it persists, if anything becomes larger, among small trades (Appendix Table A2, specification (C)), which are least likely to mechanically walk up a thinner limit order book.

The second group concerns who executes. If different dealers route to the two venues, the gap could reflect differences across dealers rather than execution conditions on the venues. Unlike for the account-type comparison, dealer fixed effects can here be added

on top of the full controls, because dealers routinely trade the same security on both venues within the same hour: with symbol × trade date × account-type × hour × dealer fixed effects, the venue gap remains at 0.004 for purchases and −0.004 for sales, precisely estimated (specification (D)).

The third group concerns the exchanges themselves: institutional differences unrelated to competition. These are limited, though not eliminated, by design, as TSX and Alpha are owned by the same parent company and operate under similar rule books and execution protocols.

To the extent that execution-related frictions orthogonal to competition remain, they mainly affect the measured price gap across the two exchanges; I return to this when calibrating the model in Appendix B.2.

B.2 Calibration

This appendix develops the calibration sketched in the main text: I first explain how I map the model to the data, and then derive the identifying moments, describe the estimation procedure, and discuss the estimates.

B.2.1 Mapping the model to the data

My goal is to rationalize aggregate market prices and the forces that drive them, such as risk tolerance and return volatility, rather than limit-order-book microstructure or order-level execution dynamics (which I do not observe).

Model-data mapping. To this end, I discretize the trading day into Γ periods and approximate the *average* price in each period *as if* it were the clearing price of a batch auction (Budish et al., 2015), treating active orders as net demand and passive resting orders as supply. I apply this approximation separately by exchange, account type, trade side, and security. Separate client and inventory trades give the observed price differences across accounts a model counterpart—the separate equilibrium prices p^*_{CL} and p^*_{IN} of the extended model in Appendix A.1.5. Keeping the buy- and sell-side separate avoids averaging over bid and ask prices, and treating each security separately avoids averaging across securities with different price levels, market capitalizations, or liquidity.

In practice, all trades clear through a common limit order book (within a security), so the separations by account type and trade side (unlike that by security) are part of the approximation rather than features of the market. Two considerations support this abstraction. First, in a continuous book, incoming orders execute against resting orders,

not against one another. Each flow (active buys, active sells, each account type) therefore forms the demand of its own 'auction', with the resting orders it consumes as supply. The shared price level and liquidity conditions, through which the flows interact, are absorbed into day–security fundamentals common to all auctions. Second, and specific to the account dimension, the separation rules out joint optimization across accounts, i.e., a dealer shading client orders to benefit its own book. Ring-fenced client and inventory books, client-priority rules, and distinct risk budgets are designed to prevent exactly this behavior. Modeling it would moreover render the trading game asymmetric, and thus intractable, since each dealer would act as two types of bidders with different objective functions.

Observables. I focus on liquid S&P/TSX 60 securities to ensure sufficient trading activity over short intervals. 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 use 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 averages reflecting the (hypothetical) auction prices $p^{el}_{tsk\tau}$, and aggregate signed traded quantities to obtain the total amount traded in each auction, $A^{el}_{tsk\tau}$. It is positive in buy-side and negative in sell-side auctions, so the price-impact term in (20") correctly flips sign across sides.

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}. The data record executed trades rather than bids, so I observe the dealers that trade, not all potential bidders. Counting over the full trading day assumes that a dealer who is active in a given cell executes at least one trade during that day. I drop observations when a security is not bought (or sold) by at least three dealers during all periods in a day: this ensures that daily averages provide a reasonable approximation to the expected auction price for a given account type and trade side, and it is consistent with the model's requirement of more than two dealers. The robustness checks at the end of this section validate that the choice of dealer count does not drive the main results qualitatively.

I approximate the return variance, Σ_{ts}, using the intraday variance of the national mid-quote for each security and day; as robustness, I also construct weekly and monthly mid-quote variances. Quote-based volatility measures are common in the related liter-

ature, 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 standard deviation is approximately C$0.194, about 38 basis points of the median stock price (C$50.33), implying a variance of ca. 0.038.[29]

B.2.2 Data moments and estimates

Data 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 share, ϕ^*.[30] Intuitively, risk tolerance governs how much prices of comparable own-account trades differ across exchanges with differing competitive conditions, while the contract share governs the price wedge between client- and own-account trades on a given exchange.

To see this formally, 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:

$$p^{el}_{t\tau sk} = \frac{1}{N^{el}_{tsk}} \sum_i \nu_{ti\tau sk} - \frac{\phi^*_k}{\rho_I} \Sigma_{ts} \left(\frac{N^{el}_{tsk} - 1}{N^{el}_{tsk} - 2} \right) \frac{A^{el}_{t\tau sk}}{N^{el}_{tsk}}, \tag{20'}$$

where $\nu_{ti\tau sk} \sim \mathcal{N}(\nu_{ts}, \Sigma_{\nu ts})$, iid across i and τ for each account type k (as specified in Appendix A.1.5), $\phi^*_{IN} = 1$ for own-accounts, and $\phi^*_{CL} = \phi^* \geq 0$ for client-accounts. Averaging over intraday periods yields the daily average price (per security, account-type, trade-side, and exchange):

$$p^{el}_{tsk} = \frac{1}{\Gamma} \sum_\tau p^{el}_{t\tau sk} \approx \nu_{ts} - \frac{\phi^*_k}{\rho_I} \Sigma_{ts} \left(\frac{N^{el}_{tsk} - 1}{N^{el}_{tsk} - 2} \right) \frac{1}{N^{el}_{tsk}} \frac{1}{\Gamma} \sum_\tau A^{el}_{t\tau sk}. \tag{20''}$$

The only unknown parameters in (20") are the dealer's risk tolerance ρ_I and ν_{ts}. Taking differences across exchanges eliminates ν_{ts}, so the cross-exchange price differences $p^{TSX,l}_{ts,IN} - p^{Alpha,l}_{ts,IN}$ identify ρ_I. Given ρ_I, the cross-account price differences $p^{TSX,l}_{ts,CL} - p^{TSX,l}_{ts,IN}$ pin down the contract share, ϕ^*.[31]

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

[29] Because the calibration restricts to liquid S&P/TSX 60 securities, this median price lies well above the full-sample median trade price of about C$22.5 reported for Table 1.

[30] 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.

[31] Because both parameters are identified from price differences for the same security and day, forces of the limit order book that affect the compared flows symmetrically difference out.

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_H^{lb}, consistent with the observed contract. Specifically, $\rho_H^{\text{lb}} = \rho_I\big(1/\phi^* - 1\big)$.

Operationally, I estimate ρ_I and ϕ^* in two steps, and then compute the implied bound ρ_H^{lb} from these estimates. For each cell (a day–security–side combination) the difference between the average inventory prices on Alpha and TSX in (20") solves for one value of ρ_I, and the estimate $\hat{\rho}_I$ is the median of these cell-level values. Given $\hat{\rho}_I$, the difference between the average client and inventory prices on TSX solves for one value of ϕ^* per cell, and $\hat{\phi}^*$ is again the median. Formally, this is a two-step GMM estimator with median-based moment conditions. Medians limit the influence of outliers, and sequential estimation is more stable than joint estimation because the first stage is relatively noisy, while remaining consistent under standard regularity conditions (Newey and McFadden, 1994).

For inference, I resample entire trading days with replacement—days are the clusters, allowing arbitrary dependence within a day but independence across days—and re-estimate ρ_I, ϕ^*, and ρ_H^{lb} with the same two-step procedure in each of 999 replications, so that first-stage uncertainty propagates into all confidence intervals. Standard errors are the standard deviations of the bootstrap draws. I report percentile 95% intervals, given by the 2.5th and 97.5th percentiles of the draws, which require no symmetry or normality and are more reliable for median-based ratio estimators; normal-approximation intervals are very similar.

Estimates. I first estimate the parameters on the full sample, before turning to how the contract varies with the number of dealers.

The full-sample estimate of risk tolerance, ρ_I, is about 670 (see Appendix Table A3). This is broadly consistent with Wittwer and Allen (2026), who infer a lower risk tolerance ($\rho_I \approx 250$), though of the same order of magnitude, for Canadian primary dealers, which account for large stock trading volume in Canada (Wittwer and Uthemann, 2024). The implied lower bound on clients' risk tolerance is around 385, which is about 57% of the dealer's risk-tolerance. Although the data do not identify clients, the most active clients are plausibly large institutional investors (such as pension funds, endowments, foundations, or fully managed investment funds), which typically hold diversified portfolios and are therefore relatively risk-tolerant.

Unobserved factors that systematically shift execution prices across exchanges for

Appendix Table A3: 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 A3 reports dealer risk tolerance, ρ_I, contract share, ϕ^*, and the implied lower bound on client risk tolerance, ρ_H^{lb}, from the main specification. Appendix Table A4 reports robustness checks. Standard errors and confidence intervals use a cluster bootstrap with 999 draws, clustering at the daily level (228 clusters).

comparable inventory trades can bias these estimates. For example, if harder-to-execute inventory orders were disproportionately routed to Alpha, for instance, ρ_I would be biased upward. Comparing average prices within narrow time windows, for the same security, across two exchanges owned by the same parent company (TMX) and operating under similar rules limits—though cannot eliminate—such concerns.

The estimated contract share, ϕ^*, is about 0.635 (Appendix Table A3).[32] This level is large, and comes from clients obtaining systematically worse prices than dealers do on their own account—recall the sizable price gap (50% of the median bid–ask spread; Table 1). Sizable shares of the surplus also arise elsewhere: intermediaries are documented to capture a substantial part of the value they create in asset management (Fama and French, 2010; French, 2008; Berk and van Binsbergen, 2015), and total manager compensation is a large share of the value generated in hedge funds and private equity (Goetzmann et al., 2003; Metrick and Yasuda, 2010; Lan et al., 2013).[33]

Client behavior unrelated to moral hazard that worsens their prices (greater urgency, say) biases the estimate $\hat{\phi}^*$ downward, because the model rationalizes larger price gaps through dealers retaining lower contract shares. However, the bias can also run upward; for example when own-account trades that hedge or offset client flow execute at prices close to the client trades that trigger them, which compresses the price gap, and inflates $\hat{\phi}^*$. Forces outside the contract that support client execution, such as best-execution requirements, or dealers' reputational concerns in repeat relationships, likewise compress

[32] My welfare exercise instead uses the contract share estimated at a specific dealer count, $\hat{\phi}^*(25) = 0.62$, which is slightly below this full-sample median.

[33] The per-trade share ϕ^* is a different object than quoted fee schedules: it is the fraction of the wealth a trade generates (i.e., the position times the return net of the purchase price) that the dealer keeps, whereas management fees are levied on assets under management and performance fees on portfolio returns above a hurdle. Moreover, quoted fees understate what intermediaries capture: in a calibrated 2-and-20 fund, total compensation absorbs approximately all of the surplus the manager creates, mostly through management fees (Lan et al., 2013).

Appendix Table A4: Robustness—Parameter estimates across specifications

Parameter	Estimate	SE (SD)	SE (Pct)	95% CI Low	95% CI High
10 minutes					
ρ_I	805.528	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 A4 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.

the gap and would be attributed to a larger contract share. Averaging execution prices smooths out trade-specific factors but not systematic differences.

In any case, I am less interested in the level of the average ϕ^* than in how it changes as more dealers compete, which is the model's main result. To recover this, I estimate ϕ^* separately for each count of active dealers, obtaining $\phi^*(N)$. This is the object plotted in Figure 3 and used in the welfare calculation; the procedure mirrors the level estimation: $\hat{\rho}_I$ is kept from the full sample, the median of the cell-level values is taken over the cells whose client auction has exactly N active dealers, and inference uses the same daily cluster bootstrap.

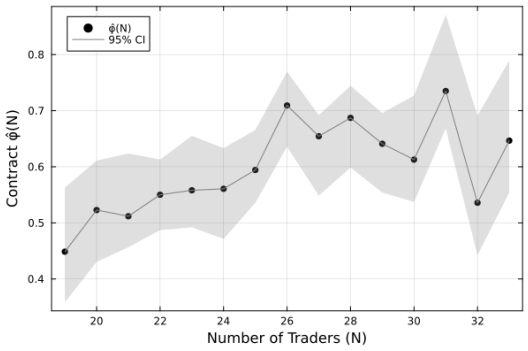
Robustness. My main object of interest, the contract share, is stable across specifications (Appendix Table A4): it stays between about 0.58 and 0.67 whether I collapse trades

into 10- to 30-minute auctions or measure return variances at daily, weekly, or monthly horizons. The dealer's risk tolerance is more sensitive, but in expected ways: it declines somewhat as the interval coarsens, with confidence intervals overlapping throughout, and it rescales mechanically with the variance horizon—since risk tolerance is measured relative to the unit of risk, a larger variance term multiplies $1/\rho_I$ while prices, dealer counts, and net supply stay fixed.

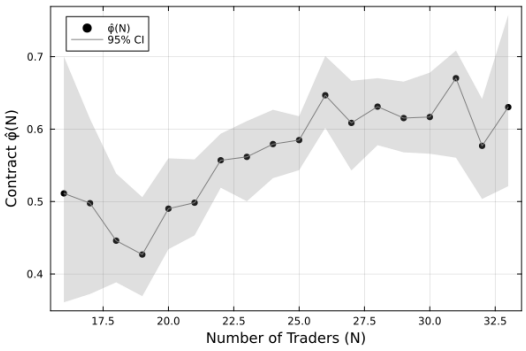
Appendix Figure A1 provides robustness checks of how the contract share varies in the number of dealers. It examines coarser auction intervals (20 and 30 minutes) and alternative dealer counts (per half-day and per two consecutive trading days) re-estimating the model under each, with the alternative counts entering the moments and the estimation of ρ_I. The magnitudes of $\hat{\phi}^*(N)$ vary across specifications, but a positive overall association between the contract share and competition persists. This qualitative pattern is what the model predicts for liquid assets, whose returns are easy to predict even without effort: the contract share rises with competition (though not necessarily monotonically). Counting dealers over windows shorter than half a day is not feasible: too few cells retain more than two dealers per account type, especially for Alpha's inventory account, consistent with executed trades understating the set of active bidders over short windows.

Appendix Figure A1: Robustness of how $\hat{\phi}^*$ changes in N

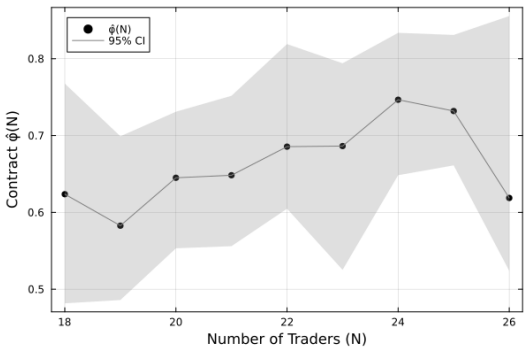
(a) 20-minute intervals



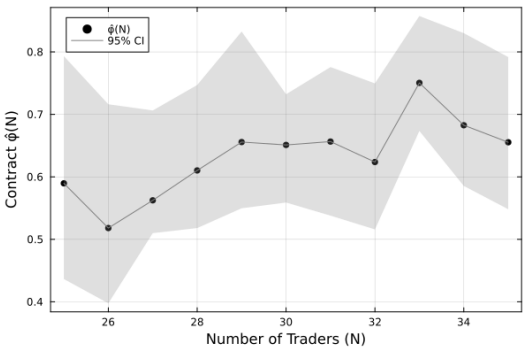
(b) 30-minute intervals



(c) N per half-day



(d) N per two trading days



Notes: All panels report $\hat{\phi}^*(N)$ and its 95% confidence interval from the daily cluster bootstrap, as in Figure 3. Panels (a) and (b) change the auction interval from 15 to 20 and 30 minutes. Panels (c) and (d) change the definition of the dealer count: distinct dealer IDs per side and security within each half-day (before/after 1 pm, averaged over the day's two halves and rounded), and over two consecutive trading days (each day inheriting the window's count). The alternative counts enter the full estimation, so the model-implied moments and ρ_I are re-estimated under each definition. Counting over windows shorter than half a day is not feasible: the model requires more than two dealers per account type, which too often fails at higher frequency for own-account trades on Alpha. All counts impose $N > 2$, and all panels restrict to dealer-count groups with at least 50 observations.

C Additional Tables and Figures

This section collects the additional appendix tables and figures referenced in the main text.

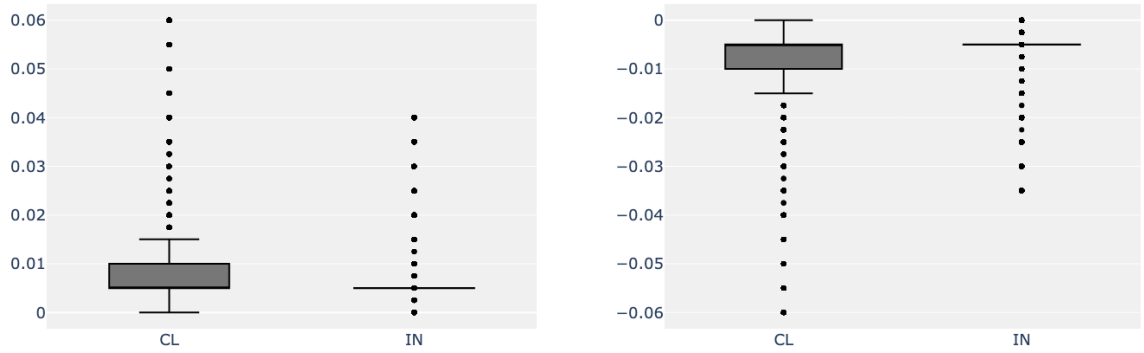
Appendix Table A5: Welfare effects for median client auction

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: Appendix Table A5 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 $\Sigma_\nu = \Sigma = 0.038$, the median return variance). Columns 1–2 list the estimated dealer risk tolerance $\hat{\rho}_I$ and the values of ρ_H, which I vary from its lower bound to values exceeding the dealer's risk tolerance. The lower bound reported here ($\rho_H^{\mathrm{lb}} = 415$) is based on the contract estimated at $N = 25$ ($\hat{\phi}^*(25) = 0.618$, shown rounded to 0.62 in the table), which is slightly below the full-sample median $\hat{\phi}^* = 0.635$; since $\rho_H^{\mathrm{lb}} = \hat{\rho}_I(1/\hat{\phi}^* - 1)$ is larger for a smaller contract, it exceeds the 385 reported in Table A3. 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 $\hat{\phi}^*(N{+}1)$. At the lower bound ($\rho_H = 415$), $\phi_{UC} = \hat{\phi}^*(25) = 0.62$, so eliminating moral hazard has no effect ($\Delta W_\phi = 0$); as ρ_H rises, ϕ_{UC} falls below $\hat{\phi}^*$ and the gain from removing moral hazard grows. Reading down the table, greater client risk tolerance (larger ρ_H) implies a smaller unconstrained contract, a wider moral-hazard gap, and a larger welfare cost of added competition (ΔW_N more negative).

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

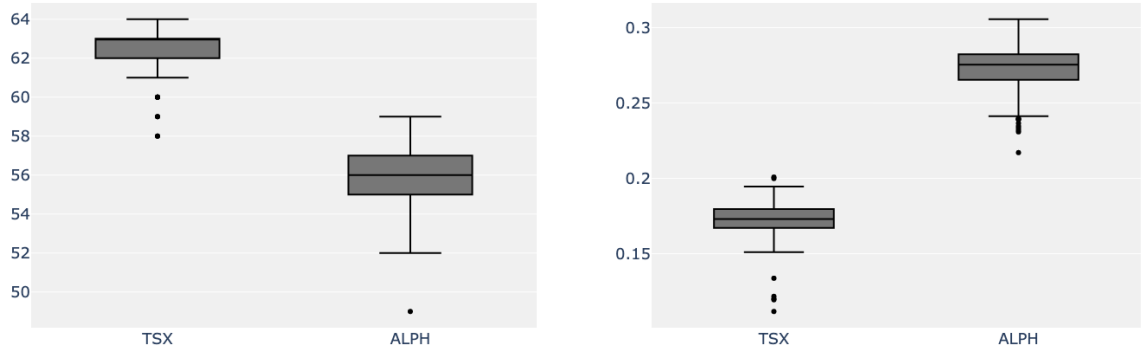
(a) Purchases (b) Sales



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.

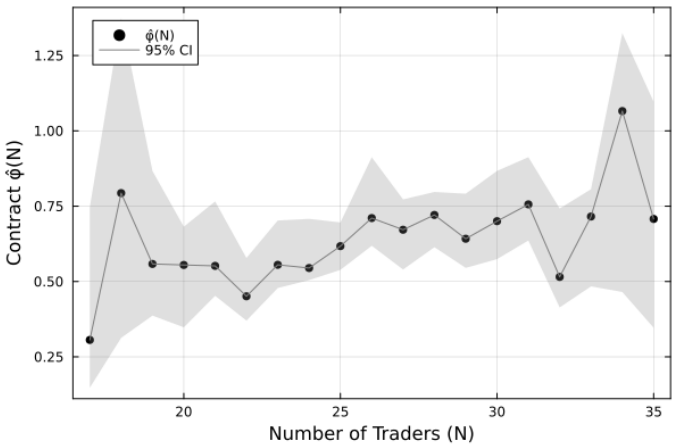
Appendix Figure A3: Market concentration is higher on Alpha

(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 $s_{ti} \in [0, 1]$.

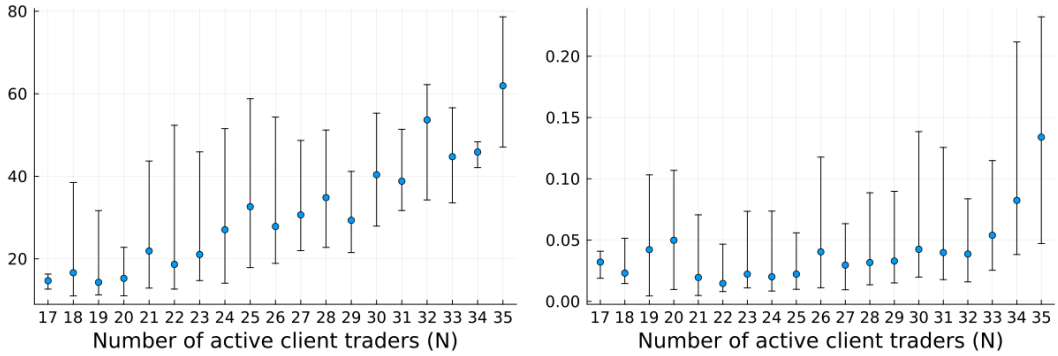
Appendix Figure A4: How ϕ^* changes in N — full, unrestricted range



Notes: $\hat{\phi}^*(N)$ and its 95% confidence interval for every dealer-count group, including the thinly populated groups at the extremes (support 17–35). The main-text Figure 3 restricts to groups with at least 50 observations (support 22–31), where the estimates are precise enough to be informative.

Appendix Figure A5: Amount traded and return variance by number of dealers

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



Notes: Appendix Figure A5a 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 A5b shows the distribution of return variances, Σ_{ts}, per N over days and securities. Supply is expressed in 1,000 shares, and return variance in Canadian dollars. The take-away is that the total amount traded tends to increase, not decrease, with the number of dealers, and return variance shows no systematic pattern across dealer counts apart from the extreme: the rise of the estimated contract share in N is unlikely driven by supply or volatility varying with N.