Market Power in the Securities Lending Market*

Shuaiyu Chen† Ron Kaniel‡ Christian C. Opp§

May 4, 2025

Abstract

We investigate market power in U.S. equity securities lending and assess its effects on distinct investor groups. Our data reveal high market concentration, non-competitive fees, and an excess supply of lendable stock inventory throughout the cross-section. Motivated by these findings, we develop a dynamic model that shows how intermediation and market power emerge as responses to short sellers' concerns over information leakage. Our quantitative analysis suggests that both short sellers and asset owners likely prefer the current intermediated market structure to a counterfactual centralized alternative—especially for smaller, less liquid stocks—since information leakages that would arise in a centralized venue undermine the very foundation of short sellers' business model. We estimate that, across stocks, non-competitive fee income increases lenders' valuations by between 1.4% and 90%.

Keywords: Short selling; Information leakage; Market power; Asymmetric information

*The authors thank Adem Atmaz (discussant), Nicola Garleanu (discussant), Sergei Glebkin (discussant), Ron Goettler, Ernst Maug (discussant), Wenxi Jiang (discussant), Ye Li (discussant), Chester Spatt (discussant), Amy Wang Huber, Chaojun Wang (discussant), Ed Van Wesep (discussant), Liyan Yang, and seminar participants at INSEAD, the University of Colorado Boulder, University of Illinois Urbana-Champaign, University of Rochester, Texas A&M, Tsinghua University, University of Wisconsin-Madison, the European Central Bank, the China International Conference in Finance, the University of Essex SENA Workshop on Financial Networks, the Annual Conference in Financial Economics Research, FIRS meetings, the Finance Theory Group Summer meeting, the MFA meetings, SFS Cavalcade, the European Finance Association Annual Meeting, and the Four Corners Index Investing Jamboree for their helpful comments.

†Purdue University. Email: chen4144@purdue.edu.

‡University of Rochester. Email: ron.kaniel@simon.rochester.edu.

§University of Rochester and NBER. Email: opp@rochester.edu.

1 Introduction

Short selling is a pervasive phenomenon in financial markets that has central implications for asset market dynamics (see, e.g., Miller, 1977; Harrison and Kreps, 1978; Scheinkman and Xiong, 2003). Institutionally, short selling is facilitated by the securities lending market, with equities on loan amounting to more than $1 trillion globally in recent years (ISLA, 2022). To establish a short position, investors such as hedge funds must first borrow shares in this market before selling them in the stock market. The fees associated with such stock loans are thus a key factor constraining short sales in practice and creating limits to arbitrage. In this paper, we empirically and theoretically examine the structure of the equity securities lending market and show that it has first-order implications for the stock valuations of both short sellers and lenders. Somewhat paradoxically, although custodian lenders charge substantial non-competitive fees under the prevailing intermediated market structure, we find that this arrangement dominates a centralized market even from the perspective of informed short sellers, particularly for illiquid stocks.

To start, we empirically document that the U.S. securities lending market is highly concentrated and subject to market power. This concentration arises from the intermediation of securities lending by a small set of custodian lenders. For a given stock, the top two lenders typically command a market share between 40% and 70%.[1] Further, we show that fees on lending contracts are non-competitive across nearly the entire universe of publicly traded stocks. Even among the stocks commanding the highest borrowing fees, lenders have significant excess inventory available for lending, consistent with strategic rationing associated with the exercise of market power.

Motivated by these findings, we develop and estimate a dynamic asymmetric-information model that reveals how intermediation and market power in securities lending emerge as responses to short sellers' concerns about information leakage. The model facilitates quantitative analyses of counterfactual market structures and yields estimates of the impact that non-competitive fees have on security lenders' stock valuations and on the distribution of abnormal returns ("alphas") that short sellers earn when targeting different segments of the cross-section of stocks.

Our results reveal a substantial impact of non-competitive fee income on security

[1] See Section 2 for details.

lenders' stock valuations, ranging from 1.4% of extra value for large-cap, low-fee stocks to increases of up to 23% for small-cap stocks and as much as 90% for nano-cap stocks. We show that accurately estimating these effects requires properly capturing the joint dynamics of stocks' dividend and fee income, which we account for via a granular state specification that reflects firms' transitions within the size and fee distributions.[2] The excess valuations we estimate do not reflect stock fundamentals in the traditional sense—that is, they are distinct from the present value of a stock's future dividends. Instead, these value wedges represent a transfer of rents from informed short sellers to security lenders via fee payments. Accordingly, our results reveal that market power in securities lending markets can have material spillover effects on valuations in stock markets, distorting the informational content of prices and thereby potentially affecting real economic outcomes (see, e.g., Hayek, 1945; Bond et al., 2012; van Binsbergen and Opp, 2019).

Turning to short-sellers, our model provides estimates of the hard-to-observe distribution of the abnormal returns these agents earn when targeting different segments of the cross-section of stocks. We find substantial variation in these net-of-loan-fee abnormal returns, both cross-sectionally and in the time series, as a function of a firm's current size (market capitalization) and loan fees. Our estimates indicate that fees reduce short sellers' profits by an average of about 54%. Despite these substantial fee payments, short sellers still generate high post-fee abnormal returns, particularly among smaller and illiquid stocks. Importantly, our estimated model predicts that, under a counterfactual centralized market structure, these abnormal returns from shorting would drop substantially. Although loan fees would decline significantly under a competitive centralized market, short sellers' ability to profit from their trading strategies would be undermined by transparency and information leakage in such a venue.

We obtain these results based on a quantitative dynamic model of trade under asymmetric information with two trading venues: a centralized limit order market akin to Glosten and Milgrom (1985) and a securities lending market facilitating short selling. For the latter, we contrast two potential market structures: an intermediated and opaque over-the-counter market that resembles the status quo, and a centralized

[2] Engelberg et al. (2018) also highlight the importance of the empirical fact that fees are not static but rather evolve stochastically over time.

market that is transparent and competitive.

Key elements of our model are informed traders' concerns about information leakages and the fact that shorting is a two-step transaction whereby securities can be sold only after they have been located with a lender committed to providing a stock loan. This requirement is due to a regulatory ban on naked short selling, which applies to market participants such as hedge funds (see SEC, 2022a).[3] Beyond this regulatory requirement, securities naturally have to be sourced by the settlement date of a sale. This is important because short sellers typically follow dynamic trading strategies that feature a gradual buildup of positions over extended periods, often several days or even weeks. Short sellers therefore aim to minimize informational leakage of their securities borrowing activity throughout this entire period.

When intermediated by a custodian lender, this activity remains localized information within the bilateral borrower-lender relationship. Intermediaries hold inventories of lendable shares, and short sellers' demand is observed only by these relationship lenders. In contrast, if securities lending takes place in a centralized venue such as a limit order book and is thus publicly observable, the resulting information leakage impedes traders' ability to profit from shorting: stock prices are immediately updated to reflect signals from the securities lending market, leading to price declines even before traders are able to fully establish their intended short positions.

Our analysis shows not only that borrowers prefer interacting with an intermediary lender that is committed to facilitating secrecy, but also that a lender's commitment itself is supported by both market concentration and non-competitive fees. This is because these forces increase the continuation value of maintaining a reputation as a committed lender that does not opportunistically use borrowing demand information for short-term trading gains. The resultant exercise of market power in our framework contributes to low inventory utilization as observed in the data and an inelastic response of the equilibrium quantity of lending with respect to before-fee shorting profits.

Our analysis shows not only that borrowers prefer interacting with an intermediary lender committed to facilitating secrecy, but also that a lender's commitment is itself crucially sustained by both market concentration and non-competitive fees.

[3]Exempted from this requirement are market makers engaged in bona fide market-making activities.

This is because future fee income provides the continuation value that supports a reputation as a committed lender–one that does not opportunistically exploit borrowing demand information for short-term trading gains. In a concentrated market, intermediaries attracting much of the market-wide securities lending can pass through a larger fraction of this fee income to the ultimate asset owners without jeopardizing their own commitment. The resulting exercise of market power in our framework contributes to low inventory utilization, as observed in the data, and to an inelastic response of the equilibrium quantity of lending to before-fee shorting profits.

Related literature. Our paper contributes to the literature seeking to understand the determinants of short selling costs and the broader implications of short selling for financial markets. First, we document the presence of market power in the equity securities lending market. Second, using a novel structural model, we examine both qualitatively and quantitatively how intermediation and market power emerge as responses to short sellers' concerns over information leakage. This focus complements prior studies of short selling in dynamic frameworks where trading is driven by belief heterogeneity (e.g., Duffie et al. (2002), Vayanos and Weill (2008), Atmaz et al. (2019), Evgeniou et al. (2019), Garleanu et al. (2021)). In these models, agents agree to disagree and therefore do not benefit from intermediated market structures that help preserve secrecy during the build-up of short positions. Our quantitative analysis suggests that both short sellers and asset owners likely prefer the current intermediated market structure to a counterfactual centralized alternative—especially for smaller, less liquid stocks—since information leakages undermine the very foundation of short sellers' borrowing demand. Accordingly, our findings shed light on why intermediation prevails in securities lending markets across major economies, including the US, UK, Germany, and France.

Our tractable structural framework yields novel estimates of how non-competitive fees affect the stock valuations of security lenders and the distribution of alphas that short sellers earn by targeting different segments of the cross-section of stocks. In contrast, much of the theoretical literature on short selling has focused on the qualitative implications of short selling constraints for asset prices; notable examples include Miller (1977), Harrison and Kreps (1978), Diamond and Verrecchia (1987), Hong and Stein (2003), Scheinkman and Xiong (2003), and Banerjee and Grave-

line (2014).[4] In modeling lending markets and assessing their impact on security prices, Blocher et al. (2013) adopt a reduced-form approach, making assumptions on properties of demand functions in the equity and lending markets. In the context of repo markets, Duffie (1996) shows that specialness raises the equilibrium price of the underlying instrument by the present value of savings in borrowing costs associated with the repo specials. Relatedly, Cherkes et al. (2013) demonstrate that the Palm-3Com spin-off puzzle can be explained by the fact that Palm's floating shares generated fee revenue for asset owners, whereas the shares held by 3Com did not.[5]

Finally, our paper is connected to the broader literature documenting and studying the implications of market power in trading environments, such as Hortaçsu and McAdams (2010), Huber (2023), An and Song (2023), Allen and Wittwer (2023a), and Allen and Wittwer (2023b).

2 Empirical Evidence of Market Power

In this section, we present evidence of market power in the U.S. securities lending market based on several key indicators. Our securities lending data are from Markit Securities Finance Data Analytics (see Online Appendix A for a detailed data description). Throughout this paper, we are interested in both aggregate measures and cross-sectional heterogeneity. To this end, we present most of our results by categorizing the cross-section of firms into 25 groups. We first sort firms by size (equity market capitalization) and then into subgroups that vary by fees. The size categories are created based on widely used Russell indices and are labeled large-cap (top 1-200), mid-cap (201-1000), small-cap (1001-2000), micro-cap (2001-4000), and nano-cap (all remaining stocks).[6] For each of those size groups, we further sort stocks according to their fee yield, where the fee yield is conceptually the equivalent of a dividend yield but capturing cash flows from securities lending rather than from dividends. Specifically, it is defined as the ratio of a stock's total annualized fee in-

[4] Other studies modeling shorting activity include the working paper version of D'Avolio (2002), Nutz and Scheinkman (2020), and Nezafat and Schroder (2021).

[5] Sikorskaya (2024) examines how institutional demand driven by investment mandates affects short-selling, finding that larger institutional holdings increase both lending supply and shorting demand, ultimately raising loan fees.

[6] An advantage of using the Russell indices is that they are widely-used US equity benchmarks for institutional investors that are major lenders and borrowers in the U.S. securities lending markets.

come to the market value of the stock's posted inventory. As we will show in our analysis below, the fee yield is a central measure entering lenders' stock valuations. Using this measure, we assign stocks in each size bin to five fee-yield bins with the percentile cutoffs 80%, 90%, 95%, and 98%. We choose these cutoffs to capture the highly non-linear relation between percentiles and fee yield levels. We provide a detailed description of our approach in Online Appendix A.1.

In the following, we discuss four metrics that jointly help us gauge whether lenders can exercise market power in the securities lending market: (1) market concentration measures, (2) prices on loan contracts, (3) marginal costs of lending, and (4) excess supply of securities available for lending. Lastly, we provide evidence on the relation between informed trading in the stock market and borrowing fees in the cross-section of stocks.

Market concentration. Tables 1 and 2 document the market shares of the top two security lenders for each of our 25 firm groups. Table 1 documents the within-group average of the market shares that are reported by Markit, weighting stock-specific market shares by each stock's relative contribution to the total value on loan in a given group. These estimates indicate that the top two lenders tend to have a large market share, ranging between 50% and 85%.

As a robustness check, we adjust our market share measures in Table 2 in two ways. First, within each group, we weight the market shares of the top lenders for a given stock by the stock's share of the total fee income generated by stocks in the group (rather than using the value on loan). This approach thus puts more weight on the market shares of stocks that are important contributors of fee income. Second, we adjust the Markit-reported market share numbers to account for the fact that Markit does not cover all securities lending activity. Markit reports that it covers approximately 85% of lending activity. We therefore scale down the Markit-reported market shares by this number. This approach assumes that any lending activity not accounted for in the market share numbers reported by Markit involves institutions other than the top two lenders identified by Markit. This more conservative approach can be viewed as providing lower-bound estimates. Yet, we still find that the top 2 lenders typically command a combined market share between 40% and 70% of the stocks on loan.

Pricing of loan contracts and potential sources of marginal costs of lending. Next, we investigate evidence from equilibrium pricing in the securities lending market. Table 3 shows that fees on loan contracts range between 0.29% per annum for large-cap, low-fee stocks to 75.14% for nano-cap, high-fee stocks. Even small-cap (mid-cap) stocks have fees ranging between 0.29% and 12.24% (0.28% and 6.88%) per annum.

When evaluating whether these prices for lending contracts are indicative of the presence of market power, it is key to benchmark them against any potential marginal costs security owners incur by lending out their securities. Such costs could theoretically emerge from channels such as (1) cash flow rights, (2) voting rights, (3) tax considertions, (4) trading needs and opportunities, and (5) operational costs related to settlement. In the following, we discuss these channels in detail.

First, lenders collect the fees we document in addition to the dividend payments of the underlying shares, that is, lenders do not give up their cash flow rights. Further, lending contracts are overnight contracts that are secured by collateral. Regulation T mandates that to protect lenders from default, agents borrowing shares have to post cash collateral in an amount equal to 102% of the value of the borrowed securities (typically cash or U.S. Government securities) plus an additional 50% in the form of other collateral. The collateral is marked to market daily.

Second, going beyond cash flow rights, one might be concerned that fees represent compensation for lenders giving up voting rights, which are passed on to the borrower during the lending period. To examine the potential role of this channel, we report in Panel A of Table A1 (see Online Appendix F) fees analogously to those presented in Table 3 but conditional on removing all observations 15 days before and after voting record dates to eliminate the potential effect of voting rights on lending fees. Comparing Tables 3 and A1 reveals that voting rights have hardly any impact on the magnitudes of loan fees. For example, whereas the highest-fee micro-cap stocks have a typical fee equal to 38.4% per annum in our baseline specification, this number is 37.6% once we exclude days around voting record dates.[7]

Third, we examine the potential role of shares transfers around ex dividend dates, for example due to concerns about a differential taxation of investors. In Panel B

[7]These results echo existing studies showing that voting record dates have quantitatively little impact on the securities lending market (see, e.g., Christoffersen et al., 2007; Aggarwal et al., 2015).

of Table A1 we report fees conditional on removing observations 15 days before and after ex-dividend dates. Just like in the case of voting record dates, the results for fees throughout the whole cross-section are hardly affected. Going back to the earlier example of the highest-fee micro-cap stocks, we find that the typical fee is now equal to 39.2% as compared to 38.4% when all dates are included.

Fourth, we consider potential marginal costs associated with trading needs and opportunities. In this context, several institutional facts are relevant. First, lending contracts are overnight contracts that allow lenders to recall their shares at a daily level. Second, large asset management companies like Vanguard are the typical lenders in this market (see the results on market concentration above). A company like Vanguard does not need permission from its fundholders to lend out the assets under management of their mutual funds and ETFs. Further, by regulation, open end funds are allowed to lend out only up to one third of their asset value (Investment Company Act, 1940; SEC, 1997, 2022b). This regulation implies that it is highly unlikely that the possibility of intra-day fund withdrawals or other trading needs generate expected marginal costs for an asset manager like Vanguard: at any point in time, at least two thirds of a fund's assets are available to be traded freely as they cannot have been lent out. Empirically, even in the most extreme scenarios, funds do not liquidate more than two thirds of their assets within a day. Moreover, if anything, funds typically liquidate their most liquid assets first. Yet, those are the ones with the lowest fees. Further, as we will highlight below, even the one third of assets that could be theoretically lent out without violating regulations are in fact not lent out. That is, there is an excess inventory supply of shares available for lending, even for the most extreme day-stock observations (e.g., the 2% of small cap stocks with the highest fees).

Fifth, and finally, any other operating costs related to the settling of lending contracts are just as small as they are for standard equity sales and purchases. Such costs are orders of magnitude below the fees we observe.

In sum, the pricing of loan contracts yields lenders significant incremental income that is orders of magnitude above possible marginal costs of lending.

Excess supply of shares available for lending. As a last step, to show the presence of market power, it is essential to also examine lending quantities, in particular, the

presence of an excess supply of shares that are readily available to be lent out. Our data set allows us to directly measure this inventory of shares, which is distinct from the total number of shares outstanding. In practice, different classes of investors face legal restrictions and reputational costs to posting their shares as available for lending, such as firm managers, founders, and some institutional investors do. As noted above, in accordance with the Investment Company Act (1940), open-end funds can lend out at most one-third of their total asset value (see SEC, 1997, 2022b). Given that a large fraction of stock market investments are intermediated by mutual funds, this rule substantially reduces the inventory of shares available for lending when compared to the total supply of shares outstanding. Importantly, what we measure as inventory in the data are shares that institutional lenders report as the subset of their shares that are available for lending.

If lending fees are set such that lenders obtain income in excess of marginal costs despite the presence of an excess supply (that is, slack inventory), this identifies the presence of market power in the securities lending market. In Panels A and B of Table 4 we document different measures of excess inventory, including a fee-income-weighted inventory slackness. This latter measure overweights the excess inventory of stock-date observations with high fee income within each of the 25 stock groups (see the table caption for details), thereby ensuring that the numbers are representative of excess inventory specifically when prices (fees) are elevated. Independent of the measure considered, we find that across all 25 stock groups, there is significant excess inventory, even for the smallest stocks and those that are in the very right tail of the fee distribution.[8]

We conclude that empirically, a lack of spare inventory is systematically not the culprit causing fees to exceed lenders' marginal costs.

Loan fees and informed trading in the stock market. As foreshadowed in the introduction, our model will capture the notion that in the presence of market power

[8]Conceptually, it is important to note that under a competitive market structure, only the aggregate inventory scarcity for given stock matters for lending contract fees. The distribution of inventory across lenders is irrelevant. That is, in a competitive market, if there is an aggregate excess supply of inventory for a given stock, then fees have to equal marginal costs (which are close to zero, as noted above). Having said that, our data set allows us to also measure the inventory specifically of the top two lenders (individual inventory levels for other lenders are not reported). Interestingly, we find that for a large fraction of stock-day observations, even these top-two lenders alone would have sufficient inventory to service all the demand we observe at prevailing fees.

in the securities lending market, fees allow lenders to extract part of hedge funds' information rents from adversely selecting other traders, in particular if information leakages in the process of establishing short positions are mitigated. Correspondingly, the model will predict a clear link between informed investors' trading activity in the stock market and fee levels in the securities lending market. Having investigated various measures related to the securities lending market across our 25 stock groups, one may wonder whether there is indeed evidence of such a relation in the data.

To address this question, we consider a useful empirical measure of informed trading in the stock market: liquidity providers' revenue from charging bid-ask spreads, as proxied by the product of the half-spread and volume. Given the competitiveness of market making activity in the stock market, this revenue captures compensation that offsets market makers' losses from getting adversely selected from better informed traders. Notably, as this measure is an increasing function of equilibrium trading volume, it does not per se capture illiquidity, but rather the equilibrium amount of informed trading in a given stock.

To address this question, we consider a useful empirical measure of informed trading in the stock market: liquidity providers' revenue from charging bid-ask spreads, proxied by the product of the half-spread and trading volume. Given the competitiveness of quote provision and the fact that market makers typically have diversified operations across hundreds, if not thousands, of stocks, this revenue reflects compensation offsetting their losses from adverse selection by better-informed traders. Notably, because this measure increases with equilibrium trading volume, it captures the amount of informed trading rather than illiquidity. Figure 1 plots fees in the securities lending market against this proxy of informed trading in the stock market, where each blue circle represents one of our 25 stock groups. The data indeed reveal a remarkably tight *positive* relation, with a correlation of 93.3%. Stocks with the highest percentage fees in the securities lending market also tend to incur the largest adverse selection costs for market makers.

Motivated by this evidence on market power and the connection between loan fees and informed trading in the stock market, we proceed to developing a dynamic model of shares lending that sheds light on why market power can be sustained in the securities lending market when trading is subject to asymmetric information.

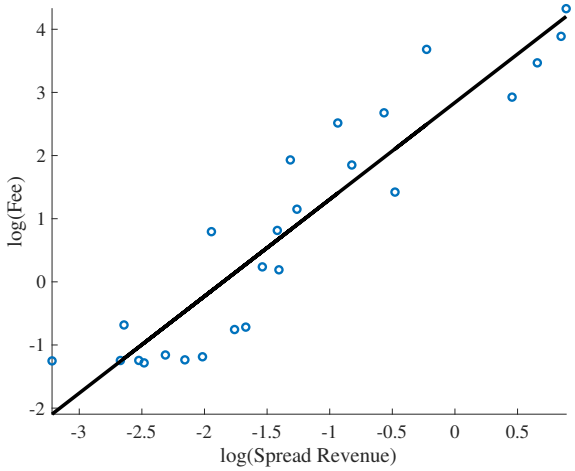


Figure 1: The figure provides a scatter plot illustrating the relation among the 25 stock groups between loan fees in the securities lending market and spread revenue in the stock market (that is, volume $\times$ 1/2 bid-ask spread; see the caption of Table A2 in Online Appendix F for details).

3 Model

Four types of financial institutions interact in a dynamic trading environment that features both a centralized limit order market and a securities lending market: (1) *hedge funds* that obtain private information about asset payoffs and may choose to short stocks, (2) *liquidity traders* that trade for liquidity reasons, (3) *regular traders* that can buy or sell shares in the centralized limit order market and can lend out their shares, and (4) *custodian security lenders* that can intermediate securities lending for participating asset owners. In Sections 5 and 6.5, we also consider an alternative structure for the securities lending market, where investors interact directly in a centralized setting rather than using the intermediation services of custodian lenders. Figure 2 provides an overview of the environment.

Stochastic discount factor. All institutions are firms that are ultimately owned by households who value cash flows according to a stochastic discount factor (SDF) m. This SDF assigns risk premia to aggregate shocks, which are publicly observable. Institutions maximize the present value of their cash flows, given the private information they may have about asset-specific (idiosyncratic) states. Liquidity traders' behavior is further affected by liquidity shocks, as detailed below. The SDF follows

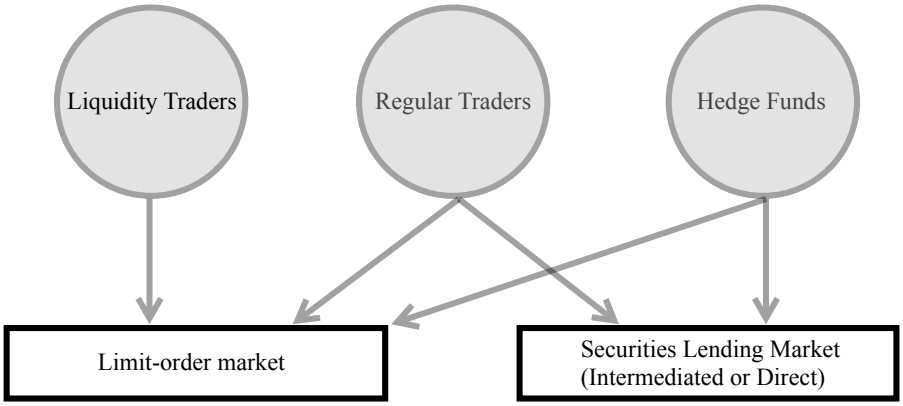


Figure 2: Investor Groups and Markets
The figure illustrates the markets and the different groups of investors in the model. We consider two distinct structures for the securities lending market, one where lending is intermediated by custodian lenders and one where investors interact directly.

the diffusion process:

$$\frac{dm_t}{m_t} = -r_f dt - \chi dB_t, \tag{1}$$

where r_f denotes the risk free rate, B_t is a Brownian motion, and χ is the price of risk for exposures to aggregate shocks dB_t.

Asset. Institutions are trading a generic asset (e.g., a stock) that generates lumpy dividends. Dividends over an interval $[t, t+dt)$ are given by $c_t dN_t$, where N_t is a Poisson process with arrival intensity λ, and where c_t can be interpreted as a productivity-adjusted measure of a firm's capital in place. The time between two innovations to N_t in our model is the stochastic equivalent of what a period would be in a discrete-time setting. However, the stochastic structure we consider has significant advantages in terms of analytical tractability.

Productivity-adjusted capital c_t follows the jump-diffusion process[9]

$$\frac{dc_t}{c_{t-}} = \mu_c dt + \sigma_c dB_t + (e^{v_{t-}} - 1) dN_t, \tag{2}$$

where v_{t-} is a random variable the realization of which is generally private infor-

[9] In the following, all processes will be right continuous with left limits. Given a process y_t, the notation y_{t-} will denote $\lim_{s \uparrow t} y_s$, whereas y_t denotes $\lim_{s \downarrow t} y_s$.

mation of some agents, specifically hedge funds. In contrast, the current value of assets in place, c_t, is publicly observable at time t and the aggregate shocks dB_t are unpredictable for all agents. The evolution of the conditional jump size v_t is given by:

$$dv_t = (z_t - v_{t-})dN_t, \tag{3}$$

where $z_t \sim \mathcal{N}(-\frac{\sigma_v^2}{2}, \sigma_v)$.[10] That is, the process for v_t stays constant between innovations to the Poisson process N_t and takes a new normally distributed value upon an innovation to N_t. We denote by $f_v(\cdot)$ and $F_v(\cdot)$ the corresponding normal PDF and CDF of v_t upon such innovations. When quantitatively estimating and evaluating the model in Section 6, we will generalize the dynamics of c_t to capture salient empirical features of the joint dynamics of fee and dividend income.

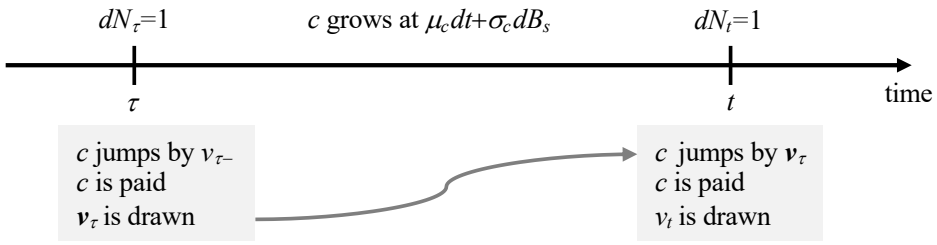


Figure 3: Timing Convention for Growth and Dividends
The figure illustrates the timing of idiosyncratic capital jumps, dividend payments, and changes in the jump size v. Hedge funds have private information about future growth by observing v_τ at time τ.

The timing convention indicated in equation (2) and Figure 3 is important. Consider some date t that features an innovation $dN_t = 1$, as illustrated in Figure 3 (marked on the right-hand side of the graph). At that time, capital c jumps by a log-change equal to v_{t-}, which in turn is still equal to the lagged value v_τ that was realized at the *previous* innovation $dN_\tau = 1$, with $\tau < t$ (marked on the left-hand side of the figure). Logically after capital has been subject to this jump of size v_τ at time t, a dividend equal to the new value of c_t is paid.

[10] Specifying $\mathbb{E}[v_t] = -\frac{\sigma_v^2}{2}$ ensures that the idiosyncratic jump term $(e^{v_{t-}} - 1)dN_t$ in equation (2) has zero mean.

While v_τ is not publicly observable at time τ, hedge funds can observe v_τ at that time, giving them private information about future dividend growth (which, in this illustrative example, is realized at time t). Once capital has jumped by v_τ at time t, agents can infer the value of v_τ by observing the realized change. Yet, at that point, this information has no relevance beyond the new level of c_t, since future growth depends on a new, independent value v_t drawn at time t.

The Poisson process N_t also governs the frequency with which new hedge funds and cohorts of liquidity traders obtain shocks and can trade in the centralized limit order market at the posted bid and ask prices.

Hedge funds. Upon an innovation to N_t, a new hedge fund arrives to the market and observes the new value of v_t. Each new hedge fund does not have any preexisting holdings of the asset and can, through trading, take positions in the set $\{-\pi, 0, \pi\}$, with $\pi \in (0,1)$. An informed hedge fund becomes a regular uninformed investor after the next shock to N_t. The assumption that the persistence of hedge funds' informational advantage is also governed by shocks to N_t has significant advantages in terms of tractability. While differential information persistence is in principle an interesting feature worth studying, it is not an essential aspect in the context of this paper's objectives.

To facilitate our quantitative analysis, we present and estimate in Section 6.5 an extended version of our model where total borrowing demand is a continuous variable and each cohort features a continuum of hedge funds that obtain noisy signals about the fundamental v. Further, in Online Appendix C.5, we extend our analysis to the case where hedge funds receive private value shocks that introduce non-informational borrowing motives.

Liquidity traders. Upon an innovation to N_t, existing liquidity traders owning $(1-\pi)$ units of the asset have to sell their holdings, and a newly arriving cohort of liquidity traders has to purchase $(1-\pi)$ units. The maximum total asset holdings across hedge funds and liquidity traders is thus normalized to one, and the variable π represents a measure of the size of informed hedge funds relative to liquidity traders.

Custodian lender. Securities lending takes place in a two-sided market where custodians can attract inventory from asset owners and lend out that inventory to hedge

funds. To streamline the exposition of our baseline model, we start by taking as given the following features of this intermediated market: (1) Securities lending is facilitated by one custodian lender who retains a fraction $\delta \in [0,1]$ of loan fee revenue. (2) The custodian is committed not to opportunistically use the information it obtains from observing hedge fund loan demand in bilateral negotiations. In Section 7.1 and Online Appendix E, we present an explicit foundation for these assumptions based on a repeated game analysis. This analysis reveals the theoretical link between custodian market power and the ability to credibly commit to non-opportunistic behavior.

Asset supply. We denote by ρ_0 the quantity of the asset that can in principle be lent out to short sellers or be used to make sales offers in the limit order market. Given that we normalized the total potential asset holdings of hedge funds and liquidity traders to one, the variable ρ_0 represents a relative measure: it is the ratio of the total quantity of the asset that can be lent out or posted at the ask price to the total size of positions taken by hedge funds and liquidity traders.

Conceptually, the total quantity of the asset that is available for lending or sales offers in the limit order market also need not be equal to the total shares outstanding. In practice, different classes of investors face legal restrictions and reputational costs to lending out their shares, such as firm managers, founders, and institutional investors do.[11] In our model, such investors follow a buy-and-hold strategy and do not lend out their shares.

Yet consistent with the empirical evidence on inventory utilization below 100% presented in Section 2, we consider the case where the lendable supply ρ_0 is large enough to ensure shares lending is not mechanically supply constrained, which corresponds to the condition that $\rho_0 > 1 + \pi$. Specifically, this assumption ensures that the supply ρ_0 is large enough to avoid mechanical constraints on the following two types of simultaneous activities: (1) The maximum possible demand (excluding hedge funds repurchasing the asset) can be offered for sale at the ask price, which is 1 units of the asset. (2) The maximum possible shorting demand, π units, can be offered in the shares lending market. We will relax this assumption in our quanti-

[11] In accordance with the Investment Company Act (1940), open-end funds can lend out at most one-third of their total asset value (see SEC, 1997, 2022b). Given that a large fraction of stock market investments are intermediated by mutual funds, this rule substantially reduces the supply for securities lending when compared to the total supply of shares outstanding.

tative analysis presented in Section 6. Note that the units of the asset that hedge funds repurchase and redeliver at the end of a lending contract period become again available as lendable shares at that time. Thus, these units do not take up additional capacity of the supply ρ_0.

Securities lending versus liquidity provision in the stock market. Upon the beginning of a new stochastic period, regular investors owning the asset can decide whether to offer their shares for sale in the limit-order market or to make them available for stock loans. Naturally, a given asset cannot be committed to both a loan and a sale simultaneously. Since shares lending generally yields fee income, asset owners therefore face an implicit opportunity cost when forgoing that income and instead acting as liquidity providers that offer their shares for sale in the stock market at an ask quote. We consider a simple maker-taker commissions model for the limit order market that compensates for these endogenous inventory costs for liquidity providers, thereby allowing both the securities lending market and the stock market to attract asset owners in equilibrium.

In practice, maker-taker fee models reward liquidity providers and charge liquidity takers: Liquidity makers (i.e., those placing limit orders that add to the book) receive a rebate, liquidity takers (i.e., those placing market orders that remove liquidity) pay a commission. Correspondingly, in our model, the limit-order market charges buyers commissions and rebates the resultant income back to ask-quote providers, using a pro-rata distribution according to the units committed for sale. In equilibrium, these rebates then render asset owners indifferent between engaging in shares lending and providing ask quotes in the stock market.

We introduce the variable ρ to denote the units of the asset that are posted as lendable inventory in equilibrium (where $\rho \in [0, \rho_0]$). The remainder of the units, $(\rho_0 - \rho)$, will be offered for sale via ask quotes in the stock market. Liquidity traders, which collectively hold $(1-\pi)$ units of the asset at each point in time, do not post their shares in either market to ensure they can sell all their units.[12] Correspondingly,

[12]Liquidity traders with positive holdings must sell their units upon an innovation $dN_t = 1$. Posting a bid quote would not guarantee the sale of all units they wish to sell. Instead, liquidity traders pick up quotes in the limit order book, which ensures a sale with probability 1 in equilibrium. This specification is also consistent with the standard characterization of *liquidity providers* in the literature, which are the ones posting bid and ask quotes, whereas investors picking up those quotes via market orders are liquidity takers.

their holdings are not part of the supply ρ_0.

Timeline. Upon an innovation $dN_t = 1$, the following logical order of events applies:

1. Dividend income realized at time t is collected by the agents who were the owners of the asset at time $t-$ (that is, t is an ex-dividend date).
2. Asset owners choose whether to deposit their shares for lending with the custodian lender or to make their units available for sale in the limit-order market. The custodian lender sets the fee ϕ_t per unit of the asset lent out to maximize expected fee income. Loan agreements start at time t and mature at the time of the next Poisson shock $dN_\tau = 1$, with $\tau > t$.[13] At maturity, borrowers must return the asset and any dividend it pays at that time.
3. A new hedge fund and a new cohort of liquidity investors arrive in the market.
4. The hedge fund can borrow shares at the fee posted by the custodian lender. The custodian lender does not reveal the total demand at this stage.
5. Bid and ask quotes are posted by market participants in the limit order market.
6. Investors submit quantities they wish to buy or sell in the limit-order market (market orders). Clearing starts with the best bid and ask offers and follows a pro-rata allocation rule.
7. Contracts are settled via delivery of the asset in accordance with all lending contracts and trades (steps 2 to 6 above). For example, assets repurchased by a borrowing hedge fund (step 6) are delivered back to the custodian lender, who may pass them on to a new hedge fund as part of a lending contract (agreed to in step 4), who in turn sold them at the bid to an investor (in step 6) and is delivering them to that investor.

4 Analysis

In this section, we analyze the presented baseline model with a delegated securities lending market. We start by characterizing the buy-and-hold value of the asset from the perspective of a regular trader who does not receive liquidity shocks or private

[13] Specifying a lending contract with stochastic maturity increases tractability and is analogous to stochastic debt maturities in corporate finance models (see, e.g., Leland, 1998).

signals. Afterwards, we describe how bid and ask quotes are set in the competitive limit order market, what payoffs hedge funds obtain from borrowing shares, and what loan fees are charged by the custodian lender.

Buy-and-hold value. An asset owner can post its shares with the custodian lender and receive both dividend and fee income. Correspondingly, the buy-and-hold value of an asset accounts for both these income streams. Denote by P_t the buy-and-hold value of one unit of the asset, given investors' prior beliefs about the distribution of the current value of v_t at time t. Upon a shock $dN_t = 1$, a hedge fund obtains private information and borrows π units of the asset to immediately sell them at the posted bid quote whenever v is below some endogenous threshold v_ϕ, that is, when $v < v_\phi$. We will characterize this threshold value v_ϕ below. The probability of shorting is therefore $F_v(v_\phi)$, and conditional on shorting, the pro-rata passed-through fee income per unit of shares posted with the custodian lender is $(1-\delta)\frac{\pi\phi_\tau}{\rho}$. The ex-dividend buy-and-hold value at time t is thus given by:

$$P_t = \mathbb{E}_t \int_t^\infty \frac{m_\tau}{m_t} \cdot \Big(c_\tau + \underbrace{(1-\delta)}_{\text{Pass-through}} \cdot \underbrace{F_v(v_\phi)\frac{\pi\phi_\tau}{\rho}}_{\text{Expected fees}} \Big) dN_\tau. \tag{4}$$

Absent signals about the current value of v, the relevant state variable for an investor's valuation is the level of productivity-adjusted capital c. Accordingly, we proceed to characterize the value function $P(c)$ for the buy-and-hold value. Since the lending fee will also be chosen without conditioning on information about v (see details below), its equilibrium value will depend only on c as well, that is, $\phi(c)$.

We conjecture and verify that $P(c)$ and $\phi(c)$ are linear functions of c. Going forward, we denote variables scaled by the expected dividend rate, λc, with a tilde. Further, we denote by $Q(c,v)$ the value of the asset conditional on knowing the current levels of c and v, which is the relevant case for hedge funds that obtain private information about v. The following proposition characterizes the value functions $P(c)$ and $Q(c,v)$.

Proposition 1. *The buy-and-hold value of the asset conditional on prior beliefs about*

v is given by:

$$P(c) = \frac{\lambda \cdot (c + (1-\delta) \cdot F_v(v_\phi) \cdot \frac{\pi\phi(c)}{\rho})}{r_f + \sigma_c \chi - \mu_c}. \tag{5}$$

The buy-and-hold value conditional on observing v is given by:

$$Q(c,v) = e^v \cdot P(c). \tag{6}$$

Proof. See Online Appendix C. □

The valuation formula (5) exhibits the familiar structure under perpetual growth, whereby the expected cash flows in the numerator (here from dividends and fees) are divided by the difference between the discount rate and the growth rate. Further, equation (6) shows the tractable feature of our setting that the valuation for privately informed agents equals the uninformed valuation $P(c)$ modulated by the factor e^v.

Next, we characterize the competitive bid and ask quotes that regular traders post in the limit order market.

Bid quotes. Bid quotes are set accounting for potential adverse selection from hedge funds. A hedge fund that observes v will borrow π units of the asset and sell them at the posted bid price B whenever $v < v_\phi$, where v_ϕ is the above-mentioned endogenous threshold value that we will characterized below. Moreover, the current cohort of liquidity traders sells their $(1-\pi)$ units at the posted bid price B irrespective of the realization of v. Anticipating this potential demand, competitive investors offer 1 unit at the bid quote, which is the maximum total quantity the two groups of traders might wish to sell at a given point in time. Competitive pricing implies that the bid price is the highest price B that satisfies liquidity providers' break-even condition:

$$\underbrace{B \cdot \big(\pi F_v(v_\phi) + (1-\pi)\big)}_{\mathbb{E}\,[\text{Payments from Purchases}]} = \underbrace{P \cdot \big(\pi F_v(v_\phi)\, \mathbb{E}[e^v | v \le v_\phi] + (1-\pi)\big)}_{\mathbb{E}\,[\text{Value of Shares Bought}]}, \tag{7}$$

that is, the bid quote B is set such that a competitive investor posting this price breaks even, accounting for the possibility that the offer is picked up by an informed hedge

fund or liquidity traders. The left-hand side of equation (7) represents the expected dollar payments made by bid-quote providers: the bid price B times the expected quantity traded. The right-hand side of the equation represents the expected total value of the shares traded at the bid. Equation (7) can be rearranged to:

$$\frac{B}{P} = \frac{\pi F_v(v_\phi)\,\mathbb{E}[e^v | v \leq v_\phi] + (1-\pi)}{\pi F_v(v_\phi) + (1-\pi)}, \tag{8}$$

where normality of v implies the following simple representation for the truncated expectation in equation (8):

$$\mathbb{E}[e^{v_t} | v_t \leq v_\phi] = \frac{F_v(v_\phi - \sigma_v^2)}{F_v(v_\phi)}. \tag{9}$$

Since providing a bid quote does not require a liquidity provider to back the quote with actual inventory, this activity does not incur an associated opportunity cost of the type present for ask quotes (see discussion below). Nevertheless, the bid quote characterized in (8) reflects the fact that an asset bought by a liquidity provider at a bid quote can be lent out only at the beginning of the next period. Accordingly, it is an ex-fee and ex-dividend quote. This is because posting a bid quote does not guarantee obtaining the asset, so there is no inventory that could be deposited with the custodian lender before a transaction has actually occurred.

Ask quotes. Whereas bid quotes do not need to be secured by existing asset inventory, ask quotes do require such inventory to ensure delivery upon settlement. Thus, ask-quote providers incur an associated opportunity cost: they cannot simultaneously offer their shares for stock loans, thereby forgoing the associated loan fee income. As noted in Section 3, the limit-order market compensates liquidity providers for these endogenous inventory costs via a maker-taker commissions model.

Ask quotes depend on existing short interest, as purchases motivated by the delivery of borrowed shares have a different informational content than other trades do. We denote by A_0 the ask price that prevails when current short interest is equal to zero and by A_1 the ask price when existing short interest is positive and equal to π in equilibrium. Further, as established in Appendix C.1, the depth of ask quotes (i.e., the units offered) exactly matches the maximum possible demand in equilib-

rium. Suppose hedge funds purchase the asset when the fundamental value exceeds a threshold v_{A_0} which we will characterize below. If short interest was zero at time $t-$, the ask price is the highest price A_0 that satisfies liquidity providers' break-even condition:

$$\underbrace{A_0 \cdot (\pi(1-F_v(v_A))+1-\pi)}_{\mathbb{E}[\text{Revenues from Sales}]} = \underbrace{P \cdot (\pi(1-F_v(v_A))\,\mathbb{E}[e^v | v \geq v_A]+1-\pi)}_{\mathbb{E}[\text{Value of Shares Sold}]}. \quad (10)$$

The left-hand side of equation (10) reflects liquidity providers' expected revenues from sales at the ask price A_0. The right-hand side of equation (10) represents the expected value of the shares sold at ask quotes. Competition in quote provision implies that ask-quote providers just break even in expectation.

As noted above, a maker-taker commissions model creates incentives for agents to provide ask quotes despite the associated opportunity cost of forgone loan fee income. Suppose that κ_0 denotes the per-unit commission buyers pay for executed market orders. At the beginning of a period, asset owners must be indifferent between the expected commission rebates they would receive from posting ask quotes and the fee income they could obtain by making their assets available for lending:

$$\underbrace{\kappa_0 \cdot (\pi(1-F_v(v_A))+1-\pi)}_{\mathbb{E}[\text{Commissions Rebates from Sales}]} = \underbrace{(1-\delta)F_v(v_\phi)\frac{\pi\phi}{\rho}}_{\mathbb{E}[\text{Forgone Loan Fees}]}. \quad (11)$$

Since a buyer has to pay both the ask price and the commission to obtain the asset, the threshold value v_{A_0} solves:

$$e^{v_{A_0}}P = A_0 + \kappa_0 \quad \Leftrightarrow \quad v_{A_0} = \log\left(\frac{A_0+\kappa_0}{P}\right). \quad (12)$$

Adding equations (10) and (11) and rearranging yields the combined condition:

$$\frac{A_0+\kappa_0}{P} = \frac{\pi(1-F_v(v_{A_0})\,\mathbb{E}[e^v | v \geq v_{A_0}]+1-\pi+(1-\delta)F_v(v_\phi)\frac{\pi}{\rho}\frac{\phi}{P}}{\pi(1-F_v(v_{A_0}))+1-\pi}, \quad (13)$$

where normality of v implies the following formula for the truncated expectation in

equation (13):

$$\mathbb{E}[e^{v_t}|v \geq v_{A_0}] = \frac{F_v(-v_{A_0})}{1-F_v(v_{A_0})}. \tag{14}$$

In contrast, if short interest was π at time $t-$, borrowing hedge funds must repurchase these units to deliver them to the custodian lender upon a realization $dN_t = 1$. In this case, the depth of the ask quote is $(1+\pi)$ units, allowing both these repurchases and additional purchases from hedge funds and liquidity traders to be accommodated. Correspondingly, uninformed demand equals 1 unit.

The cum-commissions costs of purchasing the asset $(A_1 + \kappa_1)$ now satisfy the following relation analogous to equation (13):

$$\frac{A_1+\kappa_1}{P} = \frac{\pi(1-F_v(v_{A_1}))\,\mathbb{E}[e^v|v \geq v_{A_1}] + 1 + (1-\delta)(1+\pi)F_v(v_\phi)\frac{\pi}{\rho}\frac{\phi}{P}}{\pi(1-F_v(v_{A_1})) + 1}, \tag{15}$$

where the threshold value is $v_{A_1} = \log((A_1+\kappa_1)/P)$. Equation (15) reflects the fact that there are now three types of traders potentially picking up the ask quote: (1) A new hedge fund that may have received sufficiently positive information about the asset and wishes to purchase π units, (2) a hedge fund that has to purchase π units to return borrowed assets to the custodian lender, and (3) a new cohort of liquidity traders that needs to buy $(1-\pi)$ units.

Irrespective of the level of outstanding lending contracts, market clearing across the securities lending market and the limit order market implies that the inventory available for lending is given by $\rho = \rho_0 - 1$.[14]

Hedge funds' borrowing decisions. A hedge fund observing v optimally borrows whenever the value derived from borrowing shares exceeds the loan fee, that is

$$B(c) - L(c,v) \geq \phi(c), \tag{16}$$

[14] As noted above, if hedge funds repurchase and redeliver shares at the end of a lending contract period, these shares become again available for lending shares at that time. The provision of ask quotes thus always effectively uses 1 unit of the asset.

where $L(c,v)$ denotes the present value of the liability of having to return the shares and the dividends associated with them at the end of the loan period.[15] The following proposition characterizes the value of this liability.

Proposition 2. *A hedge fund with an open short position has a liability with the following present value*

$$L(c,v) = e^{v} \frac{\lambda \cdot (c + A_1(c))}{r_f + \sigma_c \chi - \mu_c + \lambda}. \tag{17}$$

Proof. See Online Appendix C. □

The solution to the value of a hedge fund's liability (17) reflects the fact that, upon maturity of the lending contract, the fund must both provide the dividend c and repurchase the shares at the prevailing ask price A_1. Moreover, equation (17) shows that hedge funds observing negative information (lower future growth v) anticipate that fulfilling this liability will be cheaper, thereby making borrowing more profitable.

Optimal loan fee. As a last step, we derive the optimal loan fee a custodian lender charges in order to maximize the value of fee income received by its clients. The lender solves a classic monopolistic screening problem, since it does not know the hedge fund's private information (type) when posting the fee.[16] When choosing the optimal loan fee, the custodian lender faces a trade off between the probability of lending and the fee collected conditional on lending. Choosing a loan fee ϕ is equivalent to pinning down a marginal hedge fund type v_ϕ that is just indifferent between borrowing and not borrowing at the posted fee.

Proposition 3. *The custodian lender's optimal loan fee is equal to the marginal hedge fund type's value from having borrowed shares, that is,*

$$\phi(c) = B(c) - L(c, v_\phi), \tag{18}$$

[15] Recall that trades in the limit order market are ex dividend. Thus, a hedge fund purchasing shares at time t does not get a claim to the dividends those shares pay at time t. However, the lender requires to receive the dividends that will be paid at the next innovation of N_t.

[16] This is the optimal mechanism from the perspective of investors posting the inventory of lendable shares when facing a privately informed hedge fund. See also Holmstrom and Myerson (1983) for the implementation of durable decision rules. Note that the equilibrium in our model is Pareto efficient.

where the marginal type v_ϕ solves the equation:

$$\frac{f_v(v_\phi)}{F_v(v_\phi)} \cdot \left(\frac{B(c)}{L(c, v_\phi)} - 1 \right) = 1. \tag{19}$$

Proof. See Online Appendix C. □

The reversed hazard rate $\frac{f_v(v_\phi)}{F_v(v_\phi)}$ in equation (19) reflects the typical marginal trade off associated with screening problems: marginally raising fees has the cost of excluding the marginal type (which has density $f_v(v_\phi)$) and the benefit of collecting higher revenue from all inframarginal types (which have mass $F_v(v_\phi)$).

Discussion: loan fees and abnormal returns. Our model predicts that if econometricians disregard loan fees when computing returns to investors, they find negative abnormal returns in asset pricing tests, especially for assets with high fee income. This result is consistent with empirical evidence that asset pricing anomalies are concentrated in the short leg, among high-fee stocks and are often insignificant once fees are incorporated (see in particular Stambaugh et al., 2011; Muravyev et al., 2022).

In the context of our model, we define the following measure:

$$\text{fee yield} \equiv \frac{\lambda \cdot (1 - \delta) \cdot F_v(v_\phi) \frac{\pi\phi}{\rho}}{P}, \tag{20}$$

which is the analogue of the dividend yield but considers investors income from shares lending. Note that this additional income is collected by asset owners that make their shares available for lending. If an econometrician chooses the correct SDF but implements standard asset pricing tests that only account for dividends and capital gains when computing returns, she will on average estimate the following buy-and-hold alpha:

$$\alpha = -\text{fee yield}. \tag{21}$$

That is, the estimated alpha is equal to the negative of the fee yield.[17] Given this

[17] See related points made by Duffie (1996) in the context of repo markets.

alpha, we can express the buy-and-hold value as follows:

$$P(c) = \frac{\lambda \cdot c}{\underbrace{r_f + \sigma_c \chi + \alpha}_{\text{Average Return}} - \mu_c}. \tag{22}$$

In our setting where uninformed investors have unbiased beliefs, the buy-and-hold alpha (21) is directly empirically measurable, as it is purely due to the fee yield lenders obtain. In contrast, in a difference-in-opinion setting featuring fees, alphas emerge not only due to fees but also since agents' distorted beliefs affect prices (such as in Duffie et al., 2002).

Another context in which alphas would be detected empirically is when econometricians use benchmark indices to measure return performance. Suppose that the asset's cash flow c represents the cash flows generated from the firms in such an index. Then, relative to the performance of the index that excludes lending fee income, a shares lender's portfolio outperforms by $\alpha =$ fee yield, that is, a *positive* expected excess return is estimated.

In sum, standard asset pricing tests will detect abnormal returns and the magnitudes of these returns are predicted to increase with (1) loan fees and (2) the quantity of shares lent out, but decreasing in the quantity posted with the custodian lender. Thus, if fewer shares are available for lending, abnormal returns are predicted to increase. Loan fees, in turn, are predicted to increase with information asymmetry, e.g., when spread revenue is larger (bid-ask spreads times volume traded). Moreover, the quantity of borrowing should be related to information asymmetry and hedge funds' relative trading capacity (as measured by $\frac{\pi}{\rho}$).

5 Intermediated vs. Direct Securities Lending

To provide insight on the implications of shares lending being delegated to a strategic custodian lender, we analyze in this section how equilibrium outcomes would differ if the shares lending market were instead centralized and competitive. To do so, we consider the most generic form of a competitive securities lending market, that is, a market that mirrors the prevalent limit order markets that are used for regular securities trading in practice: suppliers of lendable shares submit competing offers

for loan contracts and hedge funds can pick the best offer. We analyze and discuss additional market structures in Section 7.

Contrasting the market structures reveals the key economic features of intermediation that is present in our baseline model: a custodian lender strategically chooses fees so as to maximize the value for investors lending their shares, and it does not instantaneously reveal shorting demand to all market participants (see step 4 of the description of the baseline setup). In contrast, such secrecy is absent in the centralized shares lending market, which potentially harms hedge funds' ability to extract rents from negative private information due to information leakages.

In the setting with a centralized lending market, the logical order of events upon an innovation $dN_t = 1$ is generally identical to the one we laid out for the delegated market, with differences only pertaining to steps 2 and 4. Specifically, in step 2, individual investors directly quote lending contract fees (rather than posting their shares with a custodian lender). In step 4, these posted offers can be taken up by hedge funds, with realized lending volume being publicly observable. We reiterate the detailed description of the logical order of events under this market structure in Online Appendix Section B.

Whereas the intermediated securities lending market featured a screening game, the centralized market implies a signaling game. We characterize the perfect Bayesian equilibrium that is robust to the presence of an arbitrarily small exogenous costs of contacting a custodian to receive a quote (see Online Appendix C.5 for details). The following proposition describes key equilibrium outcomes when shares lending is moved to this centralized market arrangement.

Proposition 4. *When shares lending is moved to the centralized market, the following equilibrium outcomes obtain:*

- *Shares lenders are strictly worse off if the shares lending market is centralized, since the lending fee is zero, that is,* $\phi = 0$.
- *Hedge funds may be worse off if the shares lending market is centralized, in particular when their trading capacity is sufficiently large relative to the trading volume of liquidity traders.*

Proof. See Online Appendix C.5. □

Proposition 4 confirms that in the centralized lending market, lending fees are competed down to zero, which harms shares lenders. Who are ultimately these shares lenders in practice? While large asset management companies act as lenders in today's markets, they pass on most of their securities lending income to fundholders (see Morningstar, 2018). For example, whereas Fidelity, Schwab, and TIAA, have a passthrough of around 90%, Vanguard even passes on 98% of its gross securities lending revenue from U.S. equities to its fundholders. Thus, retail and pension investors would be negatively affected by the introduction of a competitive securities lending market that leads to the elimination of lending fee income.

Despite the elimination of loan fees, the proposition reveals that hedge funds are not simply better off when securities lending is centralized and competitive. This is because securities borrowing volume is publicly observable in this case, generating a signal that leads to the updating of prices in the stock market. In the model, a mixed-strategy equilibrium obtains whereby a hedge fund borrows with probability one when it intends to short sell the asset, but also borrows with positive probability when it intends to buy the asset. In equilibrium, this latter probability is such that the ask price conditional on borrowing occurring is the same as the ask price conditional on no borrowing occurring. However, the bid price changes as a function of the signal obtained from observing borrowing volume, which limits a short sellers' ability to profit from negative private information. As noted in the proposition, this information leakage channel is particularly influential when the underlying security is illiquid in the sense that hedge funds' trading capacity is relatively large when compared to the amount of liquidity-motivated trading in the security. This is because information leakages are more consequential when securities are relatively less liquid to begin with.

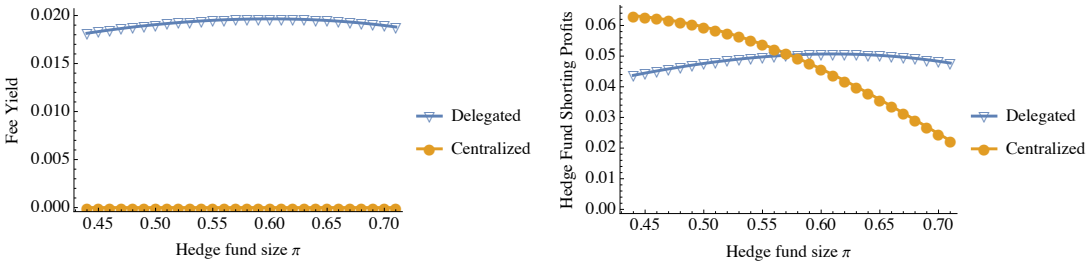
Figure 4 illustrates these results with comparative statics with respect to the parameter π, which governs the size of trades that hedge funds can take relative to those of liquidity traders. Panel (a) compares the fee yield lenders obtain conditional on shares lending being centralized versus delegated to a custodian lender. Whereas competition drives down fees to zero in the centralized securities lending market, fee yields are positive in the intermediated market. Since the sum of the maximum holdings of hedge funds (π) and liquidity traders $(1-\pi)$ is normalized to one, increasing π not only raises hedge funds' trading capacity, but also decreases liquidity trad-

ing. Higher values for the parameter π therefore reduce liquidity and can potentially reduce fee yields.

Panel (b) illustrates the key result of the proposition by comparing informed hedge funds' profits from shorting across the two market structures. Whereas informed shorting is more profitable in the centralized securities lending market when the security is more liquid, intermediated shares lending is preferable for them when there is less liquidity. That is, for values of π exceeding a threshold, hedge fund profits from shorting are higher under the intermediated market structure. When comparing market structures from the perspective of hedge funds, there is a tradeoff between lending fees (paid in the first step of establishing a short position) and the price obtained from selling immediately after borrowing shares, as represented by the bid quotes. While hedge funds do benefit from paying zero borrowing fees in a competitive securities lending market, they are harmed by information leakages and the associated repricing of securities. The market is less liquid for higher values of π, implying that bid prices respond more strongly to information leakages from the securities lending market. Therefore, centralizing securities lending can ultimately erode hedge fund profits from shorting.

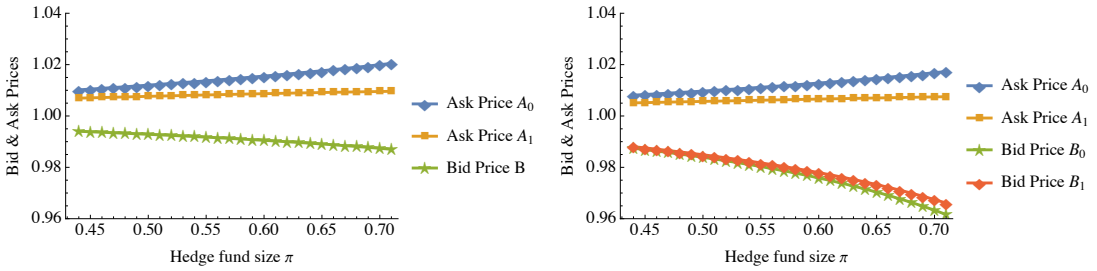
This impact on quotes in the stock market is illustrated in the bottom two panels of Figure 4. Panel (c) plots bid and ask quotes in the limit order market when securities lending is organized as an intermediated market structure, and Panel (d) considers the case where securities lending is centralized and provides quotes conditional on borrowing having occurred in the securities lending market. As noted above, in the case of centralized lending, hedge funds borrow the asset with positive probability even when they intend to buy the asset. However, despite the endogenous introduction of noise via this mixed strategy, bid prices are lowered conditional on shares borrowing having occurred, which harms hedge funds' profits from shorting. Indeed, Panel (d) shows that the conditional bid prices offered in a centralized securities lending market are substantially lower than the ones offered when shares lending is delegated (Panel (c)), in particular when market participants know that hedge funds can take larger positions relative to liquidity-motivated volume.

These insights apply also more generally when securities borrowing demand itself is potentially additionally motivated by liquidity and hedging motives. For example, in practice, a financial institution may borrow a security to offset an existing



(a) Fee yield across markets

(b) Hedge fund shorting profits across markets



(c) Limit orders with delegated securities lending

(d) Limit orders with centralized securities lending

Figure 4: The figure compares market outcomes for the cases of delegated securities lending and centralized securities lending. On the horizontal axis of each graph, we vary the relative size of the positions that hedge funds can take, π. Panels (a) and (b) plot fee yields and expected shorting profits of hedge funds. Panels (c) and (d) plot limit order market quotes in the delegated and centralized market, respectively. The subscripts on the quotes distinguish states with and without preexisting short interest (denoted as 1 and 0, respectively). The bid prices indicated in panel (d) are conditional on observing new borrowing, but differ by whether there is preexisting short interest or not. The parameters are given by $\lambda = 12$, $\sigma_v = 0.035$, $\rho_0 = 2.1$, $\mu_c = 0.042$, $\sigma_c = 0.2$, $\chi = 0.4$, and $r_f = 0.01$.

exposure for regulatory or risk management reasons. To capture this feature, we extend our analysis in Online Appendix C.5 to characterize equilibrium outcomes for the intermediated and centralized security lending markets when hedge funds additionally receive private value shocks that can generate non-informational motives for borrowing shares. While this setting features an additional source of liquidity trading, it reaffirms the prediction that hedge funds engaged in shorting are better off under an intermediated market structure when stocks are relatively less liquid. Finally, in our quantitative analysis in Section 6.5, we estimate a version of our model with a continuum of hedge funds. In this an environment, potential information leakage concerns are even stronger. This is because aggregate borrowing volume becomes

more revealing when noise shocks faced by individual hedge funds cancel out across funds.

In sum, these results consistently yield the prediction that agents borrowing securities tend to prefer the intermediated market structure for more illiquid securities, such as small stocks in practice. Moreover, investors lending their shares prefer the delegated market structure unconditionally, as it generates higher fee income. However, liquidity traders are negatively affected by an intermediated market structure, as they are effectively the source of surplus flowing to informed short sellers and security lenders.

6 Quantitative Analyses

In this section, we turn to the central quantitative questions of our paper: How do loan fees affect the stock valuations of security lenders? What alphas do short sellers earn across the cross-section of stocks? And how would these outcomes change in a counterfactual, competitive securities lending market?

Our solution for the price-dividend ratio (5) provides some initial intuition for the impact of fee revenues on security lenders' valuations. Let ψ denote the fee-to-dividend ratio. Rearranging (5) we obtain the following relationship between the price-dividend ratio that capitalizes fee income and the one that does not:

$$\text{P-D ratio with fee income} = \text{P-D ratio w/o fee income} \cdot (1+\psi)\,. \tag{23}$$

Equation (23) reveals that the fee-to-dividend ratio ψ is an essential statistic determining the price level effect of lending fee income and in fact, in our baseline model, ψ exactly measures this price-level effect, which we will also refer to as a *value wedge*.[18] Yet, more generally, the dynamic properties of fee income have to be recognized to accurately quantify these price-level effects in practice. After all, the less persistent lending fees are, the less does the current fee-to-dividend ratio matter. Consequently, as a first step, we now examine the dynamic properties of fee income in the data.

[18]This value wedge is closely related to the concept of the price wedges in van Binsbergen et al. (2023), which measure deviations of market prices from fundamental values.

6.1 Empirical Fee-Yield Dynamics

For our analysis, we are interested in both the cross-sectional and the time-series properties of lending fee income. Table 5 sheds light on the cross-sectional dispersion of lending fee income by revisiting our 25 groups of firms split based on size and fee yields. The table reveals large cross-sectional dispersion in fee yields for each size group, with a highly non-linear relation between fee yields and fee yield percentiles. This non-linearity motivated our choices for the percentile cutoffs. For the large-cap stocks in the Russell Top 200, the bottom 80% have an average fee yield less than 0.01%. In contrast, the top 2% have an average fee yield of 0.81%. Moreover, market capitalization has significant implications for fee yields: for nano-cap stocks, which are not included in the Russell indexes, the top 2% of stocks generate an average fee yield of 63.65%. For the small-cap and micro-cap stocks, this number is 8.51% and 28.96%, respectively. Table 6 further reveals significant differences in dividend yields across firms in the 25 groups (see "Data" columns) with a negative correlation between dividend yields and fee yields applying both in the fee and the size dimensions of our sorts.

If fee income and dividend income were perfectly persistent, we could immediately determine the price level implications of fees by computing the fee-to-dividend ratio, as shown in equation (23). However, as illustrated by the Markov transition matrix reported in Table A3 (see Online Appendix F), firms do migrate consistently across the 25 groups and correspondingly exhibit non-trivial fee yield dynamics. To gauge the pricing implications of fee income, it is therefore essential to account for the persistence of fee yields and their joint dynamics with dividend yields, which will be confirmed by our quantitative estimates. The Markov matrix reveals that the association with a given size group is highly persistent, whereas fee group association is persistent but less so. Moreover, persistence in fee group association varies materially by size.

Motivated by these observations, we proceed with generalizing our baseline model in order to account for these empirical dynamics. After estimating this model, we evaluate the quantitative questions of our paper.

6.2 Generalizing the Model

Corresponding to the 25 bins in our empirical analysis, we consider a 25-state Markov chain indexed by the state $s \in \Omega = \{1, ..., 25\}$. We denote the 25×25 Markov Generator matrix by Λ and by $M_t(s, s')$ a counting process keeping track of jumps from state s to s'. The capital evolution equation (previously (2)) now takes the form:

$$\frac{dc_t}{c_{t-}} = \mu_c dt + \sigma_c(s_{t-})dB_t + (e^{v_{t-}} - 1)dN_t + \sum_{s' \in \Omega} \left(e^{u(s') - u(s_{t-})} - 1 \right) dM_t(s_{t-}, s'), \tag{24}$$

where the exposure to aggregate risk $\sigma_c(s)$ and the volatility of private information, $\sigma_v(s)$, now varies with the Markov state s. Moreover, the generalized dynamics in (24) allow for capital c to jump by a log change $(u(s') - u(s))$ when the Markov state changes from s to s'. This implies that firms' expected dividend growth rates vary with the state s and take the form:

$$\mu_c + \sum_{s' \in \Omega} \left(e^{u(s') - u(s_{t-})} - 1 \right) \Lambda(s_{t-}, s'). \tag{25}$$

When considering a cross-section of a continuum of firms, this specification implies the desirable feature of a stationary size distribution around a common trend with growth rate μ_c.

To ensure that the model can quantitatively capture the empirically observed levels of inventory slackness, we additionally move from a specification with a single hedge fund per cohort to one with a continuum of hedge funds that obtain private signals. Upon an innovation $dN_t = 1$, each hedge fund i in a newly arriving cohort now obtains a noisy signal

$$w_i \equiv v + \varepsilon_i, \tag{26}$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon(s))$ and independent across hedge funds i. A hedge fund's decision-relevant type is now given by its signal w_i.

The cross-sectional dispersion in signals across hedge funds, as measured by the parameter σ_ε, implies a less volatile total borrowing demand relative to the baseline setup, where demand is binary (either 0 or π units). These different stochastic properties for the total borrowing demand ensure that when we relax the previously imposed parameter restriction that $\rho_0 > 1 + \pi$, the model can match the inventory

slackness levels observed in the data.[19] Further, the stability of inventory slackness implies that we no longer need to impose this parameter restriction to capture the empirical fact that utilization is virtually always below 100%.[20] We provide the details of the analysis of this setting in Online Appendix D.

Value wedges. For the purpose of determining valuation level effects of fee income, the key empirical metric regarding the shares lending market is the fee-to-dividend ratio $\psi(s)$ Our estimation directly pins down this ratio using the empirical counterpart: for each bin s, it is the ratio of the fee yield (Table 5) to the dividend yield (Table 6). As we will show below, the generalized version of our model can capture the empirical relation between dividend yields and fee yields both in terms of the cross-sectional distribution across the 25 bins and in terms of the persistence in a firm's assignment to any given bin (as captured by the Markov matrix in Table A3).

To determine the effect of lending fees on the level of equity valuations, we analyze the relation between two distinct present values:

$$P_t = \mathbb{E}\left[\int_t^\infty \frac{m_\tau}{m_t} c_\tau \cdot (1+\psi(s_t))\, dN_\tau\right]. \tag{27}$$

$$P_t^e = \mathbb{E}\left[\int_t^\infty \frac{m_\tau}{m_t} c_\tau dN_\tau\right]. \tag{28}$$

Whereas the valuation equation (27) capitalizes fee income, the valuation equation (28) does not. The valuation effect of fee income is then measured by the *value wedge*:

$$VW(s) \equiv \frac{P_t}{P_t^e} - 1. \tag{29}$$

This measure reflects value inflations in the sense that it captures the extra value that shares generate for shareholders due to lending fee income rather than due to firms'

[19] In the baseline model, where slackness is binary and stochastically switches between 100 percent and $1-\pi/(\rho_0-1)$, choosing parameters such that $\rho_0 < 1+\pi$ does not yield an increased average utilization relative to choosing $\rho_0 = 1+\pi$, since slackness is either 0% or 100% in either case.

[20] For example, for $\sigma_\varepsilon > 0$, consider the 98th percentile of the distribution describing the fraction of hedge funds borrowing in a given state s and denote it by $x_{98\%} \in (0,1)$. Then we can impose the less restrictive assumption that $\rho_0 > 1 + x_{98\%} \cdot \pi$ (rather than requiring our earlier assumption that $\rho_0 > 1+\pi$) while still maintaining the property that the supply virtually always is sufficient to meet the aggregate borrowing demand.

fundamental real productivity and associated dividend income. This value wedge does not, however, reflect irrationality. Rather, it arises because shares generate income not only through dividends but also through loan fees.

Valuations again scale with the level of capital c, yet now they also depend on the Markov state s. Let $\mathbf{P}(c)$ and $\mathbf{P}^e(c)$ denote the 25×1 vectors collecting the valuations $P(c,s)$ and $P^e(c,s)$ and let $\Lambda(s)$ be the s-th row of the Generator matrix Λ. Further, let $\mathbf{U}(s)$ denote a row vector collecting the values of $e^{(u(s')-u(s))}$ conditional on being in state s and ψ the vector of fee-to-dividend ratios. We obtain the following closed-form solutions for valuations with and without fee income.

Proposition 5. *The vectors of valuations with and without fee income have the solutions:*

$$\mathbf{P}(c) = -\lambda c \cdot (\Lambda \odot \mathbf{U} - diag(r_f + \sigma_c(s)\chi - \mu_c(s)))^{-1}(\mathbf{1} + \psi), \tag{30}$$

$$\mathbf{P}^e(c) = -\lambda c \cdot (\Lambda \odot \mathbf{U} - diag(r_f + \sigma_c(s)\chi - \mu_c(s)))^{-1}\mathbf{1}, \tag{31}$$

where $\odot$ denotes the Hadamard product.

Proof. See Online Appendix C. □

6.3 Estimating the Model

In this section, we discuss our estimation approach by detailing the moments in the data that identify the model's parameters. Our model estimation follows a modular approach. In a first step, we estimate the model taking the observed joint fee and dividend dynamics as given. This approach is sufficient to obtain estimates of the value added share lenders obtain from securities lending (that is, the above-discussed value wedges). In a second step, we estimate structural parameters that pin down endogenous fee dynamics and trading gains for short sellers. Estimates of these parameters are also required for our analyses comparing the intermediated securities lending market to a counterfactual centralized market.

We start by discussing the first module of the estimation approach. The key moments targeted by this estimation are listed in Tables 5, 6, and 7.

SDF parameters. We choose a price of aggregate risk $\chi = 0.40$, which is identified

by an equity premium of 8% and an equity market volatility of 20%. We pick r_f to match the historical average short-term rate of 4.3%.

Risk exposures. We directly match expected return estimates (not incorporating fee income) for firms in each bin as implied by the CAPM, which pins down the risk exposure parameters $\sigma_c(s)$ for each state s. In addition, we perform robustness tests using the Cahart 4-Factor model. The values of the risk premia across the 25 bins for these two models are listed in Table 7.

Custodian fee retention. Empirically, custodian lenders such as Fidelity, Schwab, and TIAA retain around 10% of the fee income they collect, whereas Vanguard retains only 2% (see Morningstar, 2018). In our analysis below, we set $\delta = 0.1$.[21]

Fee-to-dividend ratios. The fee-to-dividend parameters $\psi(s)$ are chosen directly to match the ratio of the fee yield (Table 5) relative to the dividend yield (Table 6) in each state. Below, we will further estimate the deep parameters determining these fee-to-dividend ratios in equilibrium.

Dividend dynamics. We choose the dividend trend growth parameter μ_c to match the general level of price-dividend ratios across all states and estimate the state-specific dividend growth process parameters $u(s)$ by targeting the dividend yields in the 25 stock groups. As shown in the "Model" columns of Table 6, the model-implied dividend yields $(1/\tilde{P}(s))$ closely match their data counterparts in all 25 states. We directly estimate the Generator matrix Λ by converting the monthly Markov transition matrix reported in Table A3 into its continuous-time counterpart.

We set $\lambda = 365$, which implies that on average, lending contracts mature after one day, consistent with the fact that security lending contracts in practice are typically overnight contracts. Next, we describe our approach for determining the additional structural parameters pinning down equilibrium loan fees and short sellers' profits.

[21] Our microfoundation in Appendix E endogenizes the equilibrium value of δ. However, this microfoundation does not aim to capture all the forces determining δ in practice. Therefore, for our estimation, we directly rely on the empirical measures of fee retention.

Joint estimation of π, σ_v, σ_ε, and ρ_0. We jointly estimate the parameters π, $\sigma_v(s)$, $\sigma_\varepsilon(s)$, and $\rho_0(s)$, by targeting several key moments in the data. In the following, we sequentially discuss the economic implications of these parameters and what moments in the data contribute to their identification.

In the model, the volatilities of learnable shocks $\sigma_v(s)$ and the signal-to-noise ratios $\frac{\sigma_v(s)}{\sigma_\varepsilon(s)}$ control how much private information potential short sellers obtain in each state s about a firm's future performance. Ceteris paribus, higher values of $\sigma_v(s)$ and $\frac{\sigma_v(s)}{\sigma_\varepsilon(s)}$ increase potential short sellers' willingness to pay, implying that observed fees are informative moments for these parameters. In our estimation, we match the empirically observed fees in each state while imposing the constraint that the annualized shock volatilities do not exceed estimates of idiosyncratic return volatility in each state (see Table A4).

Hedge funds' relative trading capacity π increases the quantity of privately informed trading in the model unconditionally (across all states s). In our joint estimation, we match the relative trading volumes of "traditional hedge funds" in the years 2010 to 2021 as documented by Martin and Wigglesworth (2021), amounting to values between 9% and 15%. The resultant estimate for π is 0.8.

The volatilities of hedge funds' signal noise shocks, $\sigma_\varepsilon(s)$ and the relative quantity of lendable inventories $\rho_0(s)$ are the primary levers that allow us to match the state-contingent inventory slackness documented in Table 4.[22] Choosing higher values of $\sigma_\varepsilon(s)$ reduces the volatility of the total borrowing demand from hedge funds conditional on a state s. The intuition for this effect can be illustrated with the following comparative statics: First, take the limiting case where the signal noise goes to zero $\sigma_\varepsilon \to 0$ and consider how the total borrowing demand evolves as a function of the shock v: the total borrowing demand is very sensitive to v around the threshold w_ϕ, and almost switches discretely between 0 and π. The overall distribution of the total borrowing demand is thus almost identical to the binary demand featured in the baseline model. Second, consider instead a parameterization with a higher value of σ_ε. In this case, in a larger vicinity of values v around the threshold w_ϕ, about 50% of hedge funds choose to short and 50% do not, because hedge funds' signal

[22] Recall that since the sum of hedge funds' and liquidity traders' maximum holdings is normalized to one, $\rho_0(s)$ represents a relative measure: it is the ratio of the total quantity of the asset that can be lent out or posted at the ask price to the total size of positions taken by hedge funds and liquidity traders.

realizations w_i are now additionally affected by the idiosyncratic noise shocks ε_i that are symmetrically distributed around zero. As a result, the total borrowing demand becomes less volatile and less sensitive to the realization of the shock v itself.

When estimating $\rho_0(s)$, we further impose the restriction that the lendable inventory is sufficient unless the shock v takes a value more than two standard deviations below zero (which happens with about 2% probability), that is,

$$\rho_0(s) > 1 + F_w(w_\phi | v = -2\sigma_v) \cdot \pi. \tag{32}$$

This lower-bound restriction on $\rho_0(s)$ limits the estimation's ability to use low inventory levels (low $\rho_0(s)$) as a means to match higher inventory utilizations and implies that $\sigma_\varepsilon(s)$ has to take sufficiently large values to ensure that empirically observed inventory slackness levels are matched by the estimated model.

It is useful to further expand on the overall intuition of the identification. Take as an example the group of small-cap stocks that empirically have a 12% fee per annum in the data (see Table 3). Rationally, only hedge funds expecting an abnormal return *exceeding* 12% (pre fee) should be willing to short the asset at such a high fee. The lender, in turn, must find it optimal to charge a fee of 12% in the first place. If hedge funds virtually never expected to generate a return exceeding 12% per annum, charging such a high fee would be suboptimal for a lender. Thus, to rationalize lenders charging a fee of 12%, they must anticipate that a sufficiently large number of hedge funds will be expecting a pre-fee alpha in excess of 12%. This, in turn, requires hedge funds to receive sufficiently precise information on a sufficiently volatile return component for the group of stocks with the highest fees. Moreover, since hedge funds are privately informed, the fees only partially extract the full alpha.

We report in Tables A5, A6, and A7 (see Online Appendix F) the parameter estimates for the volatilities of the learnable shocks, the signal-to-noise volatility ratios, and the lendable inventories. Interestingly, in high-fee states, the estimated annualized volatilities of the learnable shocks approach, and sometimes reach, the upper boundaries of our estimates of idiosyncratic return volatility. This result provides relevant information for the evaluation of the validity of the model's specification of lender market power. Our baseline model and the microfoundation of market power presented in Section 7.1 predict that lenders effectively have monopolistic market

power that is constrained by the informational advantage that hedge funds have (implying that hedge funds obtain information rents). A model specification that would instead give additional bargaining power to hedge funds would then require counterfactually high values of σ_v to still match the observed fee levels in the high-fee states. In other words, the volatility estimates suggests that the market power specification of our model is better aligned with the data in high-fee states than alternative assumptions that would move more bargaining power to hedge funds.

The estimates of the signal-to-noise volatility ratios indicate that each individual hedge fund's signal contains substantial noise, which in equilibrium causes only hedge funds with sufficiently strong signals (i.e., low realizations) to short the asset.

6.4 Value Added from Securities Lending

Equipped with the estimated model, we proceed with evaluating our first set of key quantitative questions: how does non-competitive fee income affect the value of shares in the cross-section and at the aggregate stock market level? Note that the presented estimates rely only on the first module of our estimation approach.

State-contingent value wedges $VW(s)$. Figure 5 plots the value wedges $VW(s)$ in all 25 states. To reiterate, these value wedges measure the incremental value that shares lenders assign to stocks, due to the fact that they collect lending fee income. If shares lenders are also marginal investors in these stocks, then realized transaction prices do reflect this additional value. The value wedges range from 1.40% for the large-cap stocks with the lowest fees (bottom 80%) to 90% for the group of highest-fee stocks with the smallest market capitalization (i.e., those not included in the Russell Indexes).

These results differ dramatically from those one would obtain when ignoring fee and dividend dynamics: if fee-to-dividend ratios did not exhibit time variation for firms in each given state, the value wedges would instead range from 0.14% to 54128%. Accounting for fee dynamics is thus of first-order importance for determining accurate value wedge estimates. Despite the significant persistence of group association indicated in the Markov matrix reported in Table A3, the documented transition rates play a critical role. While there is a 96% chance that a Large-Cap firm (i.e. a Russell Top 200 member) in the lowest fee yield group remains in that

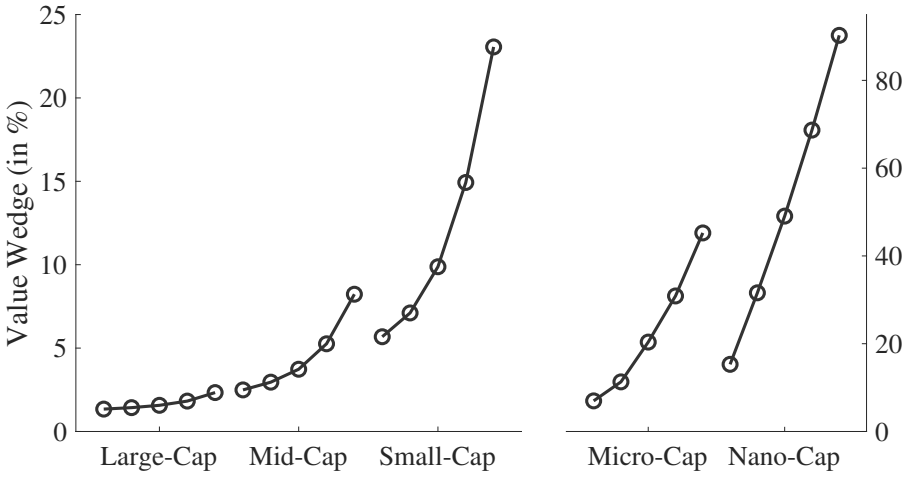


Figure 5: Value Wedges VW due to Lending Fee Income
We plot the additional percentage value that shares lenders assign to stocks due to fee income. For large-cap, mid-cap, and small-cap stocks, the left-hand side vertical axis applies; for micro-cap and nano-cap stocks, the right-hand side axis applies. If shares lenders are marginal investors, these are equilibrium value wedges over and above the prices that reflect the productive value of a firm. See Panel A of Table 8 for the numbers plotted in this graph.

group, there is a 3% chance of transitioning to the next-higher fee yield group in the Large-Cap group, which is quantitatively important. Valuations in our estimated model account for these transitions and therefore lead to significant value wedges even for those Large-Cap stocks that in their current state do not have high fee yields. These results imply that at the margin, shareholders not participating in shares lending (e.g., when investing directly in individual stocks and not making them available for lending) miss out on a significant fraction of value, in particular when investing in small stocks.

While the results in Figure 5 show the value added that shareholders obtain that actually participate in shares lending, one can also consider the following hypothetical: suppose fee income were distributed equally across all shareholders, even those that are not participating in shares lending. In this case fee yields are mechanically lower, since the same lending fee income is split across more shares. Panel B of Table 8 reports the value wedges for this counterfactual scenario. The value wedges then vary between 0.32% and 7.75%, which is still economically significant, especially for small stocks.

Aggregate value effects and their cross-sectional distribution. Next, we turn to evaluating the aggregate, market-wide value added due to fee income and the cross-sectional distribution of this incremental value. We find that the total value added from fee income, expressed as a fraction of the total market value of shares in inventory, is about 3%. That is, a representative shares lender obtains about 3% more value from its holdings than it would absent shares lending. Panel A of Table 9 further reports the distribution of that total value added across the 25 stock groups. Whereas small-cap high-fee stocks have high value wedges, they represent a relatively small fraction of the aggregate stock market capitalization. This is not surprising given that our highest-fee groups represent only 2% of the stocks in each size category. As a result, from an aggregate perspective, low-fee, large-cap stocks are still an important source of fee income. This fact is recognized by large institutional investment management companies such as Vanguard, which are able to offer particularly competitive fund management fees on index products due to significant securities lending fee income.

Finally, we again perform this evaluation under the counterfactual distribution of fee income whereby all shareholders obtain a pro-rata allocation of the total fee income. In this case, the total value added, expressed as a fraction of the total market value of all outstanding shares (not just the shares posted as inventory) is about 1%. The relative distribution of this value added across the 25 firm groups, which is reported in Panel B of Table 9, is quite similar to the one we obtained when considering only shares posted as inventory. Again, the groups of large-cap and low-fee stocks represent an important source of income at the aggregate level.

6.5 Comparing Market Outcomes Across Market Structures

Using the fully estimated model, we now quantitatively examine and compare equilibrium outcomes between the existing intermediated securities lending market and a counterfactual centralized market structure. In this analysis, we evaluate how both hedge funds and asset owners fare.

As highlighted in Section 5, a centralized market for securities lending increases transparency. In fact, in the quantitative setup we consider in this section, information leakages in the centralized market are amplified relative to our baseline setup in Section 5 because we now consider a continuum of hedge funds rather than just one

fund per cohort. While hedge funds' signals feature idiosyncratic noise, *aggregate* borrowing demand diversifies these shocks and represents an invertible function of the fundamental v absent additional frictions. In particular, depending on whether inventory is scarce in equilibrium, either the aggregate lending volume or the loan fee reveals the underlying state v. Perfect information aggregation would, however, eliminate profits from short selling, making hedge funds unwilling to pay any fee to borrow securities in the first place. In short, an issue similar to the Grossman-Stiglitz paradox (Grossman and Stiglitz, 1980) would obtain.

In practice, however, such perfect information aggregation is likely to be hindered by factors such as borrowing for non-informational reasons (see Appendix C.5), supply shocks, or general uncertainty about how prices and quantities relate to the fundamental v. To capture these frictions in a single parameter, we assume agents can only infer v with noise once the securities lending market clears; that is, they effectively observe the noisy signal:

$$\hat{v} \equiv v + b, \tag{33}$$

where $b \sim \mathcal{N}(0, \sigma_b(s))$. Frictions in information aggregation are then summarized by the volatility parameter $\sigma_b(s)$. We provide a detailed characterization of the equilibrium outcomes of this setting in Appendix D.2 and focus here, in the main text, on discussing the quantitative results.

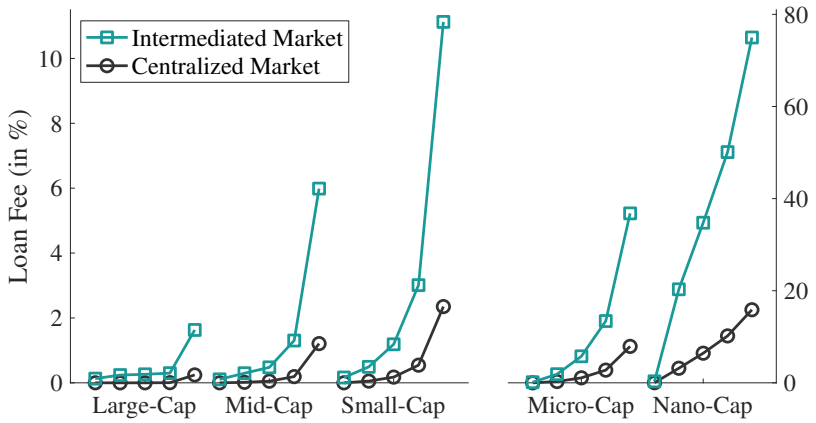
We start out by analyzing a setting where, in each state s, the noise volatility σ_b is half as large as the fundamental volatility σ_v; that is, $\sigma_b(s)/\sigma_v(s) = 0.5$. The estimated volatilities σ_v are listed in Table A5. For instance, the annualized volatility of fundamental shocks in the highest-fee groups among small-cap stocks is 61.11%, implying that the noise shocks in information aggregation are assumed to have an annualized volatility of 30.56%, which is substantial. Additionally, we perform sensitivity analyses by characterizing the ranges of values for the ratios $\sigma_b(s)/\sigma_v(s)$ (for each s) where the intermediated market dominates. For low enough values of the aggregation frictions summarized in σ_b, the intermediated market always dominates since, in the limit as $\sigma_b \to 0$, hedge funds' informational advantage relative to other market participants vanishes completely due to information leakage. As a result, both hedge funds' and asset owners' rents from the securities lending market approach zero in this limiting case.

Loan fees and fee yields. First, we examine how loan fees and fee yields vary across the intermediated and centralized market structures. Whereas loan fees represent the costs of shorting to hedge funds, fee yields measure the corresponding income flow that asset owners receive. Due to incomplete inventory utilization and fee retention by custodian lenders, fee yields are always lower than loan fees in the intermediated market. In contrast, in the centralized market, loan fees and fee yields are identical for any given stock group. This is because both measures are nil when utilization is below 100%, as competition drives fees to zero when inventory is not scarce. Further, once all inventory is lent out, fees become positive but the two measures remain identical, since each unit earns the full loan fee without any intermediary taking a cut.

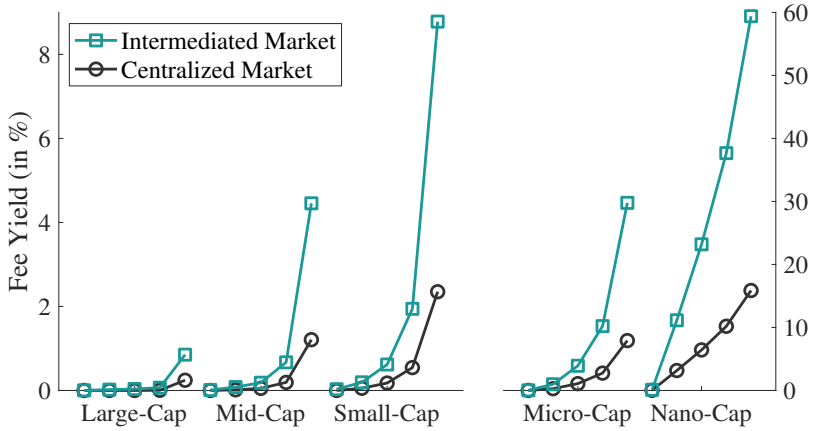
Figure 6 shows that both loan fees and fee yields are consistently lower in the centralized market. In fact, for a few stock groups in the large- and mid-cap size categories, fees are zero under the counterfactual centralized market. However, this is not the case for most stock groups because inventory utilization increases in the counterfactual economy. Whereas average inventory utilization is below 100% for all stock groups in the intermediated market, utilization increases to 100% for most stock groups in the centralized market equilibrium. Qualitatively, the variation in fees and fee yields across stock groups is similar across the two market structures, yet quantitatively, values are significantly lower in the centralized market. Whereas these changes are clearly not beneficial for asset owners, who are the recipients of fees, one might conjecture that hedge funds could benefit from the centralized market structure. We investigate this next.

Hedge funds' gains from shorting. Our qualitative analysis of the counterfactual centralized market structure in Section 5 indicated that hedge funds prefer an intermediated securities lending market when shorting sufficiently illiquid assets. We now examine this preference quantitatively across our 25 stock groups.

The panels of Figure 7 compare two measures of shorting profits across the intermediated market (blue squares) and the counterfactual centralized market (black circles). Panel (a) plots hedge funds' alphas from shorting securities net of loan fee costs. The results indicate that the intermediated market yields substantially higher net-of-fee alphas for stocks that are smaller and command higher loan fees. For ex-

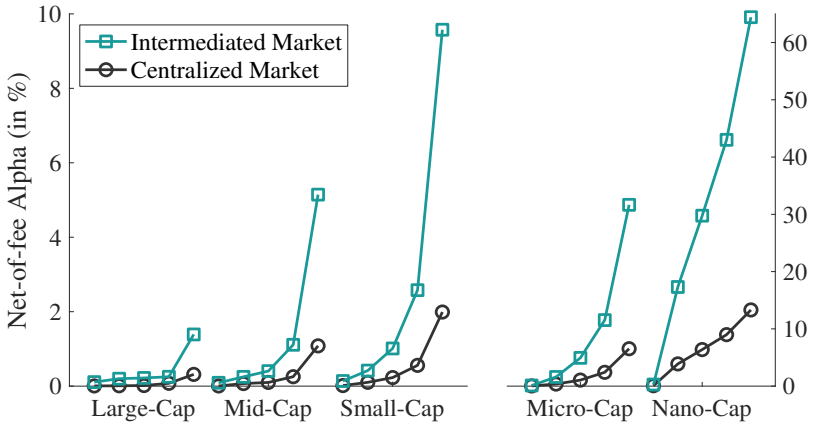


(a) Loan fees

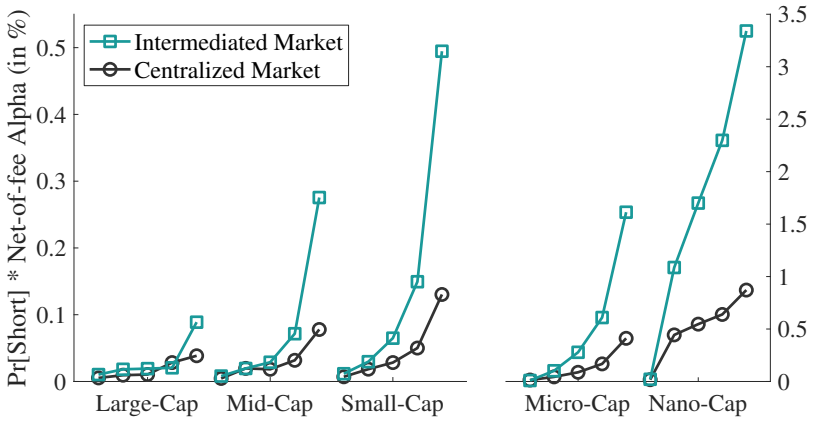


(b) Fee yields

Figure 6: Loan fees and fee yields.
The panels compare fees paid by hedge funds and fee yields earned by asset owners across the intermediated market (blue squares) and the counterfactual centralized market (black circles) for the 25 size and fee-yield groups. For large-cap, mid-cap, and small-cap stocks, the left-hand side vertical axis applies; for micro-cap and nano-cap stocks, the right-hand side axis applies. The analysis considers the case where shocks introducing noise in information aggregation have half the magnitude of fundamental shocks, $\sigma_b/\sigma_v = 1/2$.



(a) Hedge Funds' Net-of-Fee Alpha



(b) Hedge Funds' Shorting Surplus

Figure 7: Hedge funds' gains from shorting

The figure plots hedge funds' net-of-fee alphas from shorting (Panel (a)) and the product of their shorting propensity and net-of-fee alphas (Panel (b)). Each panel compares outcomes between the intermediated market (blue squares) and the counterfactual centralized market (black circles) across the 25 size and fee-yield groups. For large-cap, mid-cap, and small-cap stocks, the left-hand side vertical axis applies; for micro-cap and nano-cap stocks, the right-hand side axis applies.

ample, in the small-cap group, the alphas from shorting high-fee stocks reach around 10%, whereas they are around 2% in the centralized market. For micro-cap stocks, these numbers are 30% and 7%, respectively. On average across states, fees in the intermediated market account for 54% of the before-fee alpha. That is, fees substantially reduce short sellers abnormal returns.

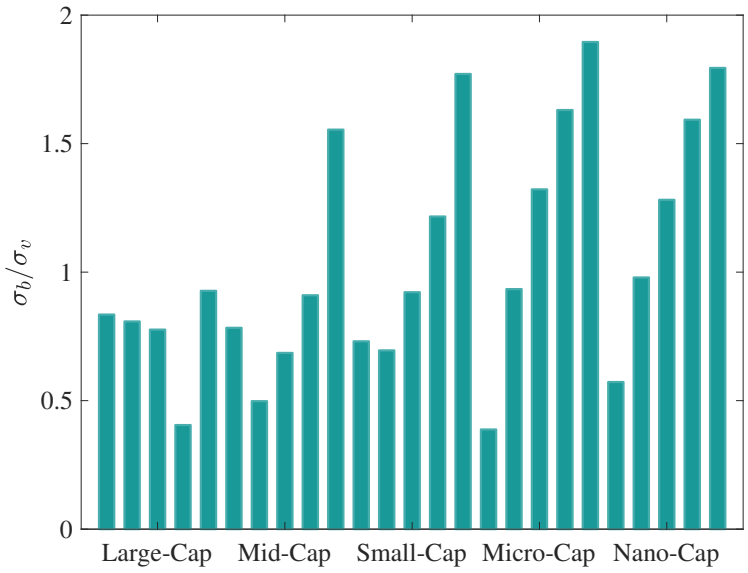


Figure 8: Ranges of aggregation noise where intermediation dominates for hedge funds. The figure plots the range of values $\sigma_b(s)/\sigma_v(s)$, as driven by changes in $\sigma_b(s)$, for which the intermediated securities lending market dominates based on the product of hedge funds' shorting propensity and the net-of-fee alpha.

Next, in Panel (b) of Figure 7, we account for both the net-of-fee alphas from shorting and the propensity to short. This latter measure is relevant if hedge funds cannot use their trading capacity for opportunities other than those considered in the model (in which case not trading leads to zero alpha). The results again indicate that the intermediated market dominates, particularly among smaller stocks with high fees, although the relative gap is somewhat reduced. This is because the frequency of shorting tends to be higher in the competitive centralized market, driven by the lower shorting fees that prevail there. As noted above, assuming that the volatility ratio of aggregation noise relative to fundamentals is one half ($\frac{\sigma_b(s)}{\sigma_v(s)} = \frac{1}{2}$) corresponds to very significant absolute aggregation noise levels, particularly for smaller stocks and high-fee stocks. This means that even with substantial aggregation noise, the

intermediated market dominates for hedge funds trading these stocks.

Further, in Figure 8, we move away from considering a particular value for the amount of noise in information aggregation and instead examine the highest levels of the noise volatility parameters $\sigma_b(s)$ at which hedge funds would still prefer an intermediated market. For this evaluation, we measure hedge fund gains based on the product of the shorting propensity and the net-of-fee alpha from shorting. The figure shows that for higher-fee stocks, the intermediated market tends to dominate across a wider range of $\sigma_b(s)/\sigma_v(s)$ values. This is particularly noteworthy given that the fundamental volatility $\sigma_v(s)$ increases monotonically with fee levels in each size group. For 21 stock groups, the volatility ratio can take values up to 0.7 without altering the conclusion that the intermediated market dominates. For the highest-fee groups among mid-cap, small-cap, micro-cap, and nano-cap stocks, these values exceed 1.5.

In all, these results suggest that under reasonable levels of information aggregation frictions, both asset owners and hedge funds fare worse under a counterfactual centralized market structure, particularly for smaller stocks that command higher fees.

7 Extensions and Discussions

In this section, we elaborate on an extension of our theoretical analysis that provides a foundation for market power and commitment by custodian lenders and discuss the implications of alternative market designs and the coexistence of options markets.

7.1 Custodian Market Power and Commitment

In our baseline model presented in Section 3, we considered a single monopolistic custodian intermediary that is committed not to opportunistically use the information revealed by hedge funds' borrowing demand. We showed both qualitatively and quantitatively in Sections 5 and 6.5 that due to concerns about information leakages, short sellers may prefer to interact with such a monopolistic lender rather than a centralized market with competitive lenders. Custodian commitment plays an important role in this analysis of the relative merits of intermediated securities lending. Absent

commitment, a custodian could opportunistically use the information it obtains from hedge funds' borrowing decisions. In particular, observing a hedge fund's borrowing demand, a custodian could front-run the hedge fund either by trading on its own book or by leaking information to affiliated traders and clients.

One may wonder how such custodian commitment is sustained and why custodians may command market power even when there is free entry to this line of business. After all, intermediation leads to a two-sided market structure where custodians interact with both beneficial asset owners and security borrowers. In this section, we address these questions based on an explicit repeated game analysis we provide in Online Appendix E. To economize on space, we relegate the detailed analysis to the appendix and discuss its main insights in this section.

The analysis shows that the very force that ensures commitment is directly linked to a custodian's ability to charge non-competitive fees. Commitment is sustained by custodians' concerns about reputation with hedge funds and the associated ability to attract future business, which is only profitable in the presence of non-competitive fees. A custodian has stronger incentives to stay committed to refraining from opportunistic behavior when future retained fee revenue is higher and the intermediary sector is most concentrated. This interaction generates endogenous scale economies: A custodian that attracts all asset owners' assets can stay committed while passing on the largest fraction of the fee revenue to them. In other words, overall, custodian agency rents are minimized, and investor rents maximized, when intermediation of securities lending is most concentrated.

This is not to say that competition is irrelevant. In fact, it is the threat that securities lending would move to another custodian in case of insufficient fee pass-through or opportunistic behavior that shapes this equilibrium outcome. Moreover, we show in Appendix E how limited percolation of information related to opportunistic behavior can lead to the existence of multiple custodians that each exercise market power in their respective relationship networks. Relatedly, in practice, hedge funds also have incentives to maintain a reputation with their security lenders that they are repeat clients. An intermediary lender is more likely to be committed not to opportunistically use the information of a specific hedge fund if this fund itself has established a reputation for sending its business exclusively to this custodian (rather than being known for shopping around). In sum, concerns about information leakages and

lenders' opportunistic behavior naturally generate an incentive for short sellers to interact primarily with one lender.

It is worth noting that in practice, the coordination on what custodians are the key intermediaries is facilitated by complementarities with the asset management business: Asset management and brokerage companies are natural suppliers of shares in the securities lending market, as they already have large and inventories of shares as a result of their core business. Moreover, even when asset managers like Vanguard pass on significant fractions of the security lending revenue to their fundholders, they benefit from generating this revenue as it increases their competitiveness in the mutual fund and ETF industry.[23]

7.2 Alternative Market Structures

In Section 5, we considered the most generic form of a competitive securities lending market, specifically a market that mirrors the prevalent limit order markets that are used for regular securities trading in practice: suppliers of lendable shares submitted their offers for loan contracts and market participants interested in borrowing shares competed for those offers. We highlighted that this centralized market structure leads to information leakages and can thereby harm short sellers' ability to profit from borrowing securities.

Alternatively, one might envision a novel market structure that is competitive while aiming to mitigate information leakage effects. Specifically, such a market structure for securities lending would need to avoid the public display of realized trades for a certain pre-specified time period, while at the same time allowing for the delivery of shares and clearing. Agents offering their shares for lending would post them with a platform and would not obtain information about whether their shares were actually lent out until the end of the trading day (in the context of our model, until clearing step 7). Only agents attempting to borrow would receive information about whether they were successful prior to the end of the day, which is necessary

[23] Companies in this space can extract rents from skills related to relative advantages in stock analysis, low-cost order execution and management, retail outreach and advertising etc., with different parts of the investor population being sensitive to different aspects. For example, Vanguard can offer particularly low-cost funds and still be profitable by taking advantage of its large scale and superior human capital and technology (including the ability to be an effective supplier of lendable securities, which generates extra securities lending income that allows Vanguard to offer particularly low-cost funds to investors).

to ensure that those agents know whether they can sell securities. However, upon settlement of those trades, parties would have to be informed about whether their shares were lent out or not, at which point information would leak to other market participants.

As explained in the previous subsection in the context of custodian lenders, commitment to secrecy is a more significant issue when security lending is competitive and does not generate profits. When introducing a platform that facilitates competitive shares lending, there would be no continuation value that could provide commitment. Having said that, new fintech developments such as smart contracts could potentially facilitate the implementation of such trading arrangements without the need to rely on an intermediary platform operator that are subject to commitment problems. However, even taking this for granted, it is useful to highlight implications and potential shortcomings of such a novel market structure.

First, liquidity externalities are important in this market, that is, platforms need sufficient inventory to be attractive for hedge funds and investors providing inventory. Yet any competitive platform that offers zero fees or low fees cannot attract inventory from investors when other players, such as custodian lenders, do offer substantial fee income. This is a major challenge for such alternative market structures.

Second, and relatedly, if regulators would simply mandate this market structure and assuming that a platform of this type could attract inventory despite offering zero fees to asset owners, retail investors would be on the losing side of such a regulation, whereas hedge funds might benefit (subject to additional caveats noted below). This is because under the current status quo large asset management companies acting as custodian lenders pass on substantial fractions of their securities lending income to their investors. As noted above, asset managers pass on most of their gross securities lending revenue to their fundholders, which are regular investors and pension investors. Moving to a competitive but intransparent securities lending market would move surplus from these regular investors to hedge funds. Such an outcome would potentially not be desirable from regulators' and policy makers' perspective.

Third and finally, implementing such a novel market structure might face additional complications as short sellers typically follow involved dynamic trading strategies that require secrecy for extended time periods of generally uncertain length. That is, a mechanical one-size-fits-all schedule of revealing securities lending volume at a

daily frequency might still harm those short sellers' profits, causing them to potentially prefer delegation to an intermediary that hides realized borrowing volume for longer time periods.[24]

In sum, there is a range of issues that limit the potential benefits and viability of a competitive, centralized market structure that aims to reduce information leakages by lowering the frequency with which transactions are publicly revealed.

7.3 Relation to Trading in Options Markets

In practice, agents wishing to bet against a security can do so not only by borrowing and selling the security but also by trading in options. For larger, more liquid stocks, options are often available and traded in centralized markets. One may wonder how this phenomenon relates to the main insights captured by our model.

If an options market were introduced in an environment like ours, hedge funds wishing to trade on negative information would use both venues (the options market and the securities lending market) in equilibrium, only if they were, at the margin, indifferent between these two venues. Indeed, Muravyev et al. (2022) find that empirically, options prices are consistent with the securities lending market and the options market being integrated in that securities loan fees are being passed on to buyers of options. A simple mechanism why such pricing may prevail in practice is the fact that options market makers face regulatory requirements to hedge their positions themselves by shorting the underlying securities and thereby incurring borrowing fees.

Since the buyer of a put option only needs to implement one trade (rather than a two-step transaction), it is the options market maker that not only faces these hedging costs but also adverse selection from informed traders. These adverse selection costs faced by options market makers are aggravated by the fact that the trades realized in the centralized options markets do become public information immediately, implying that market makers cannot borrow and sell the security before the prices of the underlying security have already changed in response to this public signal. Both these costs are then reflected in options quotes and make the options market less advantageous for agents wishing to bet against an underlying security.

However, this is not to say that only one market would be used in equilibrium.

[24] Indeed, in our data, short sellers roll over their lending contracts for about 40 days on average, but there is large variation around this average.

Rather, a key equilibriating force determining how much order flow from prospective short sellers would be directed toward the options market versus toward the securities lending market is the relative quantity of liquidity motivated trades across these two markets. If liquidity-motivated traders approach primarily the centralized market for the underlying security (rather than the options market), then more of the order flow aiming to bet against a security would be optimally implemented through a standard two-step shorting transaction that leverages secrecy and the ability to directly trade against these liquidity traders.

In sum, the presence of an options market does not take away from the relevance of the features of the securities lending market we highlight. Fundamentally, this is because (1) options market makers also need to hedge positions (that is, they cannot take unbounded naked short positions), and (2) informed traders have incentives to exploit the presence of liquidity-motivated trades in the market of the underlying stock. Thus, the securities lending market is not redundant and short sellers benefit from it being set up in a way that avoids information leakages due to the two-step nature of short sales, as highlighted by our analysis.

8 Conclusion

In this paper, we examine the presence of market power in the securities lending market and evaluate its impact on different groups of investors and valuations. We document high market concentration, non-competitive fees, and low inventory utilization in the cross-section of stocks. Motivated by this evidence, we develop a tractable dynamic model to shed light on the conditions under which the current market structure benefits informed traders and security lenders. Somewhat paradoxically, we find that informed traders such as hedge funds favor an intermediated market structure subject to market power, particularly when trading in illiquid securities, even though they are charged non-competitive fees. This is because intermediaries provide short sellers with protection against information leakages during position build-ups: when intermediated by a custodian lender, borrowing activity remains localized information within a bilateral borrower-lender relationship. In contrast, if securities lending takes place in a centralized venue, the resulting information leakage impedes traders' ability to profit from shorting.

Our model yields quantitative estimates of the impact of non-competitive fees on security lenders' stock valuations and the distribution of alphas that short sellers obtain from targeting different segments of the cross-section of stocks. The results reveal that market power in securities lending markets can have material spill-over effects on valuations in stock markets and the informational content of stock prices, with small stocks being especially impacted. These findings are particularly relevant in the context of the literature on financial market feedback effects, which argues that information aggregated by stock prices is an important input to firms' investment decisions in practice (see Bond et al., 2012; van Binsbergen and Opp, 2019). In addition to studying these real effects, future studies could shed more light on the welfare implications of the market structure of the securities lending market for those types of investors that trade in the stock market for liquidity or behavioral reasons.

References

Aggarwal, R., Saffi, P. A., Sturgess, J., 2015. The role of institutional investors in voting: Evidence from the securities lending market. The Journal of Finance 70, 2309–2346.

Allen, J., Wittwer, M., 2023a. Centralizing over-the-counter markets? Journal of Political Economy Forthcoming.

Allen, J., Wittwer, M., 2023b. Intermediary market power and capital constraints. Tech. rep., boston College Working Paper.

An, Y., Song, Z., 2023. Does the federal reserve obtain competitive and appropriate prices in monetary policy implementation? Review of Financial Studies 00, 1–45.

Atmaz, A., Basak, S., Ruan, F., 2019. Dynamic equilibrium with costly short-selling and lending market. Working paper.

Banerjee, S., Graveline, J. J., 2014. Trading in derivatives when the underlying is scarce. Journal of Financial Economics 111, 589–608.

Blocher, J., Reed, A. V., Van Wesp, E. D., 2013. Connecting two markets: An equilibrium framework for shorts, longs, and stock loans. Journal of Financial Economics 108, 302–322.

Bond, P., Edmans, A., Goldstein, I., 2012. The Real Effects of Financial Markets. Annual Review of Financial Economics 4, 339–360.

Cherkes, M., Jones, C. M., Spatt, C. S., 2013. A solution to the palm - 3com spin-off puzzles. Working paper.

Christoffersen, S. E., Geczy, C. C., Musto, D. K., Reed, A. V., 2007. Vote trading and information aggregation. The Journal of Finance 62, 2897–2929.

Diamond, D. W., Verrecchia, R. E., 1987. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics 18, 277–311.
Duffie, D., 1996. Special Repo Rates. Journal of Finance 51, 493–526.
Duffie, D., Garleanu, N., Pedersen, L. H., 2002. Securities lending, shorting, and pricing. Journal of Financial Economics 66, 307–339.
D'Avolio, G., 2002. The market for borrowing stock. Journal of Financial Economics 66, 271–306.
Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2018. Short-Selling Risk. Journal of Finance 73, 755–786.
Evgeniou, T., Hugonnier, J., Prieto, R., 2019. Costly short sales and nonlinear asset pricing. Working paper.
Fudenberg, D., Levine, D. I., Maskin, E., 1994. The Folk Theorem with Imperfect Public Information. Econometrica 62, 997–1039.
Garleanu, N. B., Panageas, S., Zheng, G. X., 2021. A long and a short leg make for a wobbly equilibrium. Working Paper 28824, National Bureau of Economic Research.
Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.
Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. The American Economic Review 70, 393–408.
Harrison, M. J., Kreps, D. M., 1978. Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations. Quarterly Journal of Economics 92, 323–336.
Hayek, F., 1945. The use of knowledge in society. American Economic Review 35, 519–530.
Holmstrom, B., Myerson, R. B., 1983. Efficient and Durable Decision Rules with Incomplete Information. Econometrica 51, 1799–1819.
Hong, H., Stein, J. C., 2003. Differences of Opinion, Short-Sales Constraints, and Market Crashes. Review of Financial Studies 16, 487–525.
Hortaçsu, A., McAdams, D., 2010. Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market. Journal of Political Economy 118, 833–865.
Huber, A. W., 2023. Market power in wholesale funding: A structural perspective from the triparty repo market. Journal of Financial Economics 149, 235–259.
Investment Company Act, 1940. Section18 (f).
ISLA, 2022. Securities lending market report. Market report.
Leland, H. E., 1998. Agency costs, risk management, and capital structure. The Journal of Finance 53, 1213–1243.
Markit, 2012. Equity buyside analytics data dictionary.
Martin, K., Wigglesworth, R., 2021. Rise of the retail army: the amateur traders transforming markets. https://www.ft.com/content/

7a91e3ea-b9ec-4611-9a03-a8dd3b8bddb5.
Miller, E. M., 1977. Risk, uncertainty, and divergence of opinion. Journal of Finance 32, 1151–1168.
Morningstar, 2018. Securities lending. an examination of the risks and rewards.
Muravyev, D., Pearson, N. D., Pollet, J. M., 2022. Anomalies and their short-sale costs. Working paper.
Nezafat, M., Schroder, M., 2021. Private Information, Securities Lending, and Asset Prices. The Review of Financial Studies 35, 1009–1063.
Nutz, M., Scheinkman, J. A., 2020. Shorting in Speculative Markets. Papers 1705.05882, arXiv.org.
Scheinkman, J. A., Xiong, W., 2003. Overconfidence and Speculative Bubbles. Journal of Political Economy 111, 1183–1220.
SEC, 1997. Response of the office of chief counsel division of investment management. https://www.sec.gov/divisions/investment/noaction/1997/brinsonfunds112597.pdf.
SEC, 2022a. Key Points About Regulation SHO. https://www.sec.gov/investor/pubs/regsho.htm.
SEC, 2022b. Securities lending by U.S. open-end and closed-end investment companies. https://www.sec.gov/divisions/investment/securities-lending-open-closed-end-investment-companies.htm.
Sikorskaya, T., 2024. Institutional investors, securities lending, and short-selling constraints, november 27, 2024.
Stambaugh, R. F., Yu, J., Yuan, Y., 2011. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.
van Binsbergen, J. H., Boons, M., Opp, C. C., Tamoni, A., 2023. Dynamic asset (mis)pricing: Build-up versus resolution anomalies. Journal of Financial Economics 147, 406–431.
van Binsbergen, J. H., Opp, C. C., 2019. Real anomalies. The Journal of Finance 74, 1659–1706.
Vayanos, D., Weill, P.-O., 2008. A Search-Based Theory of the On-the-Run Phenomenon. Journal of Finance 63, 1361–1398.

A Tables

Panel A. Market Share of the Lender with the Highest Value on Loan

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	37.09	37.82	39.73	40.78	64.06
80-90%	31.41	31.48	34.20	38.92	56.53
90-95%	30.01	30.26	31.70	37.38	55.79
95-98%	28.83	28.49	31.05	37.09	54.86
$\geq 98\%$	27.93	28.26	31.34	38.56	57.13

Panel B. Market Share of the Lender with the Second Highest Value on Loan

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	19.70	21.91	23.40	22.59	21.72
80-90%	19.13	21.49	21.75	21.08	22.94
90-95%	19.64	20.83	20.52	20.42	22.45
95-98%	19.93	19.57	19.57	20.68	22.93
$\geq 98\%$	20.24	18.02	20.26	20.70	22.64

Table 1: Concentration of Lenders' Value on Loan (Markit Data)
This table shows the concentration of lenders' value on loan for the 25 size and fee yield groups. For each stock we obtain the market share of value on loan of the top 2 lenders from Markit Securities Lending Database. We calculate the group-level market share of the top 2 lenders as

$$MarketShare_t^i = \frac{\sum_j ValueOnLoan_{j,t}^i \times MarketShare_{i,t}^i}{\sum_j ValueOnLoan_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $MarketShare_{j,t}^i$ is market share of the (second) largest lender and $ValueOnLoan_{j,t}^i$ is value of shares on loan for stock j in group i. Panel A reports the time-series average of market shares of the lender with the highest value on loan. Panel B reports the time-series average of market shares of the lender with the second highest value on loan. Market shares are reported in percent. The sample period is from 2007 through 2021.

Panel A. Market Share of the Lender with the Highest Value on Loan

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	30.14	30.51	32.79	33.21	51.42
80-90%	26.56	26.78	29.06	32.79	47.77
90-95%	25.46	25.83	26.90	31.07	46.51
95-98%	24.53	24.13	26.46	30.75	45.42
$\geq 98\%$	23.44	23.48	26.79	32.75	47.58

Panel B. Market Share of the Lender with the Second Highest Value on Loan

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	16.51	18.50	19.82	19.04	18.72
80-90%	16.18	18.24	18.45	18.09	19.28
90-95%	16.75	17.68	17.39	17.44	19.74
95-98%	16.82	16.63	16.51	17.73	19.82
$\geq 98\%$	16.18	15.61	17.49	17.87	19.77

Table 2: Concentration of Lenders' Value on Loan (Markit-coverage Adjusted)
The table reports the concentration of lenders' value on loan for the 25 size and fee yield groups, using a conservative measure that accounts for coverage of the Markit database. In a first step, for each stock, we obtain the top 2 lenders' market shares for value on loan as reported by the Markit Securities Lending Database. In a second step, we calculate the group-level market share of the top 2 lenders as the fee income-weighted average of market shares across stocks in the group,

$$MarketShare^i_t = \frac{\sum_j LendingFeeIncome^i_{j,t} \times MarketShare^i_{j,t}}{\sum_j LendingFeeIncome^i_{j,t}}, \quad i = 1,2,\ldots 25,$$

where $LendingFeeIncome^i_{j,t}$ is the sum of daily lending fee income of stock j in quarter t, and $MarketShare^i_{j,t}$ is the fee income-weighted average of daily market shares, multiplied by the Markit coverage of 85% (see Online Appendix A), for stock j in quarter t,

$$MarketShare^i_{j,t} = \frac{\sum_d MarketShare^i_{j,t,d} \times MarkitCoverage \times LendingFeeIncome^i_{j,t,d}}{\sum_d LendingFeeIncome^i_{j,t,d}}.$$

This measure is conservative in the sense that it assumes that 100% of the security lending activity not covered by Markit involves institutions that are distinct from the top 2 institutions identified by Markit.

Lending Fee (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.285	0.276	0.290	0.498	3.887
80-90%	0.287	0.305	0.487	2.188	18.223
90-95%	0.288	0.463	1.205	6.254	31.631
95-98%	0.313	1.263	3.150	14.340	48.157
$\geq 98\%$	2.204	6.880	12.242	38.425	75.141

Table 3: Lending Fee
This table reports the time-series average of lending fees for the 25 size and fee yield groups. The lending fee of group i in quarter t is computed as

$$LendingFee_t^i = \frac{\sum_j ValueOnLoan_{j,t}^i \times LendingFee_{i,t}^i}{\sum_j ValueOnLoan_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $LendingFee_{j,t}^i$ and $ValueOnLoan_{j,t}^i$ are lending fees and value of shares on loan for stock j in group i, respectively. Lending fees are annualized and reported in percent. The sample period is from 2007 through 2021.

Panel A. Excess Inventory (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	98.90	96.15	90.30	93.35	97.43
80-90%	93.31	74.04	63.06	53.91	52.62
90-95%	88.02	63.46	52.29	37.57	37.58
95-98%	79.87	53.13	41.33	27.89	27.99
$\geq 98\%$	64.69	37.29	29.34	21.30	18.08

Panel B. Excess Inventory – Alternative Measure (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	97.60	90.13	82.00	79.08	68.08
80-90%	91.88	72.57	60.66	47.37	45.15
90-95%	86.68	61.27	47.97	31.95	33.29
95-98%	78.07	48.59	35.44	23.88	24.81
$\geq 98\%$	47.69	25.60	21.08	17.70	15.17

Table 4: Excess Inventory
This table reports the time-series average of excess inventory for the 25 size and fee yield groups. In Panel A, the excess inventory of a group is computed as one minus the aggregate value on loan divided by the aggregate value of shares in inventory across stocks in the group,

$$ExcessInventory^i_t = 1 - \frac{\sum_j ValueOnLoan^i_{j,t}}{\sum_j ValueInventory^i_{j,t}}, \quad i = 1,2,\ldots 25.$$

In Panel B, the alternative measure of excess inventory of a group is computed as the fee income-weighted average of excess inventory across stocks in the group,

$$AlternativeExcessInventory^i_t = \frac{\sum_j ExcessInventory^i_{j,t} \times LendingFeeIncome^i_{j,t}}{\sum_j LendingFeeIncome^i_{j,t}}, \quad i = 1,2,\ldots 25,$$

where $LendingFeeIncome^i_{j,t}$ is the sum of daily lending fee incomes of stock j in quarter t, and $ExcessInventory^i_{j,t}$ is the fee income-weighted average of daily one minus utilizations of stock j in quarter t,

$$ExcessInventory^i_{j,t} = \frac{\sum_d (1 - Utilization^i_{j,t,d}) \times LendingFeeIncome^i_{j,t,d}}{\sum_d LendingFeeIncome^i_{j,t,d}}.$$

Excess inventory is reported in percent. The sample period is from 2007 through 2021.

Fee Yield (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.003	0.011	0.0295	0.028	0.100
80-90%	0.019	0.081	0.193	0.970	10.815
90-95%	0.035	0.184	0.616	3.873	22.803
95-98%	0.065	0.668	1.936	10.090	37.242
$\geq 98\%$	0.850	4.446	8.675	29.315	63.955

Table 5: Fee Yield
This table reports the time-series average of fee yields for the 25 size and fee yield groups. The fee yield of group i in quarter t is computed as

$$FeeYield_t^i = \frac{\sum_j LendingFeeIncome_{j,t}^i}{\sum_j InventoryValue_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $LendingFeeIncome_{j,t}^i$ is the amount of fee incomes collected by lenders from their shares on loan in quarter t and $InventoryValue_{j,t}^i$ is the value of lenders' shares in inventory. Fee yields are annualized and reported in percent. The sample period is from 2007 through 2021.

Dividend Yield (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)		Mid-Cap (2)		Small-Cap (3)		Micro-Cap (4)		Nano-Cap (5)	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
$\leq 80\%$	2.09	2.12	1.76	1.78	1.47	1.48	1.01	1.02	1.08	1.09
80-90%	2.27	2.27	1.84	1.86	0.99	0.99	1.16	1.16	0.33	0.30
90-95%	2.36	2.38	1.62	1.63	1.05	1.07	1.16	1.17	0.14	0.14
95-98%	2.12	2.13	1.81	1.83	1.59	1.60	0.71	0.71	0.07	0.06
$\geq 98\%$	2.31	2.20	1.96	1.96	1.29	1.30	0.33	0.33	0.11	0.08

Table 6: Dividend Yield (Data and Model)
This table reports dividend yields for the 25 size and fee yield groups. *Data* columns show the time-series average of groups' dividend yields. The dividend yield of group i in quarter t is computed as

$$DividendYield_t^i = \frac{\sum_j Dividend_{j,t}^i}{\sum_j MarketCapitalization_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $Dividend_{j,t}^i$ and $MarketCapitalization_{j,t}^i$ are dividend and market capitalization of stock j in group i, respectively. *Model* columns show the dividend yields calibrated from the model. The data of ordinary cash dividends are from CRSP. Dividend yields are annualized and in percentage. The sample period is from 2007 through 2021.

Panel A. CAPM Risk Premia (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	6.57	7.65	8.12	8.01	7.60
80-90%	7.34	8.44	9.44	10.25	8.68
90-95%	7.93	9.18	10.18	12.40	10.26
95-98%	9.18	10.70	10.16	10.89	8.92
$\geq 98\%$	9.87	11.63	9.97	11.75	10.40

Panel B. Carhart 4-Factor Risk Premia (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	6.47	7.73	9.74	7.19	6.73
80-90%	6.78	7.99	9.92	9.65	7.66
90-95%	7.45	9.14	11.53	9.45	9.79
95-98%	6.51	8.88	11.29	8.84	4.69
$\geq 98\%$	6.94	9.90	11.46	9.43	8.27

Table 7: Risk Premia
The table reports estimated risk premia, which are computed as the product of factor premia and a group's factor loadings. Loadings are estimated using monthly value-weighted returns from 2007 to 2021. Factor premia are estimated as the time-series average of factor returns from 1972 to 2021. Panel A considers the CAPM and Panel B the Carhart 4-factor model.

Panel A. Value Wedges due to Lending Fee Income (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	1.35	2.50	5.71	7.03	15.22
80-90%	1.44	2.97	7.13	11.08	31.24
90-95%	1.58	3.75	9.86	19.77	48.92
95-98%	1.83	5.27	14.84	30.45	68.90
$\geq 98\%$	2.35	8.25	22.82	44.35	90.07

Panel B. Value Wedges under Pro-rata Allocation of Fee Income (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.32	0.61	1.16	1.25	2.12
80-90%	0.35	0.72	1.42	1.84	3.45
90-95%	0.38	0.89	1.91	3.01	4.86
95-98%	0.43	1.23	2.82	4.19	6.24
$\geq 98\%$	0.55	1.85	4.21	5.63	7.75

Table 8: Value Wedges
The table reports value wedges for the 25 stock groups. Panel A reports the incremental value shares lenders assign to stocks because of lending fee income. Panel B reports value wedges under a counterfactual pro-rata allocation of lending fee income across all shares outstanding.

Panel A. Distribution of Aggregate Value Added from Fee Income

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	38.83	25.51	11.27	15.47	0.88
80-90%	1.21	1.16	0.88	0.41	0.04
90-95%	0.52	0.58	0.47	0.27	0.02
95-98%	0.30	0.39	0.36	0.18	0.02
$\geq 98\%$	0.27	0.43	0.32	0.18	0.02

Panel B. Distribution of Aggregate Value Added under Pro-rata Allocation of Fee Income

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	29.02	21.44	9.95	21.00	0.59
80-90%	2.11	2.41	1.49	1.55	0.11
90-95%	1.12	1.32	0.90	0.74	0.07
95-98%	0.73	1.00	0.78	0.60	0.06
$\geq 98\%$	0.71	0.95	0.78	0.52	0.05

Table 9: Distribution of Total Value Added from Lending Fee Income
The table reports the percentage of total value added from lending fee income across the 25 size and fee yield groups. The reported numbers are computed as the time-series averages of the percentages for each group. In panel A, the value added of a group is computed as its value wedge times the total market value of lenders' inventory. In panel B, this metric is computed under a counterfactual pro-rata allocation of fee income.

Short Sellers' Alphas (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.11	0.09	0.14	0.11	0.28
80-90%	0.20	0.25	0.42	1.60	17.31
90-95%	0.22	0.40	1.01	4.94	29.77
95-98%	0.25	1.11	2.58	11.53	43.01
$\geq 98\%$	1.39	5.14	9.57	31.66	64.43

Table 10: Short Sellers' Alphas
This table reports the model-implied expected excess returns short sellers obtain net of fees in each group i per annum. All numbers are annualized and reported in percent.

Online Appendix

A Data Appendix

Our sample of U.S. stocks is constructed by combining CRSP data with the FSTE Russell index membership data, as well as securities lending data from Markit. The sample period is from 2007 through 2021, during which Markit has a good coverage of CRSP stocks.

Information about prices, returns, and dividends are from CRSP. For dividend payments, we consider ordinary cash dividends paid in US dollars (CRSP distribution codes starting with 12). Because we are interested in all cash dividends that investors can earn by holding shares, we include dividend payments at any frequency.[1] As our analysis is at the quarterly frequency, we calculate the quarterly dividend of a stock as the sum of all dividends distributed in a quarter.

Our securities lending data are from Markit Securities Finance Data Analytics (formerly Data Explorers). Markit provides a variety of lending indicators, from which we select data fields that are necessary for calculating lending fee income and fee yield earned by shares lenders: shares lending supply, shares borrowing demand, and borrowing costs. Specifically, we use Beneficial Owner's Inventory Value in Markit to measure the dollar amount of shares lending supply, and Value on Loan to measure the dollar amount of shares borrowing demand. Markit provides three important fields on borrowing costs: Daily Cost of Borrow Score (DCBS), Indicative Fee, and Simple Average Fee. The DCBS is a 1–10 categorization that describes how expensive a stock is to borrow, with 1 being the cheapest and 10 being the most expensive. Markit computes DCBS based on the proprietary data of actual lending fee quotes which are received from securities dealers but are not allowed to re-distribute. Indicative Fee is a derived rate using the Markit's proprietary analytics and dataset of both contributed borrowing costs between Agent Lenders and Prime Brokers as well as contributed rates from hedge fund participants. However, Indicative Fee can not serve the our purpose of computing lending fee income because it is not the *actual* rate Prime Brokers quote or charge but rather the *expected* rate for investors such as hedge fund to borrow shares in securities lending markets. Simple Average Fee (we will call it Fee later) is the equal-weighted average of actual fees across all outstanding loan contracts on a stock. Although Fee is suitable for calculating lending fee income, its data histories are incomplete in Markit. Actually, it is not unusual to have a missing Fee for a stock-day with positive shares on loan, especially during the early sample period.

We propose a method of filling missing Fee based on three important data fea-

[1] 91.1% of all dividend observations are quarterly, 1.3% of dividends are semi-annual, 0.2% are annual, and the rest are of other, unknown, or missing frequency.

tures: 1) Fee observations of stocks with the same DCBS tend to have very similar magnitudes; 2) Fee and Indicative Fee are highly correlated with each other (the correlation is about 0.8); 3) both DCBS and Indicative Fee are well populated in Markit data. These features suggest that we can fill a missing Fee of a stock on a given day with the average of non-missing Fees of stocks with similar Indicative Fees and in the same DCBS category. Specifically, we implement our method in three steps: 1) sorting stocks with the same DCBS on a given day into 10 bins by their Indicative Fees; 2) taking the average of non-missing Fees for each combination of DCBS and Indicative Fee bin; 3) filling a missing Fee with the average Fee of stocks in the matched combination. To check the validity of our method, we simulate the "filled" Fees for non-missing Fee observations and we find the correlation between the "filled" Fees and actual Fees above 0.9 for all DCBS categories.

With lending fees data, we calculate one of our key variables, lending fee income, as value on loan times lending fees on a day. Similar to the quarterly dividend, the quarterly fee income is the sum of daily fee incomes over a quarter.

For some of our empirical market concentration measures it is useful to make adjustments for the coverage that the Markit Database has of overall securities lending market activity. Markit reports that its securities lending data is collected throughout the day from more than 85% of global securities finance practitioners, including custodial banks, agent lenders, sell-side brokers, asset managers and hedge funds (see Markit (2012)). Correspondingly, when making adjustments for Markit coverage, we use the reported 85% (as in Muravyev et al., 2022). Moreover, we confirm that Markit covers the vast majority of US stocks; the percentage of the number of stocks covered by Markit increases from 85% to 98% over the sample period from 2007 to 2021.[2]

A.1 Sorting Procedure

The Russell Indexes have become the leading US equity benchmarks and have been widely accepted by institutional investors for their academic integrity and investor usability. One key feature of the Russell indexes is that they are reconstituted purely based on the rank of market capitalization of stocks. Russell Top 200, Russell Mid-Cap, Russell 1000, Russell 2000, and Russell Micro-Cap are commonly used indexes among practitioners. In total, the Russell universe includes 4000 stocks. In our empirical analysis, we split the Russell universe into four mutually exclusive size categories (Large-Cap, Mid-Cap, Small-Cap, and Micro-Cap) as described below. The remaining US stocks are too small to be included in the Russell universe so we put them into an additional category (Nano-Cap). To be consistent with the Russell Index constituents, we consider shares listed on all US exchanges, including not only common shares of US companies (CRSP share code 10 or 11), but also shares of

[2]We merge CRSP and Markit using historical CUSIPs, which are available in both databases.

companies incorporated outside the United States (share code 12) and REITs (share code 18).

We form 25 size and fee yield groups by the following conditional sorting procedure. At the beginning of each quarter, we first classify stocks into the 5 size categories as follows

- Large-Cap: stocks in Russell Top 200, which consists of the largest 200 members in Russell 1000
- Mid-Cap: stocks in Russell Mid-Cap, which consists of the smallest 800 members in Russell 1000
- Small-Cap: the largest 1000 members in Russell 2000
- Micro-Cap: stocks in Russell Micro-Cap, which consists of the smallest 1000 members in Russell 2000 plus the largest 1000 stocks outside Russell 2000
- Nano-Cap: all remaining stocks

For each of the above size groups, we further sort stocks into five bins by the cutoffs of fee yield percentiles, 80%, 90%, 95%, and 98%. We select these cutoffs because of the strong right-skewness of fee yields in the cross-section.

B Timeline in the Centralized Lending Market

In the setting with a centralized lending market, the logical order of events upon an innovation $dN_t = 1$ is generally identical to the one we laid out for the delegated market, with differences only pertaining to steps 2 and 4:

1. Dividend income realized at time t is collected by the agents who were the owners of the asset at time $t-$ (that is, t is an ex-dividend date).
2. Asset owners choose whether to directly post fees for lendable shares or to make their units available for sale in the limit-order market. Posted shares are tied up until settlement (step 7). Loan agreements start at time t and mature at the time of the next Poisson shock $dN_\tau = 1$ with $\tau > t$. At maturity, a borrower has to return the underlying asset and its dividend at that time to the lender.
3. A new hedge fund and a new cohort of liquidity investors arrive in the market.
4. The hedge fund can pick up fee quotes for lendable shares. The realized volume of borrowed shares becomes publicly observable.
5. Bid and ask quotes are posted by market participants in the limit order market.
6. Investors submit quantities they wish to buy or sell in the limit-order market (market orders). Clearing starts with the best bid and ask offers and follows a pro-rata allocation rule.

7. Contracts are settled via delivery of the asset in accordance with all lending contracts and trades (steps 2 to 6 above).

C Analyses and Proofs Omitted from Main Text

C.1 Equilibrium Depth of Ask Quotes

We verify the conjectured equilibrium in which, in each period, the units of assets posted for sale at ask prices are equal to the maximum possible equilibrium market-order demand in the limit-order market. Any residual asset supply is made available for lending, either via an intermediary lender or in a centralized securities lending market.

For concreteness, consider the case where the maximum possible equilibrium demand via market orders is $(1+\pi)$ units. If strictly more than $(1+\pi)$ units were offered for sale in the limit-order market, the market would exhibit strictly positive excess supply in all states of the world. To ensure that market makers would still receive sufficient commission income (relative to the securities lending option), the limit-order market platform would need to charge buyers higher commissions. Such inefficiency could be eliminated by another platform offering better terms–namely, lower commissions–by capping supply at the maximum possible demand. As such, it is efficient for a platform to impose a cap on supply at $(1+\pi)$ units.

In contrast, if strictly fewer than $(1+\pi)$ units were supplied for sale on the limit-order platform, there would be a strictly positive probability that the maximum possible demand could not be met in equilibrium (at the trading stage, step 6 above). Given the pro-rata allocation rule, this would imply that, with strictly positive probability, liquidity traders would be unable to acquire their required units. Sellers in step 6 could then follow a strategy of skimming leftover liquidity traders at a price of ∞, earning unbounded expected payoffs (a positive probability of infinite revenue). This outcome would incentivize more asset owners to join the platform as sellers in the first place (step 2).

Finally, if the total supply on the platform is exactly $(1+\pi)$, the break-even ask price constitutes an equilibrium. A single atomistic supplier who deviates by setting a higher price would sell nothing, because the remaining $(1+\pi)$ units of supply can meet the entire $(1+\pi)$ demand at the break-even price. Conversely, deviating to lower ask prices than the break-even price would lead to losses. In sum, the total units available for sale on the limit-order platform will match the maximum possible demand.

C.2 Proof of Proposition 1

The buy-and-hold value $P(c)$ solves the Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} 0 = &- r_f P(c) + P'(c) c \mu_c + \frac{1}{2} P''(c) c^2 \sigma_c^2 - P'(c) c \sigma_c \chi \\ &+ \lambda \cdot \left(\mathbb{E}\left[c \cdot e^v + P\left(c \cdot e^v\right) + F_v(v_\phi) \frac{\pi}{\rho} \phi\left(c \cdot e^v\right) \right] - P(c) \right). \end{aligned} \tag{1}$$

We conjecture and verify that $P(c)$ and the fee $\phi(c)$ are linear functions of c and we define $\tilde{P} \equiv \frac{P}{\lambda c}$ and $\tilde{\phi} \equiv \frac{\phi}{\lambda c}$. Plugging $P(c) = \tilde{P} \lambda c$ and $\phi(c) = \tilde{\phi} \lambda c$ into the HJB equation yields:

$$0 = - r_f \tilde{P} \lambda c + \tilde{P} \lambda c \mu_c - \tilde{P} \lambda c \sigma_c \chi + \lambda \cdot \left(c + \tilde{P} \lambda c + F_v(v_\phi) \frac{\pi}{\rho} \tilde{\phi} \lambda c e^v - \tilde{P} \lambda c \right), \tag{2}$$

where we use the fact that $\mathbb{E}[e^v] = 1$. We obtain:

$$\tilde{P} = \frac{1 + F_v(v_\phi) \frac{\pi}{\rho} \tilde{\phi} \lambda}{r_f + \sigma_c \chi - \mu_c}, \tag{3}$$

or equivalently:

$$P(c) = \frac{\lambda \cdot \left(c + F_v(v_\phi) \frac{\pi}{\rho} \phi(c) \right)}{r_f + \sigma_c \chi - \mu_c}. \tag{4}$$

Next, we compute the buy-and-hold value given the information set of an informed hedge fund, that is, conditional on observing the current value of v_t. Let $Q(c, v)$ denote this value which solves the HJB equation:

$$\begin{aligned} 0 = &- r_f Q(c,v) + \frac{\partial Q(c,v)}{\partial c} c \mu_c + \frac{1}{2} \frac{\partial^2 Q(c,v)}{\partial c^2} c^2 \sigma_c^2 - \frac{\partial Q(c,v)}{\partial c} c \sigma_c \chi \\ &+ \lambda \cdot \left(c \cdot e^v + P\left(c \cdot e^v\right) + F_v(v_\phi) \frac{\pi}{\rho} \phi\left(c \cdot e^v\right) - Q(c,v) \right). \end{aligned} \tag{5}$$

Conjecturing the solution $Q(v, c) = e^v \lambda c \tilde{P}$ and substituting it together with $P(c \cdot e^v) = e^v \lambda c \tilde{P}$ and $\phi\left(c \cdot e^v\right) = e^v \lambda c \tilde{\phi}$ into this HJB equation yields:

$$\begin{aligned} 0 = &- r_f e^v \lambda c \tilde{P} + e^v \lambda \tilde{P} c \mu_c - e^v \lambda \tilde{P} c \sigma_c \chi \\ &+ \lambda \cdot \left(c \cdot e^v + e^v \lambda c \tilde{P} + F_v(v_\phi) \frac{\pi}{\rho} e^v \lambda c \tilde{\phi} - e^v \lambda c \tilde{P} \right), \end{aligned} \tag{6}$$

which is consistent with the earlier solution for $P(c)$ (see equation (4)). Thus, $Q(c,v) = P(c)e^v$.

C.3 Proof of Proposition 2

The liability value $L(v,c)$ solves the HJB equation:

$$0 = -(r_f + \sigma_c\chi - \mu_c)\cdot L(c,v) + \lambda(ce^v + A_1(c,v) - L(c,v)). \tag{7}$$

We conjecture that $L(v,c) = \lambda c\cdot\tilde{L}(v)$ and $A_1(c,v) = \lambda c\cdot e^v\tilde{A}_1$. Here we use the fact that the hedge fund observing v_{t-} knows that the scaled buy-and-hold value is $\tilde{Q}(v_{t-}) = \tilde{P}e^{v_{t-}}$ and similarly, the scaled ask price will be $\tilde{A}_1 e^{v_{t-}}$. Plugging in these conjectures, the HJB equation takes the form:

$$0 = -(r_f + \sigma_c\chi - \mu_c)\cdot\lambda c\cdot\tilde{L}(v) + \lambda(ce^v + \lambda c\cdot e^v\tilde{A}_1 - \lambda c\cdot\tilde{L}(v)). \tag{8}$$

which simplifies to:

$$\tilde{L}(v) = e^v\frac{1+\lambda\tilde{A}_1}{r_f + \sigma_c\chi - \mu_c + \lambda}, \tag{9}$$

or equivalently:

$$L(c,v) = e^v\frac{\lambda\cdot(c + A_1(c))}{r_f + \sigma_c\chi - \mu_c + \lambda}. \tag{10}$$

C.4 Proof of Proposition 3

The custodian knows that if it charges a loan fee $\phi(c) = \tilde{\phi}\lambda c$, all hedge fund types v satisfying

$$\tilde{\phi} \le \tilde{B} - \tilde{L}(v_\phi) \tag{11}$$

will accept. Since borrowing hedge funds have to deliver both the dividend and the shares upon maturity, a shares lender does not lose any of the cash flows from the asset. The shares lender's scaled expected payoff from extra fee income when

quoting a scaled loan fee $\tilde{\phi} > 0$ is then given by:

$$\begin{aligned}
&\pi \Pr[B(c) - L(c, v_\phi) > \phi(c)] \cdot \phi(c) && (12)\\
=&\pi \Pr\left[B(c) - e^{v} \frac{\lambda(c + A_1(c))}{r_f + \sigma_c \chi - \mu_c + \lambda} > \phi(c)\right] \cdot \phi(c)\\
=&\pi \Pr\left[v < \log\left(\frac{(B(c) - \phi(c))(r_f + \sigma_c \chi - \mu_c + \lambda)}{\lambda(c + A_1(c))}\right)\right] \cdot \phi(c). && (13)
\end{aligned}$$

Using the solutions, $A_1(c) = \lambda c \tilde{A}_1$, $B(c) = \lambda c \tilde{B}$, and $\phi(c) = \lambda c \tilde{\phi}$, we define the marginal hedge fund type v_ϕ (that is, the above-mentioned threshold value) choosing to borrow the asset:

$$v_\phi \equiv \log\left(\frac{(\tilde{B} - \tilde{\phi})(r_f + \sigma_c \chi - \mu_c + \lambda)}{1 + \lambda \tilde{A}_1}\right), \tag{14}$$

so that:

$$\tilde{\phi} = \tilde{B} - e^{v_\phi} \frac{1 + \lambda \tilde{A}_1}{r_f + \sigma_c \chi - \mu_c + \lambda}. \tag{15}$$

Using this definition, we can express the shares lender's profit as a function of the marginal hedge fund type v_ϕ:

$$\tilde{\Pi}(v_\phi) = \pi \cdot F_v(v_\phi) \cdot \left(\tilde{B} - e^{v_\phi} \frac{1 + \lambda \tilde{A}_1}{r_f + \sigma_c \chi - \mu_c + \lambda}\right). \tag{16}$$

The shares lender's marginal profit of increasing v_ϕ is

$$\begin{aligned}
\tilde{\Pi}'(v_\phi) =&\pi \cdot f_v(v_\phi) \left(\tilde{B} - \frac{1 + \lambda \tilde{A}_1}{r_f + \sigma_c \chi - \mu_c + \lambda} e^{v_\phi}\right)\\
&- \pi \cdot F_v(v_\phi) \cdot \frac{1 + \lambda \tilde{A}_1}{r_f + \sigma_c \chi - \mu_c + \lambda} e^{v_\phi}. && (17)
\end{aligned}$$

At the optimum, the marginal profit is equal to zero, that is, v_ϕ solves $\Pi'(v_\phi) = 0$, or equivalently:

$$\frac{f_v(v_\phi)}{F_v(v_\phi)} \left(\frac{\tilde{B} \cdot (r_f + \sigma_c \chi - \mu_c + \lambda)}{1 + \lambda \tilde{A}_1} e^{-v_\phi} - 1\right) = 1, \tag{18}$$

which can be rewritten as:

$$\frac{f_v(v_\phi)}{F_v(v_\phi)} \cdot \left(\frac{B(c)}{L(c, v_\phi)} - 1 \right) = 1. \tag{19}$$

C.5 Generalized Borrowing Motives and Proof of Proposition 4

In this appendix section, we generalize our baseline model to incorporate private-value shocks to hedge funds, and characterize in this more general environment the equilibria for both the delegated and centralized market structures (this generalized setting nests the baseline model).

In practice, some institutions participating in the shares lending market may do so for non-informational reasons. For example, a financial institution may wish to offset an existing exposure for regulatory or risk management reasons. To capture this feature, we extend our baseline model in this section by introducing private-value shocks affecting hedge funds.

Suppose hedge funds not only obtain information upon entry but also are subject to private-value shocks that imply that they apply an additional discount factor e^b to the future cash flows associated with the asset, where $b \sim \mathcal{N}(-\frac{\sigma_b^2}{2}, \sigma_b)$. When $b > 0$ hedge funds assign a private-value premium, otherwise a discount (for $b < 0$) to the cash flows. Apart from this private-value component, hedge funds still also observe the common value component v. A hedge fund's decision-relevant type is now given by $w \equiv v + b$, where $w \sim \mathcal{N}(-\frac{\sigma_v^2+\sigma_b^2}{2}, \sqrt{\sigma_v^2 + \sigma_b^2})$. Similar to before, there is a threshold type, which we denote by w_ϕ, that corresponds to the hedge fund type that is just indifferent between borrowing and not borrowing at the posted fee ϕ.

When buying the asset upon entry, a hedge fund now assigns the value

$$Q(c, w) = e^w \cdot P(c), \tag{20}$$

where $P(c)$ mimics our earlier solution in the baseline model (compare equation (5)):

$$P(c) = \frac{\lambda \cdot (c + F_w(w_\phi)\frac{\pi}{\rho}\phi(c))}{r_f + \sigma_c \chi - \mu_c}. \tag{21}$$

Analogously, when taking a short position upon entry, a hedge fund now assigns the following value to the cash flows associated with its liability:

$$L(c, w) = e^w \frac{\lambda \cdot (c + A_1(c))}{r_f + \sigma_c \chi - \mu_c + \lambda}. \tag{22}$$

which again mimics the solution in our baseline model (compare equation (17)). A hedge fund's maximum willingness to pay is then given by $(B(c) - L(w,c))$.

Delegated securities lending market. In the setting with private-value shocks, the equations determining the bid and ask prices conditional on a delegated market structure are given by:

$$\frac{B}{P} = \frac{\pi F_w(w_\phi)\,\mathbb{E}[e^v | w \leq w_\phi] + (1-\pi)}{\pi F_w(w_\phi) + (1-\pi)}, \tag{23}$$

$$\frac{A_0}{P} = \frac{\pi(1 - F_w(\log \frac{A_0}{P}))\,\mathbb{E}[e^v | w \geq \log \frac{A_0}{P}] + (1-\pi) + F_w(w_\phi)\frac{\pi}{\rho}\frac{\phi}{P}}{\pi(1 - F_w(\log \frac{A_0}{P})) + 1 - \pi}, \tag{24}$$

$$\frac{A_1}{P} = \frac{\pi(1 - F_w(\log \frac{A_1}{P}))\,\mathbb{E}[e^v | w \geq \log \frac{A_1}{P}] + \pi + (1-\pi) + (1+\pi)F_w(w_\phi)\frac{\pi}{\rho}\frac{\phi}{P}}{\pi(1 - F_w(\log \frac{A_1}{P})) + \pi + (1-\pi)}, \tag{25}$$

where we can rewrite the conditional expectation as follows:

$$\mathbb{E}[e^v | w < w_\phi] = \mathbb{E}[\mathbb{E}[e^v | v < w_\phi - b]] = \int_{-\infty}^{+\infty} f_b(b) \frac{F_v(w_\phi - b)}{F_w(w_\phi)}\,\mathbb{E}[e^v | v < w_\phi - b] db. \tag{26}$$

Since

$$\mathbb{E}[e^{v_t} | v_t \leq w_\phi - b] = \frac{F_v(w_\phi - b - \sigma_v^2)}{F_v(w_\phi - b)}, \tag{27}$$

we obtain:

$$\mathbb{E}[e^v | w < w_\phi] = \int_{-\infty}^{+\infty} f_b(b) \frac{F_v(w_\phi - b - \sigma_v^2)}{F_w(w_\phi)} db. \tag{28}$$

Similarly, we can compute the conditional expectation:

$$\mathbb{E}[e^v | w \geq w_\phi] = \int_{-\infty}^{+\infty} \int_{w_\phi - b}^{+\infty} \frac{f_b(b) f_v(v)}{1 - F_w(w_\phi)} e^v dv db \tag{29}$$

$$= \int_{-\infty}^{+\infty} f_b(b) \frac{1 - F_v(w_\phi - b)}{1 - F_w(w_\phi)}\,\mathbb{E}[e^v | v \geq w_\phi - b] db. \tag{30}$$

Since $\mathbb{E}[e^v|v \geq w_\phi - b] = \frac{F_v(-w_\phi+b)}{1-F_v(w_\phi-b)}$, we obtain:

$$\mathbb{E}[e^v|w \geq w_\phi] = \int_{-\infty}^{+\infty} f_b(b) \frac{F_v(-w_\phi+b)}{1-F_w(w_\phi)} db. \tag{31}$$

Analogously to the baseline model, the delegated shares lender chooses the optimal threshold type w_ϕ, which pins dow the loan fee

$$\phi(c) = B(c) - L(c, w_\phi). \tag{32}$$

The shares lender maximizes

$$\tilde{\Pi}(w_\phi) = \pi \cdot F_w(w_\phi) \cdot \left(\tilde{B} - e^{w_\phi} \frac{1+\lambda\tilde{A}_1}{r_f + \sigma_c\chi - \mu_c + \lambda} \right). \tag{33}$$

The shares lender's marginal profit of increasing w_ϕ is

$$\begin{aligned} \tilde{\Pi}'(w_\phi) =& \pi \cdot f_w(w_\phi) \left(\tilde{B} - \frac{1+\lambda\tilde{A}_1}{r_f + \sigma_c\chi - \mu_c + \lambda} e^{w_\phi} \right) \\ & - \pi \cdot F_w(w_\phi) \cdot \frac{1+\lambda\tilde{A}_1}{r_f + \sigma_c\chi - \mu_c + \lambda} e^{w_\phi}. \end{aligned} \tag{34}$$

At the optimum, the marginal profit is equal to zero, that is, w_ϕ solves $\Pi'(w_\phi) = 0$, or equivalently:

$$\frac{f_w(w_\phi)}{F_w(w_\phi)} \left(\frac{\tilde{B} \cdot (r_f + \sigma_c\chi - \mu_c + \lambda)}{1+\lambda\tilde{A}_1} e^{-w_\phi} - 1 \right) = 1, \tag{35}$$

which can be rewritten as:

$$\frac{f_w(w_\phi)}{F_w(w_\phi)} \cdot \left(\frac{B(c)}{L(c, w_\phi)} - 1 \right) = 1. \tag{36}$$

Competitive securities lending market. In the competitive securities lending market, a signaling game obtains whereby a hedge fund's borrowing choice $\{\pi, 0\}$ represents a public signal that leads to the updating of other agents' beliefs and correspondingly, generally has an impact on prices in the limit order market. We characterize the perfect Bayesian Nash equilibrium that is robust to the presence of an arbitrarily small exogenous cost for a hedge fund to take the time to borrow a security.

In the centralized securities lending market, competition between lenders implies that the lending fee is zero. Thus, even a hedge fund who does not want to short can

nonetheless costlessly borrow the shares (at a fee equal to zero), hold them for the length of the contract and then redeliver them to the lender. Conditional on having borrowed, a hedge fund can take three actions: (1) sell the security at the bid, (2) keep the borrowed security and additionally buy the security at the prevailing ask price,[3] (3) keep the security and neither buy nor sell the security. Option (3) is strictly dominated by options (1) and (2) conditional on the hedge fund actually wishing to trade at either the bid or the ask price. Moreover, conditional on not wishing to trade, the hedge fund will not borrow in the first place, since we characterize the equilibrium that is robust to the presence of an arbitrarily small exogenous cost of taking the time to borrow.

As in the analysis of the intermediated market structure, we need to distinguish the cases where there is preexisting short interest and no preexisting short interest. We use the first entry of the subscript $(\cdot|\cdot)$ to indicate whether there is preexisting short interest, and the second subscript to represent whether there is new borrowing. Note that after trading volume has been realized, it can be inferred from public information whether a hedge fund that had borrowed shares did not sell them at the bid price (but rather bought at the ask price). In that case, the preexisting short interest is 0 and agents know that there is no hedge fund in the market that will need to repurchase the asset at the end of the contract period.

Let $w_{\phi_{0|1}}$ and $w_{\phi_{1|1}}$ denote the threshold types for hedge funds wishing to short sell the asset (that is, borrow and sell at the bid). Conditional having a type below the threshold type, a hedge fund borrows the asset with probability 1 and immediately sells it at the bid price B with probability 1. The marginal type is the hedge fund that just breaks even at a zero fee when *borrowing and selling*, that is, $w_{\phi_{0|1}}$ and $w_{\phi_{1|1}}$ solve:

$$B_{0|1}(c) - L(c, w_{\phi_{0|1}}) = 0, \tag{37}$$

$$B_{1|1}(c) - L(c, w_{\phi_{1|1}}) = 0. \tag{38}$$

Bid prices. Consider liquidity providers' beliefs regarding the probability that, conditional on observing new borrowing, a hedge fund will pick up the ask quote. Let $\nu_{0|1}$ and $\nu_{1|1}$ denote the probabilities that hedge funds *buy* at the ask price after having borrowed, conditional on no preexisting short interest and existing short interest, respectively. Since option (3) is not used in equilibrium, the conditional probability that a hedge fund *sells* at the bid after having borrowed is then given by $(1-\nu_{0|1})$ and $(1-\nu_{1|1})$, respectively.

Conditional on new borrowing having occurred, the bid prices then solve (conditional on no preexisting short interest and conditional on preexisting short interest,

[3] When a hedge fund borrows the security without selling it, this action does not affect the hedge fund's trading capacity, as the fund has both the asset and the liability to redeliver it, which exactly net out to zero. Thus, the hedge fund can still buy an additional π units of the asset.

respectively):

$$\frac{B_{0|1}}{P} = \frac{(1-\nu_{0|1})\pi\,\mathbb{E}[e^{v}|w \leq w_{\phi_{0|1}}] + (1-\pi)}{(1-\nu_{0|1})\pi + 1 - \pi}, \tag{39}$$

$$\frac{B_{1|1}}{P} = \frac{(1-\nu_{1|1})\pi\,\mathbb{E}[e^{v}|w \leq w_{\phi_{1|1}}] + (1-\pi)}{(1-\nu_{1|1})\pi + 1 - \pi}. \tag{40}$$

In contrast, conditional on no new borrowing having occurred, the bid is priced to reflect that only liquidity traders might pick up the bid. That is, there is no adverse selection adjustment:

$$\frac{B_{0|0}}{P} = \frac{B_{1|0}}{P} = 1. \tag{41}$$

Ask prices. Again, we need to distinguish cases by whether there is preexisting short interest and by whether new borrowing has occurred.

First, consider the case where there is no preexisting short interest. Let $A_{0|0}$ denote the ask price conditional on the state where neither old nor new borrowing demand exists. Further let $A_{0|1}$ denote the ask price conditional no old borrowing demand but new borrowing demand. The conditional ask prices satisfy:

$$\frac{A_{0|0}}{P} = \frac{\nu_{0|0}\cdot\pi\cdot\mathbb{E}[e^{v}|w \geq \log\frac{A_{0|0}}{P}] + (1-\pi)}{\nu_{0|0}\pi + (1-\pi)}, \tag{42}$$

$$\frac{A_{0|1}}{P} = \frac{\nu_{0|1}\cdot\pi\cdot\mathbb{E}[e^{v}|w \geq \log\frac{A_{0|1}}{P}] + (1-\pi)}{\nu_{0|1}\pi + (1-\pi)}. \tag{43}$$

In principle, either conditional ask price could be advantageous for a hedge fund, depending on liquidity providers beliefs about the likelihood with which a hedge fund with positive news borrows versus does not borrow. A pure-strategy equilibrium is not sustainable since conditional on the market believing that the hedge fund always chooses one option (say a purchasing hedge fund is believed to never simultaneously borrow), the ask price conditional on borrowing occurring would reflect the belief that only uninformed traders pick up that ask quote. Yet conditional on this belief, the ask quote would feature no adverse selection adjustment, causing the hedge fund to switch to borrowing securities while simultaneously buying them. In equilibrium, the ask prices $A_{0|0}$ and $A_{0|1}$ therefore must be identical, such that a hedge fund intending to buy at the ask price does not have a strict preference over whether to borrow or not to borrow. The condition that $A_{0|0} = A_{0|1}$ implies that

$$\nu_{0|0} = \nu_{0|1}. \tag{44}$$

We define $A_0 \equiv A_{0|0} = A_{0|1}$ and $\nu_0 \equiv \nu_{0|0} = \nu_{0|1}$. We will characterize ν_0 below.

Second, consider the case where there is preexisting short interest. The conditional ask prices then account for the additional uninformed demand from hedge funds having to redeliver π units of shares. This yields the relations:

$$\frac{A_{1|0}}{P} = \frac{\pi + \nu_{1|0} \cdot \pi \cdot \mathbb{E}[e^v | w \geq \log \frac{A_{1|0}}{P}] + (1-\pi)}{\pi + \nu_{1|0}\pi + (1-\pi)}, \tag{45}$$

$$\frac{A_{1|1}}{P} = \frac{\pi + \nu_{1|1} \cdot \pi \cdot \mathbb{E}[e^v | w \geq \log \frac{A_{1|1}}{P}] + (1-\pi)}{\pi + \nu_{1|1}\pi + (1-\pi)}. \tag{46}$$

Since it is costless to borrow, the ask prices $A_{1|0}$ and $A_{1|1}$ must also be identical in equilibrium, such that a hedge fund intending to buy at the ask price does not have a strict preference over whether to borrow or not to borrow. The condition that $A_{1|0} = A_{1|1}$ implies that

$$\nu_{1|0} = \nu_{1|1}. \tag{47}$$

We define $A_1 \equiv A_{1|0} = A_{1|1}$ and $\nu_1 \equiv \nu_{1|0} = \nu_{1|1}$.

Next, we solve for ν_0 and ν_1. Conditional on there being preexisting short interest, the probability that a hedge fund buys the asset at the ask price is given by:

$$\begin{aligned} &\Pr[\text{new borrowing}|\text{existing borrowing}] \cdot \nu_{1|1} \\ &+(1 - \Pr[\text{new borrowing}|\text{existing borrowing}]) \cdot \nu_{1|0} \\ &=1 - F_w(\log \frac{A_1}{P}). \end{aligned} \tag{48}$$

Using $\nu_1 \equiv \nu_{1|0} = \nu_{1|1}$ yields:

$$\nu_1 = 1 - F_w(\log \frac{A_1}{P}). \tag{49}$$

Analogously, we obtain:

$$\nu_0 = 1 - F_w(\log \frac{A_0}{P}). \tag{50}$$

We use these results to derive the final equations for the ask prices:

$$\frac{A_0}{P} = \frac{(1 - F_w(\log \frac{A_0}{P})) \cdot \pi \cdot \mathbb{E}[e^v | w \geq \log \frac{A_0}{P}] + (1-\pi)}{(1 - F_w(\log \frac{A_0}{P}))\pi + (1-\pi)}, \tag{51}$$

$$\frac{A_1}{P} = \frac{\pi + (1 - F_w(\log \frac{A_1}{P})) \cdot \pi \cdot \mathbb{E}[e^v | w \geq \log \frac{A_1}{P}] + (1-\pi)}{\pi + (1 - F_w(\log \frac{A_1}{P}))\pi + (1-\pi)}. \tag{52}$$

Absent private value shocks, we have the following solution for truncated expectation:

$$\mathbb{E}[e^{v_t} | v \geq \log(A_0/P)] = \frac{F_v(-\log \frac{A_0}{P})}{1 - F_v(\log \frac{A_0}{P})}. \tag{53}$$

In this case, the expressions simplify to:

$$\frac{A_0}{P} = \frac{\pi \cdot F_v(-\log \frac{A_0}{P}) + (1-\pi)}{(1 - F_v(\log \frac{A_0}{P}))\pi + (1-\pi)}, \tag{54}$$

$$\frac{A_1}{P} = \frac{\pi + \pi \cdot F_v(-\log \frac{A_1}{P}) + (1-\pi)}{\pi + (1 - F_v(\log \frac{A_1}{P}))\pi + (1-\pi)}. \tag{55}$$

Finally, we determine the probability of borrowing, which occurs either by hedge funds wishing to short or by hedge funds wishing to buy the asset. Conditional on preexisting short interest we obtain the relation:

$$\begin{aligned} &\Pr[\text{new borrowing} | \text{existing borrowing}] \\ &= F_w(w_{\phi_{1|1}}) + \Pr[\text{new borrowing} | \text{existing borrowing}] \cdot \nu_{1|1}. \end{aligned} \tag{56}$$

Which yields

$$\Pr[\text{new borrowing} | \text{existing borrowing}] = \frac{F_w(w_{\phi_{1|1}})}{(1 - \nu_{1|1})} = \frac{F_w(w_{\phi_{1|1}})}{F_w(\log \frac{A_1}{P})}. \tag{57}$$

Analogously, we obtain for the case without preexisting borrowing:

$$\Pr[\text{new borrowing} | \text{no existing borrowing}] = \frac{F_w(w_{\phi_{0|1}})}{(1 - \nu_{0|1})} = \frac{F_w(w_{\phi_{0|1}})}{F_w(\log \frac{A_0}{P})}. \tag{58}$$

C.6 Proof of Proposition 5

The HJB equation corresponding to (27), scaled by λc, is given by:

$$0 = -(r_f + \sigma_c(s)\chi - \mu_c(s)) \cdot \tilde{P}(s) + (1 + \psi(s)) + (\Lambda(s) \odot \mathbf{U}(s)) \cdot \tilde{\mathbf{P}} \quad \forall s, \tag{59}$$

where $\odot$ denotes the Hadamard product and where $\psi(s)$ is the fee-to-dividend ratio in state s. Moreover, the HJB equation associated with the present value without fee income (28) is given by:

$$0 = -(r_f + \sigma_c(s)\chi - \mu_c(s)) \cdot \tilde{P}^e(s) + 1 + (\Lambda(s) \odot \mathbf{U}(s)) \cdot \tilde{\mathbf{P}}^e \quad \forall s. \tag{60}$$

These sets of linear equations yield closed-form solutions for the vectors of price-dividend ratios:

$$\tilde{\mathbf{P}} = -(\Lambda \odot \mathbf{U} - diag(r_f + \sigma_c(s)\chi - \mu_c(s)))^{-1}(\mathbf{1} + \psi), \tag{61}$$

$$\tilde{\mathbf{P}}^e = -(\Lambda \odot \mathbf{U} - diag(r_f + \sigma_c(s)\chi - \mu_c(s)))^{-1}\mathbf{1}. \tag{62}$$

D Continuum of Hedge Funds with Noisy Signals

In this appendix section, we present and analyze the adjusted model we estimate for the quantitative evaluation presented in Section 6.5. This setup considers a continuum of hedge funds that obtain noisy signals about the growth shock v. A continuum of hedge funds of measure 1 enters upon each innovation $dN_t = 1$. Each hedge fund i obtains a noisy signal

$$w_i = v + \varepsilon_i, \tag{63}$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon(s))$ and independent across i. A hedge fund's decision-relevant type is now given by w_i, where $w_i \sim \mathcal{N}(-\frac{\sigma_v(s)^2}{2}, \sqrt{\sigma_v(s)^2 + \sigma_\varepsilon(s)^2})$.

Analogously to the baseline setup, there is a threshold type, which we denote by $w_\phi(s)$, that corresponds to the hedge fund signal that renders a fund just indifferent between borrowing and not borrowing at the equilibrium fee ϕ in state s. Such a borrowing threshold type exists in both the economy with intermediated securities lending and in the counterfactual economy with a centralized securities lending market. Conditional on such a threshold, due to cross-sectional diversification of the idiosyncratic signal noise terms ε_i, the fraction of hedge funds shorting at the beginning of a contract period is a function of the fundamental v alone. Let $x(v) \equiv F_w(w_\phi|v)$ denote this fraction, where

$$F_w(w_\phi|v) = F_\varepsilon(w_\phi - v). \tag{64}$$

To ensure the tractability of this setting with a continuum of hedge funds, we reverse engineer the process by which liquidity traders arrive and purchase shares at the ask price so as to keep the overall liquidity demand constant across periods. Absent this assumption, there would be a continuum of potential ask prices corresponding to the continuum of potential fractions of hedge funds that may need to redeliver shares at the beginning of a given period, which would render obtaining exact solutions and estimating the model intractable. Specifically, we assume that apart from the $(1-\pi)$ units purchased by regular liquidity traders in our baseline model, there is an additional state-dependent quantity of *short-term liquidity traders* that buy at the ask price and turn into regular investors at the end of the period.

This additional liquidity demand is a function of the existing short interest (which can also be inferred from the previous period's realized trading volume at the bid price), with the additional short-term liquidity traders buying:

$$\pi - \text{Short Interest} \tag{65}$$

units at the ask price. That is, the additional liquidity demand is negatively related to the previous period's short interest. The total liquidity demand at the ask price is then exactly constant and equal to 1 unit in every period. As the total liquidity demand is constant, we then do not need to condition the ask price on last period's short interest and can simply denote by A the single equilibrium ask price.

In the following, we first analyze the extended economy with an intermediated securities lending market. Afterwards, we consider the counterfactual economy with a centralized securities lending market.

D.1 Intermediated Securities Lending Market

When deciding whether to buy or sell shares, a hedge fund in the intermediated securities lending market only observes its private signal w_i (which contrasts with the centralized market we will discuss below). A hedge fund observing the signal w_i assigns the following buy-and-hold value to the asset

$$Q(c,w_i,s) = \mathbb{E}[e^v \cdot P(c,s)|w_i,s] = e^{\mathbb{E}[v|w_i,s]+\frac{1}{2}Var[v|w_i,s]}P(c,s) = e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_i} \cdot P(c,s), \tag{66}$$

where $P(c,s) = \lambda c \tilde{P}(s)$.

Analogously to the analysis of our baseline setup, we conjecture and verify that given a hedge fund's private signal w_i, the value of the hedge fund's liability to redeliver shares at the end of the period takes the form

$$L(w_i,c,s) = \lambda c \cdot \tilde{L}(w_i,s). \tag{67}$$

Let $\kappa(c,s)$ denote the commission a buyer has to pay per unit purchased at the ask price. The conditional expected cost at which the shares will be repurchased is given by

$$\begin{aligned}\mathbb{E}[A(c',s)+\kappa(c',s)|w_i,s,c] =& \lambda c\,\mathbb{E}[e^v|w_i,s](\tilde{A}(s)+\tilde{\kappa}(s)) \\ =& \lambda c\cdot e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_i}\cdot(\tilde{A}(s)+\tilde{\kappa}(s)).\end{aligned} \tag{68}$$

Using these relations, the HJB equation for the scaled liability $\tilde{L}(w_i,s)$ incurred from borrowing shares takes the form:

$$\begin{aligned}0 =& -(r_f+\sigma_c(s)\chi-\mu_c(s))\cdot\lambda c\cdot\tilde{L}(w_i,s)+\lambda\cdot\{ce^v+\lambda c\cdot e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_i}(\tilde{A}(s)+\tilde{\kappa}(s)) \\ &-\lambda c\cdot\tilde{L}(w_i,s)\}+\lambda c\cdot(\Lambda(s)\odot\mathbf{U}(s))\cdot\tilde{\mathbf{L}}(w_i),\end{aligned} \tag{69}$$

which simplifies to:

$$\tilde{\mathbf{L}}(w_i)=e^{\frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2}\cdot w_i}\cdot(\Lambda\odot\mathbf{U}-diag(r_f+\sigma_c(s)\chi-\mu_c(s))+\lambda)^{-1}(\mathbf{1}+\lambda(\tilde{\mathbf{A}}+\tilde{\kappa})). \tag{70}$$

This solution mimics the one in our baseline model (compare equation (17)). The solution implies that

$$\tilde{L}(w_\phi,s)=e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_\phi}\cdot\mathbb{E}[\tilde{L}(w_i,s)|s]. \tag{71}$$

The delegated shares lender chooses the optimal threshold type $w_\phi(s)$, which pins down the loan fee:

$$\phi(c,s)=B(c,s)-L(c,w_\phi(s),s). \tag{72}$$

The shares lender maximizes its profits, which we again express as a function of the marginal hedge fund type $w_\phi(s)$:

$$\tilde{\Pi}(w_\phi(s))=\pi\cdot F_{w,s}(w_\phi(s))\cdot\left(\tilde{B}(s)-e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_\phi(s)}\cdot\mathbb{E}[\tilde{L}(w,s)|s]\right), \tag{73}$$

where we use equation (71). The shares lender's marginal profit of increasing $w_\phi(s)$

is

$$\tilde{\Pi}'(w_\phi(s)) = \pi \cdot f_{w,s}(w_\phi(s)) \left(\tilde{B}(s) - e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2} \cdot w_\phi(s)} \cdot \mathbb{E}[\tilde{L}(w,s)|s] \right)$$
$$- \pi \cdot F_{w,s}(w_\phi(s)) \cdot e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2} \cdot w_\phi(s)} \cdot \mathbb{E}[\tilde{L}(w,s)|s] \cdot \frac{\sigma_v(s)^2}{\sigma_v(s)^2 + \sigma_\varepsilon(s)^2}. \tag{74}$$

At the optimum, the marginal profit is equal to zero, that is, $w_\phi(s)$ solves $\Pi'(w_\phi(s)) = 0$, or equivalently:

$$\frac{f_{w,s}(w_\phi(s))}{F_{w,s}(w_\phi(s))} \cdot \frac{\sigma_v(s)^2 + \sigma_\varepsilon(s)^2}{\sigma_v(s)^2} \cdot \left(\frac{\tilde{B}(s)}{\tilde{L}(w_\phi(s),s)} - 1 \right) = 1. \tag{75}$$

The fee-to-dividend ratio $\psi(s)$ is given by:

$$\psi(s) = \frac{\lambda F_{w,s}(w_\phi(s)) \frac{\pi}{\rho(s)} \phi(s,c)}{\lambda c} = F_{w,s}(w_\phi(s)) \frac{\pi}{\rho(s)} \tilde{\phi}(s), \tag{76}$$

since $\phi(s,c)$ again scales linearly with c.

Commissions. In equilibrium, if asset owners provide shares for both lending and for ask quote provision in the limit-order market, these investors have to be indifferent between these two activities. When deciding to make their shares available for lending, investors expect to receive fee income equal to

$$F_{w,s}(w_\phi(s)) \frac{\pi}{\rho(s)} \tilde{\phi}(s) \cdot c\lambda \tag{77}$$

per unit of the asset that they post as inventory available for lending.

In equilibrium, given that $(1+\pi)$ units are committed for sale in the limit-order market, total *expected* commissions income must be:

$$F_{w,s}(w_\phi(s)) \frac{\pi}{\rho(s)} \tilde{\phi}(s) \cdot c\lambda \cdot (1+\pi) \tag{78}$$

and the total lending fee income is

$$F_{w,s}(w_\phi(s)) \pi \tilde{\phi}(s) \cdot c\lambda. \tag{79}$$

Suppose that in equilibrium all hedge fund types $w_i > w_A$ purchase the asset at the ask price, where w_A will be characterized below. Then the expected volume generating

commissions is:

$$\pi \cdot (1 - F_w(w_A)) + 1. \tag{80}$$

As a result, the commission per unit purchased by a buyer is:

$$\kappa(c,s) = \frac{F_{w,s}(w_\phi(s)) \frac{\pi}{\rho(s)} \tilde{\phi}(s) \cdot c \cdot \lambda \cdot (1+\pi)}{\pi \cdot (1 - F_w(w_A)) + 1}. \tag{81}$$

Bid and ask quotes. For notational simplicity, we omit in the following the dependence of variables on the Markov state s. The marginal hedge fund type *buying* at the ask price, which we denote by w_A, solves the equation:

$$Q(c,w_A) = e^{\frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2} \cdot w_A} \cdot P(c) = A(c) + \kappa(c), \tag{82}$$

as the hedge fund anticipates paying both the ask price A and the commission κ. Rearranging this relation yields:

$$w_A = \frac{\sigma_v^2 + \sigma_\varepsilon^2}{\sigma_v^2} \log \left[\frac{A(c) + \kappa(c)}{P(c)} \right]. \tag{83}$$

The equations determining the bid and ask prices conditional on a delegated market structure are given by:

$$\frac{B}{P} = \frac{\pi F_w(w_\phi)\, \mathbb{E}[e^v | w \le w_\phi] + (1-\pi)}{\pi F_w(w_\phi) + (1-\pi)}, \tag{84}$$

$$\frac{A}{P} = \frac{\pi (1 - F_w(w_A))\, \mathbb{E}[e^v | w \ge w_A] + 1}{\pi (1 - F_w(w_A)) + 1}. \tag{85}$$

We can then also the define the cum-commissions ask price:

$$\frac{A+\kappa}{P} = \frac{\pi (1 - F_w(w_A))\, \mathbb{E}[e^v | w \ge w_A] + 1 + (1+\pi) F_w(w_\phi) \frac{\pi}{\rho} \frac{\phi}{P}}{\pi (1 - F_w(w_A)) + 1} \tag{86}$$

We can rewrite the conditional expectation as follows:

$$\mathbb{E}[e^v|v+\varepsilon<w_\phi] = \int_{-\infty}^{+\infty}\int_{-\infty}^{w_\phi-\varepsilon}\frac{f_\varepsilon(\varepsilon)f_v(v)}{F_w(w_\phi)}e^v dv d\varepsilon \tag{87}$$

$$= \int_{-\infty}^{+\infty} f_\varepsilon(\varepsilon)\frac{F_v(w_\phi-\varepsilon)}{F_w(w_\phi)}\int_{-\infty}^{w_\phi-\varepsilon}\frac{f_v(v)}{F_v(w_\phi-\varepsilon)}e^v dv d\varepsilon \tag{88}$$

$$= \int_{-\infty}^{+\infty} f_\varepsilon(\varepsilon)\frac{F_v(w_\phi-\varepsilon)}{F_w(w_\phi)}\mathbb{E}[e^v|v<w_\phi-\varepsilon,\varepsilon]d\varepsilon. \tag{89}$$

Since the moment generating function for the truncated normal distributions implies that

$$\mathbb{E}[e^{v_t}|v_t \le w_\phi-\varepsilon,\varepsilon] = \frac{F_v(w_\phi-\varepsilon-\sigma_v^2)}{F_v(w_\phi-\varepsilon)}, \tag{90}$$

we obtain:

$$F_w(w_\phi)\,\mathbb{E}[e^v|w<w_\phi] = \int_{-\infty}^{+\infty} f_\varepsilon(\varepsilon)F_v(w_\phi-\varepsilon-\sigma_v^2)d\varepsilon. \tag{91}$$

Similarly, we can compute the conditional expectation:

$$\mathbb{E}[e^v|w\ge w_A] = \int_{-\infty}^{+\infty}\int_{w_A-\varepsilon}^{+\infty}\frac{f_\varepsilon(\varepsilon)f_v(v)}{1-F_w(w_A)}e^v dv d\varepsilon \tag{92}$$

$$= \int_{-\infty}^{+\infty} f_\varepsilon(\varepsilon)\frac{1-F_v(w_A-\varepsilon)}{1-F_w(w_A)}\mathbb{E}[e^v|v\ge w_A-\varepsilon]d\varepsilon. \tag{93}$$

Since the moment generating function of the truncated normal distribution implies that $\mathbb{E}[e^v|v\ge w_A-\varepsilon] = \frac{F_v(-w_A+\varepsilon)}{1-F_v(w_A-\varepsilon)}$, we obtain:

$$(1-F_w(w_A))\cdot\mathbb{E}[e^v|w\ge w_A] = \int_{-\infty}^{+\infty} f_\varepsilon(\varepsilon)F_v(-w_A+\varepsilon)d\varepsilon. \tag{94}$$

Hedge fund alphas and surplus. The alpha a hedge fund makes on average, conditional on borrowing is given by

$$\alpha = \lambda\cdot\left(\frac{B(c,s)}{P(c,s)} - \frac{\mathbb{E}[L(c,v,s)|w_i<w_\phi(s)]}{P(c,s)}\right) \tag{95}$$

The chance that a hedge fund belonging to the measure π is going to borrow is: / In the delegated market we have the following annualized expected net-of-fee surplus

yield from shorting:

$$F_{w,s}(w_\phi(s)) \cdot \left(\alpha - \lambda \frac{\phi(c,s)}{P(c,s)} \right). \tag{96}$$

D.2 Competitive Securities Lending Market

As in our baseline analysis, we focus on equilibria that are robust to the presence of arbitrarily small exogenous costs of borrowing securities, which we denote by ϕ_0. Just like the delegated securities lending market, the centralized lending market features a threshold equilibrium where hedge funds with signals w_i below a threshold value w_ϕ borrow.

Conditional on such a threshold, the aggregate borrowing demand is analogously given by

$$x = F_w(w_\phi | v) \cdot \pi = F_\varepsilon(w_\phi - v) \cdot \pi. \tag{97}$$

In the centralized market, agents can observe this quantity x and the equilibrium fee ϕ. As detailed below, observing these two aggregate measures would allow inferring the underlying state of the economy v perfectly. This is the case since agent-specific noise shocks ε_i diversify in the presence of a continuum of hedge funds. However, once there would be such perfect inference, hedge funds would not be willing to pay any positive cost to borrow, since there would be no residual information asymmetry to profit from borrowing securities.

To address this issue and to capture the fact that perfect inference would typically not be feasible in practice for a variety of reasons (such as non-informational demand, supply shocks, uncertainty about equilibrium relations, etc.), we assume that agents can only infer v with noise once the securities lending market clears (where volume x and the fee are observed). Concretely, whereas agents are assumed to be able to observe the realized borrowing volume

$$x = \pi \cdot F_\varepsilon(w_\phi(v) - v) \tag{98}$$

and can back out $(w_\phi(v) - v)$ by inverting the CDF F_ε, they cannot perfectly determine v based on that, that is, they do not exactly know what the state v is if a hedge fund observing $w_i = w_\phi(v)$ is just indifferent between borrowing and not borrowing. In other words, agents don't know exactly how $(w_\phi(v) - v)$ relates to the realization of v (e.g., since this requires more equilibrium thinking). Rather, all agents obtain the noisy signal $(\hat{v})$:

$$\hat{v} \equiv v + b, \tag{99}$$

where $b \sim \mathcal{N}(0, \sigma_b)$.

When there is excess inventory available for lending, $\pi < \rho$, the quantity x is pinned down by the marginal hedge fund type that is willing to pay the arbitrarily small exogenous cost of borrowing ϕ_0. In this case, the borrowing market features a threshold equilibrium where only those hedge fund types that have signals letting them expect to recoup these vanishingly small exogenous costs will borrow. Since this cost is arbitrarily small, the measure of hedge funds borrowing will be only slightly below the total measure of hedge funds, implying that the total borrowing demand will be almost exactly π. In contrast, when inventory is scarce, $\pi \geq \rho$, fees have to rise to the point where the market clears, $x = \rho$. In this latter case, the equilibrium fee is a monotone function of v, since the aggregate demand for borrowing is a monotone function of v and the supply is fixed at ρ.

In the presence of a centralized securities lending market, there are two distinct thresholds related to short sales: one for borrowing securities and one for actually selling a borrowed security. The threshold type changes between the borrowing stage and the trading stage, as hedge funds themselves obtain new information when the securities lending market clears. By backward induction, we start our analysis with the trading stage.

D.2.1 Trading Stage

Bid and ask quotes are provided conditional on the public signal $\hat{v}$ and the quantity of borrowing x that occurred in the securities lending market. Hedge funds additionally observe their private signal $w_i = v + \varepsilon_i$.

Ask price. Agents posting ask quotes anticipate that hedge funds with signals w_i above a threshold $w_A(\hat{v})$ will purchase the asset, where $w_A(\hat{v})$ is determined by the indifference condition:

$$A(c, \hat{v}) + \kappa(c, \hat{v}) = \mathbb{E}[e^v | \hat{v}, w_i = w_A(\hat{v})] \cdot P(c) \tag{100}$$

We define:

$$AoP(\hat{v}) \equiv \frac{A(c, \hat{v})}{P(c, \hat{v})}, \tag{101}$$

$$\kappa oP \equiv \frac{\kappa(c, \hat{v})}{P(c, \hat{v})}. \tag{102}$$

Note that we drop the dependence κoP to $\hat{v}$. Below, we describe the commission arrangement, which ensures that the ratio κoP is independent of the realized value $\hat{v}$, which, as we will see, obtains when the commission a buyer pays is an ex-ante known percentage of the price they pay (the ask price).

Scaling the indifference condition by $P(c,\hat{v}) = \mathbb{E}[e^v|\hat{v}] \cdot P(c)$ yields:

$$\frac{\mathbb{E}[e^v|\hat{v}, w_i = w_A(\hat{v})]}{\mathbb{E}[e^v|\hat{v}]} = AoP(\hat{v}) + \kappa oP, \tag{103}$$

or equivalently:

$$\exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_A(\hat{v})\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2} - \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right) = AoP(\hat{v}) + \kappa oP. \tag{104}$$

where we use the results derived in Section ... for the the following conditional expectations:

$$\mathbb{E}[e^v|\hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right), \tag{105}$$

$$\mathbb{E}\left[e^v \,\middle|\, \hat{v}, w_i\right] = \exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_i\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2}\right). \tag{106}$$

We can then solve equation (104) for $w_A(\hat{v})$:

$$w_A(\hat{v}) \;=\; \frac{\sigma_v^2}{\sigma_v^2 \,+\, \sigma_b^2}\,\hat{v} \;+\; \frac{\sigma_v^2\big(\sigma_b^2+\sigma_\varepsilon^2\big) \,+\, \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\,\ln\big[AoP(\hat{v}) + \kappa oP\big].$$

Commissions income compensates ask-quote providers (market makers) for the ex ante expected forgone fee revenue they could collect on their inventory if they offered their shares for lending. In contrast, for every realization $\hat{v}$, the ask quote $A(c,\hat{v})$ itself allows market makers to break even on the sales of their securities, accounting for adverse selection. The following equation equates the expected revenue from selling at the ask quote $A(c,\hat{v})$ to the expected buy-and-hold value of securities sold, conditional on observing $\hat{v}$ and conditional on trade occurring with those hedge funds that receive signals $w_i > w_A(\hat{v})$:

$$\begin{aligned} &A(c,\hat{v}) \cdot \left(\int_{-\infty}^{\infty} \pi \cdot (1 - F_w(w_A(\hat{v})|v,\hat{v}))f_v(v|\hat{v})dv + 1\right) \\ &= \left[\int_{-\infty}^{\infty} e^v \pi \cdot (1 - F_w(w_A(\hat{v})|v,\hat{v}))f_v(v|\hat{v})dv + \mathbb{E}[e^v|\hat{v}]\right] \cdot P(c), \end{aligned} \tag{107}$$

which accounts for a constant liquidity demand of 1 unit at the ask price and which

leads to the sale of securities with expected value $\mathbb{E}[e^v|\hat{v}]P(c)$. Since

$$F_w(w_A|v,\hat{v}) = F_w(w_A|v) = F_\varepsilon(w_A - v) = \Phi\left(\frac{w_A - v}{\sigma_\varepsilon}\right), \tag{108}$$

where $\Phi(\cdot)$ is the standard normal CDF, we can rewrite equation (107) as follows:

$$\begin{aligned} &A(c,\hat{v}) \cdot \left(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + 1\right) \\ &= \left[\int_{-\infty}^{\infty} e^v \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + \mathbb{E}[e^v|\hat{v}]\right] \cdot P(c). \end{aligned} \tag{109}$$

where the conditional pdf $f_v(v|\hat{v})$ has the following mean and variance:

$$\mu_{v|\hat{v}} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_b^2}\,\hat{v} \;+\; \frac{\sigma_b^2}{\sigma_v^2 + \sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right) \tag{110}$$

$$\sigma_{v|\hat{v}}^2 = \frac{\sigma_v^2\,\sigma_b^2}{\sigma_v^2 + \sigma_b^2} \tag{111}$$

Finally, scaling by $P(c,\hat{v}) = \mathbb{E}[e^v|\hat{v}]P(c)$ and rearranging we obtain the relation:

$$AoP(\hat{v}) \cdot \mathbb{E}[e^v|\hat{v}] = \frac{\left[\int_{-\infty}^{\infty} e^v \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + \mathbb{E}[e^v|\hat{v}]\right]}{\left(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + 1\right)}. \tag{112}$$

Let's define

$$\begin{aligned} \bar{w}_A &\equiv w_A(\hat{v}) - \mu_{v|\hat{v}} \\ &= \frac{\sigma_v^2\left(\sigma_b^2 + \sigma_\varepsilon^2\right) \;+\; \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\,\ln\left[AoP(\hat{v}) + \kappa oP\right] - \left(\frac{\sigma_b^2}{\sigma_v^2 + \sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right)\right). \end{aligned} \tag{113}$$

Note that $\bar{w}_A$ does not depend on $\hat{v}$ directly and would depend on $\hat{v}$ only if either AoP or κoP depend on $\hat{v}$. We drop the functional dependence of $\bar{w}_A$ on $\hat{v}$ to indicate that we conjecture, and verify below, that AoP and κoP are not functions of $\hat{v}$. Further, we define:

$$z \equiv v - \mu_{v|\hat{v}} \tag{114}$$

Then we obtain

$$\Phi\Big(\tfrac{w_A(\hat{v})-v}{\sigma_\varepsilon}\Big) = \Phi\Big(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon}\Big) \tag{115}$$

$$v = \mu_{v|\hat{v}} + z \tag{116}$$

$$f_v(v \mid \hat{v})\, dv = f_z(z)\, dz \quad (\text{with } z \sim \mathcal{N}(0, \frac{\sigma_v^2\, \sigma_b^2}{\sigma_v^2+\sigma_b^2})) \tag{117}$$

$$e^v \;=\; e^{\mu_{v|\hat{v}}}\, e^z = \mathbb{E}[e^v|\hat{v}] \cdot e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)} \tag{118}$$

Then we obtain

$$AoP(\hat{v}) \cdot \mathbb{E}[e^v|\hat{v}] \;=\; \frac{\mathbb{E}[e^v|\hat{v}] \int \pi\, e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)} \left[1-\Phi(\frac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\, dz \;+\; \mathbb{E}[e^v|\hat{v}]}{\int \pi \left[1-\Phi(\frac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\, dz \;+\; 1}, \tag{119}$$

which further simplifies to:

$$AoP \;=\; \frac{\int \pi\, e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)} \left[1-\Phi(\frac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\, dz \;+\; 1}{\int \pi \left[1-\Phi(\frac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\, dz \;+\; 1}, \tag{120}$$

that is, AoP is indeed independent of $\hat{v}$. These equations hold irrespective of the relative magnitudes of ρ and π. However, for $\rho > \pi$, we have $\kappa oP = 0$.

Bid quotes. Suppose that conditional on having borrowed the asset, all hedge fund types $w_i < w_B(\hat{v})$ would want to sell, where $w_B(\hat{v})$ is a function of the public signal $\hat{v}$. If it were feasible for all hedge funds to sell, the measure of them selling would always be $F_w(w_B(\hat{v})|\hat{v})$. However, since only a measure $F_\varepsilon(w_\phi(v)-v)$ of hedge funds borrowed the asset in the first place, the measure that actually sells the asset is:

$$\min[F_\varepsilon(w_\phi(v)-v), F_w(w_B(\hat{v})|\hat{v})]. \tag{121}$$

Correspondingly, after x units have been borrowed in the securities lending market, the quantity sold by informed hedge funds is:

$$\min[x, \pi \cdot F_w(w_B(\hat{v})|\hat{v})] = \min[x, \pi \cdot \Pr(w_i < w_B(\hat{v})|\hat{v})], \tag{122}$$

which is a random object from the perspective of an agent observing $\hat{v}$ since $w_i = v+\varepsilon_i$ is still random conditional on observing $\hat{v} = v+b$. However, if we condition

on both $\hat{v}$ and v we obtain the quantity:

$$\begin{aligned}\min[x, \pi \cdot \Pr[w_i < w_B(\hat{v})|\hat{v}, v]] &= \min[x, \pi \cdot \Pr[v + \varepsilon_i < w_B(\hat{v})|\hat{v}, v]] \\ &= \min[x, \pi \cdot \Pr[\varepsilon_i < w_B(\hat{v}) - v|\hat{v}, v]] \\ &= \min[x, \pi \cdot F_\varepsilon(w_B(\hat{v}) - v)] \\ &= \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] \qquad (123)\end{aligned}$$

which is deterministic given $\hat{v}$ and b. Conditional on observing $\hat{v}$, agents posting a bid quote account for these possible quantities and compute the total conditional expected value of shares traded by informed hedge funds at the bid (volume times conditional value):

$$\begin{aligned}&\int_{-\infty}^{\infty} e^v P(c) \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv \\ &= P(c) \int_{-\infty}^{\infty} e^v \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv. \qquad (124)\end{aligned}$$

The expected volume traded by informed hedge funds at the bid is:

$$\int_{-\infty}^{\infty} \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv. \qquad (125)$$

Let $B(c, \hat{v}, x)$ denote the bid price offered given that agents observe $\hat{v}$. The bid price solves the break-even condition:

$$\begin{aligned}&B(c, \hat{v}, x) \cdot \left(\int_{-\infty}^{\infty} \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv + (1 - \pi)\right) \\ &= \left[\int_{-\infty}^{\infty} e^v \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv + (1 - \pi)\, \mathbb{E}[e^v|\hat{v}]\right] \cdot P(c), \qquad (126)\end{aligned}$$

where we use the fact that a quantity $(1 - \pi)$ is sold by uninformed liquidity traders and where the average value of the asset across those liquidity trades is:

$$P(c, \hat{v}) \equiv \mathbb{E}[e^v|\hat{v}] \cdot P(c). \qquad (127)$$

Define the ratio of B over P, after conditioning on both the signal $(\hat{v})$ and the observed borrowing demand x, as $BoP(\hat{v},x) \equiv B(c,\hat{v},x)/P(c,\hat{v})$. We obtain the relation:

$$BoP(\hat{v},x)\,\mathbb{E}[e^v|\hat{v}] = \frac{\int_{-\infty}^{\infty} e^v \min\left[x, \pi\cdot\Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v})dv + (1-\pi)\,\mathbb{E}[e^v|\hat{v}]}{\int_{-\infty}^{\infty} \min\left[x, \pi\cdot\Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v})dv + (1-\pi)}. \tag{128}$$

Let's define

$$\bar{w}_B \equiv w_B(\hat{v}) - \mu_{v|\hat{v}} \tag{129}$$

where we again conjecture, and later verify, that $\bar{w}_B$ is independent of $\hat{v}$. Then we obtain the equalities:

$$\Phi\Big(\tfrac{w_B(\hat{v})-v}{\sigma_\varepsilon}\Big) \;=\; \Phi\Big(\tfrac{\bar{w}_B-z}{\sigma_\varepsilon}\Big) \tag{130}$$

$$v \;=\; \mu_{v|\hat{v}} + z \tag{131}$$

$$f_v(v \mid \hat{v})\,dv \;=\; f_z(z)\,dz \quad (\text{with } z\sim\mathcal{N}(0, \frac{\sigma_v^2\,\sigma_b^2}{\sigma_v^2+\sigma_b^2})) \tag{132}$$

$$e^v \;=\; e^{\mu_{v|\hat{v}}}\,e^z = \mathbb{E}[e^v|\hat{v}]\cdot e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)}. \tag{133}$$

Subsituting those into .. .yields:

$$BoP(\hat{v}) \;\cdot\; \mathbb{E}[e^v|\hat{v}] \;=\; \frac{\mathbb{E}[e^v|\hat{v}]\int e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)}\;\min\Big[x,\;\pi\,\Phi\big(\frac{\bar{w}_B-z}{\sigma_\varepsilon}\big)\Big] f_z(z)\,dz \;+\; (1-\pi)\,\mathbb{E}[e^v|\hat{v}]}{\int \min\Big[x,\;\pi\,\Phi\big(\frac{\bar{w}_B-z}{\sigma_\varepsilon}\big)\Big] f_z(z)\,dz \;+\; (1-\pi)}, \tag{134}$$

and after canceling out $\mathbb{E}[e^v|\hat{v}]$ we obtain:

$$BoP \;=\; \frac{\int_{-\infty}^{\infty} e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)}\;\min\Big[x,\;\pi\,\Phi\big(\frac{\bar{w}_B-z}{\sigma\varepsilon}\big)\Big] f_z(z)\,dz \;+\; (1-\pi)}{\int_{-\infty}^{\infty} \min\Big[x,\;\pi\,\Phi\big(\frac{\bar{w}_B-z}{\sigma\varepsilon}\big)\Big] f_z(z)\,dz \;+\; (1-\pi)}, \tag{135}$$

$$\bar{w}_B \;=\; w_B(\hat{v}) - \mu_{v|\hat{v}} \tag{136}$$

That is, conditional on the conjecture that $\bar{w}_B$ is independent of $\hat{v}$, BoP is also independent of $\hat{v}$. We will derive $\bar{w}_B$ below to verify this.

If inventory is scarce, $\rho < \pi$, the equilibrium borrowing quantity is $x = \rho$. In contrast, if inventory is abundant, $\rho > \pi$, then the endogenous fee is zero and almost all hedge funds borrow, implying that $x \approx \pi$ (for $\phi_0 \to 0$ we have $x(v) \to \pi$ for all v). That is, for the purpose of solving the model quantitatively we can use:

$$x \approx \begin{cases} \pi & \text{if } \pi < \rho \\ \rho & \text{if } \pi > \rho \end{cases} = \min[\pi, \rho], \tag{137}$$

that is, x, the borrowed amount, is either very close to π (when inventory is abundant and borrowing costs close to zero) or exactly equal to ρ (when inventory is scarce). We provide an exact characterization of x for the case of abundant inventory below.

When are all borrowed shares x sold at the bid? Given a realization $\hat{v}$, the *highest* fundamental value v at which all borrowed shares x are sold at the bid solves the equation:

$$x = \pi \cdot \Pr[w_i < w_B(\hat{v}) | \hat{v}, v] = \pi \cdot \Pr[\varepsilon_i < w_B(\hat{v}) - v | \hat{v}, v] = \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right) \tag{138}$$

Equivalently, the highest value $z = v - \mu_{v|\hat{v}}$ at which all borrowed shares x are sold solves:

$$x = \pi \cdot \Phi\left(\frac{\bar{w}_B - z}{\sigma_\varepsilon}\right). \tag{139}$$

Rearranging this equation yields the following threshold value for z:

$$\bar{w_B} - \Phi^{-1}\left(\frac{x}{\pi}\right)\sigma_\varepsilon. \tag{140}$$

For z-values below this threshold value, all units x that were borrowed are also sold. Yet, if z takes higher values, only a fraction of all borrowed units x are sold.

Note that z captures the deviation of the true fundamental value v from the public market's estimate of the fundamental value v, based on the public signal $\hat{v}$. If this deviation is sufficiently positive, then the market is underpricing the asset. Whereas the public signal $\hat{v}$ is affected by the *aggregate* noise term b, hedge funds' signals are affected by *idiosyncratic* noise shocks ε. That is, the distribution of signals w_i is always centered around the true v and not shifted by an aggregate mispricing shock. As a result, when the market underprices the asset sufficiently, due to a sufficiently low realization of $\hat{v}$ corresponding to a low value of the noise term b, there will be some hedge funds that borrowed the asset, but in light of the realized value of $\hat{v}$ and the corresponding underpricing, no longer wish to sell at the bid.

Determining the selling threshold $w_B(\hat{v})$. We now discuss how the conditional threshold level $w_B(\hat{v})$ at which hedge funds sell (conditional on having borrowed the asset) is determined.

Given $\hat{v}$ and given the bid price $B(c,\hat{v},x)$, hedge funds anticipate the following profit from selling a security they already borrowed:

$$\text{Profit}(c,\hat{v},x,w_i,s) \ = B(c,\hat{v},x) - L(c,\hat{v},w_i) \tag{141}$$

where $L(c,\hat{v},w_i)$ is the liability of having to redeliver the asset. To characterize this liability, we generally have to bring back our state index s and account for transition dynamics. The HJB equation for the scaled liability $\tilde{L}(\hat{v},w_i,s)$ incurred from borrowing shares takes the form:

$$\begin{aligned} 0 = &- (r_f + \sigma_c(s)\chi - \mu_c(s)) \cdot \lambda c \cdot \tilde{L}(\hat{v},w_i,s) \\ &+ \lambda \cdot \{ce^v + \lambda c \cdot \mathbb{E}[e^v|w_i,\hat{v}]\,\mathbb{E}[\tilde{A}(\hat{v}',s) + \tilde{\kappa}(\hat{v}',s)] \\ &- \lambda c \cdot \tilde{L}(\hat{v},w_i,s)\} + \lambda c \cdot (\Lambda(s) \odot \mathbf{U}(s)) \cdot \tilde{\mathbf{L}}(\hat{v},w_i), \end{aligned} \tag{142}$$

where we use the fact that the conditional expectation of the future ask and commission is:

$$\mathbb{E}[A(c',\hat{v}') + \kappa(c',\hat{v}')|w_i,\hat{v}] = \mathbb{E}[e^v|w_i,\hat{v}] \cdot \mathbb{E}[A(c,\hat{v}') + \kappa(c,\hat{v}')], \tag{143}$$

since this period's $(w_i,\hat{v})$ have predictive power for c', but they have no predictive power for $\hat{v}'$. This simplifies to:

$$\begin{aligned} \tilde{\mathbf{L}}(\hat{v},w_i) = &\mathbb{E}[e^v|w_i,\hat{v}] \cdot (\Lambda \odot \mathbf{U} - diag(r_f + \sigma_c(s)\chi - \mu_c(s)) + \lambda)^{-1} \\ &\cdot (\mathbf{1} + \lambda\,\mathbb{E}[\tilde{\mathbf{A}}(\hat{v}') + \tilde{\kappa}(\hat{v}')]). \end{aligned} \tag{144}$$

To solve the quantitative model we consider the limiting case of $\lambda \to \infty$. This approach significantly simplifies the solution of the model as it separate the solution of the ratios *AoP*, *BoP* etc. from the HJB equations determining the level of P, which accounts for future Markov state transitions. In this case, the solution simplifies to:

$$\tilde{L}(\hat{v},w_i,s) = \mathbb{E}[e^v|w_i,\hat{v}] \cdot \mathbb{E}[\tilde{A}(\hat{v}',s) + \tilde{\kappa}(\hat{v}',s)] \tag{145}$$

or equivalently the non-scaled version:

$$L(c,\hat{v},w_i) = \mathbb{E}[e^v|w_i,\hat{v}] \cdot \mathbb{E}[A(c,\hat{v}') + \kappa(c,\hat{v}')], \tag{146}$$

where we drop the state s again for notational convenience, since we no longer have

to keep track of transitions of the state s. Scaling by $P(c,\hat{v}) = P(c)\,\mathbb{E}[e^v|\hat{v}]$ yields:

$$\begin{aligned}\frac{L(c,\hat{v},w_i)}{P(c,\hat{v})} =& \frac{\mathbb{E}[e^v|w_i,\hat{v}]\cdot\mathbb{E}[A(c,\hat{v}')+\kappa(c,\hat{v}')]}{P(c)\,\mathbb{E}[e^v|\hat{v}]}\\ &= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{P(c)\,\mathbb{E}[e^v|\hat{v}]}\,\mathbb{E}[\frac{A(c,\hat{v}')+\kappa(c,\hat{v}')}{P(c,\hat{v}')}P(c)\,\mathbb{E}[e^{v'}|\hat{v}']]\\ &= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot\mathbb{E}[\frac{A(c,\hat{v}')+\kappa(c,\hat{v}')}{P(c,\hat{v}')}\,\mathbb{E}[e^{v'}|\hat{v}']]\\ &= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot(AoP+\kappa oP), \end{aligned} \tag{147}$$

where the last equality holds since both AoP and κoP are independent of $\hat{v}$. Scaling the profit equation by $P(c,\hat{v}) = P(c)\,\mathbb{E}[e^v|\hat{v}]$ then yields

$$\frac{\text{Profit}(c,\hat{v},x,w_i)}{P(c,\hat{v})} = BoP(x) - \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot LoP. \tag{148}$$

where we define:

$$LoP \equiv AoP + \kappa oP. \tag{149}$$

The threshold type that sells at the bid, $w_B(\hat{v})$, solves

$$BoP(x) - \frac{\mathbb{E}\left[e^v\,\middle|\,\hat{v}, w_B(\hat{v})\right]}{\mathbb{E}\left[e^v\,\middle|\,\hat{v}\right]}\cdot LoP = 0, \tag{150}$$

with

$$\mathbb{E}[e^v|\hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right), \tag{151}$$

$$\mathbb{E}\left[e^v\,\middle|\,\hat{v}, w_i\right] = \exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2}\right), \tag{152}$$

or equivalently, solving for $w_B(\hat{v})$ yields:

$$w_B(\hat{v}) \;=\; \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\,\hat{v} \;+\; \frac{\sigma_v^2\left(\sigma_b^2+\sigma_\varepsilon^2\right) \;+\; \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\,\ln\!\Big(\frac{BoP(x)}{LoP}\Big). \tag{153}$$

We can now verify that $\bar{w}_B$ is indeed independent of $\hat{v}$, conditional on the remaining conjecture that κoP is independent of $\hat{v}$:

$$\begin{aligned}\bar{w}_B &= w_B(\hat{v}) - \mu_{v|\hat{v}} \\ &= \frac{\sigma_v^2\left(\sigma_b^2+\sigma_\varepsilon^2\right) + \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\ln\!\left(\frac{BoP(x)}{LoP}\right) - \left(\frac{\sigma_b^2}{\sigma_v^2+\sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right)\right). \end{aligned} \tag{154}$$

D.2.2 Borrowing Stage

As derived above, the expected trading-stage profit from selling the asset, conditional on an ex post realization of $\hat{v}$ and w_i is given by, scaled by the *unconditional* value $P(c)$ is given by:

$$BoP(x)\cdot\mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}]\cdot LoP. \tag{155}$$

At the borrowing stage, an agent observes w_i but not $\hat{v}$. Their willingness to pay, scaled by $P(c)$, is given by the expected scaled profit, given w_i, accounting for the fact that the agent does not sell if the expected trading profit would be negative:

$$\begin{aligned} SoP(w_i,x) &\equiv \mathbb{E}[(BoP(x)\cdot\mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}]\cdot LoP)^+|w_i] \\ &= \int_{-\infty}^{+\infty}(BoP(x)\cdot\mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}]\cdot LoP)^+ f(\hat{v}|w_i)d\hat{v}, \end{aligned} \tag{156}$$

where $f(\hat{v}|w_i)$ is the PDF of $\hat{v}$ conditional on observing w_i. We have already characterized $\mathbb{E}[e^v|\hat{v}]$ and $\mathbb{E}[e^v|w_i,\hat{v}]$ in equations (151) and (152). The pdf $f(\hat{v}|w_i)$ is normal with the following mean and variance:

$$\mathbb{E}[\hat{v}\mid w_i] = -\frac{\sigma_v^2}{2} + \frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2}\left(w_i - \left(-\tfrac{\sigma_v^2}{2}\right)\right), \tag{157}$$

$$\mathrm{Var}(\hat{v}\mid w_i) = \sigma_b^2 + \frac{\sigma_v^2\,\sigma_\varepsilon^2}{\sigma_v^2+\sigma_\varepsilon^2}. \tag{158}$$

Equilibrium loan fees. For completeness, we note that in the presence of abundant inventory ($\rho > \pi$), the threshold type $w_\phi(v)$ solves the break even condition, thereby also pinning down the exact borrowing quantity $x(v) = \pi F_w(w_\phi(v))$:

$$\frac{\phi_0}{P} = \int_{w_B^{-1}(w_\phi(v))}^{\infty}(BoP(\hat{v},x(v))\cdot\mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i = w_\phi(v),\hat{v}]\cdot LoP) f(\hat{v}|w_i = w_\phi(v))d\hat{v}. \tag{159}$$

For $\pi < \rho$, endogenous fees are zero, that is, $\phi oP = 0$.

In contrast, in the case of scarce inventory ($\pi > \rho$), the equilibrium borrowing volume is determine by market clearing $x = \rho$, which pins down the marginal type $w_\phi(v)$:

$$F_w(w_\phi(v)|v) \cdot \pi = \rho \tag{160}$$

where $F_w(\cdot|v)$ is the CDF of w_i given conditional on v. Note that

$$F_w(w_i|v) = F_w(v + \varepsilon_i|v) = F_\varepsilon(w_i - v) \tag{161}$$

where F_ε is the CDF of ε. Thus, we can rewrite the market clearing condition as:

$$F_\varepsilon(w_\phi(v) - v) \cdot \pi = \rho \tag{162}$$

which yields the explicit result:

$$w_\phi(v) = v + F_\varepsilon^{-1}\left(\frac{\rho}{\pi}\right). \tag{163}$$

For $\pi > \rho$, the equilibrium fee is then equal to the willingness to pay of a hedge fund observing the threshold signal $w_i = w_\phi(v)$, that is:

$$\frac{\phi(c,v)}{P(c)} = \phi oP(v) = SoP(w_i = w_\phi(v), x = \rho). \tag{164}$$

The average fee and fee yield is then given by

$$\lambda\, \mathbb{E}[\phi oP(v)]. \tag{165}$$

Note that this equation does not explicitly feature utilization for the following reason: in the case of abundant inventory ($\rho > \pi$) fees are zero, so utilization π/ρ is irrelevant. In the case of scarce inventory ($\pi > \rho$) utilization of the inventory is 100%, implying that every unit of inventory earns the fee.

Commissions κ. When there is abundant inventory $\rho > \pi$ the fee is zero and so are commissions κ. In contrast, when inventory is scarce, $\rho < \pi$, an agent committing their asset to provide ask quotes (before any market activity) expects to receive a fee $\mathbb{E}[\phi(c,v)]$. As there are $(1+\pi)$ units posted at the ask price in equilibrium, the total expected commission income must be:

$$(1+\pi)\,\mathbb{E}[\phi(c,v)], \tag{166}$$

where this commissions income is distributed across ask quote providers. The buyers picking up as quotes thus have to pay total commissions equal to that amount.

$$
\begin{aligned}
&(1+\pi)\,\mathbb{E}[\phi(c,v)] \\
&= \int_{-\infty}^{+\infty} \Big(\underbrace{\int_{-\infty}^{\infty} \pi \cdot (1-\Phi(\frac{w_A(\hat{v})-v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1}_{\text{Average volume given } \hat{v}} \Big) \cdot \kappa(c,\hat{v}) f(\hat{v}) d\hat{v} \\
&= \int_{-\infty}^{+\infty} \Big(\underbrace{\int_{-\infty}^{\infty} \pi \cdot (1-\Phi(\frac{w_A(\hat{v})-v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1}_{\text{Average volume given } \hat{v}} \Big) \cdot \kappa oP \cdot \mathbb{E}[e^v|\hat{v}] \cdot P(c) f(\hat{v}) d\hat{v}
\end{aligned}
\tag{167}
$$

This yields

$$
\kappa oP = \frac{(1+\pi)\,\mathbb{E}[\phi oP(v)]}{\int_{-\infty}^{+\infty}(\int_{-\infty}^{\infty} \pi \cdot (1-\Phi(\frac{w_A(\hat{v})-v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1) \cdot \mathbb{E}[e^v|\hat{v}] \cdot f(\hat{v}) d\hat{v}}. \tag{168}
$$

where

$$
w_A(\hat{v}) = \bar{w}_A + \mu_{v|\hat{v}} \tag{169}
$$

and where we have the pdf $f_v(v|\hat{v})$ with the following:

$$
v \,\Big|\, \hat{v} \;\sim\; \mathcal{N}\Big(\frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\,\hat{v} \;+\; \frac{\sigma_b^2}{\sigma_v^2+\sigma_b^2}\,(-\tfrac{\sigma_v^2}{2}),\; \frac{\sigma_v^2\,\sigma_b^2}{\sigma_v^2+\sigma_b^2} \Big).
$$

This constant ratio κoP together with a constant AoP implies that commissions are a constant fraction of the dollar amount paid by buyers for assets posted at the ask price.

Hedge fund alphas. We have the following average alpha among borrowers, conditional on a given state realization v (note that we are dropping the dependence on the Markov state s here), for $\rho < \pi$:

$$
\alpha(v) = \lambda \int_{-\infty}^{F_\varepsilon^{-1}(\frac{\rho}{\pi})} SoP(v+\varepsilon_i, x=\rho) \frac{f_\varepsilon(\varepsilon)}{\frac{\rho}{\pi}} d\varepsilon. \tag{170}
$$

In contrast, for $\rho > \pi$ we compute

$$\alpha(v) = \lambda \int_{-\infty}^{+\infty} SoP(v+\varepsilon_i, x = \pi) f_\varepsilon(\varepsilon) d\varepsilon \tag{171}$$

since (almost) all ε types have borrowed (not just a fraction ρ/π) and can now optimally decide whether to sell, where the benefit of that selling (conditional on having borrowed) is computed via *SoP*.

Then we compare the average alpha among borrowers:

$$\mathbb{E}[\alpha(v)] = \int_{-\infty}^{+\infty} \alpha(v) f_v(v) dv \tag{172}$$

against the average fee:

$$\lambda \, \mathbb{E}[\phi oP(v)]. \tag{173}$$

That is, hedge funds' average net alpha from borrowing is:

$$\mathbb{E}[\alpha(v)] - \lambda \, \mathbb{E}[\phi oP(v)]. \tag{174}$$

The hedge fund surplus from shorting, for $\rho < \pi$ is given by:

$$(\mathbb{E}[\alpha(v)] - \lambda \, \mathbb{E}[\phi oP(v)]) \cdot \frac{\rho}{\pi} \tag{175}$$

The hedge fund surplus, for $\rho > \pi$ is given by:

$$\mathbb{E}[\alpha(v)] \tag{176}$$

since the endogenous fee is zero (and exogenous costs from shorting approach zero) and since almost all hedge funds borrow the asset.

D.3 Derivation of Conditional Expectations

Using the fact that

$$\mathbb{E}[e^v | \hat{v}] = \exp\left(\mathbb{E}[v \mid \hat{v}] + \frac{1}{2} \mathrm{Var}(v \mid \hat{v}) \right). \tag{177}$$

we obtain the result that

$$\mathbb{E}[e^v | \hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_b^2} \hat{v} \right). \tag{178}$$

since:

$$\mathbb{E}[v \mid \hat{v}] = \mathbb{E}[v] + \frac{\mathrm{Cov}(v,\hat{v})}{\mathrm{Var}(\hat{v})}(\hat{v} - \mathbb{E}[\hat{v}]) = -\frac{\sigma_v^2}{2} + \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\left(\hat{v} + \frac{\sigma_v^2}{2}\right)$$
$$= \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v} - \frac{1}{2}\frac{\sigma_v^2\sigma_b^2}{\sigma_v^2+\sigma_b^2} \tag{179}$$

$$\mathrm{Var}(v \mid \hat{v}) = \sigma_v^2 \frac{\sigma_b^2}{\sigma_v^2+\sigma_b^2} \tag{180}$$

Next, we compute

$$\mathbb{E}\left[e^v \,\middle|\, \hat{v}, w_i\right] = \exp\left(\mathbb{E}[v|\hat{v},w_i] + \frac{1}{2}\mathrm{Var}(v|\hat{v},w_i)\right). \tag{181}$$

Define the vector of conditioning variables:

$$\mathbf{Y} = \begin{bmatrix} \hat{v} \\ w_i \end{bmatrix}. \tag{182}$$

Then we obtain the moments:

$$\mathbb{E}[v|\hat{v},w_i] = \mu_v + \Sigma_{v\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1}(\mathbf{Y} - \mu_{\mathbf{Y}}),$$

and

$$\mathrm{Var}(v|\hat{v},w_i) = \sigma_v^2 - \Sigma_{v\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{Y}v},$$

where

$$\Sigma_{v\mathbf{Y}} = \begin{bmatrix} \sigma_v^2 & \sigma_v^2 \end{bmatrix} \tag{183}$$

$$\Sigma_{\mathbf{YY}} = \begin{bmatrix} \sigma_v^2+\sigma_b^2 & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2+\sigma_\varepsilon^2. \end{bmatrix} \tag{184}$$

Computing the determinant of $\Sigma_{\mathbf{YY}}$

$$D = (\sigma_v^2+\sigma_b^2)(\sigma_v^2+\sigma_\varepsilon^2) - \sigma_v^4 = \sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2) + \sigma_b^2\sigma_\varepsilon^2,$$

we obtain the inverse matrix:

$$\Sigma_{\mathbf{YY}}^{-1} = \frac{1}{D}\begin{bmatrix} \sigma_v^2+\sigma_\varepsilon^2 & -\sigma_v^2 \\ -\sigma_v^2 & \sigma_v^2+\sigma_b^2 \end{bmatrix}.$$

Then we obtain:

$$\Sigma_{v\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1} = \frac{\sigma_v^2}{D}\begin{bmatrix} \sigma_\varepsilon^2 & \sigma_b^2 \end{bmatrix},$$

and

$$\begin{aligned}\mathbb{E}[v|\hat{v},w_i] &= -\frac{1}{2}\sigma_v^2 + \frac{\sigma_v^2}{D}\left[\sigma_\varepsilon^2(\hat{v}+\frac{1}{2}\sigma_v^2)+\sigma_b^2(w_i+\frac{1}{2}\sigma_v^2)\right] \\ &= \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right) - \frac{1}{2}\sigma_v^2 + \frac{\sigma_v^2}{D}(\sigma_\varepsilon^2\frac{1}{2}\sigma_v^2) + \frac{\sigma_v^2}{D}(\sigma_b^2\frac{1}{2}\sigma_v^2) \\ &= \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right) - \frac{1}{2}\left(\sigma_v^2 - \frac{\sigma_v^4}{D}(\sigma_\varepsilon^2+\sigma_b^2)\right). \end{aligned} \tag{185}$$

Given that

$$\Sigma_{v\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{Y}v} = \frac{\sigma_v^4(\sigma_b^2+\sigma_\varepsilon^2)}{D},$$

we obtain:

$$\mathrm{Var}(v|\hat{v},w_i) = \sigma_v^2 - \frac{\sigma_v^4(\sigma_b^2+\sigma_\varepsilon^2)}{D}.$$

Then we obtain

$$\mathbb{E}[v|\hat{v},w_i] + \frac{1}{2}\mathrm{Var}(v|\hat{v},w_i) = \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right), \tag{186}$$

and thus:

$$\mathbb{E}\left[e^v \,\middle|\, \hat{v},w_i\right] = \exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2}\right).$$

E Custodian Commitment and Market Power

In this section, we explicitly consider the two-sided market structure of the securities lending market where custodians interact with both beneficial asset owners and securities borrowers (hedge funds). This analysis endogenizes our baseline model's assumption of custodian commitment and market power and show how the two channels are economically linked. Commitment plays an important role in our analysis of the intermediated securities lending market. It is a custodian's commitment to secrecy and non-opportunistic behavior that is one of the key distinguishing features of this market structure, relative to the direct lending market where information leakages occur in equilibrium (see Section 5). Absent commitment, a custodian could opportunistically use the information it obtains from hedge funds' borrowing decisions. In particular, observing hedge funds' borrowing demand, a custodian could front-run a hedge fund's trades or share information with clients or affiliated traders.

Regarding the economic link between custodian commitment and market power, we show that a custodian has stronger incentives to stay committed to refraining from opportunistic behavior when future retained fee revenue is higher and the in-

termediary sector is most concentrated. This interaction generates endogenous scale economies: A custodian that attracts all asset owners' assets can stay committed while passing on the largest fraction of the fee revenue to them. In other words, overall, custodian agency rents are minimized, and investor rents maximized, when intermediation of securities lending is most concentrated.

We further show how limited percolation of information related to opportunistic behavior can lead to the existence of multiple custodians that each exercise market power.

Setup. The setup for our analysis of the two-sided market structure is motivated by empirical evidence that beneficial owners such as mutual funds maintain very stable relationships to typically one custodian lender and sign long-term revenue sharing agreements. Custodians then interact with the demand side and set fees at higher frequencies. Our setup captures these different frequencies of how these two sides of the market interact with the intermediaries: longer-term contracts signed at date $t = 0$ pin down the delegation of stock lending to intermediaries who then interact with borrowers at higher frequencies.

Suppose that at date $t = 0$, there are $I \geq 3$ potential custodians who initially do not have inventory but can potentially attract inventory from asset owners. To streamline the analysis, we assume that a custodian can front run a hedge fund that demands shares it wishes to borrow, that is, a custodian's sale order is filled logically before the hedge fund can sell.[4] This assumption implies that if (1) it is not incentive compatible for a custodian to refrain from opportunistically front running, and (2) a hedge fund's borrowing reveals its intention to short sell, the hedge fund cannot profit from short-selling attempts. This is because the custodian would front run a hedge fund whenever the hedge fund borrows shares, leaving the latter unable to sell the borrowed shares. Moreover, we maintain the tie-breaking rule introduced in the direct lending market whereby a hedge fund does not borrow if it is indifferent (that is, if it does not strictly benefit from borrowing). This assumption rules out the possibility of an equilibrium where hedge funds always borrow (and thus do not reveal information) if the equilibrium loan fee is zero.

Jointly, these facts imply that the shares lending market requires custodian incentive compatibility to hold (that is, no incentive for custodians to front run) in order for there to be positive lending volume in equilibrium.

Distinguishing the intermediated market from the direct lending market. The direct-lending and intermediated-lending markets structurally differ in the channels through which information leakages may occur. In the case of the direct lending mar-

[4]When orders are filled based on a pro-rata allocation rule, this is consistent with a custodian having unrestricted order sizes, whereby an order of infinite size crowds out all other traders and leads existing offers in the limit order book to be allocated to the custodian.

ket considered in Section 5, aggregate lending volume becomes publicly observable and allows regular investors that post limit orders to condition their quotes on that (aggregate) information before a hedge fund can sell its shares.

In contrast, in the case of the custodian-intermediated lending market, a custodian can use the information obtained from observing bilateral hedge fund demand to trade on its own book, while bid and ask quotes are not updated since regular investors do not obtain that information. As such, this latter market structure also rules out the possibility that an atomistic asset owner can become a custodian itself. That is, direct lending is infeasible under this custodian-intermediated market structure. If an asset owner could also be its own custodian, it would potentially also adjust the bid and ask quotes it posts as a liquidity provider as a function of the hedge fund demand it observes. In that case, the distinction between the two market structures would no longer apply. In this context, it is also worth noting that in practice, agents and institutions incur fixed operational costs to gain access to the OTC market. Even an arbitrarily small fixed cost of this type would make it unprofitable for an individual, atomistic investor to become a player in this market.

Structure of the date-0 enrollment game. The date-0 custodian enrollment game features the following logical order of events:

1. Custodians $i = 1, ..., I$ make simultaneous contractual offers specifying the fraction δ_i of the fee revenue each retains in future periods. The contract specifies that an asset owner enrolling with a custodian will always deposit shares with that custodian unless the custodian acts opportunistically and forfeits future business from hedge funds (see details below). After enrolling at date 0, an agent is not required to always make shares available for lending. Rather, the contract stipulates that conditional on wishing to lend, an agent is committed to depositing the shares with the custodian. Alternatively, an agent might not have shares in a given period or might want to use the shares to offer them for sale via an ask quote in the limit order market.

2. All agents that can potentially become asset owners at some time $t \in [0, \infty)$ decide whether to accept one of the custodians' offers.

Note that the date-0 contract does not need to directly stipulate what fees a custodian will charge hedge funds in the future. Given the linear fee sharing rule, there is effectively no divergence in the objective of the custodian and the investor depositing shares on a period-by-period level.

Payoffs from the date-0 enrollment game. In the custodian enrollment game we consider, each asset owner's payoff is the continuation payoff that is consistent with

the equilibrium outcome(s) of the game described in our baseline model, subject to the modification that there may be multiple custodians. In particular, consider an asset owner that has signed up with custodian i and chooses to deposit shares with that custodian at a given point in time t when there is a Poisson shock $dN_t = 1$ (as opposed to making the shares available for sale by posting an ask quote in the limit-order market). If a hedge fund borrows π units from that custodian, the asset owner receives fee income equal to

$$(1-\delta_i)\frac{\pi\phi_{i,t}}{\rho_i}, \tag{187}$$

where $\phi_{i,t}$ is the fee offer from custodian i and ρ_i is the inventory of custodian i. In the above equation, we assume that $\pi \leq \rho_i$, that is, the custodian has enough inventory to service the full hedge fund demand. Otherwise, utilization is at most 100%.

Definition 1 (Equilibrium of the date-0 enrollment game)**.** *Custodians' offers* $\{\delta_i\}_{i=1}^I$ *and asset owner enrollment decisions constitute an equilibrium of the custodian-enrollment game if no asset owner can improve its continuation payoff by deviating to another custodian (or no custodian), and if there is no custodian that can be better off by changing its offer* δ_i*.*

Proposition 6. *The date-0 enrollment game features an equilibrium in which all asset owners enroll with one custodian* $i \in \{1,...,I\}$*, and this custodian* i *retains the minimum fraction of fee income ensuring that its incentive compatibility (IC) constraint is satisfied in all future periods,* $\delta_i = \delta_{min}$*. All other custodians* $j \neq i$ *make the same offer,* $\delta_j = \delta_{min}$*. Starting at date* $t = 0$*, custodian* i *chooses the monopolistic lending fee characterized in analysis of the baseline model.*

Proof of Proposition 6. To verify the proposed equilibrium, we proceed in three steps: First, we characterize the conditions under which a custodian i that successfully enrolled all asset owners at date 0 has proper incentives to refrain from opportunistic front running. Second, we determine an asset owner's payoff from deviating unilaterally to another custodian $j \neq i$ who doesn't have any other enrollees. We will show that this other custodian cannot satisfy the incentive compatibility (IC) constraint and thus cannot attract hedge fund demand. As a result, the payoff from this deviation for an asset owner is zero. Finally, we show that anticipating asset owners' behavior, custodians at date $t = 0$ cannot choose offers $\delta_j \neq \delta_{min}$ that would make them better off.

Step 1: Custodian incentive compatability. Our setup naturally lends itself to analyzing custodians' commitment through reputation concerns by resorting to a version of

the Folk Theorem (see, e.g., Fudenberg et al., 1994). Specifically, the series of stage games occurring upon innovations to the process N_t represent a repeated game with infinite horizon.

Opportunistic deviations and their detection. As noted above, a custodian's potential opportunistic action is to front run a hedge fund: While a custodian has to comply with its promise to deliver the shares to the borrowing hedge fund (implied by the take-it-or-leave-it-offer for a lending contract), it can itself sell (buy) additional units at the bid (ask).[5] The underlying idea is that custodians are often large asset management companies or prime brokers that have additional shares they can sell or buy. Alternatively, they could share the information with clients and receive corresponding kickbacks in return.

However, such opportunistic behavior can be detected ex post, after a stage game: other investors can observe after a trading date whether the custodian traded upon the information revealed by hedge funds' borrowing demand. This is because such trades would crowd out the hedge fund's and liquidity traders' allocations relative to the case without custodian demand. For example, suppose the bid quote in the limit order book is for one unit of the asset (which is the case in the equilibrium of the baseline model). Further, suppose that a hedge fund attempts to sell π units (and thus has borrowed π units) and liquidity traders try to sell $(1-\pi)$ units. Then, if a custodian effectuates a market order for a sale of a positive quantity, the hedge fund and the liquidity traders get crowded out (by the pro-rata rule for allocations) and do not receive the full quantity they expected, since only 1 unit is available in total at the bid. Thus, both hedge funds and liquidity traders then observe ex post that they were not able to sell the anticipated π and $(1-\pi)$ units respectively, and can detect the custodian's opportunistic behavior. Note that the liquidity traders always want to sell $(1-\pi)$ units and the custodian optimally wants to sell exactly in those states of the world where the hedge fund would also want to sell (that is, the custodian aims to mimic the hedge fund's intended trade). Thus, opportunistic front running for sales would involve the crowding out of the hedge fund's and liquidity traders' trades. As detailed below, this detection mechanism further operates analogously (albeit probabilistically) for any positive quantity opportunistically purchased by the custodian (which could in principle occur if a hedge fund does not borrow shares, thereby revealing positive news).

In our baseline analysis of this market we presume that, at the end of each stage game, hedge funds truthfully and publicly report if their trades have been crowded out, which is an incentive compatible action by indifference.[6]

[5] Similar implications would obtain if the custodian could monetize trades equivalent to those positions by sharing information with clients or affiliated traders.

[6] For simplicity, we assume that liquidity traders do not report when getting crowded out.

Penalty associated with detected opportunistic behavior. If a custodian is reported to have front-run a hedge fund, a new enrollment game akin to the date-0 enrollment game is initiated, with the difference that this game effectively excludes the custodian that acted opportunistically (this custodian does not receive any enrollees since hedge funds no longer "trust" the custodian that acted opportunistically). This switch to an equilibrium where other custodians attract inventory going forward yields the worst possible subgame perfect equilibrium for a custodian. We use this subgame as the threat that incentivizes commitment by a custodian.

Outside options for required trades. The above described allocations in the limit order market imply that, when considering off-equilibrium deviations by the custodian that lead to the crowding out of other agents' trades, hedge funds having to return their shares and liquidity traders would not be able to execute their desired trades at the regular posted bid and ask quotes. To address this off-equilibrium issue, we assume that additional outside options are available to sell at a quote $B_f \ll B$ and buy at a price $A_f \gg A$.

Monopolistic custodian's continuation value under commitment. It is useful to define the continuation value of a custodian i conditional on it receiving all future securities lending business (as assumed in our baseline model). Given the contract signed at date $t = 0$, the custodian obtains a fraction $\delta_i \in (0,1)$ of the fee revenue. For this monopolistic custodian the continuation value is then given by:

$$\delta_i \cdot \mathbb{E}_t \int_t^\infty \frac{m_\tau}{m_t} F_v(v_\phi) \frac{\pi \phi_\tau}{\rho} dN_\tau = \delta_i \cdot \frac{\lambda \cdot F_v(v_\phi) \frac{\pi \phi(c)}{\rho}}{r_f + \sigma_c \chi - \mu_c}. \tag{188}$$

In addition to this retained fee income, a custodian might also benefit from fee pass-through if it is affiliated with an open end fund. For example, an asset management company like Vanguard might indirectly benefit from the fraction of fee income they pass on to their own fund holders, as doing so improves their funds' performance and allows attracting more clients. That is, the pass-through of fees can generate positive spillover effects to other parts of the asset management business. Having said that, going forward, we focus on the explicit fee retention benefit captured in equation (188).

The value of opportunistic one-shot deviations. As noted above, we consider the case where a custodian does not face position limits when trading on its own books. As such, a custodian can, in the limit, effectuate an infinitely large market order, which by the pro-rata rule for allocations yields it the *finite* quantity posted at the bid or ask (depending on a sale or purchase order). In a given stage game, a

custodian's expected trading gain from opportunistically selling the asset conditional on observing a hedge fund borrowing ($v < v_\phi$) is then given by:

$$B(c) - \mathbb{E}[e^v | v < v_\phi] P(c), \tag{189}$$

which accounts for the fact that the custodian can sell one unit at the bid, since only one unit is available in equilibrium. Conditional on observing that a hedge fund borrows, the custodian can infer that $v < v_\phi$ (on the equilibrium path where the hedge fund expects that the custodian is committed, the hedge fund follows the same threshold strategy as in our baseline model). Conditional on this signal, a custodian values keeping the asset at $\mathbb{E}[e^v | v < v_\phi] P(c)$ but can sell it at the bid quote $B(c)$.

Analogously, the custodian's expected trading gains from opportunistically buying assets at the ask price conditional on observing no hedge fund borrowing (which signals that $v > v_\phi$) are given by the following two expressions, which condition on the cases with and without open interest, respectively:

$$\mathbb{E}[e^v | v > v_\phi] P(c) - A_0(c), \tag{190}$$

$$(1+\pi) \cdot (\mathbb{E}[e^v | v > v_\phi] P(c) - A_1(c)). \tag{191}$$

Custodian incentive compatibility with full enrollment. The current-period fee of a stage game is collected irrespective of whether the custodian engages in opportunistic behavior or not, implying that this current-period fee payment does not affect incentives. Custodian commitment is sustainable if the continuation value from future business outweighs the short-run temptation to front run a hedge fund. The worst possible subgame perfect equilibrium from the perspective of a custodian is to not receive any future business conditional on hedge funds' trades getting crowded out, which occurs when a custodian chooses to trade opportunistically.

Conditional on observing a hedge fund borrowing, it is then incentive compatible for a custodian i that currently has all asset owners enrolled to refrain from opportunistic behavior when the following IC constraint is satisfied:

$$\underbrace{B(c) - \mathbb{E}[e^v | v < v_\phi] P(c)}_{\text{One-shot trading profit}} \leq \underbrace{\delta_i \cdot \frac{\lambda \cdot F_v(v_\phi) \frac{\pi \phi(c)}{\rho}}{r_f + \sigma_c \chi - \mu_c}}_{\text{Continuation value}}. \tag{IC 1}$$

Note that conditional on hedge fund borrowing occurring, hedge funds get crowded out by a custodian's trade with probability one, implying that detection occurs with probability one.

Next, consider the IC constraint conditional on observing that a hedge fund is *not* borrowing. In this case, the custodian might find it optimal to buy the asset (since the expectations of v are truncated from below). Conditional on not borrowing the asset,

a hedge fund will want to buy shares with the probabilities:

$$\frac{\Pr[v \geq \log(A_0/P)]}{\Pr[v > v_\phi]} \text{ and } \frac{\Pr[v \geq \log(A_1/P)]}{\Pr[v > v_\phi]} \tag{192}$$

in the cases of no existing short interest and positive short interest, respectively. If a custodian opportunistically buys the stock after observing no borrowing demand from the hedge fund, these probabilities also represent the probabilities of detection. Conditional on observing a hedge fund not borrowing, it is then incentive compatible for a custodian to refrain from opportunistic behavior (buying the asset) when the following relations hold conditional on no short interest and positive short interest, respectively:

$$\underbrace{(\mathbb{E}[e^v|v > v_\phi]P(c) - A_0(c))}_{\text{One-shot trading profit}} \leq \underbrace{\frac{\Pr[v \geq \log(A_0/P)]}{\Pr[v > v_\phi]} \cdot \delta_i \cdot \frac{\lambda \cdot F_v(v_\phi)\frac{\pi\phi(c)}{\rho}}{r_f + \sigma_c\chi - \mu_c}}_{\text{Expected Continuation Value Loss}}, \tag{IC 2}$$

$$(1+\pi)(\mathbb{E}[e^v|v > v_\phi]P(c) - A_1(c)) \leq \frac{\Pr[v \geq \log(A_1/P)]}{\Pr[v > v_\phi]} \cdot \delta_i \cdot \frac{\lambda \cdot F_v(v_\phi)\frac{\pi\phi(c)}{\rho}}{r_f + \sigma_c\chi - \mu_c}. \tag{IC 3}$$

A custodian that enrolled all asset owners has proper incentives as long as all three IC constraints are satisfied. Maximum passthrough of fees to investors is achieved if δ_i is chosen to be the minimum δ such that (IC 1), (IC 2), and (IC 3) are satisfied:

$$\delta_{min} = \min\{\delta_i \in [0,1] : \text{IC 1, IC 2, IC 3}\}. \tag{193}$$

Note that in principle there might be no $\delta_i \in [0,1]$ satisfying these conditions, in particular, if the present value of *all* future fee income is smaller than the profits from a one-shot deviation. However, the following proposition highlights that the conditions can always be satisfied provided that the expected period length is sufficiently short, since trading gains scale with period length.

Proposition 7. *(Folk Theorem) Keeping the annual volatility of hedge funds' private information, $(\lambda \cdot (e^{\sigma_v(s)^2} - 1))^{1/2}$, fixed, if the expected period length $1/\lambda$ is below a threshold value, the incentive compatibility conditions (IC 1), (IC 2), and (IC 3) are satisfied.*

The proposition follows from the fact that holding the annual volatility of hedge funds' private information constant, the trading gains (189),(190), and (191) approach zero as the expected period length $1/\lambda$ goes to zero (that is, for $\lambda \to \infty$).

In contrast, the continuation values in (IC 1) through (IC 3) are strictly positive and do not approach zero.

Step 2: Payoff from an asset owner's unilateral deviation. We now examine what happens if in the date-0 enrollment game, an atomistic asset owner deviated from custodian i (who attracts all other asset owners) to enrolling with a different custodian $j \neq i$. This custodian j can attract hedge fund demand at a positive fee only if hedge funds know that custodian j has proper incentives. This is because a hedge fund cannot make any profits from trading when borrowing from a custodian that engages in front running (due to the above-described crowding-out effect). As a result, paying a positive borrowing fee to a custodian without incentives is never optimal.

Correspondingly, an asset owner can only expect to receive positive fee income from a custodian j if that custodian's IC constraints are satisfied. However, when enrolling only one atomistic asset owner, this custodian's continuation value from future fee income is always infinitesimal, for all $\delta_j \in [0,1]$. In contrast, if a hedge fund were to borrow from custodian j with the intention to short sell, the benefit of front running this hedge fund would be a value of positive and discrete magnitude, as expressed in (189). As a result, custodian j cannot have proper incentives, implying that it will not be able to attract any hedge fund demand at positive fees in the future. As a result, deviating to a custodian j will yield an asset owner zero fee income in the future. An individual atomistic investor is best off posting its inventory with custodian i as well, as this is the only custodian who has the critical mass of inventory needed to have proper incentives.

Coordination on a symmetric pure-strategy Nash equilibrium. In a symmetric pure strategy Nash equilibrium, all asset owners choose the same custodian. As shown in steps 1 and 2 above, conditional on expecting that all other asset owners play a pure strategy of selecting custodian i, it is optimal for an atomistic asset owner to also select that custodian.

Several formal mechanisms for coordination on a pure strategy are typically used in the literature: (1) Pre-play communication, that is, allowing asset owners to communicate before making their decisions can help them coordinate on a single custodian. This communication does not need to be binding but can help establish a focal point. (2) Focal points: one custodian may stand out as a natural choice, for example due to its existing reputation or visibility in the asset management industry. These focal points can help agents converge on the same custodian without explicit communication. (3) Mechanism design: A central authority or mechanism can be designed to recommend or enforce a particular choice. For instance, a regulator might suggest a default custodian to ensure coordination.

We consider the following focal point ranking convention: It is common knowledge that asset owners choose the custodian offering the lowest δ_j subject to the con-

straint that $\delta_j \geq \delta_{min}$. In case of ties in the offers, asset owner's choose the custodian with the lowest index j (e.g., custodian 1 if $\delta_1 = \delta_2 = \delta_{min}$).

This focal point ranking is motivated by the following observations: First, choosing any δ_j that is lower than δ_{min} characterized in equation (193) is too low to ensure that a custodian satisfies the IC constraints even in the best-case scenario where it attracts all enrollees and acts as a monopolist in the lending market after date 0. Thus, it is common knowledge that a custodian offering a $\delta_j < \delta_{min}$ will lack proper incentives and will thus cause the custodian to generate no fee income in the future. As such, asset owners are better off not enrolling with any custodian choosing $\delta_j < \delta_{min}$. Second, a custodian choosing $\delta_j > \delta_{min}$ retains more rents than strictly necessary to ensure proper incentives. In sum, this focal point ranking allows coordination on the best equilibrium from the perspective of maximizing the total surplus for asset owners.

Step 3: Custodians' Deviations. Given the focal point ranking described above, it is weakly optimal for each custodian j to choose $\delta_j = \delta_{min}$, conditional on expecting that other custodians will make that choice. For custodian 1, this choice is strictly optimal, as deviating from this value will cause custodian 2 to attract all enrollees.

Naturally, absent the considered focal point ranking, other equilibria exist in this setting. However, the scope of the analysis presented here is to highlight economic forces that sustain commitment and market power as considered in our baseline model.

Coexistence of multiple custodians with market power. While empirically, the intermediary sector in the securities lending market is highly concentrated (see Section 2), it is still the case that several lending agents frequently have inventory available for lending. One may wonder why competition does not emerge in this case.

In this section, we therefore formally discuss how limited percolation of information related to opportunistic intermediary behavior can sustain multiple custodians that exercise market power simultaneously. In particular, differences in geographic location and personal networks may cause information about a custodian's opportunistic behavior to percolate only among subgroups of hedge funds. This channel adds to other mechanisms that can in principle sustain market power in the presence of multiple intermediaries, such as tacit collusion.

Formally, suppose that every time there is an innovation to N_t, the newly arriving hedge fund representing a current cohort belongs to one of two groups, determined by nature with equal probability. Information about opportunistic deviations by a custodian percolates only within each group. That is, if a custodian front runs a hedge fund from group 1, only hedge funds from that group learn about this opportunistic action.

Further, suppose that the focal point ranking convention is such that conditional on a tie in offers, $\delta_1 = \delta_2$, asset owners rank custodians 1 and 2 equally.

Then there exists an equilibrium where asset owners mix in their enrollment decisions between custodians 1 and 2, enrolling with 50% probability with each custodian (symmetry ensures indifference due to identical continuation payoffs). Given the atomistic nature of asset owners, custodians 1 and 2 can then attract exactly 50% of the available inventory, which is sufficiently large to service a given hedge fund's demand π provided that ρ_0 is sufficiently large.

Absent limits to information percolation, two custodians would compete for the demand of every arriving hedge fund. Yet with limited information percolation this does not necessarily obtain. In particular, suppose that hedge funds from group 1 are expected to always approach custodian 1 in the future, and hedge funds from group 2 approach custodian 2. Given these expectations, in the current period, a hedge fund from group 1 knows that custodian 2 would not have proper incentives, since front running the hedge fund would be information that percolates to group-1 hedge funds only, but these hedge funds are not expected to be future customers of custodian 2. As a result, the hedge fund from group 1 can only make profits by approaching custodian 1. Since custodian 1 knows that it is the only committed lender for the hedge fund from group 1, it can charge monopolistic fees. The expectations that group-1 hedge funds borrow exclusively from custodian 1 and group-2 hedge funds borrow exclusively from custodian 2 are thus self-sustaining.

By introducing groups of hedge funds, this extension to our model captures a mechanism that is similar to two-sided reputation in the presence of long-lived players: in such a setting, a hedge fund would realize that it has to repeatedly approach a single custodian to be viewed as a repeat customer that is associate with continuation value from the perspective of a custodian. In this case, reputation is also "local" in nature: If the hedge fund is not going to custodian 1, then this specifically harms the reputation in that one relationship.

F Additional Tables

Panel A. Removing Observations Around Voting Record Dates

Lending Fee (annual, in percent)					
Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.261	0.252	0.264	0.491	3.897
80-90%	0.263	0.281	0.452	2.125	17.067
90-95%	0.266	0.435	1.109	5.992	29.834
95-98%	0.291	1.144	2.844	13.473	46.027
$\geq 98\%$	2.061	6.263	11.380	37.578	73.728

Panel B. Removing Observations Around Ex-Dividend Dates

Lending Fee (annual, in percent)					
Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.206	0.211	0.245	0.411	3.913
80-90%	0.214	0.253	0.438	2.075	18.113
90-95%	0.222	0.396	1.064	6.057	31.060
95-98%	0.245	1.040	2.779	13.985	47.841
$\geq 98\%$	1.860	6.172	11.583	39.177	74.410

Table A1: Lending Fee

This table reports the time-series average of lending fees for the 25 size and fee yield groups. In panel A, we remove observations 15 days before and after voting record dates. Following Aggarwal et al. (2015), we collect the data of voting record dates from ISS. In panel B, we remove observations 15 days before and after dividend ex-dividend dates. The lending fee of group i in quarter t is computed as

$$LendingFee_t^i = \frac{\sum_j ValueOnLoan_{j,t}^i \times LendingFee_{i,t}^i}{\sum_j ValueOnLoan_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $LendingFee_{j,t}^i$ and $ValueOnLoan_{j,t}^i$ are lending fees and value of shares on loan for stock j in group i, respectively. Lending fees are annualized and in percentage. The sample period is from 2007 through 2021.

Panel A. Spread Yield (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.04	0.08	0.12	0.07	0.62
80-90%	0.07	0.13	0.19	0.24	1.58
90-95%	0.08	0.17	0.25	0.44	1.93
95-98%	0.10	0.21	0.28	0.57	2.33
$\geq 98\%$	0.14	0.27	0.39	0.80	2.43

Panel B. Alternative Spread Yield (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.07	0.14	0.23	0.23	1.95
80-90%	0.23	0.45	0.63	2.28	29.85
90-95%	0.30	0.67	0.95	3.85	41.43
95-98%	0.45	0.95	1.24	6.01	53.56
$\geq 98\%$	0.78	1.37	2.08	10.74	75.49

Table A2: Spread Yield

This table reports the time-series average of spread yields for the 25 size and fee yield groups. The spread yield of group i in quarter t is computed as

$$SpreadYield_t^i = \frac{\sum_j SpreadRevenue_{j,t}^i}{\sum_j MarketCapitalization_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $SpreadRevenue_{j,t}^i$ is the sum of daily spread revenue of stock j in group i over quarter t. Spread revenue on a given day is the share trading volume times the bid-ask spread at the market close. An alternative measure of spread yield is defined as spread revenue scaled by the dollar value of share inventory,

$$SpreadYield_t^i = \frac{\sum_j SpreadRevenue_{j,t}^i}{\sum_j InventoryValue_{j,t}^i}, \quad i = 1,2,\ldots 25,$$

where $InventoryValue_{j,t}^i$ is the value of lenders' shares in inventory. Spread yields are annualized and in percentage. The sample period is from 2007 through 2021.

Markov Matrix across the 25 Size and Fee Yield Groups

From/To		Large-Cap					Mid-Cap					Small-Cap					Micro-Cap					Nano-Cap				
Fee Yield Percentiles		≤ 80%	80-90%	90-95%	95-98%	≥ 98%	≤ 80%	80-90%	90-95%	95-98%	≥ 98%	≤ 80%	80-90%	90-95%	95-98%	≥ 98%	≤ 80%	80-90%	90-95%	95-98%	≥ 98%	≤ 80%	80-90%	90-95%	95-98%	≥ 98%
Large-Cap	≤ 80%	0.96	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80-90%	0.45	0.39	0.12	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	90-95%	0.11	0.28	0.42	0.15	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	95-98%	0.06	0.07	0.23	0.49	0.13	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	≥ 98%	0.07	0.01	0.06	0.17	0.67	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mid-Cap	≤ 80%	0.01	0.00	0.00	0.00	0.00	0.95	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80-90%	0.00	0.00	0.00	0.00	0.00	0.32	0.53	0.12	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	90-95%	0.00	0.00	0.00	0.00	0.00	0.07	0.28	0.49	0.13	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	95-98%	0.00	0.00	0.00	0.00	0.01	0.04	0.06	0.22	0.53	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	≥ 98%	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.01	0.22	0.69	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small-Cap	≤ 80%	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.92	0.03	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80-90%	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.28	0.54	0.10	0.01	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	90-95%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.06	0.27	0.45	0.12	0.02	0.03	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	95-98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.04	0.26	0.45	0.14	0.01	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	≥ 98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.04	0.22	0.59	0.00	0.01	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00
Micro-Cap	≤ 80%	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.95	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80-90%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.33	0.51	0.07	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	90-95%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.02	0.06	0.26	0.43	0.13	0.03	0.01	0.01	0.00	0.00	0.00
	95-98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.04	0.01	0.07	0.28	0.39	0.15	0.01	0.01	0.01	0.01	0.00
	≥ 98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.05	0.00	0.03	0.05	0.24	0.55	0.00	0.01	0.01	0.01	0.01
Nano-Cap	≤ 80%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.01	0.01	0.00	0.00	0.85	0.03	0.00	0.00	0.00
	80-90%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.02	0.02	0.01	0.29	0.47	0.11	0.02	0.01
	90-95%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.04	0.02	0.07	0.30	0.39	0.11	0.03
	95-98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.04	0.02	0.09	0.26	0.40	0.13
	≥ 98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.01	0.04	0.07	0.24	0.52

Table A3: Markov Matrix

This table reports the Markov matrix of quarterly transition rates of firms across the 25 size and fee yield groups. The quarter-t transition rate from group i to group j is computed as $p_t^{i\to j} = \frac{TotalMarketValue_t^{i\to j}}{\sum_{j=1}^{25} TotalMarketValue_t^{i\to j}}, i, j = 1, 2, \ldots, 25$, where $TotalMarketValue_t^{i\to j}$ is the sum of market capitalizations of stocks moving from group i to group j over quarter t. The Markov matrix presents the time-series average of transition rates, which are standardized such that the sum of rates in each row equals one.

Idiosyncratic Volatility (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	19.75	25.52	32.99	26.88	59.57
80-90%	23.13	30.75	43.25	42.57	93.27
90-95%	24.41	34.78	47.39	51.93	108.54
95-98%	26.51	39.22	48.66	64.97	128.20
$\geq 98\%$	27.61	42.62	61.11	73.77	135.27

Table A4: Idiosyncratic Volatility
This table reports the time-series average of value-weighted stock idiosyncratic volatilities across the 25 stock groups. A stock's idiosyncratic volatility is estimated as the standard deviation of residuals from regressions of the stock's daily returns on the CRSP value-weighted market returns over a one-year rolling window. All numbers are annualized and reported in percent.

Volatilities of Hedge Funds' Signals (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.60	0.49	0.72	0.65	1.23
80-90%	1.08	1.22	2.29	5.56	77.12
90-95%	1.24	1.06	5.50	22.67	97.10
95-98%	1.31	4.56	15.77	56.46	95.14
$\geq 98\%$	8.34	28.68	54.08	73.77	135.27

Table A5: Estimated Volatilities of Informed Traders' Signals
This table reports the estimated volatilities of informed traders' signals in each group i. Specifically, we compute for each state s: $\sqrt{\lambda\left(e^{\sigma_v(s)^2}-1\right)} \approx \sqrt{\lambda}\sigma_v(s)$. All numbers are annualized and reported in percent.

Hedge Funds' Signal-to-Noise Volatility Ratios (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	2.00	2.07	2.20	2.01	2.66
80-90%	2.08	2.37	2.15	3.46	2.71
90-95%	2.02	4.50	2.21	2.68	3.77
95-98%	2.21	2.93	2.00	2.54	5.63
$\geq 98\%$	2.00	2.22	2.21	5.38	5.98

Table A6: Estimated Signal-to-Noise Volatility Ratios of Hedge Funds
This table reports the estimated signal-to-noise volatility ratios $\sigma_v/\sigma_\varepsilon$ in each group i. All numbers are reported in percent.

Inventory of Lendable Shares

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	3.18	0.73	0.37	0.31	0.19
80-90%	0.90	0.23	0.14	0.10	0.09
90-95%	0.53	0.15	0.10	0.07	0.07
95-98%	0.30	0.10	0.07	0.06	0.06
$\geq 98\%$	0.10	0.06	0.05	0.05	0.05

Table A7: Estimated Inventories of Lendable Shares
This table reports the estimated inventories available for lending $\rho = \rho_0 - 1$ in each group i.