Market Power in the Securities Lending Market

Shuaiyu Chen[*] Ron Kaniel[†] Christian C. Opp[‡]

May 30, 2026

Abstract

We document market power in U.S. equity securities lending and examine its origins and consequences. We develop and estimate a dynamic model showing that the OTC custodian-intermediated structure, which prevails across the world's largest financial markets, emerges as an equilibrium response to short sellers' information leakage concerns. Despite paying non-competitive fees that raise asset owners' valuations by up to 100%, short sellers prefer this prevailing intermediated structure over a centralized alternative, particularly for smaller stocks. This is because securities lending is not a standard product market: the demand side prefers opacity. Our results inform market design and transparency regulation.

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

[*]University of Virginia, Darden School of Business, Charlottesville, VA 22906. Email: chens@darden.virginia.edu.
[†]University of Rochester, Simon School of Business, Rochester, NY 14627. Email: ron.kaniel@simon.rochester.edu.
[‡]University of Rochester, Simon School of Business, Rochester, NY 14627. 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 market structure of securities lending, which predominantly operates through opaque, custodian-intermediated over-the-counter (OTC) arrangements across the largest financial markets world-wide. In particular, we evaluate the quantitative implications of this structure for market outcomes such as fees and informed traders' rents and contrast these with those that would emerge under a counterfactual centralized competitive market. Somewhat paradoxically, although custodian lenders charge substantial non-competitive fees under the prevailing market structure, we find that this arrangement dominates a centralized competitive 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%. Further, we provide evidence 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 consistently have significant excess inventory available for lending, consistent with strategic rationing associated with the exercise of market power.

Motivated by these findings, we develop a dynamic model that reveals how intermediation and market power in securities lending generically emerge as responses to short sellers' concerns about information leakage. That is, a core function of this market structure is to keep short sellers' private information from leaking to other traders, which is essential for the profitability of their strategies. Our tractable framework is estimated on the full cross-section of U.S. equities, jointly capturing the dynamics of stock prices and the securities-lending market across the entire fee distribution. Using this quantitative framework we can perform counterfactual analyses of alternative market structures and obtain estimates of the impact of non-competitive fees on securities lenders' valuations and on short sellers' rents. Our findings shed light on the somewhat puzzling fact that securities lending is predominantly an opaque, custodian-intermediated OTC market, despite the fact that transparent exchanges are the norm for many other financial transactions. Supplementing the insights from our framework, we highlight several sources of empirical evidence establishing the quantitative importance of information-leakage concerns in securities lending, including evidence from hedge fund groups' commentary and legal actions in the context of SEC rulemaking processes that attempted to increase transparency in this market.

Our results reveal a substantial impact of non-competitive fee income on security lenders' stock valuations, ranging from 2.00% of extra value for large-cap, low-fee stocks to increases of up to 27% for small-cap stocks and as much as 193% for micro-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.[1] 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 across firms, 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 stocks that feature substantial amounts of informed trading.

Importantly, our estimated model predicts that, under a counterfactual centralized market structure, these abnormal returns from shorting would drop substantially, even though loan fees would decline significantly. Short sellers' ability to profit from their trading strategies would be undermined by transparency and information leakage in such a venue, with this effect being particularly pronounced for smaller stocks commanding higher fees. Asset owners would similarly be worse off: competition would eliminate the fee income that in the current intermediated market flows from short sellers to security lenders. These findings are consistent with a seemingly puzzling feature of recent regulatory debates. The SEC has proposed rules that aim to increase transparency in the securities lending market and thereby lower fees. Nonetheless, industry groups representing short sellers have consistently and forcefully opposed these SEC efforts, a stance our framework helps explain.

More broadly, our paper sheds light on the prevalence of the custodian-intermediated OTC structure of securities lending markets across the largest financial markets in the world, including the US, UK, Canada, Western European countries, South Korea, Australia, Israel, and New Zealand. According to the International Securities Lending Association (ISLA, 2026), the securities lending market is overwhelmingly an OTC market. For example, as of 2024, only 8% of securities lending transactions in EU countries were conducted on exchanges (ESMA, 2024).[2] As such, our paper addresses the seemingly puzzling phenomenon that, despite the technological feasibility of centralized, competitive, and transparent securities lending, the prevailing market structure across the world relies on OTC transactions that ensure that borrowing information remains local to the involved counterparties and is intermediated by agent lenders.

We obtain these results based on a quantitative dynamic model of trade under asymmetric information

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

[2] In the United States, a centralized securities lending platform was introduced around 2001 (Quadriserv), but it was ultimately unsuccessful. Its successor, NGT by Equilend, is still operating, but it is not a centralized public exchange. Rather, Equilend's NGT uses targeted requests for quote, whereby information about borrowing quotes and quantities remains local to the negotiating parties, with counterparties knowing each other's identities. That is, it facilitates OTC transactions by streamlining the request-for-quote process.

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 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. Custodian lenders 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 framework captures the two-sided nature of the securities lending market. On one side, beneficial owners establish long-term relationships with custodian lending agents, agreeing upfront on a revenue-sharing rule that specifies a fixed proportional split of the fee income earned from borrowers. We provide novel evidence from Form N-CEN data confirming the stability of these relationships in practice: more than 95% of U.S. mutual funds maintained the same lending agent throughout 2019–2024. On the other side, lending agents interact with short-selling borrowers at high frequency, quoting fees on an ongoing basis. Our analysis shows that borrowers prefer a custodian lender committed to facilitating secrecy, and that this commitment is sustained by both market concentration and non-competitive fees. The key link between the two sides is that the lending agent must retain sufficient fee income, after the revenue split with beneficial owners, to remain incentivized not to exploit borrowers' information opportunistically. In a concentrated market, lending agents can pass through a larger share of fee income to beneficial owners while still retaining enough to support this commitment. The resulting exercise of market power contributes to low inventory utilization, as observed in the data, and to an inelastic response of equilibrium 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. One strand of this literature examines short selling in dynamic frameworks where trading is driven by belief heterogeneity (e.g., Duffie

[3] Exempted from this requirement are market makers engaged in bona fide market-making activities. The locate requirement of Regulation SHO's Rule 203(b)(1) implies that, before accepting a short sale order, broker-dealers must document that a security can be borrowed and delivered by settlement (see SEC, 2022a). For "special" stocks, those commanding elevated fees, brokers must directly contact lenders to locate shares, which in itself reveals information about clients' borrowing demand.

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.

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 Graveline (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. Sikorskaya (2025) finds that stocks with greater institutional ownership face higher short-selling costs despite having larger lending supply, pointing to a mechanism driven by index-inclusion-induced overpricing. 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.

Finally, our paper is connected to the broader literature documenting and studying the origins and implications of market power in other trading environments, such as Hortaçsu and McAdams (2010), Glode and Opp (2016), Glode and Opp (2019), Huber (2023), An and Song (2023), Allen and Wittwer (2023a), and Allen and Wittwer (2023b). This literature does not examine the economics of information leakage in securities lending.

We contribute to these strands of the literature in several ways. 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 endogenously as responses to short sellers' concerns over information leakage. Third, our tractable structural framework is estimated on the full cross-section of U.S. equities, jointly capturing the dynamics of stock prices and the securities-lending market for the full range of lowest- to highest-fee stocks. Using this framework, we provide novel estimates of how non-competitive fees affect security lenders' stock valuations and the alphas earned by short sellers, and we evaluate how counterfactual centralized market structures would affect the surpluses of lenders and short sellers, with direct implications for ongoing policy debates.

2 Market Power and Informed Trading: Empirical Evidence

In this section, we present evidence of market power in the U.S. securities lending market based on several key indicators, and document a tight empirical link between loan fees and informed trading in the stock market. 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

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

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).[5] 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 income 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 as 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, whose identities may vary across stocks and over time, tend to have a large market share, ranging between 47% and 84%. In Table 2, we further consider two adjustments to these measures to obtain conservative estimates. 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 document fee levels in the securities lending market and show that these fees far exceed any plausible marginal cost of lending, particularly for smaller and higher-fee stocks. Table 3 shows that fees on loan contracts range between 0.30% per annum for large-cap, low-fee-yield stocks to 122% for micro-cap, high-fee-yield stocks.

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

Even small-cap (mid-cap) stocks have fees ranging between 0.32% and 13.44% (0.30% and 7.64%) per annum.

When evaluating whether these fees for lending contracts are indicative of the presence of market power, it is key to benchmark them against any potential marginal costs beneficial owners and agent lenders incur when lending out securities. Such costs could theoretically emerge from channels such as (1) cash flow rights, (2) voting rights, (3) tax considerations, (4) liquidity needs and trading 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 and do not bear material incremental downside risk from lending: lending contracts are overnight contracts that allow lenders to recall their shares on a daily basis and are secured by 102% cash collateral that is marked to market daily, effectively eliminating default risk. Direct evidence confirms that default risk is immaterial in practice: BlackRock (2021) reports that over 40 years of its lending program (since 1981), only three borrowers defaulted, and in each case collateral proceeds fully covered the shortfall without any loss to beneficial owners. Corroborating this, using SEC Form N-CEN filings from 2019 to 2024, we find that across all U.S. equity mutual funds in our sample, funds report no instance of a borrower failure adversely impacting the fund.[6]

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 G) 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 57% 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 share transfers around ex-dividend dates, for example due to concerns about a differential taxation of investors. In Panel B 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 57% when all dates are included.

Fourth, we consider potential marginal costs associated with liquidity needs and with trading opportunities. Regarding liquidity needs, several institutional facts are relevant. First, lending contracts are generally overnight contracts that allow lenders to recall their shares at a daily level, so lenders always retain the option to reclaim their shares.[8] Second, large asset management companies are the typical beneficial owners in this

[6]Specifically, Form N-CEN asks: “During the report period, did any borrower fail to return the loaned securities by the contractual deadline with the result that the Fund was otherwise adversely impacted?” The answer is “No” for 100% of fund–year observations in our sample.

[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).

[8]Lenders may optimally choose not to recall their shares, since maintaining the lending relationship generates more fee income than recalling them. Importantly, this does not represent a marginal cost of lending but rather a higher benefit of lending in a

market. These institutions do not need permission from their fundholders to lend out the assets under management of their mutual funds and ETFs. Third, by regulation, open-end funds are allowed to lend out only up to one third of their overall asset value (Investment Company Act, 1940; SEC, 1997, 2022b). Moreover, using SEC Form N-PORT filings across all U.S. equity mutual funds, we show that the regulatory one-third cap is not even binding: the histogram of the fund-level fraction of asset value on loan consistently lies below 33% (see Figure 9 in Online Appendix G). Intra-day fund withdrawals or other aggregate liquidity needs are therefore unlikely to generate marginal costs for typical beneficial owners like mutual funds: at any point in time, more than two thirds of a fund's assets are freely tradable, and funds typically liquidate their most liquid assets first, which are generally those with the lowest borrowing demand and the lowest fees. Beyond liquidity needs, an additional question is whether fees might be elevated since lending forecloses lenders' own stock-specific trading opportunities. Apart from the daily recall option, it is important to recognize that beneficial owners that make their shares available for lending are predominantly passive funds such as index funds that are not in the business of stock picking. For these institutions, providing inventory to agent lenders does not interfere with any stock-specific information-driven trading. These types of institutions are the key players in the market, and we show below that they hold substantial excess inventory even for the highest-fee stocks.[9]

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 amount to a few basis points, orders of magnitude below the fees we observe.

We note that agent lenders with large inventories may also benefit from diversification of idiosyncratic borrower demand risks, which for example stabilizes investments in cash collateral. Such diversification benefits, if anything, lower agent lenders' marginal costs and cannot explain the observed fee levels, since the presence of multiple diversified lenders with excess inventory would, through competition, bring fees down to these low marginal costs. Moreover, borrowers may have a preference to interact with few counterparties to facilitate renegotiation. This preference, however, is itself rooted in information-leakage concerns, the central channel of our analysis: absent these, a borrower would simply borrow day-to-day from the cheapest lender at the spot rate, with no reason to value a persistent lending relationship. Neither channel therefore provides an independent explanation for fees above marginal cost.

In sum, the pricing of loan contracts yields lenders substantial income that is orders of magnitude above possible marginal costs of lending. The fees we document are those collected by custodian lenders from borrowers. Industry reports indicate that custodians pass approximately 85–97% of this fee income to beneficial owners (BlackRock, 2021; Morningstar, 2023; State Street, 2025), retaining only a small fraction as compensation. We further analyze the economics of this revenue-split in the context of our model in Section 7.1, where the stream of future compensation provides the incentive not to exploit borrower information opportunistically.

multi-period context.

[9] Separate from beneficial owners' own trading concerns, our model highlights that agent lenders do acquire an opportunity to trade on borrower demand information once they observe hedge fund short-selling activity. This opportunity arises as a side-product of being an agent lender. Our analysis shows that the future fee income custodian lenders retain provides commitment to maintain a reputation as an agent lender that does not opportunistically front-run its borrower clients.

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, lendable inventory falls short of total shares outstanding because many shareholders do not participate in lending programs: retail investors are often unaware of the existence and benefits of opt-in programs offered by their brokers, while firm managers, founders, and some institutional investors face reputational or signaling concerns, or are subject to policy, contractual, or legal restrictions, that prevent them from facilitating short selling of their own holdings. Importantly, the inventory we measure consists of shares that beneficial owners have made fully available for lending by depositing them with their custodian.

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.[10]

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 analysis indicates that information-leakage concerns are a central force giving rise to the prevailing opaque OTC market structure that is subject to market power. This market power allows agent lenders to charge elevated fees that extract part of hedge funds' information rents. Correspondingly, our framework predicts a clear link between informed investors' trading activity in the stock market and fee levels in the securities lending market. We now ask whether there is indeed evidence of such a relation in the data.

To address this question, we consider the stock-market measure of informed trading suggested by our model: liquidity providers' revenue from charging bid-ask spreads, defined as the product of the half bid-ask spread and trading volume. Given the competitiveness of market making and the fact that market makers typically operate across hundreds or thousands of stocks, this spread revenue reflects compensation for adverse selection losses from better-informed traders, not idiosyncratic inventory-holding costs, which diversify away across stocks.

Crucially, it is the product of spread and volume, not the spread alone, that captures this channel. A wide

[10] 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 (see Table A12 in Online Appendix G).

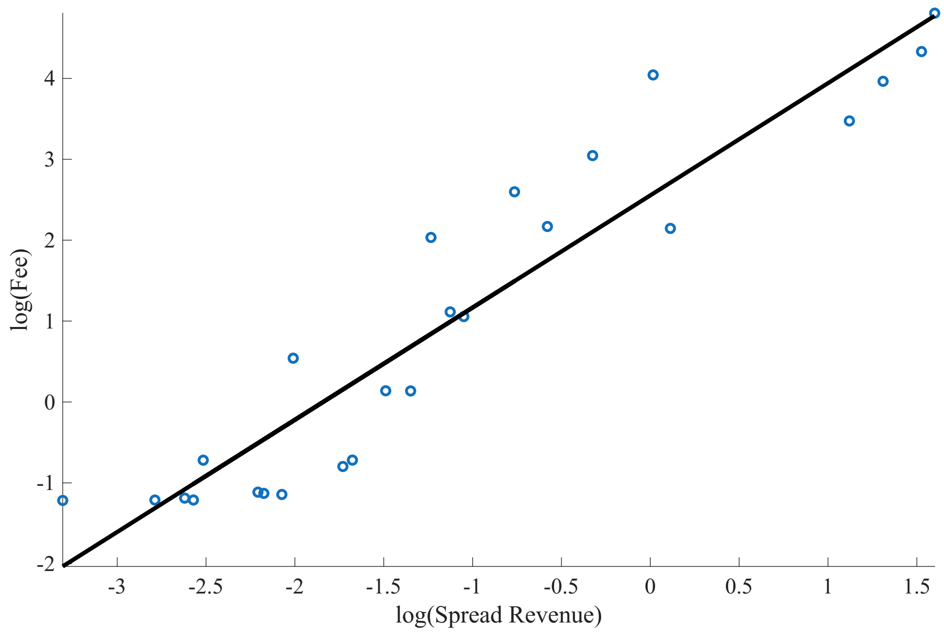


Figure 1: The figure provides a scatter plot illustrating the relation among the 25 stock groups between loan fees in the stock lending market and spread revenue in the stock market (that is, volume × 1/2 bid-ask spread; see the caption of Table A2 in Online Appendix G for details).

bid-ask spread in a thinly traded stock generates little spread revenue and may reflect inventory-holding costs rather than adverse selection. High spread revenue, by contrast, requires both elevated spreads and active trading, identifying stocks where market makers are genuinely exposed to informed order flow in equilibrium. Consistent with this, Table A10 shows that trading volume is monotonically increasing with lending fee yields across all size groups. Figure 1 plots fees in the stock 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 across groups, with a correlation of 94%. Stocks with the highest fees in the securities lending market also exhibit the largest adverse selection costs for market makers.

Table A15 in Online Appendix G further investigates this relation at the stock level by presenting pooled OLS regressions of loan fees on spread yield, inventory utilization, and market capitalization across our stock–year–quarter sample from 2007–2024. Spread yield alone generates an R^2 of 29% and is by far the most important explanatory variable. Adding inventory utilization, a securities-lending-market measure, increases R^2 by only 5%, while further including market capitalization adds just 0.6%. These results provide several insights on the relevance of alternative hypotheses:

First, they are inconsistent with stories in which inventory scarcity drives fees: utilization contributes far less to fees than spread yield, which has no direct connection to lending supply.[11] Time-series evidence reinforces this: Tables A11 and A14 in Online Appendix G document that in recent years (2020–2024), fees have risen sharply for the smallest stocks (micro- and nano-cap) and remained substantial overall, while

[11]This is consistent with Kaplan et al. (2013), who find that exogenous increases in lendable supply do not measurably affect stock returns.

excess inventory has generally increased across most groups, a pattern inconsistent with inventory scarcity driving fees. Second, while smaller stocks exhibit greater information asymmetry, size is largely redundant once the informational environment is directly captured by spread yield: market capitalization adds little, confirming it is the degree of informed trading, not firm size per se, that determines fees.

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 arises in the securities lending market when trading is subject to asymmetric information.

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* (also referred to as lending agents) 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.

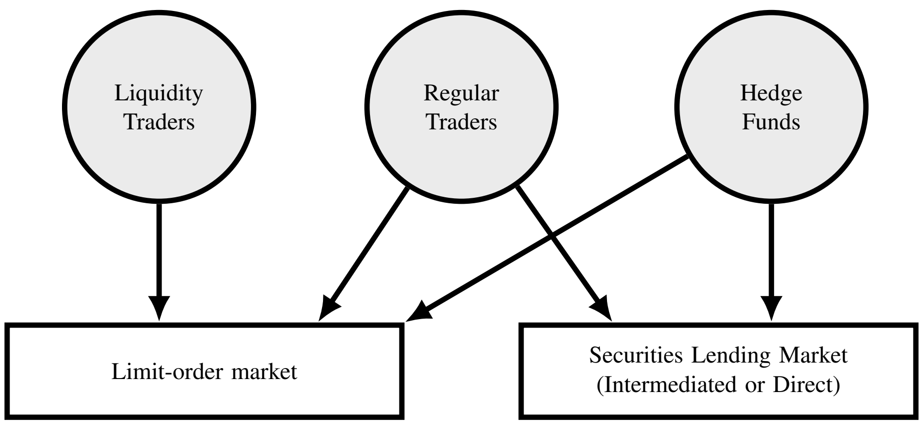


Figure 2: Investor Groups and Markets
The figure illustrates the markets and the different groups of investors in the model. We consider two distinct structures for the securities lending market, one where lending is intermediated by custodian lenders and one where investors interact directly in a centralized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

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 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.

Assets. Institutions are trading stocks that generate 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[12]

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

where v_{t-} is a random variable the realization of which is generally private information 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 $z_t \sim \mathcal{N}(-\frac{\sigma_v^2}{2}, \sigma_v)$.[13] 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 in the cross-section of stocks.

The timing convention indicated in equation (2) and Figure 3 is important. Consider a date τ_{n+1} that features an innovation $dN_{\tau_{n+1}} = 1$, 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_{τ_n}, which was realized at the *previous* innovation $dN_{\tau_n} = 1$, with $\tau_n < \tau_{n+1}$ (marked on the left-hand side of the figure). Logically after capital has been subject to this jump of size v_{τ_n} at time τ_{n+1}, a dividend equal to the new value of $c_{\tau_{n+1}}$ is paid.

While v_{τ_n} is not publicly observable at time τ_n, hedge funds can observe v_{τ_n} at that time, giving them private information about future dividend growth (which, in this illustrative example, is realized at time τ_{n+1}). Once capital has jumped by v_{τ_n} at time τ_{n+1}, agents can infer the value of v_{τ_n} by observing the realized change. Yet, at that point, this information has no relevance beyond the new level of $c_{\tau_{n+1}}$, since

[12] 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 $\lim_{s \downarrow t} y_s$.

[13] 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.

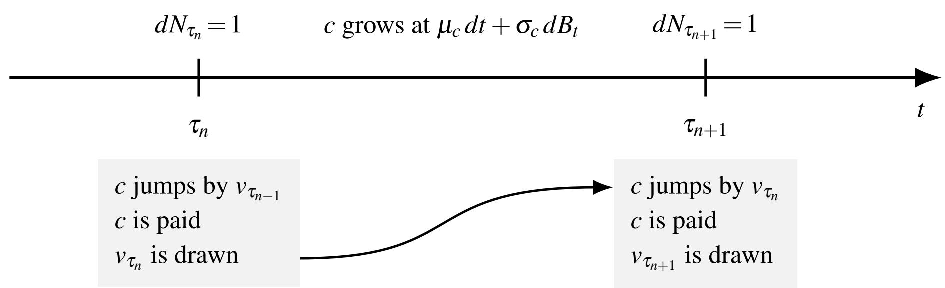


Figure 3: Timing Convention for Growth and Dividends
The figure illustrates the timing of idiosyncratic capital jumps, dividend payments, and changes in the jump size v. Hedge funds have private information about future growth by observing v_{τ_n} at time τ_n.

future growth depends on a new, independent value $v_{\tau_{n+1}}$ drawn at time τ_{n+1}.

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 D.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 F, 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. The two-sided market structure captured by our model also matches our empirical evidence (see Section 7.1 and Table A9): beneficial owners such as mutual funds maintain stable, long-term relationships with a single custodian lender—pinning down revenue-sharing arrangements at the outset—while custodian lenders then interact with securities borrowers at higher frequencies.

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, not all shareholders make their shares available for lending: retail investors frequently lack awareness of opt-in lending programs offered by their brokers, and firm managers, founders, and certain institutional investors face reputational or signaling concerns or are subject to policy, contractual, or legal restrictions that preclude facilitating short selling of their own stakes. In our model, such investors follow a buy-and-hold strategy and do not lend out their shares.

Among the shares that beneficial owners do make available for lending — referred to as lendable inventory — a further fraction remains unlent, as documented in Section 2. We therefore 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 unit 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 quantitative 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.[14] Correspondingly, 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 $\tau > t$.[15] 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.

[14] 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.

[15] 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).

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 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 $(1-\delta)\frac{\pi\phi_t}{\rho}$. The ex-dividend buy-and-hold value at time t is thus given by:

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

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:*

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

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 D. □

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:

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

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 D.1, the depth of ask quotes (i.e., the units offered) exactly matches the maximum possible demand in equilibrium. 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:

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

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:

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

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

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

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

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

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 $v_{A_1} = \log\left((A_1+\kappa_1)/P\right)$. 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 $\rho = \rho_0 - 1$.[16]

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

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

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.[17] 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 D. □

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.

[16] 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.

[17] 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.

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.[18] 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,*

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

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 D. □

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:

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

[18] 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.

That is, the estimated alpha is equal to the negative of the fee yield.[19] Given this 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

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

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 D.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 D.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. Pass-through rates are high across all major lending agents: Morningstar (2023) reports that Vanguard passes on approximately 97% of fee income to fund holders and Fidelity approximately 90%; BlackRock (2021) reports pass-through rates of 77–85%; and State Street (2025) reports rates of 85–90%. 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. Liquidity traders, in contrast, are worse off under the intermediated market structure: they are effectively the source of the surplus that flows to informed short sellers and security lenders. Online Appendix C provides comparative statics illustrating these results, including the impact on limit-order market quotes.

These insights apply more generally when borrowing demand is also motivated by liquidity and hedging motives. For example, in practice, a financial institution may borrow a security to offset an existing exposure for regulatory or risk management reasons. This is potentially relevant for larger stocks, for which derivative contracts exist, and where market makers in these derivatives may need to short the underlying stock to hedge their exposure. To capture this feature, we extend our analysis in Online Appendix D.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.

Furthermore, in our quantitative analysis in Section 6, we estimate a version of our model with a continuum of hedge funds. In this environment, potential information leakage concerns are even stronger, because aggregate borrowing volume becomes more revealing when noise shocks faced by individual hedge funds cancel out across funds. We next move to this quantitative analysis, evaluating how hedge funds and lenders fare under the prevailing OTC market structure and a counterfactual centralized alternative.

6 Quantitative Analyses

In this section, we turn to the central quantitative questions of our paper: How do loan fees affect the valuations that security lenders assign to stocks? 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? Answering these questions quantitatively for the cross-section of stocks requires a generalization of our dynamic framework to accommodate richer joint dynamics of firm fundamentals and market characteristics.

Our solution for the price-dividend ratio (5) provides some initial intuition for the impact of fee revenues on the valuations lenders assign. 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*.[20] 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.

[20] 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 small-cap and micro-cap stocks, the top 2% of stocks generate average fee yields of 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 the fee-to-dividend ratio were constant, we could immediately determine the price level implications of fees, as shown in equation (23). However, as illustrated by the Markov transition matrix reported in Table A3 (see Online Appendix G), 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.[21] 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%.[22] We provide the details of the analysis of this setting in Online Appendix E.

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:

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

$$P_t^e = \mathbb{E}\left[\int_t^\infty \frac{m_u}{m_t} c_u dN_u\right]. \tag{28}$$

Whereas the valuation equation (27) capitalizes fee income, the valuation equation (28) does not. The

[21] 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.

[22] 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.

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' 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:*

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

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

where $\odot$ denotes the Hadamard product.

Proof. See Online Appendix D. □

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 Carhart 4-factor model, which pins down the risk exposure parameters $\sigma_c(s)$ for each state s. The values of the risk premia across the 25 bins 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 $\delta = 0.1$.[23]

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 estimate the common dividend growth parameter μ_c and the state-specific parameters $u(s)$ by targeting the dividend yields in the 25 stock groups. The estimated values are reported in Table A4. 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 using the approach by Israel et al. (2001). We set $\lambda = 365$, which implies that on average, periods last for one day, consistent with the fact that securities lending contracts in practice are typically overnight contracts. This value of λ is more than a hundred times larger than the Markov transition rates between states s, which, as explained in Appendix E, implies that the solution for short sellers' price-scaled redelivery liability L/P is approximately separable across states s, and hence so are the price-scaled bid, ask, and fee quotes. We take advantage of this separability to make model solution and estimation across the full cross-section of 25 states computationally feasible.

Next, we describe our approach for determining the additional structural parameters pinning down equilibrium loan fees and short sellers' profits.

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 fee-to-dividend ratios in each state.

Hedge funds' relative trading capacity π increases the quantity of privately informed trading in the model unconditionally (across all states s). With $\pi = 0.8$, hedge fund trades account for approximately 25% of overall trading volume in our estimated model, consistent with the 20% to 30% combined trading volume share of traditional and quant hedge funds in U.S. equities over 2010 to 2021 documented in the data by Martin and Wigglesworth (2021).

[23] Our microfoundation in Appendix F 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.

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.[24] 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 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 13% fee per annum in the data (see Table 3). Rationally, only hedge funds expecting an abnormal return *exceeding* 13% (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 13% in the first place. If hedge funds virtually never expected to generate a return exceeding 13% per annum, charging such a high fee would be suboptimal for a lender. Thus, to rationalize lenders charging a fee of 13%, they must anticipate that a sufficiently large number of hedge funds will be expecting a pre-fee alpha in excess of 13%. 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 A6, A7, and A8 (see Online Appendix G) 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 in 2 states exceed, our estimates of idiosyncratic return volatility. This result speaks to the validity of the model's specification of lender market power. Our baseline model and the microfoundation of market power presented in

[24]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.

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). Under an alternative specification in which hedge funds had additional bargaining power and retained a larger share of their trading profits, matching the same observed high fees would require counterfactually higher values of σ_v. The fact that our estimated σ_v values already approach, and in a few cases exceed, the idiosyncratic return volatilities in the same states, thus supports our market power specification over alternatives that allocate 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 now address our first key quantitative question: how does non-competitive fee income affect the value of shares in the cross-section and at the aggregate stock market level? The estimates addressing this question rely only on the first module of our estimation approach.

State-contingent value wedges $VW(s)$. Figure 4 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 marginal investors, then transaction prices reflect this additional value. The value wedges range from 2.00% for the large-cap stocks with the lowest fees (bottom 80%) to 193% for the highest-fee group of micro-cap stocks. These results imply that at the margin, shareholders not participating in shares lending — for instance, those investing directly in individual stocks outside of lending programs — miss out on significant value, in particular when investing in small stocks.

The value wedges differ dramatically from those one would obtain when ignoring fee and dividend dynamics: if one were to assume that fee-to-dividend ratios remain constant within each state, the value wedges would instead range from 0.18% to 418818%. 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 group, there is a 3% chance of transitioning to the next-higher fee yield group within the Large-Cap category, which is quantitatively important. Even a small probability of transitioning to a high-fee state contributes meaningfully to the present value of expected fee income. 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.

While the results in Figure 4 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 currently 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

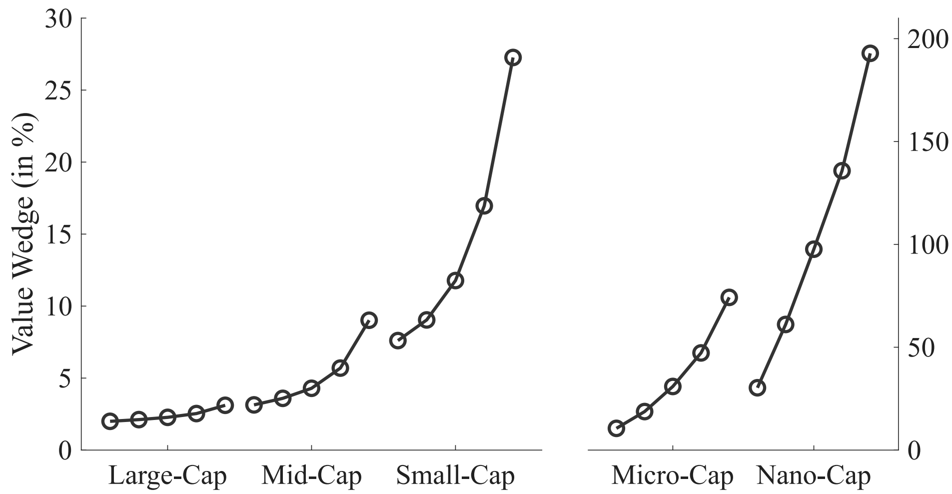


Figure 4: Value Wedges VW due to Lending Fee Income
We plot the additional percentage value that shares lenders assign to stocks due to fee income. For large-cap, mid-cap, and small-cap stocks, the left-hand side vertical axis applies; for micro-cap and nano-cap stocks, the right-hand side axis applies. 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.

0.45% and 13.89%, 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 about 2.13% 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%. This number is about one-third of our baseline estimate, reflecting that inventory itself represents about one-third of total market capitalization when value-weighting across stock groups (see Table A13 in Online Appendix G). The relative distribution of 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 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. As a result, absent additional frictions, either the equilibrium loan volume or the equilibrium loan fee reveals the fundamental v. In particular, when inventory is not fully utilized and competition drives fees to zero, aggregate borrowing volume becomes a revealing signal of v; when inventory is scarce, the market-clearing fee itself reveals 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, a paradox akin to 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 D.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 E.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 volatility of aggregation noise σ_b is half as large as the fundamental volatility σ_v; that is, $\sigma_b(s)/\sigma_v(s) = 0.5$, where the estimated fundamental volatilities σ_v are listed in Table A6. For instance, the annualized volatility of fundamental shocks in the highest-fee groups among small-cap stocks is 50.88%, implying that the aggregate noise shocks in information aggregation are assumed to have an annualized volatility of 25.44%. Crucially, unlike the idiosyncratic noise ε_i that averages out across the continuum of hedge funds, b is a fully aggregate shock impeding information revelation. The calibrated ratio $\sigma_b/\sigma_v = 0.5$ should therefore not be interpreted as a standard ratio of total noise volatility to fundamental volatility; rather, it conservatively assumes that the aggregate noise alone (the component that does not diversify away) has this much volatility relative to fundamentals.

Below, we will assess the robustness of this calibration by characterizing the ranges of values for the ratios $\sigma_b(s)/\sigma_v(s)$ (for each s) where the intermediated market dominates. To gain intuition for the effect of

σ_b, it is helpful to recall the limiting case of no aggregation noise, $\sigma_b \to 0$. Here, hedge funds' informational advantage vanishes completely in the centralized market due to information leakage, leaving zero rents for both hedge funds and lenders. As a result, for sufficiently low values of σ_b, the intermediated market always dominates the centralized one for both sides of the securities lending market.

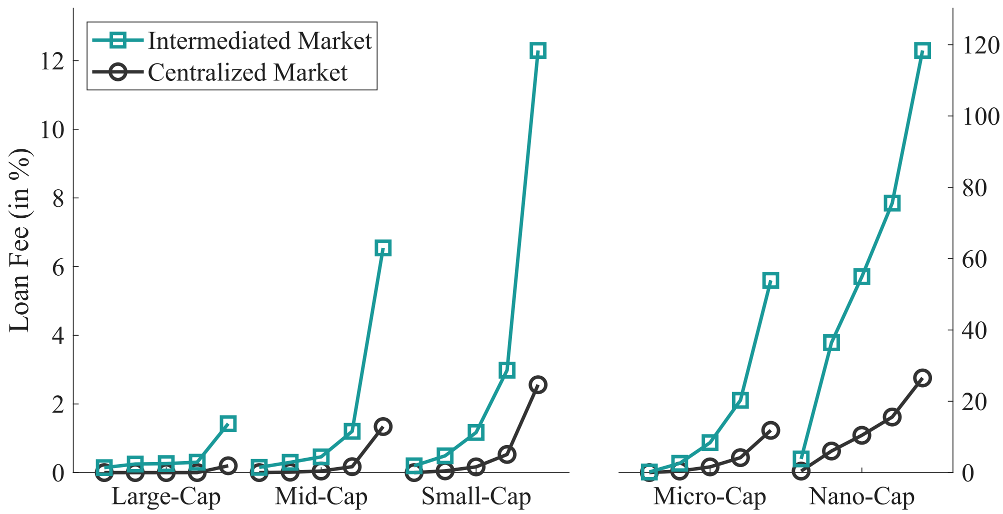
Loan fees and fee yields. First, we examine how loan fees and fee yields vary across the intermediated and centralized market structures. Whereas loan fees represent the costs of shorting to hedge funds, fee yields measure the corresponding income flow that asset owners receive. Due to incomplete inventory 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 custodian lender taking a cut.

Figure 5 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.

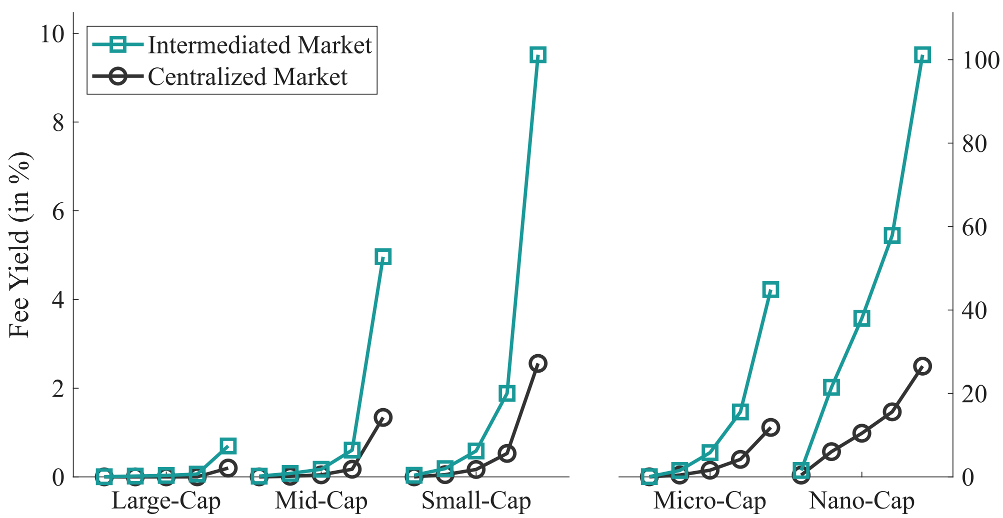
Hedge funds' gains from shorting. Our qualitative analysis of the counterfactual centralized market structure in Section 5 indicated that hedge funds prefer an intermediated securities lending market when shorting sufficiently illiquid assets. We now examine this preference quantitatively across our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

The panels of Figure 6 compare two measures of shorting profits between the intermediated market (blue squares) and the counterfactual centralized market (black circles). Panel (a) plots hedge funds' net-of-fee alphas from shorting securities. 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 example, in the small-cap group, the net-of-fee alphas from shorting high-fee stocks reach around 11%, whereas they are only around 2% in the centralized market. For micro-cap stocks, these numbers are 46% and 10%, respectively. These results arise despite the fact that, 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 in the intermediated market. However, information leakage in the centralized market is more detrimental to hedge funds' net-of-fee alphas than the high fees in the intermediated market. As a result, hedge funds' net-of-fee alphas are higher in the intermediated market for all 25 stock groups.

Next, in Panel (b) of Figure 6, we account for both the net-of-fee alphas from shorting and the propensity

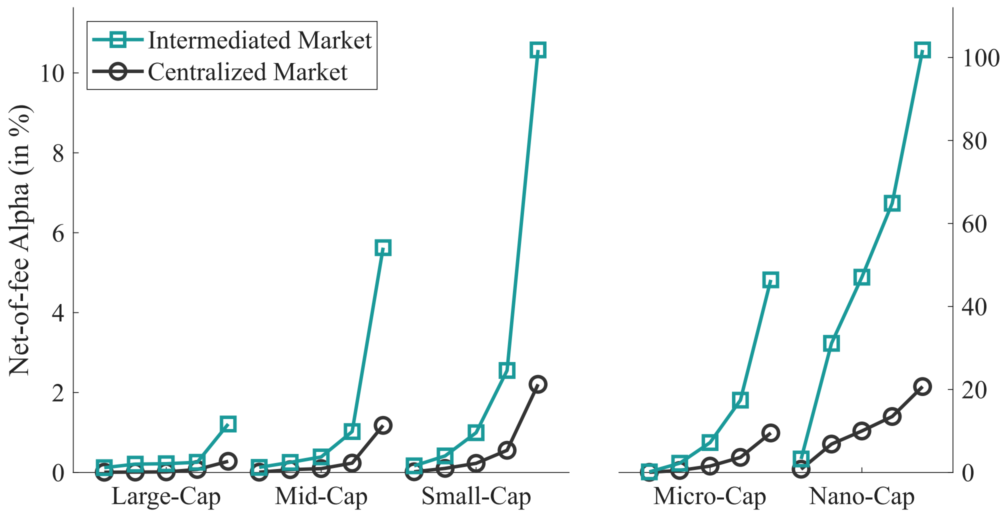


(a) Loan fees

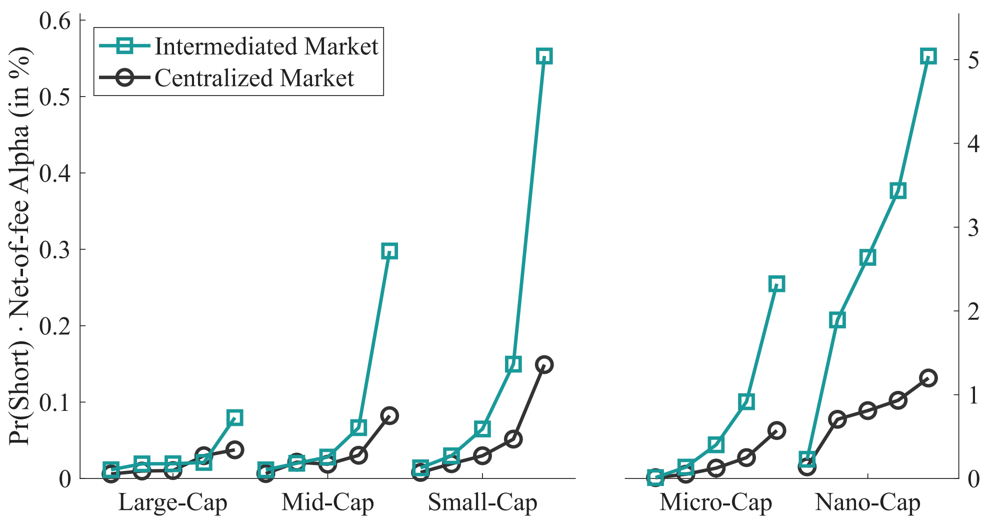


(b) Fee yields

Figure 5: 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 6: 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.

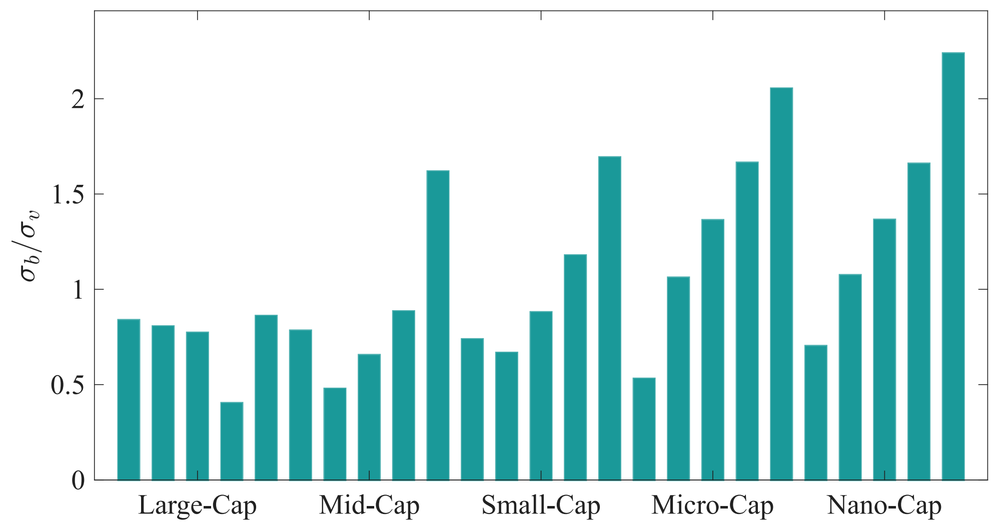


Figure 7: Ranges of aggregation noise where intermediation dominates for hedge funds.
The figure plots the range of values $\sigma_b(s)/\sigma_v(s)$, as driven by changes in $\sigma_b(s)$, for which the intermediated securities lending market dominates based on the product of hedge funds' shorting propensity and the net-of-fee alpha.

to short. This measure is conservative in that it assumes hedge funds have no alternative use of their trading capacity. That is, when not shorting a stock in this group, a hedge fund obtains zero alpha on its available capital. The results indicate that the intermediated market dominates for almost all stock groups (23 out of 25), particularly among smaller stocks where hedge fund trades are most profitable, although the relative gap is somewhat reduced. This narrowing of the gap reflects the fact that the frequency of shorting is higher in the competitive centralized market, driven by lower shorting fees; this higher frequency partly offsets the lower net-of-fee alphas. As noted above, a volatility ratio of aggregation noise to fundamentals equal to one half ($\frac{\sigma_b(s)}{\sigma_v(s)} = \frac{1}{2}$) implies large absolute levels of aggregation noise, particularly for smaller and high-fee stocks. This means that even with substantial aggregation noise, the intermediated market dominates for these stocks.

To evaluate the robustness of these results, we examine the highest levels of the noise volatility parameters $\sigma_b(s)$ at which hedge funds would still prefer an intermediated market (Figure 7). For this evaluation, we measure hedge fund gains based on the conservative measure discussed above: 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, as would be expected in light of information leakage concerns. This is particularly noteworthy given that the fundamental volatility $\sigma_v(s)$ increases monotonically with fee levels in each size group. For 20 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 other words, for centralized lending to be preferable for hedge funds trading these stocks, the volatility of aggregation noise by itself would have to be implausibly large, substantially exceeding fundamental volatility.

In all, these results suggest that both asset owners and hedge funds fare worse under a counterfactual centralized market structure across the wide range of information aggregation frictions that might exist in this market, particularly for smaller stocks that command higher fees under the prevailing intermediated market structure.

7 Extensions and Discussions

In this section, we present extensions and discussions on four topics: the theoretical foundation for custodian market power and commitment; empirical evidence corroborating short sellers' information leakage concerns; the implications and limitations of alternative market designs; and the relation between securities lending and options markets.

7.1 Custodian Market Power and Commitment

In our baseline model presented in Section 3, we considered a single monopolistic custodian lender 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 prefer to interact with such a monopolistic lender rather than a centralized market with competitive lenders, particularly for smaller stocks. 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 despite free entry to this line of business. After all, in a two-sided market with multiple potential custodians, one might expect competition to eliminate market power and the associated rents required to sustain commitment. In this section, we address these questions based on an explicit repeated game analysis we provide in Online Appendix F. To economize on space, we relegate the detailed analysis to this appendix and discuss its main insights in this section.

Empirically, the two-sided structure of the securities lending market is characterized by a fundamental asymmetry in interaction frequencies. On the beneficial-owner side, mutual funds establish long-term contracts with agent lenders upfront, specifying a fixed proportional revenue-sharing rule that does not vary with the fees ultimately collected. Beneficial owners delegate full command over the committed inventory to the agent lender, who works on their behalf without seeking approval for individual loans. Using SEC Form N-CEN filings, which require mutual funds to annually disclose their lending agents, Table A9 (Online Appendix G) shows that approximately 85% of mutual funds used a single lending agent at any given time, and more than 95% maintained the same lending agent throughout 2019–2024. On the borrower side, agent lenders interact at high frequency, providing loan quotes as borrowing demand arrives throughout each day. Our model captures this structure: at date 0, custodians compete for beneficial owners via the revenue-sharing split δ; thereafter, they interact with potential borrowers at higher frequencies.

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 custodian 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: rather, it is the threat that asset owners would move their securities lending to another custodian, in case of insufficient fee pass-through or opportunistic behavior, that shapes this equilibrium outcome. Moreover, we show in Appendix F how limited percolation of information related to opportunistic behavior can lead to the existence of multiple custodians operating, as observed in practice, with each exercising market power in its own relationship network. Relatedly, in practice, hedge funds also have incentives to maintain a reputation for loyalty with their security lenders. A custodian 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 to this custodian in the future. In sum, both custodians and hedge funds have incentives that reinforce concentrated, bilateral lending relationships.

It is worth noting that in practice there are complementarities with the asset management business that are beyond the scope of our model, but that reinforce coordination on specific custodians: asset management and brokerage companies are natural suppliers of shares in the securities lending market, as they already have large 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.[25]

7.2 Evidence about Short-Sellers' Information Leakage Concerns

A fundamental premise of our model is that information leakage is a first-order concern for informed traders in the process of establishing a short position. In this section, we provide three sources of empirical evidence corroborating this premise.

SEC rulemaking evidence. The most direct evidence that short sellers perceive information leakage from securities lending as a first-order concern comes from recent regulatory rulemaking processes. The SEC formally adopted two rules aimed at increasing transparency in securities lending and short selling, both of which faced strong industry opposition during rulemaking and subsequent legal challenges. As detailed below, these challenges resulted in neither rule ever being implemented in practice, with the SEC now back to reviewing both.

[25] Companies in this industry can extract rents from skills related to relative advantages in stock analysis, low-cost order execution and management, retail outreach and advertising, with different parts of the investor population being sensitive to different aspects. Generating securities lending income allows asset managers to offer more competitive, lower-cost funds to investors, reinforcing their position in the industry.

Rule 13f-2. Section 929X of the Dodd–Frank Act mandated the SEC to increase transparency in short selling. The SEC adopted Rule 13f-2 in October 2023 (SEC, 2023b), requiring institutional investment managers to report short positions monthly using Form SHO, though as discussed below, the rule has not been implemented to date. The original proposal called for more frequent and granular public disclosure, but industry groups—including the Alternative Investment Management Association (AIMA) and the Committee on Capital Markets Regulation (CCMR)—raised concerns that public disclosure would enable other market participants to identify and front-run short-selling strategies. In response, the SEC modified the rule to require only monthly, aggregate, confidential reporting.

Rule 10c-1a. The SEC also adopted Rule 10c-1a (SEC, 2023a), a separate rulemaking concerning securities lending transparency. The original proposal required that full loan-level data—including fee rates and loan amounts—be reported to FINRA within 15 minutes of each transaction, with FINRA making the data publicly available as soon as practicable. Industry groups strongly opposed the proposal, arguing that near-real-time public disclosure of borrowing volume would reveal short sellers' position build-ups to the market. After this pushback, the SEC retained end-of-day reporting to FINRA but incorporated a 20-business-day delay for the public disclosure of individual loan amounts in the final rule. The SEC itself argued that this modification "should significantly reduce the novelty of the information disseminated under the final rule regarding short sellers' positions, such that it is less timely than pre-existing sources of short-selling transparency, such as FINRA's bimonthly short-interest data."

Legal challenge and Fifth Circuit ruling. Despite these adjustments during the rulemaking process, industry groups representing short sellers challenged both rules in court, arguing that Rule 10c-1a still provided information that could allow observers to identify short-selling strategies that Rule 13f-2 was specifically designed to keep confidential. In August 2025, the Fifth Circuit agreed, remanding both rules back to the SEC for failing to analyze how one rule's transparency could cancel out the other rule's protections. The SEC subsequently stayed compliance deadlines for both rules while it conducts further analysis, and has indicated it may propose amendments or re-propose the rules entirely. Neither rule is currently being implemented.

In sum, short sellers actively resisted both rules, revealing that the transparency of securities lending activity is a first-order concern for them. In the context of standard product markets, such resistance to increased transparency from the demand side would be surprising; after all, increased transparency would be expected to foster competition on the supply side, thereby lowering borrowing fees for short sellers. However, securities lending is not a standard product market: rather than welcoming the increased transparency proposed by the SEC, the demand side of this market prefers opacity, as it is essential to its business model built on private information.

Opacity and hedge funds' information-based business model. The notion that informed traders care about concealing their trading strategies is well established in the literature. A central feature of hedge funds' business model is that they actively avoid sharing information about their trading strategies and portfolio positions, even with their own investors (see, e.g., DiBlasi, 2010; Aragon et al., 2013; Agarwal et al., 2013; Yu, 2019; Cao et al., 2022). Position-level disclosure, even with a lag, is widely understood to allow other traders to reverse-engineer and front-run hedge funds' strategies, eroding their trading profits.

Short sellers as informed traders. Chague et al. (2023) provide direct evidence that short sellers possess genuinely private information that can be profitably exploited. Using granular data from the Brazilian securities lending market and stock exchange over 2015–2018, the authors establish three key facts. First, short sellers' borrowing activity predicts subsequent stock underperformance, confirming that short sellers are privately informed traders. Second, broker–dealers observe short sellers' positions in the OTC lending market before the broader market does, and selectively share information about skilled short sellers' bets with preferred clients. Third, clients who receive this shared information trade profitably and stocks subject to information sharing experience faster price declines. This evidence establishes that short sellers hold private information with real economic value, reflecting genuine informational advantages rather than mere differences in opinion.

The timing of this information sharing is key. Crucially, the information sharing documented here need not harm the original borrower if it occurs after the short position has already been established. The concern central to our model is therefore distinct: it is leakage that occurs during or before completing the build-up of the position that short sellers seek to avoid. Leakage thereafter can in fact benefit them through faster price declines that allow profitably closing their positions.

7.3 Alternative Market Structures

In Section 5, we considered the most common form of a competitive securities lending market, namely one that mirrors the limit order markets typically used for regular securities trading in practice: suppliers of lendable shares submit offers for loan contracts, and market participants interested in borrowing shares compete for those offers. We highlighted that this centralized market structure leads to information leakage, thereby harming short sellers' ability to profit from their positions.

Alternatively, one might seek to design a competitive market structure that mitigates information leakage by delaying the public disclosure of realized trades for a pre-specified period. Under such an arrangement, lenders might, for example, post their shares on a platform and not learn whether those shares were actually lent out until the end of the trading day, thereby delaying information leakage to them. Only borrowers would receive timely confirmation of whether their requests had been filled, since they must know how many shares they can actually sell short. New fintech developments such as smart contracts could potentially facilitate the implementation of such arrangements. However, even such a market structure would still face two fundamental challenges related to information leakage.

First, a one-day delay in disclosing lending activity to lenders is insufficient to protect short sellers who build positions dynamically over extended periods.[26] A mechanical daily disclosure schedule would therefore still reveal position build-ups as they accumulate. Moreover, lenders who know that they will

[26]In our data, short sellers roll over their lending contracts for approximately 40 days on average, with substantial variation around this mean.

learn about realized demand only with a delay have an incentive to withhold supply strategically, waiting to see how demand clears before committing additional shares. This strategic behavior would spread the borrowing of any desired quantity over a prolonged period, further undermining the short seller's ability to establish a position discreetly.

Second, beyond the fact that disclosure delays have limited value in preventing information leakage to lenders, they do not address leakage to agents approaching the demand side of the market. A competitive market creates incentives for agents to participate on the demand side purely to extract information, submitting exploratory borrowing requests not to establish short positions, but to elicit real-time signals about aggregate borrowing demand. Such agents effectively free-ride on the clearing mechanism: both the market-clearing fee and the quantity allocated to them convey information about aggregate demand, which in turn is informative about the fundamental v (see Section 6.5). Both pieces of information must be communicated to borrowers in real time: the allocated quantity tells them how many shares they can sell short, and the clearing fee tells them the cost they will incur. Information leakage to participants on the borrowing side is therefore an inherent feature of any centralized lending process.

Beyond these fundamental problems related to information leakage, such alternative centralized market structures may also face an inventory problem and raise distributional concerns. First, any competitive platform offering zero or near-zero fees may be unable to attract lendable shares from investors when custodian lenders offer substantial fee income, leaving the platform without the inventory needed to attract borrowers. Second, even if regulators were to mandate such a structure, thereby ensuring inventory, retail investors would bear the cost: under the current regime, large asset managers pass on most of their gross securities lending revenue to fundholders. Moving to a competitive market would transfer this surplus from retail investors to hedge funds, an outcome that regulators and policymakers may find undesirable.

7.4 Relation to Trading in Options Markets

In practice, agents wishing to bet against a security may be able to do so either by borrowing and short selling the security or, when available, by trading options. While options are typically not available for small stocks, they are traded in centralized markets for many larger stocks. One may therefore wonder how the availability of options trading relates to the main insights of 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. 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 diminish the relevance of the features of the securities lending market that we highlight. Fundamentally, this is because (1) options market makers also need to hedge their positions, that is, they cannot take unbounded naked short positions, and (2) informed traders have incentives to exploit the presence of liquidity-motivated trading in the market for the underlying stock. In addition, options markets do not exist for smaller stocks. Thus, the securities lending market is not redundant, and, as our analysis highlights, short sellers benefit from an opaque, intermediated market structure that mitigates information leakage.

8 Conclusion

In this paper, we document market power in the U.S. equity securities lending market and examine its origins and consequences. We show that the current OTC custodian-intermediated structure, which prevails across the world's largest financial markets, arises as an equilibrium response to short sellers' concerns about information leakage. Somewhat paradoxically, informed short sellers 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 custodian lending agents provide short sellers with protection against information leakage during position build-ups: when intermediation occurs through a custodian lending agent, information about borrowing activity remains localized 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 quantitative counterfactual analyses show that, under a competitive centralized market structure, loan fees would decline substantially and, for some large- and mid-cap stock groups, fall to zero. However, despite lower fees, short sellers would be worse off under such a regime, particularly for smaller stocks, because the transparency of a centralized venue causes information leakage that erodes their informational advantage by more than they would gain from the lower fees. For instance, in the small-cap group, our estimates indicate that net-of-fee alphas from shorting high-fee stocks reach around 11% in the intermediated market versus only 2% under a centralized counterfactual; for the highest-fee micro-cap stocks, the corresponding alphas are 46% versus only 10%. Asset owners participating in lending would also be worse off, since the current substantial fee income they receive would be reduced significantly. These findings speak directly to ongoing policy debates over regulatory proposals affecting the transparency of securities lending markets.

Our quantitative analyses also indicate substantial value to participating in securities lending for asset owners themselves: for shareholders who lend out their shares, fee income adds as much as 27% of value to holding the highest-fee small-cap stocks, with even larger magnitudes for micro-cap stocks. Retail investors who hold such stocks but do not participate in lending forgo these present-value gains, highlighting the benefits of educating retail investors about the option of making their shares available for lending.

These present-value gains from securities lending can also have material spillover effects on the informational content of prices, with small stocks being especially affected. Specifically, when lenders are marginal investors, these value gains are reflected in equilibrium stock prices. 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). Future studies could shed more light on the real implications of these spillover effects.

References

Agarwal, V., Aragon, G. O., Shi, Z., 2013. Uncovering hedge fund skill from the portfolio holdings they hide. Journal of Finance 68, 739–783.

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.

Aragon, G. O., Hertzel, M., Shi, Z., 2013. Why do hedge funds avoid disclosure? evidence from confidential 13f filings. Journal of Financial and Quantitative Analysis 48, 1499–1518.

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.

BlackRock, 2021. Blackrock securities lending: Unlocking the potential of portfolios. White Paper L&LH0221U/S-1538534, BlackRock Investments, LLC, prepared by BlackRock Investments, LLC, member FINRA.

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.

Cao, S., Da, Z., Jiang, D., Yang, B., 2022. Do hedge funds strategically misreport their holdings? evidence from 13f restatements. Working paper (September 12, 2022) .

Chague, F., Giovannetti, B., Herskovic, B., 2023. Information leakage in securities lending, working paper.

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.

DiBlasi, P., 2010. Information leaks cause headline risk. The Hedge Fund Journal, industry commentary on hedge fund information leakage.

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.

ESMA, 2024. Eu securities financing transactions markets 2024. Tech. Rep. ESMA50-524821-3147, European Securities and Markets Authority, Paris, eSMA Market Report.

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.

Glode, V., Opp, C., 2016. Asymmetric Information and Intermediation Chains. American Economic Review 106, 2699–2721.

Glode, V., Opp, C. C., 2019. Over-the-Counter versus Limit-Order Markets: The Role of Traders' Expertise. The Review of Financial Studies 33, 866–915.

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.

ISLA, 2026. Securities lending market regulations. https://www.islaemea.org/securities-lending-borrowing-regulation/, accessed: 2026-02-25.

Israel, R. B., Rosenthal, J. S., Wei, J. Z., 2001. Finding generators for markov chains via empirical transition matrices, with applications to credit ratings. Mathematical Finance 11, 245–265.

Kaplan, S. N., Moskowitz, T. J., Sensoy, B. A., 2013. The effects of stock lending on security prices: An experiment. Journal of Finance 68, 1891–1936.

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, [Online; accessed 18-January-2023].

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.

Morningstar, 2023. An inside look at securities lending: An additional source of return (and risk) for fundholders. Research report, Morningstar Manager Research, Chicago, IL, morningstar Manager Research Report.

Muravyev, D., Pearson, N. D., Pollet, J. M., 2022. Is there a risk premium in the stock lending market? evidence from equity options. The Journal of Finance 77, 1787–1828.

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, [Online; accessed 14-April-2022].

SEC, 2022a. Key Points About Regulation SHO. https://www.sec.gov/investor/pubs/regsho.htm, [Online; accessed 15-November-2022].

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, [Online; accessed 14-April-2022].

SEC, 2023a. Reporting of securities loans. Exchange Act Release No. 34-98737, 88 Fed. Reg. 75644, adopting release for SEC Rule 10c-1a.

SEC, 2023b. Short position and short activity reporting by institutional investment managers. Exchange Act Release No. 34-98738, 88 Fed. Reg. 75100, adopting release for SEC Rule 13f-2 and Form SHO.

Sikorskaya, T., 2025. Institutional investor mandates, securities lending, and short-selling constraints, working Paper.

Stambaugh, R. F., Yu, J., Yuan, Y., 2011. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

State Street, 2025. Unlocking portfolio potential with securities lending. Insights / Marketing Communication ID2617450-5933575.3.1.EMEA.INST, State Street Global Advisors, for professional investor use only.

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.

Yu, X., 2019. Information disclosure and hedge fund performance. Review of Financial Studies 32, 4304–4345.

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%	35.37	35.22	38.31	42.57	59.40
80-90%	30.96	31.22	34.00	37.35	57.69
90-95%	29.40	29.79	31.41	37.17	59.61
95-98%	28.19	28.20	31.04	36.94	57.25
$\geq$ 98%	27.43	28.41	31.63	39.97	60.09

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%	19.46	21.48	23.10	22.56	22.25
80-90%	18.88	21.16	21.56	21.02	23.02
90-95%	19.47	20.44	20.27	20.92	22.87
95-98%	19.68	19.18	19.70	21.01	24.19
$\geq$ 98%	19.43	18.74	20.50	21.72	23.41

Table 1: Concentration of Custodian Lenders' Value on Loan (Markit Data)
This table reports the concentration of custodian lenders in the U.S. stock lending market across 25 groups formed by stock size and lending fee yield. For each stock j on day d, the Markit Securities Finance database reports the shares of value on loan accounted for by the two largest custodian lenders. We first aggregate these daily observations to the quarterly level by computing quarter-t averages of both value on loan and lender market shares for each stock. Using these quarterly quantities, we construct the group-level market share of lender $k \in \{1,2\}$ (ranked by value on loan) as

$$\text{MarketShare}^i_{k,t} = \frac{\sum_{j\in i} \text{MarketShare}^i_{k,j,t} \times \text{ValueOnLoan}^i_{j,t}}{\sum_{j\in i} \text{ValueOnLoan}^i_{j,t}}, \qquad i = 1,\ldots,25,$$

where $\text{MarketShare}^i_{k,j,t}$ is lender k's average daily share of stock j's total loan volume in quarter t, and $\text{ValueOnLoan}^i_{j,t}$ is the corresponding average daily dollar value of stock j'shares on loan. Panel A reports the time-series average of $\text{MarketShare}^i_{1,t}$, the market share of the largest lender in each group, and Panel B reports the corresponding averages for the second-largest lender. Market shares are expressed in percent. The sample period is 2007–2024.

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%	30.25	30.07	32.74	35.96	51.29
80-90%	26.34	26.70	29.14	32.80	49.76
90-95%	24.99	25.51	26.90	32.09	51.07
95-98%	24.11	24.09	26.85	31.83	48.64
$\geq$ 98%	22.98	24.44	27.28	34.92	51.05

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%	16.56	18.28	19.64	18.90	18.82
80-90%	16.04	18.01	18.34	17.93	19.70
90-95%	16.53	17.40	17.22	17.86	19.57
95-98%	16.71	16.32	16.77	17.89	20.74
$\geq$ 98%	15.79	15.95	17.63	18.57	20.05

Table 2: Concentration of Lenders' Value on Loan (Markit-Coverage Adjusted)
This table reports custodian lender concentration across the 25 size–fee-yield groups using a conservative measure that adjusts for the incomplete coverage of the Markit Securities Finance database. We first construct, for each stock, a quarter-level market share of lender $k \in \{1,2\}$ as a lending-fee-income-weighted average of its daily market shares, multiplied by Markit's reported coverage rate of 85% of the securities lending market,

$$\text{MarketShare}^i_{k,j,t} = \frac{\sum_d \text{MarketShare}^i_{k,j,t,d} \times \text{LendingFeeIncome}^i_{j,t,d} \times 0.85}{\sum_d \text{LendingFeeIncome}^i_{j,t,d}},$$

where $\text{MarketShare}^i_{k,j,t,d}$ is lender k's market share and $\text{LendingFeeIncome}^i_{j,t,d}$ is the total lending fee income (lending fee times value on loan) for stock j on day d. We then aggregate to the group level by computing a lending-fee-income-weighted average across stocks,

$$\text{MarketShare}^i_{k,t} = \frac{\sum_{j\in i} \text{MarketShare}^i_{k,j,t} \times \text{LendingFeeIncome}^i_{j,t}}{\sum_{j\in i} \text{LendingFeeIncome}^i_{j,t}}, \qquad i = 1,\ldots,25,$$

where $\text{LendingFeeIncome}^i_{j,t}$ denotes the total lending fee income of stock j in quarter t. This adjustment is conservative in the sense that it assumes all securities-lending activity not covered by Markit is intermediated by institutions other than the two largest lenders identified by Markit (see Online Appendix A).

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%	0.296	0.298	0.323	0.487	8.546
80-90%	0.298	0.319	0.488	2.873	32.268
90-95%	0.305	0.450	1.146	8.750	52.664
95-98%	0.328	1.148	3.045	21.027	76.059
$\geq$ 98%	1.718	7.640	13.440	56.980	122.224

Table 3: Lending Fee
This table reports time-series averages of lending fees for the 25 size–fee-yield groups. We first aggregate daily lending fees for each stock to the quarterly level as a value-on-loan–weighted average of its daily lending fees,

$$\text{LendingFee}^i_{j,t} = \frac{\sum_d \text{LendingFee}^i_{j,t,d} \times \text{ValueOnLoan}^i_{j,t,d}}{\sum_d \text{ValueOnLoan}^i_{j,t,d}}.$$

The lending fee for group i in quarter t is then computed as a value-on-loan–weighted average across stocks,

$$\text{LendingFee}^i_t = \frac{\sum_{j\in i} \text{LendingFee}^i_{j,t} \times \text{ValueOnLoan}^i_{j,t}}{\sum_{j\in i} \text{ValueOnLoan}^i_{j,t}}, \qquad i = 1, \ldots, 25,$$

where $\text{ValueOnLoan}^i_{j,t}$ denotes the average value of stock j's shares on loan in quarter t. Lending fees are annualized and reported in percent. The sample period is 2007–2024.

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%	98.68	94.32	88.36	89.34	87.13
80-90%	93.69	75.29	63.75	46.15	46.73
90-95%	88.84	64.36	52.59	33.39	35.45
95-98%	81.26	53.38	40.10	25.44	26.88
$\geq$ 98%	65.24	34.31	28.11	19.42	18.24

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%	97.48	89.43	81.65	70.50	59.84
80-90%	92.72	74.00	62.01	41.52	41.09
90-95%	87.53	62.73	50.02	30.76	30.79
95-98%	79.24	49.78	36.88	23.19	23.41
$\geq$ 98%	51.05	24.16	22.59	16.72	14.52

Table 4: Excess Inventory for Stock Lending

This table reports time-series averages of excess inventory for the 25 size–fee-yield groups. In Panel A, group-level excess inventory in quarter t is defined as one minus the ratio of aggregate value on loan to aggregate value of lendable inventory across stocks in the group,

$$\text{ExcessInventory}_t^i = 1 - \frac{\sum_{j\in i} \text{ValueOnLoan}_{j,t}^i}{\sum_{j\in i} \text{ValueInventory}_{j,t}^i}, \qquad i = 1, \ldots, 25,$$

where $\text{ValueOnLoan}_{j,t}^i$ is the average daily dollar value of stock j's shares on loan and $\text{ValueInventory}_{j,t}^i$ is the corresponding average daily dollar value of lendable inventory.

In Panel B, the alternative measure of group-level excess inventory is constructed as a lending-fee-income–weighted average across stocks,

$$\text{AlternativeExcessInventory}_t^i = \frac{\sum_{j\in i} \text{ExcessInventory}_{j,t}^i \times \text{LendingFeeIncome}_{j,t}^i}{\sum_{j\in i} \text{LendingFeeIncome}_{j,t}^i}, \quad i = 1, \ldots, 25,$$

where $\text{LendingFeeIncome}_{j,t}^i$ is the total lending fee income of stock j in quarter t. The stock-level excess inventory used in this aggregation is computed as a lending-fee-income–weighted average of daily excess inventory,

$$\text{ExcessInventory}_{j,t}^i = \frac{\sum_d \left(1 - \text{Utilization}_{j,t,d}^i\right) \times \text{LendingFeeIncome}_{j,t,d}^i}{\sum_d \text{LendingFeeIncome}_{j,t,d}^i},$$

where $\text{Utilization}_{j,t,d}^i$ is the fraction of lendable inventory on loan for stock j on day d. Excess inventory is reported in percent. The sample period is 2007–2024.

Lending Fee Yield (annual, in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq 80\%$	0.004	0.016	0.036	0.050	1.540
80-90%	0.018	0.078	0.186	1.528	21.477
90-95%	0.033	0.171	0.584	5.819	38.032
95-98%	0.063	0.604	1.884	15.573	57.896
$\geq 98\%$	0.698	4.965	9.521	44.913	101.216

Table 5: Lending Fee Yield
This table reports time-series averages of lending fee yields for the 25 size–fee-yield groups. For group i in quarter t, the lending fee yield is calculated as

$$\text{LendingFeeYield}_t^i = \frac{\sum_{j\in i} \text{LendingFeeIncome}_{j,t}^i}{\sum_{j\in i} \text{ValueInventory}_{j,t}^i}, \qquad i = 1,\dots,25,$$

where $\text{LendingFeeIncome}_{j,t}^i$ is the total lending fee income collected by lenders on stock j in quarter t, and $\text{ValueInventory}_{j,t}^i$ is the corresponding average daily dollar value of lendable inventory. Lending fee yields are annualized and reported in percent. The sample period is 2007–2024.

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
$\leq 80\%$	1.98	1.98	1.76	1.75	1.46	1.46	0.91	0.91	0.90	0.90
80-90%	2.24	2.24	1.74	1.74	1.01	1.01	1.17	1.17	0.29	0.29
90-95%	2.44	2.44	1.61	1.61	1.07	1.07	1.05	1.05	0.16	0.16
95-98%	2.10	2.09	1.69	1.69	1.48	1.48	0.52	0.52	0.13	0.13
$\geq 98\%$	2.06	2.06	1.67	1.67	1.14	1.14	0.25	0.25	0.02	0.02

Table 6: Dividend Yield (Data and Model)
This table reports dividend yields for the 25 size–fee-yield groups. The *Data* columns report time-series averages of group-level dividend yields. For group i in quarter t, the dividend yield is computed as

$$\text{DividendYield}_t^i = \frac{\sum_{j\in i} \text{Dividend}_{j,t}^i}{\sum_{j\in i} \text{MarketCapitalization}_{j,t}^i}, \qquad i = 1,\dots,25,$$

where $\text{Dividend}_{j,t}^i$ denotes the total cash dividends paid by stock j in quarter t, and $\text{MarketCapitalization}_{j,t}^i$ denotes the corresponding average daily market capitalization. The *Model* columns report the dividend yields implied by the calibrated model. Ordinary cash dividend data are from CRSP. Dividend yields are annualized and reported in percent. The sample period is 2007–2024.

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%	6.77	8.07	9.86	7.16	4.82
80-90%	6.79	7.26	10.04	8.30	2.39
90-95%	7.68	8.37	10.29	8.20	2.93
95-98%	6.93	7.90	10.15	6.63	5.18
$\geq$ 98%	6.74	7.43	9.53	7.77	8.35

Table 7: Risk Premia
This table reports estimated risk premia for the 25 size–fee-yield groups. Group-level risk premia are computed as the product of each group's factor loadings and the corresponding factor premia. Factor loadings are estimated by regressing monthly value-weighted group returns on the Carhart four factors over the period 2007–2024. Factor premia are computed as the time-series averages of the factor returns over 1972–2024.

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%	2.00	3.14	7.61	10.57	30.30
80-90%	2.12	3.59	9.04	18.73	61.08
90-95%	2.28	4.29	11.77	30.93	97.66
95-98%	2.54	5.70	16.96	47.24	135.80
$\geq$ 98%	3.11	9.02	27.27	74.24	192.87

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%	0.45	0.70	1.27	1.51	3.19
80-90%	0.48	0.80	1.49	2.41	5.54
90-95%	0.51	0.95	1.90	3.70	7.98
95-98%	0.57	1.25	2.68	5.22	10.44
$\geq$ 98%	0.71	1.97	4.20	7.25	13.89

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%	35.67	25.23	15.47	7.86	0.18
80-90%	2.44	2.47	1.88	0.55	0.03
90-95%	1.24	1.18	0.89	0.38	0.02
95-98%	0.73	0.70	0.65	0.30	0.02
$\geq$ 98%	0.80	0.50	0.56	0.25	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%	34.61	20.68	9.07	18.61	0.36
80-90%	2.43	2.17	1.27	1.21	0.07
90-95%	1.29	1.14	0.73	0.63	0.05
95-98%	0.82	0.82	0.60	0.55	0.04
$\geq$ 98%	0.83	0.81	0.62	0.54	0.04

Table 9: Distribution of Total Value Added from Lending Fee Income
This table reports the distribution of total value added from lending fee income across the 25 size–fee-yield groups. The reported numbers are time-series averages of each group's share of aggregate value added. In Panel A, a group's value added is computed as its value wedge multiplied by the total market value of lenders' shares in inventory in that group. In Panel B, the same quantity is computed under a counterfactual in which lending fee income is allocated pro rata across groups.

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%	0.12	0.13	0.17	0.14	3.25
80-90%	0.21	0.25	0.41	2.23	31.12
90-95%	0.22	0.39	1.00	7.21	47.06
95-98%	0.25	1.02	2.55	17.43	64.87
$\geq$ 98%	1.21	5.63	10.57	46.41	101.82

Table 10: Short Sellers' Alphas
This table reports model-implied expected excess returns, net of lending fees, earned by short sellers in each group *i*. All values 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 2024, 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 U.S. 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 features: 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

[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.

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 U.S. 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 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.

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

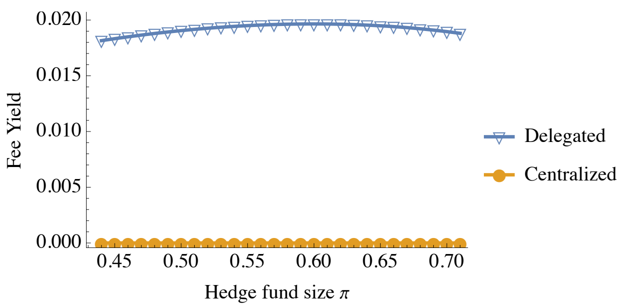
3. A new hedge fund and a new cohort of liquidity investors arrive in the market.
4. The hedge fund can pick up fee quotes for lendable shares. The realized volume of borrowed shares becomes publicly observable.
5. Bid and ask quotes are posted by market participants in the limit order market.
6. Investors 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 Comparative Statics: Intermediated vs. Centralized Securities Lending

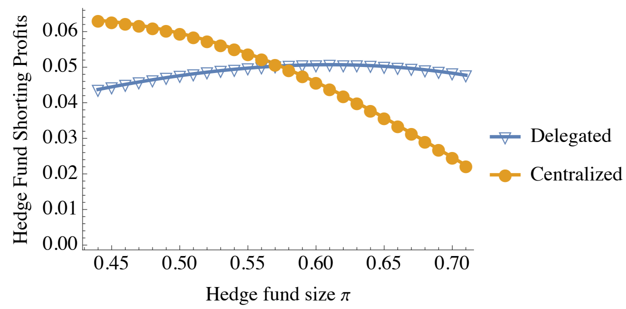
Figure 8 illustrates the results of Proposition 4 with comparative statics with respect to the parameter π, which governs the size of trades that hedge funds can take relative to those of liquidity traders. Panel (a) compares the fee yield lenders obtain conditional on shares lending being centralized 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 trading. 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.

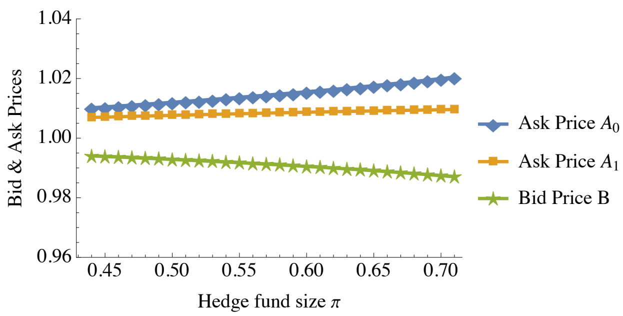
This impact on quotes in the stock market is illustrated in the bottom two panels of Figure 8. Panel (c) plots bid and ask quotes in the limit order market when securities lending is organized as an intermediated market structure, and Panel (d) considers the case where securities lending is centralized 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.



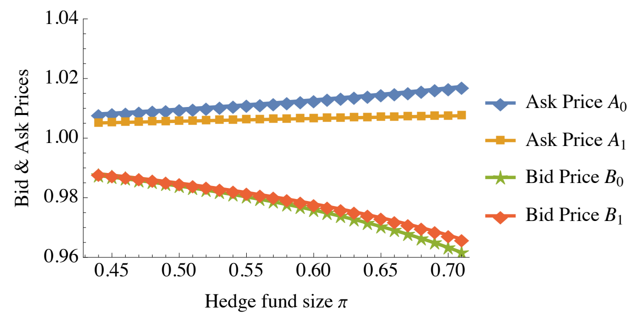
(a) Fee yield across markets



(b) Hedge fund shorting profits across markets



(c) Limit orders with delegated securities lending



(d) Limit orders with centralized securities lending

Figure 8: 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 $r_f = 0.01$.

D Analyses and Proofs Omitted from Main Text

D.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 a custodian 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.

D.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 $\phi(c \cdot e^v) = e^v \lambda c \tilde{\phi}$ 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, $Q(c,v) = P(c)e^v$.

D.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:

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

D.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

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

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

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

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:

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

D.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, respectively):

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

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

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

[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.

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 $v_1 \equiv v_{1|0} = v_{1|1}$ yields:

$$v_1 = 1 - F_w(\log \frac{A_1}{P}). \tag{49}$$

Analogously, we obtain:

$$v_0 = 1 - F_w(\log \frac{A_0}{P}). \tag{50}$$

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:

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

Which yields

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

Analogously, we obtain for the case without preexisting borrowing:

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

D.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:

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

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

E 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.

E.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:

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

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

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

where we used the fact that $\mathbb{E}[e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_i}|s] = 1$ given the normalization of w_i specified above.

To solve the quantitative model numerically, when computing $\tilde{L}$, we exploit the fact that the solution substantially simplifies in the large-λ limit (short periods) to:

$$\tilde{L}(w_i, s) = e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_i} \cdot (\tilde{A}(s) + \tilde{\kappa}(s)). \tag{72}$$

That is, the solution decouples across states s rather than being coupled through the HJB equation. Intuitively, when λ is large, period ends arrive much more frequently than Markov state transitions, so the state typically remains fixed within a period; the liability is then dominated by the cost of redelivery at the end of the period (the expected next-period ask), and the cross-state coupling becomes negligible. Our calibrated value $\lambda = 365$ is more than a hundred times larger than the Markov transition rates between states s, which motivates this approach. We employ this approximation because numerical model solution and estimation across the full cross-section of 25 states would otherwise not be computationally feasible.

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

$$\phi(c,s) = B(c,s) - L(c, w_\phi(s), s). \tag{73}$$

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

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

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

$$\begin{aligned}\tilde{\Pi}'(w_\phi(s)) =& \pi \cdot f_{w|s}(w_\phi(s)) \left(\tilde{B}(s) - e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_\phi(s)} \cdot \mathbb{E}[\tilde{L}(w,s)|s] \right) \\ &- \pi \cdot F_{w|s}(w_\phi(s)) \cdot e^{\frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}\cdot w_\phi(s)} \cdot \mathbb{E}[\tilde{L}(w,s)|s] \cdot \frac{\sigma_v(s)^2}{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}.\end{aligned} \tag{75}$$

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

$$\frac{f_{w|s}(w_\phi(s))}{F_{w|s}(w_\phi(s))} \cdot \frac{\sigma_v(s)^2+\sigma_\varepsilon(s)^2}{\sigma_v(s)^2} \cdot \left(\frac{\tilde{B}(s)}{\tilde{L}(w_\phi(s),s)} - 1 \right) = 1. \tag{76}$$

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

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

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

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

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:

$$F_{w|s}(w_\phi(s))\frac{\pi}{\rho(s)}\tilde{\phi}(s)\cdot c\lambda\cdot(1+\pi) \tag{79}$$

and the total lending fee income is

$$F_{w|s}(w_\phi(s))\pi\tilde{\phi}(s)\cdot c\lambda. \tag{80}$$

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:

$$\pi\cdot(1-F_w(w_A))+1. \tag{81}$$

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

$$\kappa(c,s) = \frac{F_{w|s}(w_\phi(s))\frac{\pi}{\rho(s)}\tilde{\phi}(s)\cdot c\cdot\lambda\cdot(1+\pi)}{\pi\cdot(1-F_w(w_A))+1}. \tag{82}$$

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:

$$Q(c,w_A) = e^{\frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2}\cdot w_A}\cdot P(c) = A(c)+\kappa(c), \tag{83}$$

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

$$w_A = \frac{\sigma_v^2+\sigma_\varepsilon^2}{\sigma_v^2}\log\left[\frac{A(c)+\kappa(c)}{P(c)}\right]. \tag{84}$$

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

$$\frac{B}{P} = \frac{\pi F_w(w_\phi)\,\mathbb{E}[e^v|w\leq w_\phi]+(1-\pi)}{\pi F_w(w_\phi)+(1-\pi)}, \tag{85}$$

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

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

$$\frac{A+\kappa}{P} = \frac{\pi(1-F_w(w_A))\,\mathbb{E}[e^v|w\geq w_A]+1+(1+\pi)F_w(w_\phi)\frac{\pi}{\rho}\frac{\phi}{P}}{\pi(1-F_w(w_A))+1} \tag{87}$$

We can rewrite the conditional expectation as follows:

$$\mathbb{E}[e^v|v+\varepsilon<w_\phi]=\int_{-\infty}^{+\infty}\int_{-\infty}^{w_\phi-\varepsilon}\frac{f_\varepsilon(\varepsilon)f_v(v)}{F_w(w_\phi)}e^v dv d\varepsilon \tag{88}$$

$$=\int_{-\infty}^{+\infty}f_\varepsilon(\varepsilon)\frac{F_v(w_\phi-\varepsilon)}{F_w(w_\phi)}\int_{-\infty}^{w_\phi-\varepsilon}\frac{f_v(v)}{F_v(w_\phi-\varepsilon)}e^v dv d\varepsilon \tag{89}$$

$$=\int_{-\infty}^{+\infty}f_\varepsilon(\varepsilon)\frac{F_v(w_\phi-\varepsilon)}{F_w(w_\phi)}\mathbb{E}[e^v|v<w_\phi-\varepsilon,\varepsilon]d\varepsilon. \tag{90}$$

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

$$\mathbb{E}[e^{v_t}|v_t\leq w_\phi-\varepsilon,\varepsilon]=\frac{F_v(w_\phi-\varepsilon-\sigma_v^2)}{F_v(w_\phi-\varepsilon)}, \tag{91}$$

we obtain:

$$F_w(w_\phi)\,\mathbb{E}[e^v|w<w_\phi]=\int_{-\infty}^{+\infty}f_\varepsilon(\varepsilon)F_v(w_\phi-\varepsilon-\sigma_v^2)d\varepsilon. \tag{92}$$

Similarly, we can compute the conditional expectation:

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

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

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

$$(1-F_w(w_A))\cdot\mathbb{E}[e^v|w\geq w_A]=\int_{-\infty}^{+\infty}f_\varepsilon(\varepsilon)F_v(-w_A+\varepsilon)d\varepsilon. \tag{95}$$

Fraction of hedge fund trading volume. The market-wide fraction of trading volume by informed hedge funds, aggregated across states using the stationary state probabilities $\Pr[s]$, is given by

$$\frac{\sum_s \Pr[s]\cdot\pi\cdot\left[F_{w|s}(w_\phi(s))+(1-F_{w|s}(w_A(s)))\right]}{(2-\pi)+\sum_s \Pr[s]\cdot\pi\cdot\left[F_{w|s}(w_\phi(s))+(1-F_{w|s}(w_A(s)))\right]}, \tag{96}$$

where the numerator captures volume from hedge funds (selling at the bid for $w_i\leq w_\phi$ and buying at the ask for $w_i\geq w_A$), and the denominator additionally includes liquidity-trader volume.

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

$$\alpha(s)=\lambda\cdot\left(\frac{B(c,s)}{P(c,s)}-\frac{\mathbb{E}[L(c,v,s)|w_i<w_\phi(s)]}{P(c,s)}\right) \tag{97}$$

The probability that a hedge fund belonging to the measure π is going to borrow is: $F_{w|s}(w_\phi(s))$. We then have the following annualized expected net-of-fee surplus yield from shorting:

$$F_{w|s}(w_\phi(s)) \cdot \left(\alpha(s) - \lambda \frac{\phi(c,s)}{P(c,s)}\right). \tag{98}$$

E.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

$$x = F_w(w_\phi|v) \cdot \pi = F_\varepsilon(w_\phi - v) \cdot \pi. \tag{99}$$

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

$$x = \pi \cdot F_\varepsilon(w_\phi(v) - v) \tag{100}$$

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 $(\hat{v})$:

$$\hat{v} \equiv v + b, \tag{101}$$

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.

E.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:

$$A(c,\hat{v}) + \kappa(c,\hat{v}) = \mathbb{E}[e^v|\hat{v}, w_i = w_A(\hat{v})] \cdot P(c) \tag{102}$$

We define:

$$AoP(\hat{v}) \equiv \frac{A(c,\hat{v})}{P(c,\hat{v})}, \tag{103}$$

$$\kappa oP \equiv \frac{\kappa(c,\hat{v})}{P(c,\hat{v})}. \tag{104}$$

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:

$$\frac{\mathbb{E}[e^v|\hat{v}, w_i = w_A(\hat{v})]}{\mathbb{E}[e^v|\hat{v}]} = AoP(\hat{v}) + \kappa oP, \tag{105}$$

or equivalently:

$$\exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_A(\hat{v})\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2} - \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right) = AoP(\hat{v}) + \kappa oP. \tag{106}$$

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

$$\mathbb{E}[e^v|\hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right), \tag{107}$$

$$\mathbb{E}\left[e^v \,\middle|\, \hat{v}, w_i\right] = \exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_i\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2}\right). \tag{108}$$

We can then solve equation (106) for $w_A(\hat{v})$:

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

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 $w_i > w_A(\hat{v})$:

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

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

$$F_w(w_A|v,\hat{v}) = F_w(w_A|v) = F_\varepsilon(w_A - v) = \Phi\left(\frac{w_A - v}{\sigma_\varepsilon}\right), \tag{110}$$

where $\Phi(\cdot)$ is the standard normal CDF, we can rewrite equation (109) as follows:

$$\begin{aligned} & A(c,\hat{v}) \cdot \left(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + 1 \right) \\ & = \left[\int_{-\infty}^{\infty} e^v \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + \mathbb{E}[e^v|\hat{v}] \right] \cdot P(c). \end{aligned} \tag{111}$$

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

$$\mu_{v|\hat{v}} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_b^2}\,\hat{v} \,+\, \frac{\sigma_b^2}{\sigma_v^2 + \sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right) \tag{112}$$

$$\sigma_{v|\hat{v}}^2 = \frac{\sigma_v^2\,\sigma_b^2}{\sigma_v^2 + \sigma_b^2} \tag{113}$$

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

$$AoP(\hat{v}) \cdot \mathbb{E}[e^v|\hat{v}] = \frac{\left[\int_{-\infty}^{\infty} e^v \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + \mathbb{E}[e^v|\hat{v}] \right]}{\left(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon})) f_v(v|\hat{v}) dv + 1 \right)}. \tag{114}$$

Let's define

$$\begin{aligned} \bar{w}_A & \equiv w_A(\hat{v}) - \mu_{v|\hat{v}} \\ & = \frac{\sigma_v^2\left(\sigma_b^2 + \sigma_\varepsilon^2\right) \,+\, \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\ \ln\!\left[AoP(\hat{v}) + \kappa oP\right] - \left(\frac{\sigma_b^2}{\sigma_v^2 + \sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right)\right). \end{aligned} \tag{115}$$

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:

$$z \equiv v - \mu_{v|\hat{v}} \tag{116}$$

Then we obtain

$$\Phi\left(\tfrac{w_A(\hat{v})-v}{\sigma_\varepsilon}\right) = \Phi\left(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon}\right) \tag{117}$$

$$v = \mu_{v|\hat{v}} + z \tag{118}$$

$$f_v(v \mid \hat{v})\,dv = f_z(z)\,dz \quad (\text{with } z \sim \mathcal{N}(0, \frac{\sigma_v^2\,\sigma_b^2}{\sigma_v^2+\sigma_b^2})) \tag{119}$$

$$e^v \;=\; e^{\mu_{v|\hat{v}}}\,e^z = \mathbb{E}[e^v|\hat{v}]\cdot e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)} \tag{120}$$

Then we obtain

$$AoP(\hat{v})\cdot\mathbb{E}[e^v|\hat{v}] \;=\; \frac{\mathbb{E}[e^v|\hat{v}]\displaystyle\int \pi\, e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)}\left[1-\Phi(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\,dz \;+\; \mathbb{E}[e^v|\hat{v}]}{\displaystyle\int \pi\left[1-\Phi(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\,dz \;+\; 1}, \tag{121}$$

which further simplifies to:

$$AoP \;=\; \frac{\displaystyle\int \pi\, e^{\left(z-\frac{1}{2}\,\sigma^2_{v|\hat{v}}\right)}\left[1-\Phi(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\,dz \;+\; 1}{\displaystyle\int \pi\left[1-\Phi(\tfrac{\bar{w}_A-z}{\sigma_\varepsilon})\right] f_z(z)\,dz \;+\; 1}, \tag{122}$$

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:

$$\min[F_\varepsilon(w_\phi(v)-v), F_w(w_B(\hat{v})|\hat{v})]. \tag{123}$$

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

$$\min[x, \pi\cdot F_w(w_B(\hat{v})|\hat{v})] = \min[x, \pi\cdot \Pr(w_i < w_B(\hat{v})|\hat{v})], \tag{124}$$

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:

$$\begin{aligned}\min[x, \pi\cdot \Pr[w_i < w_B(\hat{v})|\hat{v}, v]] &= \min[x, \pi\cdot \Pr[v+\varepsilon_i < w_B(\hat{v})|\hat{v}, v]]\\ &= \min[x, \pi\cdot \Pr[\varepsilon_i < w_B(\hat{v})-v|\hat{v}, v]]\\ &= \min[x, \pi\cdot F_\varepsilon(w_B(\hat{v})-v)]\\ &= \min\left[x, \pi\cdot\Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] \end{aligned} \tag{125}$$

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):

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

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

$$
\int_{-\infty}^{\infty} \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv. \tag{127}
$$

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:

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

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:

$$
P(c,\hat{v}) \equiv \mathbb{E}[e^{v}|\hat{v}] \cdot P(c). \tag{129}
$$

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:

$$
BoP(\hat{v},x)\, \mathbb{E}[e^{v}|\hat{v}] = \frac{\int_{-\infty}^{\infty} e^{v} \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv + (1-\pi)\, \mathbb{E}[e^{v}|\hat{v}]}{\int_{-\infty}^{\infty} \min\left[x, \pi \cdot \Phi\left(\frac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right)\right] f_v(v|\hat{v}) dv + (1-\pi)}. \tag{130}
$$

Let's define

$$
\bar{w}_B \equiv w_B(\hat{v}) - \mu_{v|\hat{v}} \tag{131}
$$

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

$$
\Phi\left(\tfrac{w_B(\hat{v})-v}{\sigma_\varepsilon}\right) = \Phi\left(\tfrac{\bar{w}_B - z}{\sigma_\varepsilon}\right) \tag{132}
$$

$$
v = \mu_{v|\hat{v}} + z \tag{133}
$$

$$
f_v(v \mid \hat{v})\, dv = f_z(z)\, dz \quad (\text{with } z \sim \mathcal{N}(0, \frac{\sigma_v^2 \sigma_b^2}{\sigma_v^2 + \sigma_b^2})) \tag{134}
$$

$$
e^{v} = e^{\mu_{v|\hat{v}}} e^{z} = \mathbb{E}[e^{v}|\hat{v}] \cdot e^{\left(z - \frac{1}{2}\sigma^2_{v|\hat{v}}\right)}. \tag{135}
$$

Subsituting those results into (130) yields

$$BoP(\hat{v}) \cdot \mathbb{E}[e^{v}|\hat{v}] \tag{136}$$

$$= \frac{\mathbb{E}[e^{v}|\hat{v}] \int e^{\left(z-\frac{1}{2}\sigma^2_{v|\hat{v}}\right)} \min\left[x,\ \pi\Phi\left(\frac{\bar{w}_B-z}{\sigma_\varepsilon}\right)\right] f_z(z)\,dz \ +\ (1-\pi)\,\mathbb{E}[e^{v}|\hat{v}]}{\int \min\left[x,\ \pi\Phi\left(\frac{\bar{w}_B-z}{\sigma_\varepsilon}\right)\right] f_z(z)\,dz \ +\ (1-\pi)}, \tag{137}$$

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

$$BoP \ = \ \frac{\int_{-\infty}^{\infty} e^{\left(z-\frac{1}{2}\sigma^2_{v|\hat{v}}\right)} \min\left[x,\ \pi\Phi\left(\frac{\bar{w}_B-z}{\sigma\varepsilon}\right)\right] f_z(z)\,dz + (1-\pi)}{\int_{-\infty}^{\infty} \min\left[x,\ \pi\Phi\left(\frac{\bar{w}_B-z}{\sigma\varepsilon}\right)\right] f_z(z)\,dz + (1-\pi)}, \tag{138}$$

$$\bar{w}_B \ = \ w_B(\hat{v}) - \mu_{v|\hat{v}} \tag{139}$$

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:

$$x \approx \begin{cases} \pi & \text{if } \pi < \rho \\ \rho & \text{if } \pi > \rho \end{cases} \ = \ \min[\pi, \rho], \tag{140}$$

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:

$$x \ = \pi \cdot \Pr[w_i < w_B(\hat{v})|\hat{v}, v] = \pi \cdot \Pr[\varepsilon_i < w_B(\hat{v}) - v|\hat{v}, v] = \pi \cdot \Phi\left(\frac{w_B(\hat{v}) - v}{\sigma_\varepsilon}\right) \tag{141}$$

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

$$x \ = \pi \cdot \Phi\left(\frac{\bar{w}_B - z}{\sigma_\varepsilon}\right). \tag{142}$$

Rearranging this equation yields the following threshold value for z:

$$\bar{w}_B - \Phi^{-1}\left(\frac{x}{\pi}\right)\sigma_\varepsilon. \tag{143}$$

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:

$$\text{Profit}(c,\hat{v},x,w_i,s) \ = B(c,\hat{v},x) - L(c,\hat{v},w_i) \tag{144}$$

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:

$$\begin{aligned}
0 = &-(r_f+\sigma_c(s)\chi-\mu_c(s))\cdot\lambda c\cdot\tilde{L}(\hat{v},w_i,s) \\
&+\lambda\cdot\{ce^v+\lambda c\cdot\mathbb{E}[e^v|w_i,\hat{v}]\,\mathbb{E}[\tilde{A}(\hat{v}',s)+\tilde{\kappa}(\hat{v}',s)] \\
&-\lambda c\cdot\tilde{L}(\hat{v},w_i,s)\}+\lambda c\cdot(\Lambda(s)\odot\mathbf{U}(s))\cdot\tilde{\mathbf{L}}(\hat{v},w_i),
\end{aligned} \tag{145}$$

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

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

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:

$$\begin{aligned}
\tilde{\mathbf{L}}(\hat{v},w_i) =&\text{diag}(\mathbb{E}[e^v|w_i,\hat{v},s])\cdot(\Lambda\odot\mathbf{U}-\text{diag}(r_f+\sigma_c(s)\chi-\mu_c(s))+\lambda)^{-1} \\
&\cdot(\mathbf{1}+\lambda\,\mathbb{E}[\tilde{\mathbf{A}}(\hat{v}')+\tilde{\kappa}(\hat{v}')]).
\end{aligned} \tag{147}$$

Analogously to the intermediated marketcase discussed above, we exploit the large-λ approximation to obtain the state-by-state solution:

$$\tilde{L}(\hat{v},w_i,s) = \mathbb{E}[e^v|w_i,\hat{v},s]\cdot\mathbb{E}[\tilde{A}(\hat{v}',s)+\tilde{\kappa}(\hat{v}',s)], \tag{148}$$

or equivalently the non-scaled version:

$$L(c,\hat{v},w_i,s) = \mathbb{E}[e^v|w_i,\hat{v},s]\cdot\mathbb{E}[A(c,\hat{v}',s)+\kappa(c,\hat{v}',s)]. \tag{149}$$

Going forward, 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:

$$\begin{aligned}
\frac{L(c,\hat{v},w_i)}{P(c,\hat{v})} &= \frac{\mathbb{E}[e^v|w_i,\hat{v}]\cdot\mathbb{E}[A(c,\hat{v}')+\kappa(c,\hat{v}')]}{P(c)\,\mathbb{E}[e^v|\hat{v}]} \\
&= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{P(c)\,\mathbb{E}[e^v|\hat{v}]}\,\mathbb{E}[\frac{A(c,\hat{v}')+\kappa(c,\hat{v}')}{P(c,\hat{v}')}P(c)\,\mathbb{E}[e^{v'}|\hat{v}']] \\
&= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot\mathbb{E}[\frac{A(c,\hat{v}')+\kappa(c,\hat{v}')}{P(c,\hat{v}')}\,\mathbb{E}[e^{v'}|\hat{v}']] \\
&= \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot(AoP+\kappa oP), \qquad (150)
\end{aligned}$$

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

$$\frac{\text{Profit}(c,\hat{v},x,w_i)}{P(c,\hat{v})} = BoP(x) - \frac{\mathbb{E}[e^v|w_i,\hat{v}]}{\mathbb{E}[e^v|\hat{v}]}\cdot LoP. \qquad (151)$$

where we define:

$$LoP \equiv AoP + \kappa oP. \qquad (152)$$

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

$$BoP(x) - \frac{\mathbb{E}\left[e^v \,\middle|\, \hat{v}, w_B(\hat{v})\right]}{\mathbb{E}\left[e^v \,\middle|\, \hat{v}\right]}\cdot LoP = 0, \qquad (153)$$

with

$$\mathbb{E}[e^v|\hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v}\right), \qquad (154)$$

$$\mathbb{E}\left[e^v \,\middle|\, \hat{v}, w_i\right] = \exp\left(\frac{\sigma_v^2\left(\sigma_\varepsilon^2\hat{v}+\sigma_b^2 w_i\right)}{\sigma_v^2(\sigma_b^2+\sigma_\varepsilon^2)+\sigma_b^2\sigma_\varepsilon^2}\right), \qquad (155)$$

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

$$w_B(\hat{v}) \;=\; \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\,\hat{v} \;+\; \frac{\sigma_v^2\left(\sigma_b^2+\sigma_\varepsilon^2\right) \;+\; \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\,\ln\!\Big(\frac{BoP(x)}{LoP}\Big). \qquad (156)$$

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

$$\begin{aligned}
\bar{w}_B \;&=\; w_B(\hat{v}) - \mu_{v|\hat{v}} \\
&=\; \frac{\sigma_v^2\left(\sigma_b^2+\sigma_\varepsilon^2\right) \;+\; \sigma_b^2\,\sigma_\varepsilon^2}{\sigma_v^2\,\sigma_b^2}\,\ln\!\Big(\frac{BoP(x)}{LoP}\Big) - \left(\frac{\sigma_b^2}{\sigma_v^2+\sigma_b^2}\left(-\frac{\sigma_v^2}{2}\right)\right). \qquad (157)
\end{aligned}$$

E.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:

$$BoP(x) \cdot \mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}] \cdot LoP. \tag{158}$$

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:

$$\begin{aligned} SoP(w_i,x) &\equiv \mathbb{E}[(BoP(x) \cdot \mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}] \cdot LoP)^+|w_i] \\ &= \int_{-\infty}^{+\infty} (BoP(x) \cdot \mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i,\hat{v}] \cdot LoP)^+ f(\hat{v}|w_i) d\hat{v}, \end{aligned} \tag{159}$$

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 (107) and (108). The pdf $f(\hat{v}|w_i)$ is normal with the following mean and variance:

$$\mathbb{E}[\hat{v} \mid w_i] \; = -\frac{\sigma_v^2}{2} + \frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2}\big(w_i - \big(-\tfrac{\sigma_v^2}{2}\big)\big), \tag{160}$$

$$\mathrm{Var}(\hat{v} \mid w_i) \; = \sigma_b^2 + \frac{\sigma_v^2\,\sigma_\varepsilon^2}{\sigma_v^2+\sigma_\varepsilon^2}. \tag{161}$$

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 $x(v) = \pi F_w(w_\phi(v))$:

$$\frac{\phi_0}{P} = \int_{w_B^{-1}(w_\phi(v))}^{\infty} (BoP(\hat{v},x(v)) \cdot \mathbb{E}[e^v|\hat{v}] - \mathbb{E}[e^v|w_i = w_\phi(v),\hat{v}] \cdot LoP) f(\hat{v}|w_i = w_\phi(v)) d\hat{v}. \tag{162}$$

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 $w_\phi(v)$:

$$F_w(w_\phi(v)|v) \cdot \pi = \rho \tag{163}$$

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

$$F_w(w_i|v) = F_w(v+\varepsilon_i|v) = F_\varepsilon(w_i - v) \tag{164}$$

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

$$F_\varepsilon(w_\phi(v) - v) \cdot \pi = \rho \tag{165}$$

which yields the explicit result:

$$w_\phi(v) = v + F_\varepsilon^{-1}\left(\frac{\rho}{\pi}\right). \tag{166}$$

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:

$$\frac{\phi(c,v)}{P(c)} = \phi oP(v) = SoP(w_i = w_\phi(v), x = \rho). \tag{167}$$

The average fee and fee yield is then given by

$$\lambda\, \mathbb{E}[\phi oP(v)]. \tag{168}$$

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:

$$(1+\pi)\,\mathbb{E}[\phi(c,v)], \tag{169}$$

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.

$$\begin{aligned}
&(1+\pi)\,\mathbb{E}[\phi(c,v)]\\
&= \int_{-\infty}^{+\infty} \underbrace{\Big(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1\Big)}_{\text{Average volume given } \hat{v}} \cdot \kappa(c,\hat{v}) f(\hat{v}) d\hat{v}\\
&= \int_{-\infty}^{+\infty} \underbrace{\Big(\int_{-\infty}^{\infty} \pi \cdot (1 - \Phi(\frac{w_A(\hat{v}) - v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1\Big)}_{\text{Average volume given } \hat{v}} \cdot \kappa oP \cdot \mathbb{E}[e^v|\hat{v}] \cdot P(c) f(\hat{v}) d\hat{v}
\end{aligned} \tag{170}$$

This yields

$$\kappa oP = \frac{(1+\pi)\,\mathbb{E}[\phi oP(v)]}{\int_{-\infty}^{+\infty}(\int_{-\infty}^{\infty}\pi \cdot (1-\Phi(\frac{w_A(\hat{v})-v}{\sigma_\varepsilon}))f_v(v|\hat{v})dv + 1)\cdot \mathbb{E}[e^v|\hat{v}] \cdot f(\hat{v})d\hat{v}}. \tag{171}$$

where

$$w_A(\hat{v}) = \bar{w}_A + \mu_{v|\hat{v}} \tag{172}$$

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 $\rho < \pi$:

$$\alpha(v) = \lambda \int_{-\infty}^{F_\varepsilon^{-1}\left(\frac{\rho}{\pi}\right)} SoP(v + \varepsilon_i, x = \rho) \frac{f_\varepsilon(\varepsilon)}{\frac{\rho}{\pi}} d\varepsilon. \tag{173}$$

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

$$\alpha(v) = \lambda \int_{-\infty}^{+\infty} SoP(v + \varepsilon_i, x = \pi) f_\varepsilon(\varepsilon) d\varepsilon \tag{174}$$

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:

$$\mathbb{E}[\alpha(v)] = \int_{-\infty}^{+\infty} \alpha(v) f_v(v) dv \tag{175}$$

against the average fee:

$$\lambda \, \mathbb{E}[\phi o P(v)]. \tag{176}$$

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

$$\mathbb{E}[\alpha(v)] - \lambda \, \mathbb{E}[\phi o P(v)]. \tag{177}$$

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

$$(\mathbb{E}[\alpha(v)] - \lambda \, \mathbb{E}[\phi o P(v)]) \cdot \frac{\rho}{\pi} \tag{178}$$

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

$$\mathbb{E}[\alpha(v)] \tag{179}$$

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

E.3 Derivation of Conditional Expectations

Using the fact that

$$\mathbb{E}[e^v | \hat{v}] = \exp\left(\mathbb{E}[v \mid \hat{v}] + \frac{1}{2} \mathrm{Var}(v \mid \hat{v})\right). \tag{180}$$

we obtain the result that

$$\mathbb{E}[e^v | \hat{v}] = \exp\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_b^2} \hat{v}\right). \tag{181}$$

since:

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

$$= \frac{\sigma_v^2}{\sigma_v^2+\sigma_b^2}\hat{v} - \frac{1}{2}\frac{\sigma_v^2\sigma_b^2}{\sigma_v^2+\sigma_b^2} \tag{182}$$

$$\mathrm{Var}(v \mid \hat{v}) = \sigma_v^2 \frac{\sigma_b^2}{\sigma_v^2+\sigma_b^2} \tag{183}$$

Next, we compute

$$\mathbb{E}\left[e^v \middle| \hat{v}, w_i\right] = \exp\left(\mathbb{E}[v|\hat{v},w_i] + \frac{1}{2}\mathrm{Var}(v|\hat{v},w_i)\right). \tag{184}$$

Define the vector of conditioning variables:

$$\mathbf{Y} = \begin{bmatrix} \hat{v} \\ w_i \end{bmatrix}. \tag{185}$$

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

$$\Sigma_{v\mathbf{Y}} = \begin{bmatrix} \sigma_v^2 & \sigma_v^2 \end{bmatrix} \tag{186}$$

$$\Sigma_{\mathbf{YY}} = \begin{bmatrix} \sigma_v^2+\sigma_b^2 & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2+\sigma_\varepsilon^2. \end{bmatrix} \tag{187}$$

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

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

$$= \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_i\right) - \frac{1}{2}\sigma_v^2 + \frac{\sigma_v^2}{D}(\sigma_\varepsilon^2\frac{1}{2}\sigma_v^2) + \frac{\sigma_v^2}{D}(\sigma_b^2\frac{1}{2}\sigma_v^2)$$

$$= \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_i\right) - \frac{1}{2}\left(\sigma_v^2 - \frac{\sigma_v^4}{D}(\sigma_\varepsilon^2+\sigma_b^2)\right). \tag{188}$$

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

$$\mathbb{E}[v|\hat{v}, w_i] + \frac{1}{2}\mathrm{Var}(v|\hat{v}, w_i) = \frac{\sigma_v^2}{D}\left(\sigma_\varepsilon^2\hat{v} + \sigma_b^2 w_i\right), \tag{189}$$

and thus:

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

F 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 custodian 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 our empirical results in Section 7.1, which reveal that beneficial owners such as mutual funds maintain very stable relationships to typically one custodian lender and sign long-term contracts specifying a fixed proportional revenue-sharing rule. 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 custodian lenders: longer-term contracts signed at date $t = 0$ pin down the delegation of stock lending to custodian lenders 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

[4]When orders are filled based on a pro-rata allocation rule, this is consistent with a custodian having unrestricted order sizes,

for a custodian to refrain from opportunistically front running, and (2) a hedge fund's borrowing reveals its intention to short sell, then 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 market 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, who provide these quotes, do not obtain that information.

Implicit in this distinction between market structures is that an atomistic asset owner cannot simultaneously act as a custodian; that is, direct lending is precluded under the intermediated market structure. Otherwise, an asset owner serving as its own custodian 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.

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.

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

$$(1-\delta_i)\frac{\pi\phi_{i,t}}{\rho_i}, \tag{190}$$

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 $i \in \{1,\ldots,I\}$, 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.

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

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

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]

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:

$$\delta_i \cdot \mathbb{E}_t \int_t^{\infty} \frac{m_u}{m_t} F_v(v_\phi) \frac{\pi \phi_u}{\rho} dN_u = \delta_i \cdot \frac{\lambda \cdot F_v(v_\phi) \frac{\pi \phi(c)}{\rho}}{r_f + \sigma_c \chi - \mu_c}. \tag{191}$$

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

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:

$$B(c) - \mathbb{E}[e^v | v < v_\phi] P(c), \tag{192}$$

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:

$$\mathbb{E}[e^v | v > v_\phi] P(c) - A_0(c), \tag{193}$$

$$(1+\pi) \cdot (\mathbb{E}[e^v | v > v_\phi] P(c) - A_1(c)). \tag{194}$$

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:

$$\frac{\Pr[v \geq \log(A_0/P)]}{\Pr[v > v_\phi]} \text{ and } \frac{\Pr[v \geq \log(A_1/P)]}{\Pr[v > v_\phi]} \tag{195}$$

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:

$$\delta_{min} = \min\{\delta_i\in[0,1]: \text{IC 1, IC 2, IC 3}\}. \tag{196}$$

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 (192),(193), and (194) 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.

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 (192). 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 constraint 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 (196) 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.

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 custodian 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 custodian 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 custodian lenders, 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.

G Additional Tables

Panel A. Removing Observations Around Voting Record Dates

	Lending Fee (annual, in percent)				
	Large-Cap	Mid-Cap	Small-Cap	Micro-Cap	Nano-Cap
Fee Yield Percentiles	(1)	(2)	(3)	(4)	(5)
$\leq$ 80%	0.270	0.272	0.295	0.461	8.050
80-90%	0.276	0.295	0.457	2.680	30.538
90-95%	0.279	0.417	1.064	8.268	49.610
95-98%	0.300	1.050	2.774	19.959	73.123
$\geq$ 98%	1.633	7.050	12.422	55.039	117.903

Panel B. Removing Observations Around Ex-Dividend Dates

	Lending Fee (annual, in percent)				
	Large-Cap	Mid-Cap	Small-Cap	Micro-Cap	Nano-Cap
Fee Yield Percentiles	(1)	(2)	(3)	(4)	(5)
$\leq$ 80%	0.208	0.222	0.268	0.440	8.479
80-90%	0.223	0.262	0.440	2.754	32.083
90-95%	0.228	0.383	1.026	8.477	52.893
95-98%	0.252	0.960	2.713	20.876	75.849
$\geq$ 98%	1.496	6.928	12.618	57.073	122.467

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 2024.

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	0.11	0.08	1.12
80-90%	0.06	0.13	0.19	0.35	3.07
90-95%	0.07	0.18	0.26	0.56	3.71
95-98%	0.11	0.23	0.32	0.72	4.60
$\geq$ 98%	0.13	0.29	0.47	1.02	4.96

Table A2: Spread Yield

This table reports time-series averages of spread yields for the 25 size–fee-yield groups. For group i in quarter t, the primary measure of spread yield is defined as

$$\text{SpreadYield}_t^i = \frac{\sum_{j\in i}\text{SpreadRevenue}_{j,t}^i}{\sum_{j\in i}\text{MarketCapitalization}_{j,t}^i}, \qquad i = 1,\ldots,25,$$

where $\text{SpreadRevenue}_{j,t}^i$ is the total spread revenue of stock j in quarter t, computed as daily share trading volume multiplied by one-half of the closing bid–ask spread and summed over days in the quarter.

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	$\leq$ 80%	80-90%	90-95%	95-98%	$\geq$ 98%	$\leq$ 80%	80-90%	90-95%	95-98%	$\geq$ 98%	$\leq$ 80%	80-90%	90-95%	95-98%	$\geq$ 98%	$\leq$ 80%	80-90%	90-95%	95-98%	$\geq$ 98%	$\leq$ 80%	80-90%	90-95%	95-98%	$\geq$ 98%
Large-Cap	$\leq$ 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%	0.44	0.38	0.14	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%	0.13	0.28	0.41	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	0.08	0.23	0.48	0.12	0.01	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
	$\geq$ 98%	0.06	0.01	0.05	0.16	0.71	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	$\leq$ 80%	0.01	0.00	0.00	0.00	0.00	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
	80-90%	0.00	0.00	0.00	0.00	0.00	0.31	0.54	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	0.00
	90-95%	0.00	0.00	0.00	0.00	0.00	0.07	0.28	0.50	0.12	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	0.05	0.24	0.53	0.11	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
	$\geq$ 98%	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.02	0.20	0.70	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	$\leq$ 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	0.00	0.01	0.00	0.00	0.29	0.54	0.09	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	0.47	0.11	0.01	0.04	0.02	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	0.13	0.02	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\geq$ 98%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.03	0.22	0.59	0.01	0.02	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00
Micro-Cap	$\leq$ 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	0.01	0.01	0.00	0.01	0.02	0.02	0.01	0.30	0.50	0.08	0.02	0.01	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	0.03	0.05	0.26	0.44	0.12	0.03	0.01	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	0.00	0.02	0.00	0.00	0.00	0.01	0.04	0.02	0.09	0.24	0.41	0.14	0.01	0.01	0.00	0.00	0.00
	$\geq$ 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	0.03	0.01	0.02	0.07	0.26	0.55	0.00	0.01	0.00	0.01	0.01
Nano-Cap	$\leq$ 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	0.11	0.02	0.01	0.01	0.00	0.81	0.03	0.01	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	0.02	0.04	0.03	0.03	0.02	0.31	0.43	0.10	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	0.01	0.02	0.03	0.03	0.03	0.11	0.27	0.35	0.11	0.04
	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	0.01	0.04	0.05	0.06	0.11	0.26	0.34	0.12
	$\geq$ 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	0.01	0.09	0.03	0.05	0.11	0.26	0.45

Table A3: Markov Matrix

This table reports the Markov transition matrix of quarterly movements of stocks across the 25 size–fee-yield groups. For quarter t, the transition probability from group i to group j is defined as $p_t^{i \to j} = \text{TotalMarketValue}_t^{i \to j} / \sum_{j=1}^{25} \text{TotalMarketValue}_t^{i \to j}$, for $i, j = 1, \ldots, 25$, where $\text{TotalMarketValue}_t^{i \to j}$ is the sum of market capitalizations of stocks that move from group i to group j over quarter t. The reported Markov matrix contains time-series averages of these transition probabilities. Each row is normalized so that transition probabilities from a given origin group sum to one.

Parameters Governing State-Contingent Growth

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	0.00	-0.33	-0.68	-1.05	-0.87
80-90%	0.11	-0.34	-1.03	-0.73	-1.77
90-95%	0.19	-0.42	-0.94	-0.74	-2.14
95-98%	0.04	-0.35	-0.56	-1.31	-2.20
$\geq$ 98%	0.02	-0.33	-0.73	-1.87	-3.65

Table A4: Estimated Dividend Growth Parameters
This table reports the estimated state-contingent parameter values $u(s)$, where the value for the first state is normalized to zero, $u(1) = 0$. In addition, our estimation yields $\mu_c = 0.0884$.

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%	19.70	24.95	32.57	27.68	73.66
80-90%	23.59	31.03	43.09	53.32	110.79
90-95%	24.63	36.59	49.74	69.93	123.35
95-98%	28.04	42.07	56.51	80.07	146.34
$\geq$ 98%	28.54	48.96	67.53	108.84	154.47

Table A5: Idiosyncratic Volatility
This table reports time-series averages of idiosyncratic volatility for the 25 size–fee-yield groups. Stock-level idiosyncratic volatility is measured as the standard deviation of residuals from rolling regressions of a stock's daily returns on the CRSP value-weighted market return using a one-year rolling window; group values are value-weighted averages across stocks. All values are annualized and reported in percent.

Volatilities of Fundamentals Reflected in 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%	0.51	0.58	0.64	0.44	14.45
80-90%	0.94	0.42	1.95	10.67	53.65
90-95%	0.98	1.53	4.77	26.33	98.74
95-98%	0.80	4.66	12.41	79.98	146.32
$\geq$ 98%	4.49	27.66	50.88	144.36	389.48

Table A6: Estimated Volatilities of Fundamentals Reflected in Hedge Funds' Signals
This table reports the estimated volatilities of fundamentals implied by hedge fund signals for the 25 size–fee-yield groups. For each group, the reported quantity is $\sqrt{\lambda\left(e^{\sigma_v(s)^2}-1\right)}$. All values 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%	2.62	2.50	2.99	3.79	2.66
80-90%	2.50	7.04	2.51	2.54	7.07
90-95%	2.54	2.98	2.50	3.37	5.90
95-98%	3.67	2.63	2.51	2.71	5.55
$\geq$ 98%	3.23	2.53	2.59	4.05	3.34

Table A7: Estimated Signal-to-Noise Volatility Ratios of Hedge Funds
This table reports the estimated signal-to-noise volatility ratios $\sigma_v/\sigma_\varepsilon$ for the 25 size–fee-yield groups. All values 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%	3.03	0.68	0.37	0.21	0.14
80-90%	1.01	0.24	0.15	0.08	0.08
90-95%	0.57	0.15	0.10	0.06	0.06
95-98%	0.32	0.10	0.07	0.05	0.06
$\geq$ 98%	0.11	0.06	0.05	0.05	0.05

Table A8: Estimated Inventories of Lendable Shares
This table reports the estimated inventories available for lending $\rho = \rho_0 - 1$ for the 25 size–fee-yield groups.

Panel A. Usage of Distinct Lending Agents by A Mutual Fund from 2018 through 2025

	One lending agent	**Two lending agents**
Number of Funds	1739	323
Distribution	84.3%	15.7%

Panel B. Usage of Distinct Lending Agents by A Mutual Fund in Each Year

	One lending agent	**Two lending agents**
Number of Fund-Year Pairs	7284	1253
Distribution	85.3%	14.7%

Panel C. Fund-Agent Relationships

	Constant	**Temporary**
Number of Fund-Agent Pairs	2166	107
Distribution	95.3%	4.7%

Table A9: Usage of Securities Lending Agents by U.S. Equity Mutual Funds.
This table reports the distribution of U.S. equity mutual funds by the number of securities lending agents employed by the mutual funds. Panel A reports the distribution of mutual funds by the number of distinct lending agents used during the period of 2018 to 2024. We find a fund ever used at most 2 lending agents. Panel B reports the distribution of mutual funds over time by the usage of lending agents. Specifically, in our pooled fund-year sample, we count the fund-year observations where a fund uses one lending agent, or two lending agents. Panel C reports the distribution of fund-agent relationships, where we count the relationships for the constant case in which the same agent serves a fund over the 2018–2024 period, or the temporary case where a disruption exists in the relationship. Our sample consists of actively-managed funds, passive index funds, and exchange-traded funds focused on the U.S. equity market, with investment styles classified by Morningstar into the nine categories: Large-cap Value, Large-cap Blend, Large-cap Growth, Mid-cap Value, Mid-cap Blend, Mid-cap Growth, Small-cap Value, Small-cap Blend, Small-cap Growth. Our data is sourced from Form N-CEN, a regulatory SEC form that registered investment companies in the United States must file to report certain census-type information about the funds. N-CEN became effective for funds with fiscal years ending on or after June 1, 2018, and registered investment companies have been required to file Form N-CEN annually since then. The form collects a wide range of information about the fund, if applicable, including: [1] general identifying information; [2] details on service providers such as the identity of its securities lending agents.

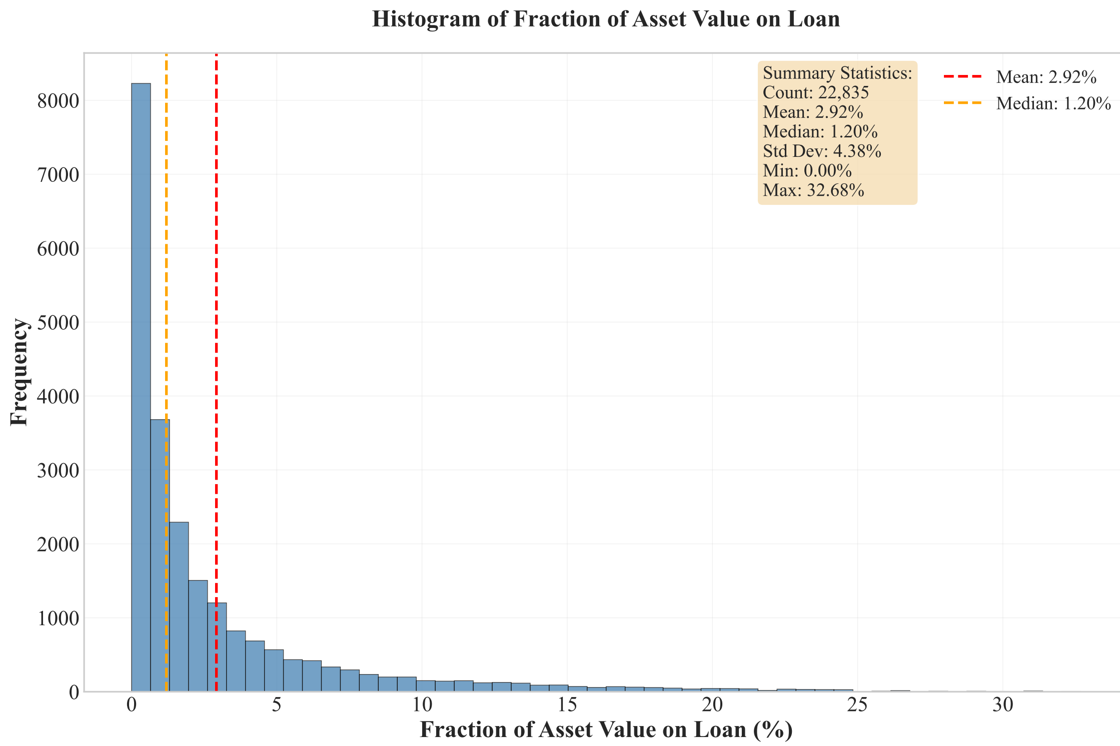


Figure 9: Histogram of the Fraction of Asset Value on Loan at the Fund and Quarter Level.
The figure plots the distribution of the fraction of asset value on loan at the fund-quarter level, computed as the sum of the value on loan relative to the sum of dollar values of all holdings reported in Form N-PORT in a given quarter by a fund engaging in securities lending. The figure confirms that the regulatory one-third constraint on lending (Investment Company Act of 1940) is never approached in practice.

Trading Volume Turnover

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	1.70	2.53	2.43	1.86	1.35
80-90%	2.44	3.64	3.85	3.11	4.25
90-95%	2.66	4.38	4.42	3.56	5.53
95-98%	3.26	5.26	5.47	4.29	6.20
$\geq$ 98%	3.29	5.55	6.36	5.90	7.42

Table A10: Trading Volume Turnover

This table reports the time-series average of trading volume turnover for the 25 size and fee yield groups. The turnover of a group is computed as the value-weighted average of traded share turnover across stocks in the group:

$$Turnover_t^i = \frac{\sum_j Turnover_{j,t}^i \times MarketCapitalization_{j,t}^i}{\sum_j MarketCapitalization_{j,t}^i}, \quad i = 1, 2, \ldots 25,$$

where $Turnover_{j,t}^i$ is the sum of daily traded shares volume against total shares outstanding for stock j in quarter t,

$$Turnover_{j,t}^i = \sum_{d=1}^{N} \frac{SharesVolume_{j,t,d}^i}{SharesOutstanding_{j,t,d}^i},$$

where N represents the number of trading days in the quarter. All numbers are annualized.

Panel A. Excess Inventory (in percent), 2007-2019

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	98.07	92.68	85.76	86.70	87.06
80-90%	91.46	70.84	58.61	43.41	45.18
90-95%	85.51	58.83	46.99	32.20	31.17
95-98%	76.78	47.05	36.35	23.86	22.04
$\geq$ 98%	59.66	31.71	26.26	17.26	12.76

Panel B. Excess Inventory (in percent), 2020-2024

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	99.41	96.67	93.21	91.70	79.34
80-90%	96.85	84.46	74.03	44.23	35.23
90-95%	93.98	75.48	61.85	28.95	32.04
95-98%	88.01	62.35	43.52	22.63	28.42
$\geq$ 98%	70.14	31.66	26.88	18.37	22.31

Table A11: Excess Inventory Before vs. After 2020.

This table reports the excess inventory across the 25 size–fee-yield groups. The reported numbers are time-series averages of each group's excess inventory. Panel A covers 2007–2019; Panel B covers 2020–2024.

Panel A. The Largest Two Lenders' Excess Inventory (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	98.33	92.58	84.57	86.63	86.51
80-90%	92.52	70.81	57.46	45.03	45.55
90-95%	86.87	60.33	50.23	32.90	33.59
95-98%	78.64	52.78	39.73	24.78	24.87
$\geq$ 98%	60.42	35.64	27.39	18.32	16.43

Panel B. The Largest Two Lenders' 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%	96.86	86.57	76.05	66.29	57.43
80-90%	91.09	68.71	54.81	39.25	38.60
90-95%	85.00	57.89	46.95	29.04	28.37
95-98%	76.28	48.74	35.66	21.70	20.82
$\geq$ 98%	48.47	24.23	21.15	14.94	12.57

Table A12: The Largest Two Lenders' Excess Inventory.
This table reports time-series averages of excess inventory held by the largest two lenders across the 25 size–fee-yield groups. In Panel A, group-level excess inventory of the largest two lenders in quarter t is defined as one minus the ratio of aggregate value on loan of the largest two lenders to their aggregate value of lendable inventory across stocks in the group,

$$\text{LargestTwoExcessInventory}_t^i = 1 - \frac{\sum_{j\in i} \text{LargestTwoValueOnLoan}_{j,t}}{\sum_{j\in i} \text{LargestTwoValueInventory}_{j,t}}, \qquad i = 1,\ldots,25,$$

where $\text{LargestTwoValueOnLoan}_{j,t}$ is the average daily dollar value of stock j's shares on loan times the market share of shares on loan by the largest two lenders and $\text{ValueInventory}_{j,t}$ is the corresponding average daily dollar value of lendable inventory times the market share of inventory by the largest two lenders.

In Panel B, the alternative measure of group-level excess inventory is constructed as a lending-fee-income–weighted average across stocks,

$$\text{AlternativeLargestTwoExcessInventory}_t^i = \frac{\sum_{j\in i} \text{LargestTwoExcessInventory}_{j,t} \times \text{LendingFeeIncome}_{j,t}}{\sum_{j\in i} \text{LendingFeeIncome}_{j,t}}, \quad i = 1,\ldots,25,$$

where $\text{LendingFeeIncome}_{j,t}$ is the total lending fee income of stock j in quarter t. The stock-level excess inventory used in this aggregation is computed as a lending-fee-income–weighted average of daily excess inventory,

$$\text{LargestTwoExcessInventory}_{j,t} = \frac{\sum_d \left(1 - \text{LargestTwoUtilization}_{j,t,d}\right) \times \text{LendingFeeIncome}_{j,t,d}}{\sum_d \text{LendingFeeIncome}_{j,t,d}},$$

where $\text{LargestTwoUtilization}_{j,t,d}$ is the fraction of lendable inventory on loan for stock j on day d. Excess inventory is reported in percent. The sample period is 2007–2024.

Loan Supply (in percent)

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	30.79	36.05	38.17	28.13	7.77
80-90%	29.95	33.47	32.74	18.22	5.49
90-95%	28.63	30.14	25.97	16.05	5.02
95-98%	27.31	24.68	21.63	15.74	4.78
$\geq$ 98%	28.40	18.26	17.50	14.29	3.82

Table A13: Loan Supply
This table reports the time-series average of loan supply for the 25 size and fee yield groups. The loan supply of a group is computed as the sum of value of shares in inventory across stocks against the sum of market capitalization in the group:

$$LoanSupply_t^i = \frac{\sum_j ValueInventory_{j,t}^i}{\sum_j MarketCapitalization_{j,t}^i}, \quad i = 1,2,\ldots 25.$$

Lending Fee (annual, in percent), 2007-2019

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	0.267	0.267	0.282	0.460	6.733
80-90%	0.276	0.298	0.501	2.527	25.152
90-95%	0.283	0.473	1.333	7.759	39.300
95-98%	0.310	1.380	3.563	17.310	54.110
$\geq$ 98%	2.072	7.590	13.441	43.020	80.970

Lending Fee (annual, in percent), 2020-2024

Fee Yield Percentiles	Large-Cap (1)	Mid-Cap (2)	Small-Cap (3)	Micro-Cap (4)	Nano-Cap (5)
$\leq$ 80%	0.364	0.378	0.433	0.616	11.043
80-90%	0.352	0.374	0.447	3.820	51.382
90-95%	0.362	0.387	0.657	11.100	91.198
95-98%	0.369	0.530	1.667	30.154	138.084
$\geq$ 98%	0.436	6.668	12.503	92.939	242.728

Table A14: Lending Fee Before vs. After 2020.
This table reports the lending fees across the 25 size–fee-yield groups. The reported numbers are time-series averages of each group's lending fees. Panel A covers 2007–2019; Panel B covers 2020–2024.

Panel A: Spread Yield, Utilization, and Size

	Lending Fees			
	(1)	(2)	(3)	(4)
Spread Yield	5.02***		4.17***	4.09***
	(0.411)		(0.338)	(0.333)
Utilization		0.30***	0.18***	0.22***
		(0.034)	(0.022)	(0.024)
Log(Market Cap)				−0.32***
				(0.028)
Observations	326,595	326,595	326,595	326,595
R-squared	0.287	0.168	0.338	0.344

Panel B: Spread, Turnover, and Spread Yield

	Lending Fees				
	(1)	(2)	(3)	(4)	(5)
Spread	6.96***			5.25***	3.03***
	(1.531)			(1.179)	(1.022)
Turnover		2.02***		1.88***	0.37***
		(0.211)		(0.193)	(0.104)
Spread Yield			5.02***		4.29***
			(0.411)		(0.444)
Observations	326,598	326,598	326,598	326,598	326,598
R-squared	0.057	0.160	0.287	0.192	0.300

Table A15: Determinants of Lending Fees
This table reports results from pooled OLS regressions of lending fees on bid-ask spread yield, utilization, stock size (log market capitalization), bid-ask spread, and turnover. Panel A examines spread yield, utilization, and size. Panel B examines bid-ask spread, turnover, and spread yield, both individually and jointly. The sample is at the stock–year–quarter level and covers 2007 through 2024. Standard errors, reported in parentheses, are clustered by stock and time. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.