Dynamic Equilibrium with Costly Short-Selling and Lending Market*

Adem Atmaz
Purdue University

Suleyman Basak
London Business School
and CEPR

Fangcheng Ruan
Purdue University

This Version: March 2021

Abstract

We develop a tractable model of costly stock short-selling and lending market within a familiar dynamic asset pricing framework. The model addresses the vast empirical literature in this market and generates implications that support many of the empirical regularities. In the model, investors' belief disagreement leads to the presence of stock lenders and short-sellers. To borrow stock shares, short-sellers pay shorting fees to lenders. Our main results for a costly-to-short stock in equilibrium are as follows. The shorting fee increases in belief disagreement. The stock price is positively, while its risk premium is negatively related to its shorting fee. The stock volatility is increased due to costly short-selling. More notably, the short interest increases in shorting fee and predicts future stock returns negatively. Higher short-selling risk can be associated with lower stock returns and less short-selling activity.

JEL Classifications: G11, G12.
Keywords: Short-selling, stock lending, belief disagreement, shorting fee, short interest, stock price, stock risk premium, volatility, short-selling risk, short-selling activity.

*Email addresses: aatmaz@purdue.edu, sbasak@london.edu, and ruan13@purdue.edu. We thank Snehal Banerjee, Adam Reed, Pedro Saffi, Edward Van Wesep, and our discussants Adrian Buss, Lawrence Jin, and Paymon Khorrami, as well as the seminar participants at the 2021 AFA meetings, 2021 MFA meetings, 2020 SFS Finance Cavalcade, and Purdue University for helpful comments. All errors are our responsibility.

Electronic copy available at: https://ssrn.com/abstract=3516969

1 Introduction

A significant fraction of stock trades in markets nowadays emerge from the activities in the stock short-selling and lending market.[1] The key imperfection in this market is that to borrow a stock, a short-seller must pay a shorting fee to the stock lender. Corresponding to the growth in the stock short-selling and lending market, a vast empirical literature (elaborated below) has developed investigating the effects of costly short-selling on the shorting fee, stock price, stock risk premium and volatility, short interest and its predictive power, short-selling risk (uncertainty about future shorting fees) and short-selling activity (volume). The existing theoretical studies (discussed below), however, are not suitable to address many of the empirical regularities documented in this literature. This is because, static models cannot meaningfully shed light on the findings of stock volatility, the predictive power of short interest, and the short-selling risk which all require a dynamic model at the minimum. On the other hand, the shorting fee and the short interest are typically specified exogenously or are deterministic in the extant dynamic models, which limit their ability to explain some of these empirical regularities.

In this paper, we fill this void and provide a comprehensive analysis of the costly stock short-selling and lending market within a familiar dynamic asset pricing framework. Our model generates rich implications that support the extensive empirical evidence on the behavior of the shorting fee, stock price, its risk premium and volatility, short interest and its predictive power, and short-selling risk and activity. We also endogenize the optimal size of stock lenders by introducing a cost of setting up a lending facility in our economy. We provide new insights and predictions, and offer simple, straightforward intuitions for all our findings.

Specifically, we develop a tractable dynamic equilibrium model by incorporating costly stock short-selling and lending following the usual market practices into an otherwise fairly standard economy. Our model builds on the fact that even though most stocks can be shorted cheaply in today's markets, for many other stocks short-selling costs can be fairly significant (D'Avolio (2002), Drechsler and Drechsler (2016)). To capture these distinct differences in short-selling costs in the cross-section, we consider a simple two-stock economy in which one

[1] For instance, Diether, Lee, and Werner (2009) find that roughly 30% of the trading volume in NYSE and NASDAQ is due to short-selling, while Hanson and Sunderam (2014) report that the average short interest ratios for NYSE and AMEX stocks have more than quadrupled from 1988 to 2011. On the other hand, Saffi and Sigurdsson (2010) find that the amount of global stock lending supply in December 2008 was $15 trillion (about 20% of the total market capitalization) and $3 trillion of this amount was lent out to short-sellers.

Electronic copy available at: https://ssrn.com/abstract=3516969

stock represents a typical costly-to-short stock, while the other represents a typical cheap-to-short stock, whose shorting fee, for clarity, is constructed to be zero in equilibrium. In this economy, investors' belief disagreement on the stock payoffs generates demand for short-selling and supply of lendable shares. The (pessimistic) short-sellers borrow stock shares from the (optimistic) lenders by paying a *shorting fee*, which along with the stock price and short interest, are determined endogenously in equilibrium. In the ensuing equilibrium, the costly-to-short stock shorting fee and short interest arise as time-varying processes with stationary steady-state distributions, enabling our model to shed light on some of the regularities that the extant works cannot. We first demonstrate that investors' subjective stock risk premia differ since in addition to the usual capital gains/losses and dividends for non-lending investors, holding the stock long yields additional income for lenders, whereas having a short position in the stock generates additional costs for short-sellers.

We then determine the equilibrium and find that the costly-to-short stock shorting fee is risky since it is driven by a time-varying, mean-reverting disagreement process so that a higher belief disagreement leads to a higher shorting fee, consistent with the empirical findings of D'Avolio (2002). This result occurs because not all optimistic investors are lenders and those who are lenders can only lend a part of their stock holdings, *partial lending*. Hence, a higher disagreement leads to a lesser increase in lending supply than the increase in shorting demand, and so for the stock lending market to clear the equilibrium shorting fee increases.

We next show that the costly-to-short stock price is positively while its risk premium negatively related to its shorting fee, a result well-supported by empirical evidence (e.g., Jones and Lamont (2002), Ofek, Richardson, and Whitelaw (2004), Cohen, Diether, and Malloy (2007), Blocher, Reed, and Van Wesep (2013), Prado (2015), Drechsler and Drechsler (2016), Duong, Huszár, Tan, and Zhang (2017)). This result can also be viewed as confirming in a dynamic setting the classic Miller (1977) argument that with higher short-selling costs, the stock price is inflated since it reflects the views of the optimistic investors more relative to those of the pessimistic investors. In our model this behavior occurs because the stock is ultimately held by optimistic investors and among them there are lenders, who increase their stock demand and reflect their view more, due to the additional stock lending income as the shorting fee increases, which in turn leads to a higher current stock price and lower subsequent returns.

Electronic copy available at: https://ssrn.com/abstract=3516969

We further find that the costly-to-short stock volatility is greater in our economy than that in an economy with costless short-selling, even though the stock has the same time-varying disagreement in both economies. This result is supported by the empirical evidence in Drechsler and Drechsler (2016), who find that expensive-to-short stocks have higher stock return volatility as compared to almost-costless-to-short stocks. Saffi and Sigurdsson (2010) also find that stocks with high shorting fees are associated with high volatility. In our model, this result arises because costly-to-short stock price is additionally driven by the shorting fee and the fluctuations in the shorting fee due to disagreement shocks leads to more volatile stock price changes. This is again consistent with the classic Miller (1977) argument that in the absence of short-selling costs (and wealth transfer effects) the belief disagreement itself does not affect stock prices, leading to relatively lower stock price volatilities.

We also investigate how the key economic quantities are affected by the lenders' size in our economy, a quantity we later endogenize. We find that the costly-to-short stock equilibrium shorting fee and price decrease while its risk premium increases in the lenders' size. These relations arise because as more investors become lenders, the costly-to-short stock lending supply increases, leading to a lower shorting fee in the stock lending market, which in turn leads to a lower stock price and higher subsequent returns on average. The empirical support for these effects are provided by Prado, Saffi, and Sturgess (2016), who show that the stock ownership composition matters for the stock lending market and stock prices. In particular, they find that stocks with more concentrated institutional ownership, i.e., stocks that are held by fewer institutions but with larger stock holdings, corresponding to lower lenders' size in our model, have lower lending supply, higher shorting fees, and lower future returns on average. Likewise, Nagel (2005) finds that shorting fee is negatively associated while future stock returns are positively associated with institutional ownership, a plausible proxy for the lenders' size in our model.

We next turn to the stock *short interest*, a widely used measure to infer the amount of short-selling for a stock. We first determine the costly-to-short stock short interest and show that it increases in its shorting fee, consistent with the empirical evidence (D'Avolio (2002), Beneish, Lee, and Nichols (2015), Drechsler and Drechsler (2016)). This result arises because, a higher current shorting fee corresponds to a higher current disagreement and a higher stock price, and hence the short-sellers are relatively more pessimistic now and increase their shorting demand, leading to the positive relation between the short interest and the shorting fee of the stock.

Electronic copy available at: https://ssrn.com/abstract=3516969

We then obtain our key implication that the costly-to-short stock current short interest predicts its future returns negatively. This predictability arises in our model because, as discussed above, a higher short interest corresponds to a higher current shorting fee, which is now expected to be lower in the future due to mean-reversion, leading to lower future stock prices on average as compared to the relatively high current stock prices. This finding is strongly supported by the vast evidence both at the individual stock level and at the aggregate market level (Seneca (1967), Figlewski (1981), Senchack and Starks (1993), Desai, Ramesh, Thiagarajan, and Balachandran (2002), Asquith, Pathak, and Ritter (2005), Boehmer, Huszar, and Jordan (2010), Beneish, Lee, and Nichols (2015), Rapach, Ringgenberg, and Zhou (2016)). Among them Beneish, Lee, and Nichols (2015) show that this predictability effect is stronger for costly-to-short stocks, which lends further support to our underlying mechanism, since we also find that the predictive ability of short interest is solely due to the presence of shorting fees (i.e., a costless-to-short stock short interest can still be time-varying but does not predict future returns). Therefore, our model reveals that the current short interest is an "informative" signal for future returns for costly-to-short stocks but not for stocks that are costless-to-short.[2]

We next shed light on the recent empirical findings of Engelberg, Reed, and Ringgenberg (2018), who show that in the cross-section, stocks with higher *short-selling risk* (higher uncertainty about their future shorting fees) have lower returns and less *short-selling activity* (volume), suggesting that the short-selling risk is a significant source of limits to arbitrage. In our model, due to the time-variation in disagreement, short-sellers' demand and lenders' supply of the costly-to-short stock shares also fluctuate, which in turn lead to a time-variation, and hence uncertainty, in the shorting fee, resulting with short-selling risk. We find that short-selling risk matters in equilibrium, and show that a higher short-selling risk can be associated with lower stock returns and less short-selling activity as in Engelberg, Reed, and Ringgenberg due to the differences in stock partial lending.

The negative relation between the short-selling risk and risk premium arises because the shorting fee of a stock with a low degree of partial lending is more sensitive to the disagreement shocks. This is because such a stock has a lower lending supply, which can only absorb the short-selling demand by increasing the shorting fee, leading to a higher shorting fee variance,

[2] Moreover, we also find that, in addition to the short interest (a stock variable), a positive change in short interest (a flow variable) also predicts lower future returns, consistently with the findings in Boehmer, Jones, and Zhang (2008) and Diether, Lee, and Werner (2009), as well as the current short selling being positively related to past stock performance, as shown in Diether, Lee, and Werner.

Electronic copy available at: https://ssrn.com/abstract=3516969

as well as a higher stock price and a lower risk premium. Similarly, the negative relation between the short-selling risk and short-selling activity occurs because a lower partial lending also leads to a lower short-selling activity (volume) since short-sellers now short less for any given disagreement level. In sum, our theory suggests that for two similar costly-to-short stocks, the one lent to short-sellers in greater proportion by lenders will have less short-selling risk, higher future returns, and more short-selling activity.

Given that lenders earn additional income in our economy, a natural question arises whether all optimistic investors can be lenders. Towards that, we endogenously determine the optimal size of lenders by introducing a cost of setting up a lending facility. We show that even when the entry is costless not all optimistic investors become lenders and the lenders' optimal size decreases in the cost of entry, and in fact when the cost is too high no investor would become a lender. We also find that the optimal size is non-monotonic, first zero, then increasing, and then decreasing, in partial lending. That is, very low levels of stock partial lending corresponds to no lenders, and as the partial lending, and hence the expected future income from lending increases, so does the optimal size of lenders. However, for sufficiently high levels of partial lending, the optimal lenders' size starts decreasing in partial lending, since an increase in partial lending now leads to too low of a shorting fee and a lower future lending income. Finally, we show that the costly-to-short stock shorting fee and the price are both increasing in the cost of entry since a higher cost of entry leads to a lower lenders' optimal size.

Taken together, our paper makes several contributions. First, our stock price and risk premium results demonstrate that the classic Miller (1977) intuition also arises in our fairly standard dynamic setup. Second, to our best knowledge, ours is the first theory work to reconcile the extensive empirical evidence on the predictive power of short interest, as well as our results on the short-selling risk, stock volatility and the effects of lenders' size. Third, we endogenize the lenders' optimal size and identify its economic determinants. Finally, our methodological contribution and the tractability of our model is due in large part to our ability to identify the short-sellers' and lenders' subjective risk premia in the evolution of their financial wealth (Lemma 1 of Section 2.4). This in turn allows us to employ familiar stochastic dynamic programming techniques in the determination of equilibrium, in which investors face different market imperfections.

Electronic copy available at: https://ssrn.com/abstract=3516969

Our paper is related to the large theoretical literature studying the effects of short-sale constraints and restrictions in the stock market. The works in this literature typically examine the effects of investors not being able hold short (i.e., negative) stock positions, which in static settings include Miller (1977), Diamond and Verrecchia (1987), Hong and Stein (2003), Nezafat, Schroder, and Wang (2017), and in dynamic settings include Harrison and Kreps (1978), Detemple and Murthy (1997), Scheinkman and Xiong (2003), Gallmeyer and Hollifield (2008), Chabakauri (2015). More closely related papers in this literature are those in which investors can hold short positions by paying a stock shorting fee, as in our model. However, these works are typically fairly stylized and primarily focus on the effects of the shorting fee on the stock price and risk premium only.[3] In this strand, under static settings Blocher, Reed, and Van Wesep (2013), Banerjee and Graveline (2014), and under dynamic settings in which the stock is a claim to a single payoff Duffie, Gârleanu, and Pedersen (2002) (search and bargaining-based model with risk-neutral investors) and Daniel, Klos, and Rottke (2018) (a series of one-period optimization and exogenous short interest) find that a higher stock shorting fee leads to a higher stock price and lower future expected returns, as in our paper. Our paper, however, differs from these works in terms of its methodology and results. Differently from the dynamic works above, our setting is fairly standard with risk averse investors, the stock being a claim to a dividend flow rather than a single payoff, featuring no search and bargaining considerations, and is cast in a fully dynamic economy in which the short interest is determined endogenously. Moreover, due to the stock price and the shorting fee being deterministic in Duffie, Gârleanu, and Pedersen (2002), and the short interest being exogenous in Daniel, Klos, and Rottke (2018), in these works, it is not possible to address issues related to the stock volatility, predictive power of the short interest, and short-selling risk, as we do. Similarly, the other more recent dynamic models of Evgeniou, Hugonnier, and Prieto (2019) in which the shorting fee is assumed to be proportional to stock volatility, and Nutz and Scheinkman (2020) in which the shorting fee is specified exogenously, do not have our key results on the predictive power of the short interest, short-selling risk and activity, as well as our analysis on the lenders' optimal size.

Finally, this paper is also related to the vast literature on heterogeneous beliefs in financial markets (Detemple and Murthy (1994), Zapatero (1998), Basak (2000, 2005), Johnson (2004), David (2008), Yan (2008), Dumas, Kurshev, and Uppal (2009), Cvitanić and Malamud (2011),

[3]Relatedly, in a partial equilibrium setting, Atmaz and Basak (2019) provide an analysis of how option prices are affected by the shorting fee of the underlying stock.

Electronic copy available at: https://ssrn.com/abstract=3516969

Banerjee (2011), Bhamra and Uppal (2014), Ehling, Graniero, and Heyerdahl-Larsen (2017), Atmaz and Basak (2018), Andrei, Carlin, and Hasler (2019)). However, none of these works consider the costly stock short-selling and lending market as we do, hence do not have our main mechanisms and results.

The remainder of the paper is organized as follows. Section 2 presents our model with costly short-selling. Section 3 provides our results on the shorting fee, stock price, risk premium and volatility, Section 4 on the short interest and short-selling risk, and Section 5 on the optimal size of lenders. Section 6 concludes. Appendix A contains all the proofs, and Appendix B discusses the parameter values employed in our figures.

2 Economy with Costly Short-Selling

In this Section, we develop a tractable model of costly stock short-selling and lending market within a familiar dynamic asset pricing framework. Investors' belief disagreement leads to the presence of stock lenders and short-sellers in the economy. Our model builds on the fact that even though most stocks can be shorted cheaply in today's markets, for many other stocks short-selling costs can be fairly significant (D'Avolio (2002), Drechsler and Drechsler (2016)). For instance, Drechsler and Drechsler (2016) find the average shorting fee for a typical stock in the cheapest-to-short and expensive-to-short deciles to be 2 basis points (bps) per annum and 696 bps, respectively. They also find that the average shorting fee for half of the stocks in their sample is no more than 13 bps. To capture these stark differences in short-selling costs in the cross-section, we consider a simple two-stock economy in which the first stock represents a typical costly-to-short stock, while the second stock represents a typical cheap-to-short stock, whose shorting fee, for clarity, is constructed to be zero in equilibrium (Section 3).

2.1 Securities Market

We consider an economy with an infinite horizon evolving in continuous time. There are three securities available for trading, two risky stocks and a riskless asset. The risky stock n, $n = 1, 2$, is in fixed supply of Q_n units and is a claim to the dividend flow D_n with dynamics

$$dD_{nt} = \mu_{D_n} dt + \sigma_{D_n} d\omega_{nt}, \tag{1}$$

Electronic copy available at: https://ssrn.com/abstract=3516969

where the constants μ_{D_n} and σ_{D_n} are the mean and volatility of the dividend (changes), respectively, and ω_n is a standard Brownian motion under the objective probability measure, $\mathbb{P}$, with $d\omega_{1t}d\omega_{2t} = 0$. The individual stock prices S_n, $n = 1, 2$, are to be determined endogenously in equilibrium. The riskless asset is in perfectly elastic supply and pays a constant interest rate r.[4] In what follows, we refer to stock 1 as the costly-to-short stock, and stock 2 as the zero-fee stock, representing a costless-to-short stock that approximates a cheap-to-short stock in reality.

2.2 Investors' Beliefs

The economy is populated by optimistic and pessimistic investors. All investors commonly observe the dividend processes D_n, $n = 1, 2$, and agree on their volatilities but have different beliefs about their means. The optimistic investors, with a population mass of 0.5, perceive the mean of the dividend D_n to be $\mu_{D_n} + \theta_{nt}$, while the pessimistic investors, with a population mass of 0.5, perceive it to be $\mu_{D_n} - \theta_{nt}$. The quantity θ_{nt} captures the *disagreement* among investors on stock n since it is the (population) weighted-difference between the optimistic and pessimistic investors' expectations on stock n dividend, $0.5\left(\mu_{D_n} + \theta_{nt}\right) - 0.5\left(\mu_{D_n} - \theta_{nt}\right) = \theta_{nt}$. Moreover, due to the identical masses and the symmetry in beliefs, the weighted-average expectation is unbiased, $0.5\left(\mu_{D_n} + \theta_{nt}\right) + 0.5\left(\mu_{D_n} - \theta_{nt}\right) = \mu_{D_n}$.[5]

The disagreement on stock n is assumed to follow

$$d\theta_{nt} = \kappa_n\left(\bar{\theta}_n - \theta_{nt}\right)dt + \upsilon_n d\omega_{\theta_n t}, \tag{2}$$

[4] The exogenously specified interest rate is suitable for our purposes for several reasons. First, most studies on short-selling costs are cross-sectional in nature (e.g., studying return differences for cheap-to-short vs. expensive-to-short stocks) for which the common interest rate does not play a role. Second, economically it is hard to imagine short selling costs of individual stocks to have any material effect on aggregate quantities, particularly on the market interest rate, which is determined in a relatively very large market where foreign institutions also play a significant role. In fact, to our best knowledge there is no empirical evidence on interest rates being affected by the short-selling fees of individual stocks. So, by taking the interest rate as exogenous, we avoid any potential spurious effects that may arise due to modeling choices.

[5] Neither the equal mass nor the belief symmetry is necessary for our main mechanism and results to hold, and they can be generalized in a straightforward manner. We consider this specification because it provides a clear benchmark economy in which without short-selling costs, the belief disagreement alone does not affect the ensuing equilibrium stock prices and yields the single-investor rational beliefs economy prices (as highlighted in Section 3). We note that in this economy, our disagreement measure θ_{nt} is also the standard deviation of investors' expectations on the stock dividend, a commonly employed measure of belief disagreement.

Electronic copy available at: https://ssrn.com/abstract=3516969

where the constants κ_n, $\bar{\theta}_n$, and υ_n are the speed of mean reversion, long-run mean, and volatility of the disagreement θ_n, respectively, and ω_{θ_n} is a standard Brownian motion, independent of all other Brownian motions, under the objective probability measure, $\mathbb{P}$. Without loss of generality, we set the initial value of disagreement to its long-run mean, $\theta_{n0} = \bar{\theta}_n$, for convenience. In our analysis, while the disagreement on (costly-to-short) stock 1 is a time-varying Ornstein-Uhlenbeck process as (2) illustrates, by setting $\kappa_2 = \upsilon_2 = 0$, we obtain a constant disagreement for (cheap-to-short) stock 2 (i.e., $\theta_{2t} = \bar{\theta}_2$), which enables us construct its shorting fee to be zero in equilibrium (Proposition 1). With this specification, the stock price S_n, $n = 1, 2$, has the posited dynamics

$$dS_{nt} + D_{nt}dt = \mu_{nt}dt + \sigma_{S_nt}d\omega_{nt} + \sigma_{\theta_nt}d\omega_{\theta_nt}, \tag{3}$$

where μ_n is the mean of the stock price (changes), and the diffusion terms σ_{S_n} and σ_{θ_n} determine the volatility of the stock price (changes), denoted by σ_n, through $\sigma_{nt}^2 = \sigma_{S_nt}^2 + \sigma_{\theta_nt}^2$.

Remark 1 (Further discussion of disagreement processes). In Appendix A, we discuss the microfoundations of the type of mean-reverting, Gaussian disagreement process in (2) by considering a Bayesian learning environment, in which investors are symmetrically informed but have different interpretations of signals as often considered in the dynamic differences in beliefs models with stationary disagreement (e.g., Scheinkman and Xiong (2003), Dumas, Kurshev, and Uppal (2009), Xiong and Yan (2010)). The special case of constant beliefs, as formulated on stock 2, is typically referred to as "dogmatic beliefs" in the differences of opinion literature and is also adopted by numerous works (Kogan, Ross, Wang, and Westerfield (2006), Cvitanić and Malamud (2011), among others).

Moreover, it is well-known that in a framework like ours, i.e., CARA preferences with Gaussian payoffs, the stock prices can admit negative values with positive probabilities. Due to being a Gaussian process, now additionally the disagreement θ_{1t} in (2), as well as the ensuing equilibrium shorting fee that turns out to be driven by the disagreement (Proposition 1), can take on negative values with positive probabilities. However, our model allows us to control these probabilities and obtain almost always positive shorting fee and disagreement for suitable choices of parameter values. For example, for the parameters values in Table 1 of Appendix B, the probabilities of the shorting fee and the disagreement being positive at the steady-state are $\mathbb{P}(\phi_{1t} > 0) = 0.999$ and $\mathbb{P}(\theta_{1t} > 0) = 1 - \Phi(-20.35) \simeq 1$, where Φ is the cumulative distribution function of standard normal distribution.

Electronic copy available at: https://ssrn.com/abstract=3516969

2.3 Stock Short-Selling and Lending Market

In reality, short-sellers sell stocks that they do not own with the anticipation of making a profit from future stock price declines. To do so, they (effectively) pay a shorting fee to borrow shares from lenders who are long in the stock, and sell those borrowed shares to other willing buyers, who are not necessarily lenders of the stock.[6] Accordingly, we incorporate costly short-selling and stock lending market into our economy by capturing the standard market practices briefly discussed above. Towards that, and also in anticipation of the equilibrium outcomes of Sections 3–5, we further categorize the optimists and pessimists populations into two groups with respect to their ability to participate in the stock short-selling and lending market.

The *lenders*, ℓ, with a population mass λ, are the optimistic investors who can only successfully lend a fraction $0 < \alpha_n \leq 1$ of their long position in stock n to short-sellers, where henceforth we refer to α_n as *partial lending* in stock n, $n = 1, 2$. The remaining optimistic investors are *holders*, h, with a population mass $0.5 - \lambda$, and they do not participate in the stock short-selling and lending market. As D'Avolio (2002) discusses, we need such investors to help clear the stock market as "the outstanding securities must come to rest with non-lending investors willing to hold these securities despite forgoing the loan fees capitalized into the equilibrium price".[7] While the quantity λ is taken as given in Sections 2–4, it is endogenously determined in the equilibrium of Section 5.

[6]The exact mechanics of stock short-selling are somewhat more involved but its essentials are captured by our description above (see D'Avolio (2002) and Reed (2013) for an extensive discussion of short-selling). Briefly, to protect lenders, short-sellers leave a collateral typically equals to 102% of the market value of the borrowed amount in an account, which earns the interest rate. This interest income is shared between the short-seller and the lender. The short-seller's account earns the rebate amount, while the lender's account earns the shorting fee, which is effectively what short-sellers pay to lenders for each share they short. Moreover, Kolasinski, Reed, and Ringgenberg (2013) and Chague et al. (2017) provide evidence of dispersion in shorting fees. In our model there is one type of lender leading to a unique shorting fee, which can be viewed as the average shorting fee.

[7]The partial lending feature of ℓ-type investors can also be justified on the grounds that in reality investors may not be able to lend and earn additional income from all their long stock holdings. This is evident from the presence of the excess supply of lendable shares for most stocks. For example, Saffi and Sigurdsson (2010) find that out of \$15 trillion of stocks that are available globally for short-sellers to borrow, only \$3 trillion was actually lent out in December 2008. Moreover, the partial lending feature in our model also enables us to make a distinction between short interest (fraction of outstanding shares held by short-sellers) and lending supply (fraction of outstanding shares available for lending), as in the data. On the other hand, the presence of h-type investors in our model guarantees that in equilibrium, the short interest is determined endogenously rather than exogenously implied by the market clearing conditions.

Electronic copy available at: https://ssrn.com/abstract=3516969

The *short-sellers*, s, with a population mass λ^s, are the pessimistic investors who can borrow from lenders by paying for each stock n share the *shorting fee* ϕ_{nt}, which is to be determined endogenously in equilibrium. Finally, the remaining pessimistic investors are *non-participants* with a population mass $0.5 - \lambda^s$, and they do not participate in the stock short-selling and lending market. By being pessimistic and not participating in the short-selling market, in equilibrium, non-participants do not hold any stock positions, and therefore their presence is not crucial in the determination of the endogenous stock price or the shorting fee, and hence for our main mechanism and results. Therefore, without loss of generality, we set their population mass to 0 (i.e., $\lambda^s = 0.5$), and consider an economy with three types of investors ℓ, h, and s.[8]

2.4 Investors' Preferences and Optimization

Each type of investor $i = \ell, h, s$, is endowed at time zero with the same initial wealth W_0. The investor then chooses a consumption process c^i, and an admissible portfolio strategy ψ_n^i, the number of shares in stock n, $n = 1, 2$, to maximize her subjective expected constant absolute risk aversion (CARA) preferences over her life-time consumption

$$\mathrm{E}^i \left[\int_0^\infty e^{-\rho t} \frac{e^{-\gamma c_t^i}}{-\gamma} dt \right],$$

where $\gamma > 0$ is the absolute risk aversion and $\rho > 0$ is the time discount factor. Here, E^i denotes the expectation under each i-type investor's subjective beliefs $\mathbb{P}^i$, on which is defined her perceived Brownian motions ω_n^i, given by $d\omega_{nt}^i = d\omega_{nt} - (\theta_{nt}/\sigma_{D_n})dt$, for $i = \ell, h$, and $d\omega_{nt}^i = d\omega_{nt} + (\theta_{nt}/\sigma_{D_n})dt$, for $i = s$, due to their beliefs as described above. The financial wealth of each i-type investor, W^i, evolves over time in a distinct manner, since in addition to the usual capital gains/losses and dividends, holding stocks long yields an additional income for lenders, whereas having short positions in stocks implies an additional cost for short-sellers. Lemma 1 presents the dynamics of each i-type investor's financial wealth, and identifies each investor's (shorting fee incorporated) subjective risk premium.

[8] We note that our mechanisms and results do not crucially depend on the specific choice of the short-sellers' mass $\lambda^s = 0.5$ and also go through for any lower value of λ^s. Alternatively, one could think of investors in our economy as those already participating in the stock market with half of them being pessimistic short-sellers and the other half being optimistic, either lenders or holders.

Electronic copy available at: https://ssrn.com/abstract=3516969

Lemma 1 (**Financial wealth dynamics and investor type-specific risk premia**). *In the economy with costly stock short-selling and lending market, the financial wealth of each* i*-type investor,* $i = \ell, h, s$*, evolves according to*

$$dW_t^i = W_t^i r dt + \sum_{n=1}^{2} \psi_{nt}^i \pi_{nt}^i dt + \sum_{n=1}^{2} \psi_{nt}^i \left(\sigma_{S_{nt}} d\omega_{nt}^i + \sigma_{\theta_{nt}} d\omega_{\theta_{nt}} \right) - c_t^i dt, \tag{4}$$

where π_{nt}^i *is each* i*-type investor's subjective risk premium on stock* $n = 1, 2$*, and is given by*

$$\pi_{nt}^i = \begin{cases} \pi_{nt} + \frac{\sigma_{S_{nt}}}{\sigma_{D_n}} \theta_{nt} + \alpha_n \phi_{nt} & \text{for } i = \ell, \\ \pi_{nt} + \frac{\sigma_{S_{nt}}}{\sigma_{D_n}} \theta_{nt} & \text{for } i = h, \\ \pi_{nt} - \frac{\sigma_{S_{nt}}}{\sigma_{D_n}} \theta_{nt} + \phi_{nt} & \text{for } i = s, \end{cases} \tag{5}$$

where $\pi_{nt} = \mu_{nt} - rS_{nt}$ *is the objective risk premium on stock* n*.*

The first term in the investors' subjective risk premium (5) is the standard objective risk premium. The second term is due to their subjective beliefs and adding it gives the subjective stock risk premium if there is no additional shorting fee cost or lending income for the investor. For the optimists this term is positive, while for the pessimists is negative. The third term, if exists, adjusts for the additional shorting fee cost or lending income from being a short-seller or a lender in equilibrium, respectively. If an investor turns out to be a lender in equilibrium, she must be an optimist and earning a shorting fee ϕ_{nt} for each stock n share she successfully lends out, which is a fraction α_n of her long stock position, thereby increasing her subjective stock risk premium by an amount $\alpha_n \phi_{nt}$. Similarly, if an investor turns out to be a short-seller in equilibrium, she must be a pessimist and paying a shorting fee ϕ_{nt} for each stock n share she borrows, thereby increasing her (negative) subjective stock risk premium by an amount ϕ_{nt}.[9] The subjective and objective risk premia will be determined endogenously in equilibrium.

[9] Even though the short-sellers' subjective risk premium is intuitive, the exact mechanism leading to it is a bit more involved. In particular, if an investor turns out to be a short-seller in equilibrium, $\psi_{nt}^s < 0$, her time-t wealth is given by $W_t^s = F_t + \sum_{n=1}^2 \psi_{nt}^s S_{nt} + \sum_{n=1}^2 M_{nt}$, where F_t is the amount invested in the riskless asset, and M_{nt} is the total amount of short-selling proceeds from stock n that are collateralized, and hence cannot be invested in other securities, and is given by $M_{nt} = -\psi_{nt}^s S_{nt}$. Over time, the collateral earns the interest rate rM_{nt}, which is shared between the short-seller and the lender. The lender's account earns the shorting fee, $-\phi_{nt}\psi_{nt}^s > 0$, while the short-seller's account earns the remainder rebate amount, $rM_{nt} + \phi_{nt}\psi_{nt}^s$. These imply the dynamics $dM_{nt} = (rM_{nt} + \phi_{nt}\psi_{nt}^s)dt = -\psi_{nt}^s \left(rS_{nt} - \phi_{nt}\right) dt$ in the evolution of her financial wealth $dW_t^s = dF_t + \sum_{n=1}^2 \psi_{nt}^s \left(dS_{nt} + D_{nt}dt\right) + \sum_{n=1}^2 dM_{nt} - c_t^s dt$ with $dF_t = W_t^s r dt$, leading to (4)–(5).

Electronic copy available at: https://ssrn.com/abstract=3516969

Our explicit identification of the investor type-specific risk premia above facilitates much tractability in our subsequent general equilibrium analysis in which investors face different market imperfections. This in turn allows us to employ familiar stochastic dynamic programming techniques in the determination of equilibrium.

3 Shorting Fee, Stock Price and Its Dynamics

In this Section, we solve for the equilibrium in which investors face different market imperfections, and recover the effects of costly short-selling on stocks. In particular, we find that for a costly-to-short stock, the equilibrium shorting fee is time-varying and increases in disagreement, the stock price is positively while the risk premium is negatively related to the shorting fee, and the stock volatility is increased, all consistently with empirical evidence.

The costly stock short-selling and lending market economy is said to be in *equilibrium* if the stock prices S_n, the shorting fees ϕ_n, $n = 1, 2$, and the investors' consumption and portfolio strategies $(c^i, \psi_1^i, \psi_2^i)$, $i = \ell, h, s$, are such that (i) all investors choose their optimal consumption and portfolio strategies given the stock prices, the shorting fees, and their beliefs, (ii) the stock markets and the stock short-selling and lending markets clear, i.e., $\lambda\psi_{nt}^\ell + (\frac{1}{2} - \lambda)\psi_{nt}^h + \frac{1}{2}\psi_{nt}^s = Q_n$, and $\lambda\alpha_n\psi_{nt}^\ell + \frac{1}{2}\psi_{nt}^s = 0$, for $n = 1, 2$, at all times t, respectively.

We will often make comparisons with the costless short-selling benchmark economy (i.e., $\phi_{nt} = 0$, $n = 1, 2$) denoted with an upper bar ($\bar{\ }$). Since there is no shorting fee for stocks in the benchmark economy, only the stock markets need to clear in equilibrium. With costly short-selling, the stock short-selling and lending markets need to additionally clear. In particular, the equilibrium shorting fee for a stock must be such that the total amount of its shares shorted equals to the total amount lent. Thus, in our model, the shorting fee is determined by the classic supply-and-demand channel, for which there is much evidence as we discuss later while presenting our results.[10]

To ensure that stock 2, which represents a typical cheap-to-short stock in our model, has a zero shorting fee in equilibrium, we assume the following parameter restriction throughout

[10]Existing research shows that other channels and considerations such as search costs and bargaining may also play a role in this market (Duffie, Gârleanu, and Pedersen (2002)). So as to not confound the analysis, we abstract away from other potential channels, which may also include asymmetric information, agency issues, and institutional incentives. We leave these considerations for future research since to our best knowledge ours is the first comprehensive work with the supply-and-demand channel within a standard dynamic framework.

Electronic copy available at: https://ssrn.com/abstract=3516969

our analysis

$$\gamma\sigma_{D_2}^2 Q_2 < \bar{\theta}_2 \leq \frac{1/2+\lambda\alpha_2}{1/2-\lambda\alpha_2}\gamma\sigma_{D_2}^2 Q_2. \tag{6}$$

This restriction essentially states that we need to have some level of disagreement among investors for a non-trivial short-selling activity to occur in stock 2 and this disagreement should not be too large relative to the stock lending supply so as to prevent a positive shorting fee. We further note that the upper bound is increasing in α_2 so this inequality is likely to hold when the partial lending for stock 2 is higher, such as for large and liquid stocks. The restriction (6) is not only economically plausible, but is also supported by the empirical findings. For example, D'Avolio (2002) find that a typical "general collateral" stock, a stock with a shorting fee less than 1%, has a low belief dispersion and much excess lending supply, with S&P500 constituents almost always belonging in this group.[11] Our focus, however, is on the costly-to-short stock 1, and so we model its disagreement as being time-varying as in reality. Incorporating stock 2 into our model is for generality and also realism given the cross-sectional nature of the empirical evidence we attempt to explain. Alternatively, one could study the costly-to-short stock in isolation without a second stock, and as our analysis shows, all our results would continue to hold.

In this costly stock short-selling economy, there are three sources of uncertainty and two risky securities available for trading for all investors, and hence markets are dynamically incomplete. We employ the standard stochastic dynamic programming method (Merton (1971)) to solve for each investor's optimal consumption and portfolio strategies and apply market clearing conditions to obtain the equilibrium. Proposition 1 reports the equilibrium in this economy, by presenting the equilibrium shorting fees ϕ_n, the stock prices S_n, the (objective) stock risk premia π_n, the stock volatilities σ_n, and the investors' indirect utility functions $J^i\left(W_t^i, \theta_{1t}, t\right) = \max_{(c^i,\psi_1^i,\psi_2^i)} \mathrm{E}_t^i\left[\int_t^\infty e^{-\rho u}\frac{e^{-\gamma c_u^i}}{-\gamma}du\right]$, $i=\ell,h,s$, along with their key properties.

Proposition 1 (Equilibrium shorting fee, stock price and its dynamics). *In the costly short-selling economy, for the costly-to-short stock 1, the equilibrium shorting fee and stock*

[11] That being said, we wish to highlight that the parameter restriction (6) is only for clarity and in fact our main results for costly-to-short stocks remain equally valid if we allowed disagreement levels higher than the upper bound in (6), resulting with a strictly positive shorting fee, which then can be calibrated to the observed low shorting fee of a typical cheap-to-short stock in the data.

Electronic copy available at: https://ssrn.com/abstract=3516969

price are given by

$$\phi_{1t} = a_1 + b_1\theta_{1t}, \tag{7}$$

$$S_{1t} = \bar{S}_{1t} + A_1 + B_1\phi_{1t}, \tag{8}$$

and the stock risk premium and volatility by

$$\pi_{1t} = \bar{\pi}_1 - \Big(rA_1 - (a_1 + b_1\bar{\theta}_1)\kappa_1 B_1\Big) - (\kappa_1 + r)\, B_1\phi_{1t}, \tag{9}$$

$$\sigma_1 = \sqrt{\bar{\sigma}_1^2 + \upsilon_1^2 b_1^2 B_1^2}, \tag{10}$$

where the constants a_1*,* A_1*,* $b_1 > 0$*, and* $B_1 > 0$ *solve the non-linear equations that are provided in Appendix A. For the zero-fee stock 2, the equilibrium shorting fee and stock price are given by* $\phi_{2t} = 0$ *and* $S_{2t} = \bar{S}_{2t}$*, and the stock risk premium and volatility by* $\pi_2 = \bar{\pi}_2$ *and* $\sigma_2 = \bar{\sigma}_2$*, respectively. The i-type investor's,* $i = \ell, h, s$*, indirect utility function is given by*

$$J^i\left(W_t^i, \theta_{1t}, t\right) = -e^{-\rho t} e^{-\gamma r W_t^i} e^{F^i + G^i\theta_{1t} + H^i\theta_{1t}^2}, \tag{11}$$

where the constants F^i*,* G^i*,* H^i*, and each i-type investor's optimal consumption and portfolio strategies* $(c^i, \psi_1^i, \psi_2^i)$ *are provided in Appendix A.*

In the benchmark economy with costless short-selling, for stock n*,* $n = 1, 2$*, the equilibrium stock price is given by* $\bar{S}_{nt} = (D_{nt}/r) + (\mu_{D_n}/r^2) - (\gamma\sigma_{D_n}^2 Q_n/r^2)$*, and the stock risk premium and volatility by* $\bar{\pi}_n = \gamma\sigma_{D_n}^2 Q_n/r$ *and* $\bar{\sigma}_n = \sigma_{D_n}/r$*, respectively. The i-type investor's,* $i = \ell, h, s$*, indirect utility function is given by* $\bar{J}^i\,(W_t^i, \theta_{1t}, t) = -e^{-\rho t} e^{-\gamma r W_t^i} e^{\bar{F}^i + \bar{G}^i\theta_{1t} + \bar{H}^i\theta_{1t}^2}$*, where the constants* $\bar{F}^i$*,* $\bar{G}^i$*,* $\bar{H}^i$ *and each i-type investor's optimal consumption and portfolio strategies* $(\bar{c}^i, \bar{\psi}_1^i, \bar{\psi}_2^i)$ *are provided in the Appendix A.*

Consequently, in the costly short-selling economy, the costly-to-short stock

i) shorting fee is increasing in disagreement θ_{1t}*,*

ii) price is increasing, while its risk premium is decreasing in shorting fee ϕ_{1t}*,*

iii) volatility is higher than that in the benchmark economy with costless short-selling.

Proposition 1 reveals that costly-to-short stock equilibrium shorting fee (7) is risky since it is driven by disagreement and the fluctuations in the disagreement lead to time-variation in the shorting fee so that a higher belief disagreement leads to a higher shorting fee (property

Electronic copy available at: https://ssrn.com/abstract=3516969

(i)).[12] This result arises because not all optimistic investors are lenders and those who are lenders can only lend a part of their stock holdings, and so a higher disagreement leads to a lesser increase in lending supply than the increase in shorting demand. Hence, for the stock lending market to clear the equilibrium shorting fee must increase. This result is consistent with the empirical evidence in D'Avolio (2002), who shows that the shorting fee of a stock is high when the investor disagreement is high.

Turning to the equilibrium stock 1 price and its dynamics, we first see that its price in the benchmark economy with costless short-selling is as in the standard economy with a single investor with unbiased beliefs. With costly short-selling, the equilibrium stock 1 price (8) and its risk premium (9) have simple structures and are additionally affected by the shorting fee ϕ_1. A notable implication here is that the costly-to-short stock price increases while its risk premium decreases in the shorting fee (property (ii)), as also illustrated in Figure 1, consistently with vast empirical evidence (e.g. Jones and Lamont (2002), Ofek, Richardson, and Whitelaw (2004), Cohen, Diether, and Malloy (2007), Blocher, Reed, and Van Wesep (2013), Prado (2015), Drechsler and Drechsler (2016), Duong, Huszár, Tan, and Zhang (2017)).[13] This result can be viewed as the dynamic version of the classic Miller (1977) argument that with higher short-selling costs, the stock price is inflated since it reflects the views of the optimistic investors more relative to those of the pessimistic investors, who, due to the increased cost of short-selling, reflect their views on the stock less. In our model this occurs because the stock is ultimately held by optimistic investors and among them there are lenders, who increase their stock demand due to the additional stock lending income as the shorting fee increases, which in turn leads to a higher current stock price and lower subsequent returns. As Figure 1 illustrates, for a very high shorting fee, the stock risk premium can be negative as typically observed in the data (e.g., Drechsler and Drechsler (2016)).

We also find that the costly-to-short stock volatility is higher than that in the benchmark economy with costless short-selling (property (iii)) even though the stock has the same time-varying disagreement θ_{1t} in both economies. This is again consistent with the classic Miller (1977) argument that in the absence of short-selling costs (and wealth transfer effects) the

[12] The costly-to-short stock equilibrium shorting fee (and other economic quantities) are also affected by the parameters of the disagreement process, lenders' size λ, partial lending α_1, as well as the risk adjustment term $\gamma\sigma_1^2 Q_1$, through the endogenous constants a_1, b_1, A_1, and B_1.

[13] We note that even though we focus on the effects of the shorting fee on the risk premium in Proposition 1, we find that the stock mean return μ_{1t} and the Sharpe ratio π_{1t}/σ_1 are also decreasing in the shorting fee.

Electronic copy available at: https://ssrn.com/abstract=3516969

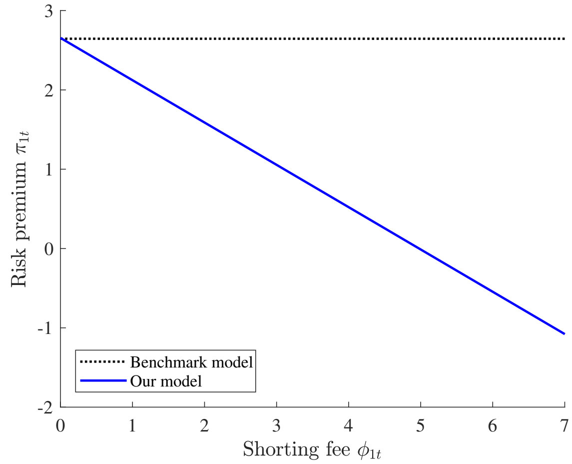


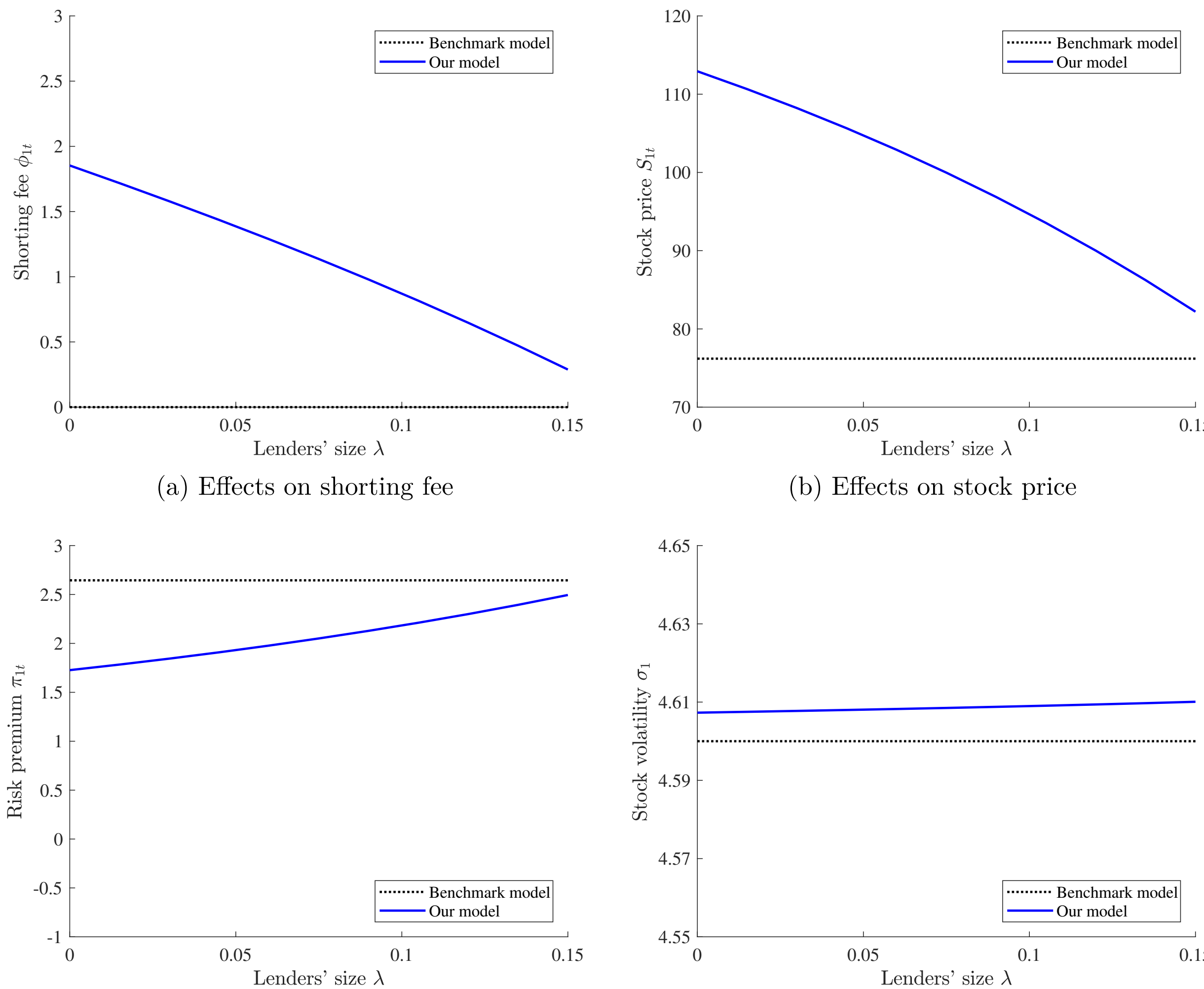
Figure 1: **Risk premium behavior.** This figure plots the costly-to-short stock equilibrium risk premium π_{1t} against the shorting fee ϕ_{1t}. The solid blue line corresponds to our economy with costly short-selling and the dotted black line corresponds to the benchmark economy with costless short-selling. The parameter values follow from Table 1 of Appendix B.

belief disagreement itself does not affect stock prices, leading to relatively lower stock price (change) volatilities. In contrast, with costly short-selling, stock prices are additionally driven by the shorting fee (Proposition 1), and the fluctuations in the shorting fee due to disagreement shocks leads to more volatile stock price changes (Figure 2 Panel (d)). This result is consistent with the empirical evidence in Drechsler and Drechsler (2016), who find that expensive-to-short stocks have higher stock return volatility compared to almost-costless-to-short stocks. Similarly, Saffi and Sigurdsson (2010) find a positive relation between high shorting fees and high volatility.

Unsurprisingly, we also see that the zero-fee stock 2 price is as in the benchmark economy with costless short-selling. As discussed earlier, this result arises because the constant disagreement on this stock is assumed to be not large relative to its stock lending supply, paralleling the reality that most large, liquid stocks have much excess lending supply, leading to almost zero shorting fee levels. On the other hand, investors' value function (11) has a familiar structure that typically arises in dynamic linear equilibrium asset pricing models with Gaussian processes.

We also look at the effect of lenders' size λ, a key quantity we endogenize in Section 5, on the equilibrium quantities in Figure 2. Panels (a)–(c) illustrate that the costly-to-short stock

Electronic copy available at: https://ssrn.com/abstract=3516969



(a) Effects on shorting fee

(b) Effects on stock price

(c) Effects on risk premium

(d) Effects on stock volatility

Figure 2: **Effects of lenders' size.** These panels plot the costly-to-short stock equilibrium shorting fee ϕ_{1t} (panel (a)), stock price S_{1t} (panel (b)), stock risk premium π_{1t} (panel (c)), and stock volatility σ_1 (panel (d)), against lenders' size λ. The solid blue lines correspond to our economy with costly short-selling, while the dotted black lines correspond to the benchmark economy with costless short-selling. The parameter values follow from Table 1 of Appendix B.

equilibrium shorting fee and price decrease while its risk premium increases in the lenders' size, respectively. These relations arise because as more investors become lenders, the costly-to-short stock lending supply increases, leading to a lower shorting fee in the stock lending market, which in turn leads to a lower stock price and higher subsequent returns on average. The empirical support for these effects are provided by Prado, Saffi, and Sturgess (2016),

Electronic copy available at: https://ssrn.com/abstract=3516969

who show that stock ownership composition matters for the stock lending market and stock prices. In particular, they find that stocks with more concentrated institutional ownership, i.e., stocks that are held by fewer institutions but with larger stock holdings, corresponding to lower lenders' size λ in our model, have lower lending supply, higher shorting fees, and lower future returns on average. Likewise, Nagel (2005) finds that shorting fee is negatively associated while future stock returns are positively associated with institutional ownership, a plausible proxy for the lenders' size in our model.[14]

As discussed in the Introduction, consistent with the classic Miller (1977) argument, a number of existing theoretical works find results similar to ours that a higher shorting fee leads to a higher stock price and a lower risk premium, as in our Proposition 1. Moreover, Duffie, Gârleanu, and Pedersen (2002), Blocher, Reed, and Van Wesep (2013), and Daniel, Klos, and Rottke (2018) also find that the shorting fee increases in disagreement. Our findings complement these works by demonstrating that the Miller intuition also arises in our fairly standard dynamic setup. However, to the best of our knowledge, our findings on volatility (Proposition 1 property (iii)), the effects of lenders' size as illustrated in Figure 2, as well as all our findings in the next Section 4 on the predictive power of short interest (Proposition 2) and short-selling risk (Proposition 3) are novel.

4 Short Interest and Short-Selling Risk

In this Section, we examine the behavior of the short interest and short-selling risk (uncertainty about future shorting fee) of costly-to-short stock. We find that the costly-to-short stock equilibrium short interest is increasing in the shorting fee and predicts future stock returns negatively, supporting the vast empirical evidence. We further find that short-selling risk matters in equilibrium, and show that a higher short-selling risk can be associated with lower stock returns and less short-selling activity, also consistently with the recent empirical evidence.

[14]As Figure 2 Panel (d) illustrates, the costly-to-short stock volatility does not vary much with lenders' size λ, thereby implying that two stocks with the same level of high shorting fee but differing in lenders' size, and hence differing in lending supply, may have similar volatility. In our model this occurs because the fluctuations in the disagreement is now mostly absorbed by the fluctuations in shorting fee, whose marginal effects on the stock price is smaller compared to fundamental dividend shocks. This finding may help us understand the somewhat surprising evidence in Saffi and Sigurdsson (2010) and Kaplan, Moskowitz, and Sensoy (2013), who do not find a significant relation between the lending supply and stock volatility.

Electronic copy available at: https://ssrn.com/abstract=3516969

4.1 Short Interest

A widely used and closely watched measure to infer the amount of short-selling for a stock is its *short interest*, which is the fraction of its outstanding shares held by short-sellers. There is vast empirical evidence documenting that a higher current short interest predicts lower future stock returns, both for individual stocks and the stock market itself. We now examine our model implication for the short interest for stock $n = 1, 2$, defined as $\mathcal{SI}_{nt} = -\frac{1}{2}\psi^s_{nt}/Q_n$, and its predictive ability for future stock n returns in the regression

$$S_{n(t+\tau)} - S_{nt} = \alpha_{\mathcal{SI}_n} + \beta_{\mathcal{SI}_n}\mathcal{SI}_{nt} + \epsilon_{n(t+\tau)}, \tag{12}$$

where $\alpha_{\mathcal{SI}_n}$ is some constant, $\beta_{\mathcal{SI}_n}$ is the slope coefficient that we are interested in, and $\epsilon_{n(t+\tau)}$ are the error terms. Towards that, Proposition 2 reports the equilibrium short interest $\mathcal{SI}_{nt}$ and the slope coefficient $\beta_{\mathcal{SI}_n}$ in the predictive regression for stock $n = 1, 2$, along with their key properties.

Proposition 2 (**Equilibrium short interest and predictability**). *In the costly short-selling economy, for the costly-to-short stock 1, the equilibrium short interest and its slope coefficient in the predictive regression* (12) *are given by*

$$\mathcal{SI}_{1t} = \overline{\mathcal{SI}}_{1t} + K_1 + \frac{1}{2r}\Big(\frac{(\kappa_1 + r - 2H^s\upsilon_1^2)\, b_1 B_1 - b_1 + 1/r}{\gamma\sigma_1^2 Q_1} - \frac{r}{\gamma\sigma_{D_1}^2 Q_1}\Big)\frac{1}{b_1}\phi_{1t}, \tag{13}$$

$$\beta_{\mathcal{SI}_1} = -2rb_1B_1\frac{\gamma\sigma_1^2 Q_1}{(\kappa + r - 2H^s\upsilon_1^2)\, b_1 B_1 - b_1 + 1/r}\left(1 - e^{-\kappa_1\tau}\right), \tag{14}$$

where the shorting fee ϕ_{1t}, the stock volatility σ_1, and the constants b_1, B_1, and H^s are as introduced in Proposition 1, and the constant K_1 is provided in Appendix A. For the zero-fee stock 2, the equilibrium short interest and its slope coefficient in the predictive regression (12) *are given by $\mathcal{SI}_{2t} = -1/2 + \theta_2/(2\gamma\sigma_{D_2}^2 Q_2)$ and $\beta_{\mathcal{SI}_2} = 0$, respectively. In the benchmark economy with costless short-selling, for stock n, $n = 1, 2$, the equilibrium short interest and its slope coefficient in the predictive regression* (12) *are given by $\overline{\mathcal{SI}}_{nt} = -1/2 + \theta_{nt}/(2\gamma\sigma_{D_n}^2 Q_n)$ and $\bar{\beta}_{\mathcal{SI}_n} = 0$, respectively.*

Consequently, in the costly short-selling economy, the costly-to-short stock

i) short interest is increasing in shorting fee ϕ_{1t},

ii) higher short interest predicts lower future stock returns.

Electronic copy available at: https://ssrn.com/abstract=3516969

Proposition 2 reveals that the costly-to-short stock short interest increases in its shorting fee (property (i)), as also illustrated in Figure 3.[15] This is because, a higher current shorting fee corresponds to a higher current disagreement and a higher stock price, and hence the short-sellers are relatively more pessimistic now and increase their shorting demand, leading to the positive relation between the short interest and the shorting fee of the stock. This result is consistent with the empirical evidence in Drechsler and Drechsler (2016), who find that expensive-to-short stocks have higher short interest compared to cheap-to-short stocks. The typical positive relation between the short interest and shorting fee is also documented by many other studies, including D'Avolio (2002) and Beneish, Lee, and Nichols (2015).

The key implication of Proposition 2 is that the costly-to-short stock short interest predicts future stock returns negatively, implying a higher current short interest tends to be followed by lower stock prices (property (ii)). This predictability arises in our model because, as discussed above, a higher short interest corresponds to a higher current shorting fee, which is now expected to be lower in the future due to mean-reversion, leading to lower future stock prices on average as compared to the relatively high current stock prices. We would like to highlight here that the predictive ability of short interest is solely due to the presence of shorting fee. This is because in the benchmark economy, the short interest is still time-varying (as it is driven by disagreement), but it does not predict future returns since stock 1 price does not depend on disagreement when the short-selling is costless. Therefore, our model shows that the current short interest is an "informative" signal for future returns for costly-to-short stocks but not for stocks that are costless-to-short.

In the literature, empirical investigations of whether the short interest predicts future returns or not date back to Seneca (1967), who finds that high short interest predicts lower future returns for the S&P500. Subsequent research also confirmed these findings for both at the individual stock levels and the aggregate market level. For the individual stock levels, for example, Desai, Ramesh, Thiagarajan, and Balachandran (2002) show that stocks with a high short interest experience lower returns and this predictability persists up to 12 months. Similarly, Boehmer, Huszar, and Jordan (2010) show that stocks with a low short interest

[15] We do not find it necessary to impose a non-negativity restriction on our short interest measure, since the reasonable parameter values that ensure the almost always positivity of costly-to-short stock disagreement, also ensure the almost always positivity of the short interest in our model. For example, for the parameters values in Table 1 of Appendix B, the probability of $\mathcal{SI}_{1t}$ being positive at the steady-state is $\mathbb{P}\left(\mathcal{SI}_{1t} > 0\right) = 1 - \Phi\left(-44.04\right) \simeq 1$, where Φ is the cumulative distribution function of standard normal distribution.

Electronic copy available at: https://ssrn.com/abstract=3516969

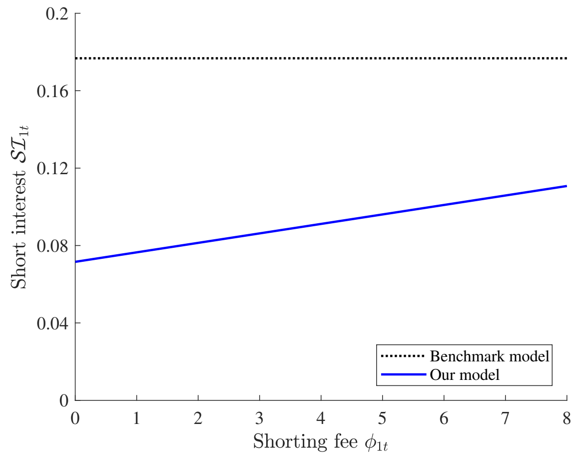


Figure 3: **Short interest behavior.** This figure plots the costly-to-short stock equilibrium short interest $\mathcal{SI}_{1t}$ against its shorting fee ϕ_{1t}. The solid blue line corresponds to our economy with costly short-selling, while the dotted black line corresponds to the benchmark economy with costless short-selling. The parameter values follow from Table 1 of Appendix B.

in the previous month experience higher returns in the next month. Similar findings are also in Figlewski (1981), Senchack and Starks (1993), Asquith, Pathak, and Ritter (2005), and Beneish, Lee, and Nichols (2015). Among them Beneish, Lee, and Nichols show that the effect is stronger for stock with higher shorting fees, which is in line with our finding that in the benchmark economy short interest is still time-varying but it does not predict future returns. For the aggregate market, Rapach, Ringgenberg, and Zhou find that a higher aggregate short interest predicts lower future stock market returns both at the monthly and at the yearly horizon, and in fact argue that the short interest is the strongest known predictor of aggregate stock market returns.[16] On the theory side however, to our knowledge, our analysis is the first to demonstrate that a higher short interest predicts lower future stock returns for various horizons, as well as obtaining an endogenous dynamic short interest with a stationary steady-state distribution.

Moreover, several empirical studies additionally show that, in addition to the short interest (a stock variable), a positive change in short interest (a flow variable) also predicts lower future returns (Boehmer, Jones, and Zhang (2008), Diether, Lee, and Werner (2009)). Our model is also consistent with these findings since for the costly-to-short stock, the slope coefficient in the regression of future stock returns on past changes in short interest $S_{1(t+\tau)} - S_{1t} =$

[16]In our model, the aggregate (equally-weighted) short interest is $\mathcal{SI}_{1t} + \mathcal{SI}_{2t}$,which also predicts the aggregate market $S_{1t}Q_1 + S_{2t}Q_2$ future returns.

Electronic copy available at: https://ssrn.com/abstract=3516969

$\alpha_{\Delta\mathcal{SI}} + \beta_{\Delta\mathcal{SI}} \left(\mathcal{SI}_{1t} - \mathcal{SI}_{1(t-\tau)}\right) + \epsilon_{t+\tau}$, is negative. Furthermore, Diether, Lee, and Werner (2009) also show that the current short selling is positively related to past stock performance. Our model is also consistent with this finding since the slope coefficient in the regression of the current short interest on past stock price changes $\mathcal{SI}_{1t} = \alpha_{\Delta S} + \beta_{\Delta S} \left(S_{1t} - S_{1(t-\tau)}\right) + \epsilon_t$, is positive.[17]

4.2 Short-Selling Risk

In recent empirical work, Engelberg, Reed, and Ringgenberg (2018) find that in the cross-section, stocks with higher short-selling risk (higher uncertainty about future shorting fee), as measured by the shorting fee variance, have lower returns and less short-selling activity (volume), as measured by the number of shares shorted each day, suggesting the short-selling risk as a significant source of limits to arbitrage.[18] Since our model generates a time-varying shorting fee with a stationary steady-state distribution, our model is well-suited to shed light on these findings. We take our measure of *short-selling risk*, V_{ϕ_n}, for stock $n = 1, 2$, as in Engelberg, Reed, and Ringgenberg, and define it to be the variance of the shorting fee (at the steady-state), $V_{\phi_n} = \lim_{t\to\infty} \mathrm{Var}\left[\phi_{nt}\right]$. For our measure of *short-selling activity* (or volume), $\sigma_{\mathcal{SI}_n}$, we consider the volatility of the short interest changes, $\sigma_{\mathcal{SI}_n} = \sqrt{\mathrm{Var}_t\left[d\mathcal{SI}_{nt}\right]/dt}$, consistently with the trading volume proxies employed by works in continuous-time settings (e.g., Xiong and Yan (2010), Longstaff and Wang (2012)).[19] Proposition 3 presents the equilibrium short-selling risk (shorting fee variance) V_{ϕ_n} and the short-selling activity (volatility of the short interest changes) $\sigma_{\mathcal{SI}_n}$ for stock $n = 1, 2$.

[17]We omit the analysis of these two results for brevity, but they can be shown in a straightforward manner following the steps of the proof of Proposition 2.

[18]Engelberg, Reed, and Ringgenberg (2018) find that even though the shorting fee of a typical stock is low on average, it experiences large fluctuations over time, creating significant short-selling risk for investors. For example, they find that for a typical stock, the 1^{st} percentile of its shorting fee is 7 bps on average, while the 99^{th} percentile is 301 bps, indicating a significant variation in shorting fees over time. Moreover, they find the mean (median) number of days for a short position to be open to be 81 (65) days, indicating a significant exposure to the short-selling risk for investors. Engelberg, Reed, and Ringgenberg further show that stocks with higher short-selling risk have less price efficiency, where their price efficiency measure is computed by regressing the weekly stock returns on the current and the lagged (value-weighted) market returns, which is something we cannot capture meaningfully in our framework.

[19]As is well recognized in the case of trading volume, employing the standard definition of short-selling volume, by taking the absolute value of the short interest changes, $|d\mathcal{SI}_{nt}|$, in a continuous-time setting is problematic since the local variation of the driving uncertainty (Brownian motion) is unbounded. The measure $\sigma_{\mathcal{SI}_n}$ does not suffer from this issue and captures the unexpected short-selling volume by not taking into account of the expected changes in the short interest.

Electronic copy available at: https://ssrn.com/abstract=3516969

Proposition 3 (Equilibrium short-selling risk and short-selling activity). *In the costly short-selling economy, for the costly-to-short stock 1, the equilibrium short-selling risk and short-selling activity are given by*

$$V_{\phi_1} = b_1^2 \frac{\upsilon_1^2}{2\kappa_1}, \tag{15}$$

$$\sigma_{\mathcal{SI}_1} = \bar{\sigma}_{\mathcal{SI}_1} - \frac{\upsilon_1}{2r}\Big(\frac{r}{\gamma\sigma_{D_1}^2 Q_1} - \frac{(\kappa_1 + r - 2H^s\upsilon_1^2)\, b_1 B_1 - b_1 + 1/r}{\gamma\sigma_1^2 Q_1}\Big), \tag{16}$$

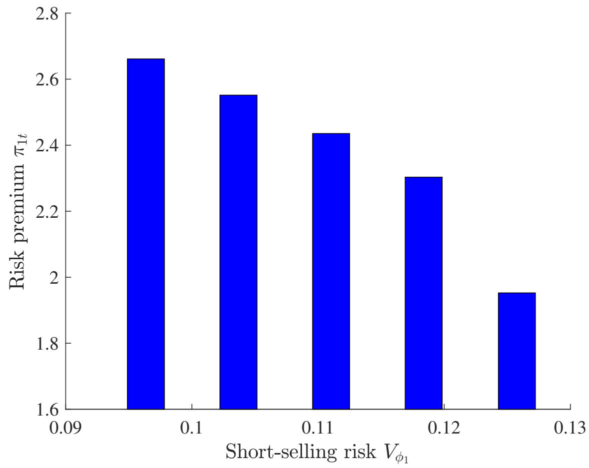
where the stock volatility σ_1, and the constants b_1, B_1, and H^s are as introduced in Proposition 1. For the zero- fee stock 2, the equilibrium short-selling risk and short-selling activity are given by $V_{\phi_2} = \sigma_{\mathcal{SI}_2} = 0$. In the benchmark economy with costless short-selling, for stock n, $n = 1, 2$, the equilibrium short-selling risk and activity are given by $\bar{V}_{\phi_n} = 0$ and $\bar{\sigma}_{\mathcal{SI}_1} = \upsilon_1/(2\gamma\sigma_{D_1}^2 Q_1)$, $\bar{\sigma}_{\mathcal{SI}_2} = 0$, respectively.

In our economy, due to the time-variation in disagreement, short-sellers' demand and the lenders' supply of costly-to-short stock shares also fluctuate, leading to a time-variation in the shorting fee, and thus to a short-selling risk (uncertainty about future shorting fee) in equilibrium. The time-variation in disagreement also leads to fluctuations in the short interest, resulting in a non-trivial short-selling activity (16). Having determined the costly-to-short stock short-selling risk and activity analytically, we here argue that the cross-sectional differences in stock partial lending α_n could very well be behind the evidence in Engelberg, Reed, and Ringgenberg (2018).

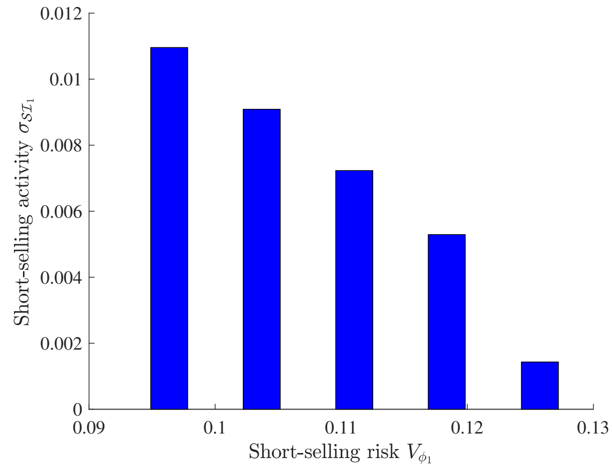
To illustrate this, Figure 4 presents how the short-selling risk is related to risk premium (Panel (a)) and the short-selling activity (Panel (b)) in equilibrium, where each bar is obtained by varying the partial lending parameter. We see that when the costly-to-short stock has a low partial lending it has a higher short-selling risk, a lower risk premium and a lower short-selling activity, as documented in Engelberg, Reed, and Ringgenberg.[20] The negative relation in Figure 4 Panel (a) occurs because the shorting fee of a stock with a low degree of partial lending is more sensitive to the disagreement shocks (higher b_1 in (7)). This is because such a stock has a lower lending supply, which can only absorb the short-selling demand by increasing the shorting fee, leading to a higher shorting fee variance, as well as a higher stock price and

[20]We also find a similar result if instead of short interest volatility (a flow measure) we use the short-interest itself (a stock measure) for the short-selling activity. We demonstrate our results using the short interest volatility to be consistent with Engelberg, Reed, and Ringgenberg,who also employ a flow measure.

Electronic copy available at: https://ssrn.com/abstract=3516969



(a) Risk premium



(b) Short-selling activity

Figure 4: **Effects of short-selling risk.** These panels illustrate the relations between the costly-to-short stock equilibrium short-selling risk V_{ϕ_1} and its risk premium π_{1t} and short-selling activity $\sigma_{\mathcal{SI}_1}$, where each bar is obtained by varying the partial lending α_1. The partial lending parameter values for each bar (from left-to-right) are 0.960, 0.865, 0.765, 0.650, 0.295, which are chosen so that the each bar is equally distanced on the x-axis and covers the range for short-selling risk. The parameter values follow from Table 1 of Appendix B.

a lower risk premium (Proposition 1). Similarly, the negative relation in Figure 4 Panel (b) occurs because a lower partial lending also leads to a lower short-selling activity since short-sellers now short less for any given disagreement level.

In sum, our model offers a possible explanation for the evidence in Engelberg, Reed, and Ringgenberg (2018) by demonstrating that the differences in partial lending across stocks may very well be behind their findings.[21] In particular, our theory suggests that for two similar costly-to-short stocks, the one lent to short-sellers in greater proportion by lenders will have less short-selling risk, higher future returns, and more short-selling activity.

[21] We would like to highlight here that our numerical analysis shows that the changes in short-selling risk due to variations in other quantities such as κ_1, $\bar{\theta}_1$, υ_1, σ_{D_1}, and λ in our model would not generate the observed relation between the short-selling risk and the risk premium and the short-selling activity simultaneously.

Electronic copy available at: https://ssrn.com/abstract=3516969

5 Lenders' Optimal Size

As clearly evident from our analysis so far and the related empirical evidence, the lenders' size and partial lending are important driving factors in the stock short-selling and lending market. In this Section, we endogenously determine the optimal size of lenders given a cost of setting up a lending facility. We show that the lenders' optimal size decreases in the cost of entry, and is non-monotonic, first increasing then decreasing, in partial lending, while the equilibrium stock price and shorting fee increase in the cost of entry.

Absent any other frictions, we can show that among optimistic investors, stock lenders have a higher time-0 indirect utility than the non-lending holders do. This is simply because the lenders partially lend their long stock positions and earn additional income, while non-lenders do not. Given that there are benefits of being a lender in our economy, a natural question arises whether all optimistic investors can be lenders. To determine the optimal size of stock lenders, we introduce an *entry cost*, ξ, of setting up a lending facility at time 0. The optimists then decide whether to pay the cost and become lenders, or not pay the cost and remain as non-lenders. In equilibrium, the lenders' optimal size λ^* is determined endogenously such that the time-0 indirect utility of both types of optimistic investors are equated.[22] That is, the lenders' optimal size λ^* solves

$$J^{\ell}\left(W_0 - \xi, \theta_0, 0; \lambda^*\right) = J^{h}\left(W_0, \theta_0, 0; \lambda^*\right), \tag{17}$$

where $J^i(\cdot)$ is the indirect utility of i-type investor, $i = \ell, h$, defined at time t as $J^i\left(W_t^i, \theta_{1t}, t\right) = \max_{(c^i, \psi_1^i, \psi_2^i)} \mathrm{E}_t^i\left[\int_t^{\infty} e^{-\rho u} \frac{e^{-\gamma c_u^i}}{-\gamma} du\right]$. Proposition 4 presents the lenders' optimal size in our costly short-selling economy.

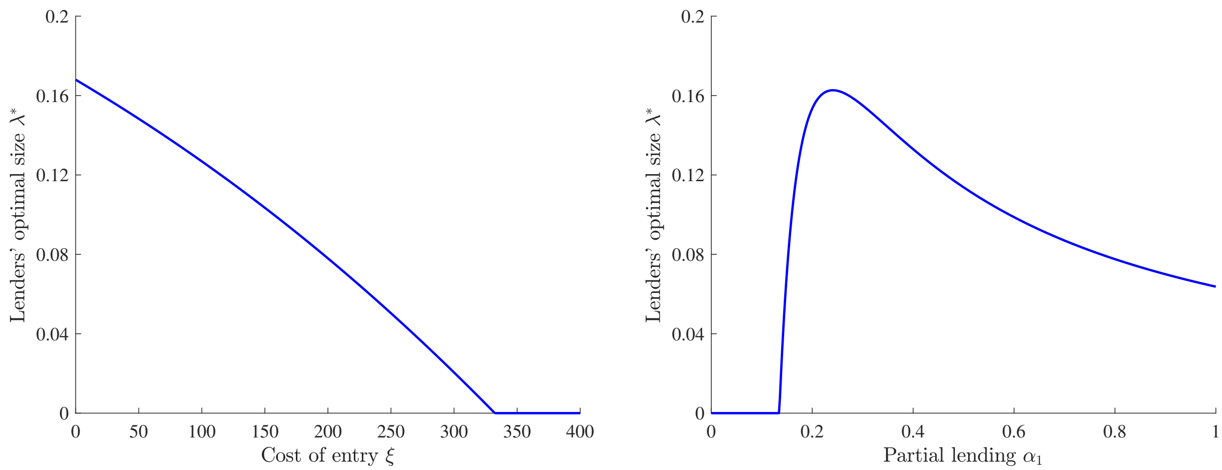
Proposition 4 (**Lenders' optimal size in equilibrium**). *In the costly short-selling economy with a cost of entry ξ, the lenders' optimal size λ^* solves*

$$\left(F^h + G^h\theta_0 + H^h\theta_0^2\right) - \left(F^{\ell} + G^{\ell}\theta_0 + H^{\ell}\theta_0^2\right) - \xi\gamma r = 0, \tag{18}$$

where the constants F^i, G^i, H^i, $i = \ell, h$, are functions of λ^* and are as introduced in Proposition 1.

[22] Since both types of optimistic investors, lenders and holders, have the same subjective beliefs in our model,

Electronic copy available at: https://ssrn.com/abstract=3516969



(a) Effects of cost of entry

(b) Effects of partial lending

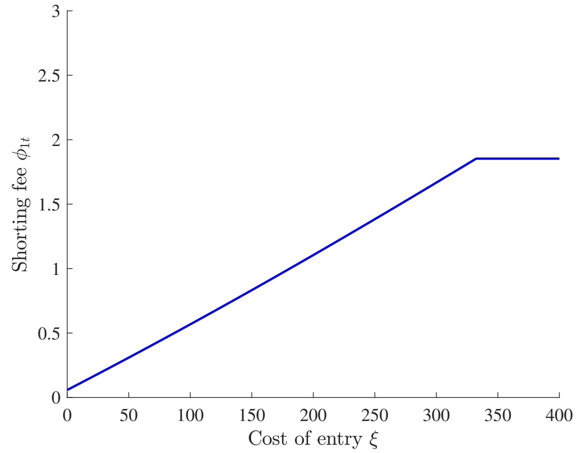
Figure 5: **Lenders' optimal size in equilibrium.** These panels plot the lenders' optimal size λ^* against the cost of entry ξ and costly-to-short stock partial lending α_1. The solid blue lines correspond to costly-to-short stock initial disagreement at its long-run mean $\bar{\theta}_1$. The parameter values follow from Table 1 of Appendix B.

Proposition 4 provides the analytical expression that the lenders' optimal size solves. We illustrate in Figure 5 Panel (a) that even when the entry is costless not all optimistic investors become lenders and the lenders' optimal size decreases in the cost of entry, and when the cost is too high no investor becomes a lender. This is fairly intuitive since an optimistic investor is reluctant to become a lender when it is more costly to do so. On the other hand, from Figure 5 Panel (b) we see that in the presence of entry costs, very low levels of costly-to-short stock partial lending corresponds to no lenders, since the low potential future income from lending does not outweigh the entry cost ξ. As the lenders can lend a higher fraction of each share held long, the future income from lending increases and so does the optimal size of lenders. However, for sufficiently high levels of partial lending, the optimal lenders' size is decreasing in partial lending, since an increase in partial lending now leads to too low of a shorting fee and a lower future lending income.

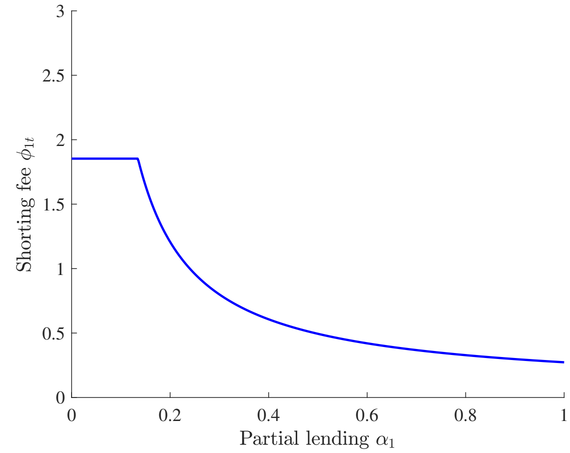
We now turn to investigating the effects of cost of entry and costly-to-short stock partial lending on its equilibrium shorting fee and stock price when the stock lenders' size is endoge-

comparing their indirect utilities is economically meaningful and does not suffer from the well-known issues that would arise if they had different, distorted, beliefs (e.g., Brunnermeier, Simsek, and Xiong (2014)).

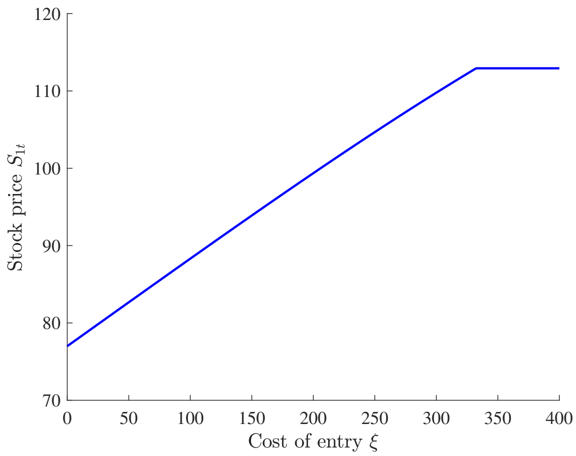
Electronic copy available at: https://ssrn.com/abstract=3516969



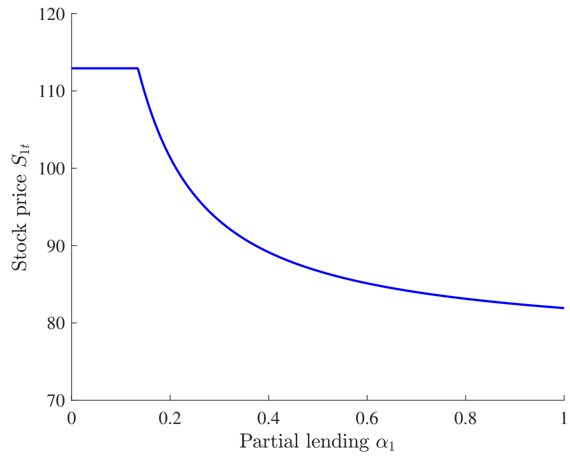
(a) Effects of cost of entry



(b) Effects of partial lending



(c) Effects of cost of entry



(d) Effects of partial lending

Figure 6: **Effects of cost of entry and partial lending on shorting fee and stock price.** These panels plot the costly-to-short stock equilibrium shorting fee ϕ_{1t} and price S_{1t} against the cost of entry ξ and partial lending α_1 under the optimal lenders' size. The solid blue lines correspond to costly-to-short stock initial disagreement at its long-run mean $\bar{\theta}_1$. The parameter values follow from Table 1 of Appendix B.

nously determined as above.[23] Figure 6 Panels (a) and (c) show that the costly-to-short stock shorting fee and the price are both increasing in cost of entry. This is because a higher cost of entry leads to a lower lenders' optimal size, which leads to a higher shorting fee and a higher

[23]We also highlight that all our earlier results based on our Propositions 1–3 continue to hold with the endogenous lenders' optimal size λ^*.

Electronic copy available at: https://ssrn.com/abstract=3516969

stock price for the reasons as discussed in Proposition 1. Similarly, Figure 6 Panels (b) and (d) show that the costly-to-short stock shorting fee and the stock price are both decreasing in its partial lending. These results are a bit more subtle and due to two effects. First effect is due to the fact that a higher partial lending itself leads to a lower shorting fee and lower stock price, since it leads to a higher lending supply. Second effect is due to the impact of partial lending on lenders' optimal size. As discussed above, for low levels of partial lending, lenders' size increases in partial lending, which reinforces the first effect, and hence leading to a steep decrease in the shorting fee and the stock price. For sufficiently high levels of partial lending, lenders' size decreases in partial lending, which partially counters the first effect, leading to a less steep decrease in the shorting fee and the stock price.

In the existing theoretical literature, in a dynamic setting Duffie, Gârleanu, and Pedersen (2002) endogenize the short-selling capital by introducing a fixed cost of entry to being a short-seller, whereas in a static setting Banerjee and Graveline (2014) endogenize the fraction of lenders' long stock position that they can lend out by introducing a convex cost function. Our findings in this section complement these works by endogenizing the lenders' size in our dynamic economy by introducing a fixed cost of entry to being a lender, and show how this cost of entry, along with partial lending, affect the equilibrium stock price and shorting fee.

6 Conclusion

In this paper, we provide a comprehensive analysis of the costly short-selling and lending market within a familiar dynamic asset pricing framework. Our model generates rich implications that support the extensive empirical evidence on the behavior of the costly-to-short stock shorting fee, price, its risk premium and volatility, short interest and its predictive power, and short-selling risk and activity, and also offer simple straightforward intuitions for them. Moreover, we endogenously determine the optimal size of lenders by introducing a fixed cost of entry to being a lender.

Our main results for the costly-to-short stock are as follows. We find that the equilibrium shorting fee increases in belief disagreement. We also find that the stock price is positively, while its risk premium is negatively related to its shorting fee, thereby demonstrating that the classic Miller (1977) intuition also arises in our fairly standard dynamic setup. We additionally show that the stock volatility is increased in the presence of costly short-selling. More notably,

Electronic copy available at: https://ssrn.com/abstract=3516969

we show that the equilibrium short interest increases in the shorting fee and predicts future stock returns negatively. Furthermore, we demonstrate that higher short-selling risk can be associated with lower stock returns and less short-selling activity. These implications of our model are all consistent with empirical evidence, and in particular, to the best of our knowledge, our results on the predictive power of the short interest, as well as on the stock volatility and the short-selling risk, are new and have not been demonstrated in the extant theoretical literature.

So as to not unnecessarily complicate our model, we do not consider some specific features of the actual stock short-selling and lending practices. For instance, in our model 100% of the short-selling proceeds are kept as collateral (see proof of Lemma 1 in Appendix A). This rate is very close to that in the US for domestic stocks, for which lenders typically require 102% of the short-selling proceeds as a collateral to help protect themselves (D'Avolio (2002)). Moreover, in our model, lenders get all of the shorting fee upon lending a share. In reality, this is true for large institutions with internal lending facilities, which directly lend to short-sellers. Other lenders typically use an agent bank/brokerage and receive only a fraction of the shorting fee, with the rest paid to the agent bank/brokerage for providing the lending service. Our framework would be able to accommodate these generalizations. Furthermore, in Section 5, we do not consider the potential costs of lending in the form of lenders' forgoing their voting rights (D'Avolio (2002)). While it may be challenging to consider such a setup with a dynamic trade-off between the benefits and costs of lending, it may generate significant implications for policymaking that is concerned with regulating these markets. Relatedly, our framework may also accommodate other regulatory interventions such as the short-selling bans and their welfare consequences as in Buss et al. (2016), benchmarking incentives of the lending institutions as in Basak and Pavlova (2013), additional trading fees as in Buss and Dumas (2019), as well as noise trader risk and short squeeze in the short selling market. We leave these considerations for future research.

Electronic copy available at: https://ssrn.com/abstract=3516969

Appendix A: Proofs

Proof of Lemma 1. The financial wealth dynamics of each i-type investor, $i = \ell, h, s$, are obtained as follows. Since the holders, $i = h$, do not participate in the stock short-selling and lending market, their stock position yields the usual capital gains/losses and dividends, and hence the evolution of their financial wealth is the familiar

$$dW_t^h = \left(W_t^h - \sum_{n=1}^{2} \psi_{nt}^h S_{nt} \right) r dt + \sum_{n=1}^{2} \psi_{nt}^h \left(dS_{nt} + D_{nt} dt \right) - c_t^h dt. \tag{A.1}$$

The stock position of the lenders, $i = \ell$, on the other hand, in addition to the capital gains/losses and dividends, yields an additional lending income of $\alpha_n \phi_{nt}$ per share of stock n, $n = 1, 2$, held, and hence the evolution of their financial wealth becomes

$$dW_t^\ell = \left(W_t^\ell - \sum_{n=1}^{2} \psi_{nt}^\ell S_{nt} \right) r dt + \sum_{n=1}^{2} \psi_{nt}^\ell \left(dS_{nt} + D_{nt} dt \right) + \sum_{n=1}^{2} \psi_{nt}^\ell \alpha_n \phi_{nt} dt - c_t^\ell dt. \tag{A.2}$$

Finally, the stock position of the short-sellers, $i = s$, in addition to the capital gains/losses and dividends, effectively incurs an additional short-selling cost of ϕ_{nt} per share of stock n, and hence the evolution of their financial wealth becomes

$$dW_t^s = \left(W_t^s - \sum_{n=1}^{2} \psi_{nt}^s S_{nt} \right) r dt + \sum_{n=1}^{2} \psi_{nt}^s \left(dS_{nt} + D_{nt} dt \right) + \sum_{n=1}^{2} \psi_{nt}^s \phi_{nt} dt - c_t^s dt. \tag{A.3}$$

Next, we note that the optimistic investors agreeing on the dividend levels of stock n, $n = 1, 2$, and hence on its changes, yields the consistency relation $dD_{nt} = \mu_{D_n} dt + \sigma_{D_n} d\omega_{nt} = (\mu_{D_n} + \theta_{nt}) dt + \sigma_{D_n} d\omega_{nt}^i$, for $i = \ell, h$, which yields the relation between the perceived and objective Brownian motions as stated in the main text, $d\omega_{nt}^i = d\omega_{nt} - (\theta_{nt}/\sigma_{D_n})dt$, for $i = \ell, h$. A similar consideration yields the relation $d\omega_{nt}^s = d\omega_{nt} + (\theta_{nt}/\sigma_{D_n})dt$, for the short-sellers. Substituting the posited stock price dynamics (3) into (A.1)–(A.3), and using the relations for the perceived and the objective Brownian motions above leads to (4). $\square$

Further discussion on disagreement process (2)**.** In this note, we demonstrate how the dynamics for the belief disagreement (2) could arise in an economy with Bayesian investors, who are symmetrically informed but have different interpretations of signals, as in the dynamic differences in beliefs models with stationary disagreement (e.g., Scheinkman and Xiong

Electronic copy available at: https://ssrn.com/abstract=3516969

(2003), Dumas, Kurshev, and Uppal (2009), Xiong and Yan (2010)). In these models, the quantity the investors disagree on (typically, the mean of a fundamental process) is a mean-reverting unobservable process, and investors observe signals but have different knowledge on the informativeness of signals, which leads to investors to react differently to the signals and therefore to have different posterior beliefs, and hence disagree.

To illustrate this disagreement in the presence of learning in our setting, we adopt the Xiong and Yan (2010) approach, and assume that the costly-to-short stock under consideration is a claim to the dividend D with dynamics $dD_t = \mu_t dt + \sigma d\omega_t$. The mean dividend change μ is now an unobservable process with dynamics $d\mu_t = \kappa_\mu (\bar{\mu} - \mu_t) dt + \sigma_\mu d\omega_{\mu t}$, where the constants κ_μ, $\bar{\mu}$, and σ_μ are the speed of mean reversion, long-run mean, and volatility of the mean dividend change process, respectively, and ω_μ is a standard Brownian motion, independent from ω. All investors observe the signal $dN_t = d\omega_{\theta t}$, where ω_θ is another standard Brownian motion independent of all other Brownian motions. Even though the signal N is "pure noise" and is not informative, each i-type investor, $i = \ell, h, s$, believes it contains useful information by having different prior knowledge of it in the form of $dN_t = \varphi^i d\omega_{\mu t} + \sqrt{1 - \varphi^{i2}} d\omega_{\theta t}$, where the parameter φ^i captures the each i-type investor's perceived correlation between the signal dN_t and $d\omega_{\mu t}$. As in Xiong and Yan (2010), we also assume that each i-type investor misperceives the volatility of μ as $k^i \sigma_\mu$ rather than σ_μ, which helps us to isolate the effects of belief disagreement, as shown below.

In this setting, each i-type investor, $i = \ell, h, s$, estimates μ from the observations of the dividend and the signal, and hence, their information is the filtration $\mathcal{G}_t = \sigma \{D_s, N_s : 0 \leq s \leq t\}$. Assuming investors view the prior distribution of μ as normal with mean m_o and variance V_o, the application of the standard Kalman filtering (e.g., Liptser and Shiryaev (2001)) yields the posterior mean $m_t^i = \mathrm{E}[\mu_t | \mathcal{G}_t]$ and the posterior variance $V_t^i = \mathrm{E}\left[(\mu_t - m_t^i)^2 | \mathcal{G}_t\right]$ as

$$
\begin{aligned}
dm_t^i &= \kappa_\mu \left(\overline{\mu} - m_t^i\right) dt + \frac{1}{\sigma} V_t d\widehat{\omega}_t^i + \varphi^i k^i \sigma_\mu dN_t, \\
dV_t^i &= -\left[\frac{1}{\sigma^2} V_t^{i^2} + 2\kappa_\mu V_t^i - \left(1 - \varphi^{i^2}\right) k^{i^2} \sigma_\mu^2\right] dt,
\end{aligned}
$$

where each i-type investor's perceived Brownian motion is given by $d\widehat{\omega}_t^i = \frac{1}{\sigma}(dD_t - m_t^i dt)$. By choosing the constants $\varphi^\ell = \varphi^h = -\varphi^s = \varphi > 0$, and $k^\ell = k^h = k^s = 1/\sqrt{1 - \varphi^2}$, similarly to Xiong and Yan (2010), we shut down the channels due to investors' overconfidence on signal precision and isolate the effects of disagreement. We also note that with this specifi-

Electronic copy available at: https://ssrn.com/abstract=3516969

cation, the ℓ-type and the h-type investors have identical beliefs, and more importantly the posterior variance of all investors become identical, and is equal to $\overline{V}^{\ell} = \overline{V}^{h} = \overline{V}^{s} = \overline{V} = \sigma^2\sqrt{\kappa_\mu^2 + \sigma_\mu^2/\sigma^2} - \sigma^2\kappa_\mu$, at the steady-state, which we base our analysis. Finally, by defining the belief disagreement simply as the population weighted difference between the optimistic and pessimistic investors' posterior means, $\theta_t = 0.5m_t^{\ell} - 0.5m_t^{s}$, (as in our model also), we obtain the disagreement dynamics, similarly to Xiong and Yan (2010), as

$$d\theta_t = \kappa\left(0 - \theta_t\right)dt + \upsilon d\omega_{\theta t}, \tag{A.4}$$

where the constants are given by $\kappa = \kappa_\mu + \overline{V}/\sigma^2$ and $\upsilon = \varphi\sigma_\mu/\sqrt{1-\varphi^2}$.

Hence, we see that the main features of our disagreement process (2) also arise in this Bayesian learning environment, namely the disagreement being a stationary mean-reverting process with independent shocks. We see that differently from (2), the disagreement process (A.4) has a zero long-run mean, whereas for generality we assume it to be non-zero in our main analysis. However, this is not crucial and our main analysis remains valid if we also take it to be zero, since the long-run mean of disagreement does not play an important role in our mechanisms and results, though it helps us ensure that the costly-to-short stock shorting fee is positive with a high probability at the steady-state. In sum, we highlight that our model can easily accommodate this more richer learning environment with unobservable stochastic dividend mean, but it would unnecessarily complicate the analysis as it would introduce an additional state process μ_{1t}, therefore, for clarity, we carry out our analysis with our disagreement in (2).

Proof of Proposition 1. To determine the equilibrium stock prices in the costly short-selling economy, we first conjecture, and later verify, that the equilibrium stock price and the shorting fee of stock n, $n = 1, 2$, take simple linear forms. Specifically, since only the costly-to short stock 1 has a time-varying disagreement θ_1, we conjecture the stock prices and shorting fees as

$$S_{1t} = \tilde{A}_1 + \tilde{B}_1\theta_{1t} + \frac{1}{r}D_{1t}, \qquad S_{2t} = \tilde{A}_2 + \frac{1}{r}D_{2t}, \tag{A.5}$$

$$\phi_{1t} = a_1 + b_1\theta_{1t}, \qquad \phi_{2t} = a_2, \tag{A.6}$$

for some constants $\tilde{A}_1$, $\tilde{B}_1$, $\tilde{A}_2$, a_1, b_1 and a_2 to be determined endogenously in equilibrium. Using the stock price form (A.5), the disagreement dynamics (2) and the posited stock price

Electronic copy available at: https://ssrn.com/abstract=3516969

dynamics (3), we obtain the mean and the diffusion terms of the stock price (changes) as, for stock 1, $\mu_{1t} = \mu_{D_1}/r + \tilde{B}_1\kappa_1\bar{\theta}_1 - \tilde{B}_1\kappa_1\theta_{1t} + D_{1t}$, $\sigma_{S_1} = \sigma_{D_1}/r$, and $\sigma_{\theta_1} = \tilde{B}_1\upsilon_1$, and for stock 2, $\mu_{2t} = \mu_{D_2}/r + D_{2t}$, $\sigma_{S_2} = \sigma_{D_2}/r$, and $\sigma_{\theta_2} = 0$, respectively. These in turn imply the volatility of the stock price (changes) as constants given by $\sigma_1 = \sqrt{\sigma_{S_1}^2 + \sigma_{\theta_1}^2} = \sqrt{(\sigma_{D_1}/r)^2 + \tilde{B}_1^2\upsilon_1^2}$ and $\sigma_2 = \sigma_{S_2} = \sigma_{D_2}/r$, respectively. With these dynamics, each i-type investor's subjective risk premium for stock 1 takes a linear form as $\pi_{1t}^i = \pi_a^i - \pi_b^i\theta_{1t}$, where

$$\pi_a^i = \begin{cases} \frac{\mu_{D_1}}{r} + \tilde{B}_1\kappa_1\bar{\theta}_1 - \tilde{A}_1 r + \alpha_1 a_1 & \text{for } i = \ell, \\ \frac{\mu_{D_1}}{r} + \tilde{B}_1\kappa_1\bar{\theta}_1 - \tilde{A}_1 r & \text{for } i = h, \\ \frac{\mu_{D_1}}{r} + \tilde{B}_1\kappa_1\bar{\theta}_1 - \tilde{A}_1 r + a_1 & \text{for } i = s, \end{cases} \qquad \pi_b^i = \begin{cases} (\kappa_1 + r)\,\tilde{B}_1 - \frac{1}{r} - \alpha_1 b_1 & \text{for } i = \ell, \\ (\kappa_1 + r)\,\tilde{B}_1 - \frac{1}{r} & \text{for } i = h, \\ (\kappa_1 + r)\,\tilde{B}_1 + \frac{1}{r} - b_1 & \text{for } i = s, \end{cases}$$

and their subjective risk premium for stock 2 becomes constant and given by

$$\pi_2^i = \begin{cases} \frac{\mu_{D_2}}{r} - \tilde{A}_2 r + \frac{1}{r}\theta_2 + \alpha_2 a_2 & \text{for } i = \ell, \\ \frac{\mu_{D_2}}{r} - \tilde{A}_2 r + \frac{1}{r}\theta_2 & \text{for } i = h, \\ \frac{\mu_{D_2}}{r} - \tilde{A}_2 r - \frac{1}{r}\theta_2 + a_2 & \text{for } i = s. \end{cases}$$

We next solve each i-type investor's optimization problem using the standard stochastic dynamic programming method. From the theory of stochastic control, investors' optimal consumption c^i and portfolio strategy ψ_n^i satisfy the Hamilton–Jacobi–Bellman equation

$$\begin{aligned} 0 = \max_{(c^i, \psi_1^i, \psi_2^i)} \frac{e^{-\rho t - \gamma c_t^i}}{-\gamma} + \frac{\partial J^i}{\partial t} + \left(W_t^i r + \psi_{1t}^i \pi_{1t}^i + \psi_{2t}^i \pi_2^i - c_t^i\right) \frac{\partial J^i}{\partial W} + \frac{1}{2}\left(\psi_{1t}^{i^2}\sigma_1^2 + \psi_{2t}^{i^2}\sigma_2^2\right) \frac{\partial^2 J^i}{\partial W^2} \\ + \kappa_1\left(\bar{\theta}_1 - \theta_{1t}\right)\frac{\partial J^i}{\partial \theta_1} + \frac{1}{2}\upsilon_1^2 \frac{\partial^2 J^i}{\partial \theta_1^2} + \psi_{1t}^i \tilde{B}_1 \upsilon_1^2 \frac{\partial^2 J^i}{\partial W \partial \theta_1}, \end{aligned} \tag{A.7}$$

where $J^i\left(W_t^i, \theta_{1t}, t\right) = \max_{(c^i, \psi_1^i, \psi_2^i)} \mathrm{E}_t^i\left[\int_t^\infty e^{-\rho u} \frac{e^{-\gamma c_u^i}}{-\gamma} du\right]$ is i-type investor's indirect utility function. We proceed by taking i-type investor's indirect utility $J^i(\cdot)$ as in (11). Taking the first-order conditions of (A.7) with respect to c^i, ψ_1^i and ψ_2^i, and substituting the indirect

Electronic copy available at: https://ssrn.com/abstract=3516969

utility (11) gives the optimal consumption and portfolio strategy, respectively, as

$$c_t^i = rW_t^i - \frac{\ln(\gamma r)}{\gamma} - \frac{F^i + G^i\theta_{1t} + H^i\theta_{1t}^2}{\gamma}, \tag{A.8}$$

$$\psi_{1t}^i = \frac{\pi_{1t}^i}{\gamma r\sigma_1^2} + \frac{\tilde{B}_1\upsilon_1^2\left(G^i + 2H^i\theta_{1t}\right)}{\gamma r\sigma_1^2}, \tag{A.9}$$

$$\psi_2^i = \frac{\pi_2^i}{\gamma r\sigma_2^2}. \tag{A.10}$$

Substituting (11), (A.8), (A.9), and (A.10) into (A.7) and rearranging gives the following quadratic equation in θ_{1t},

$$0 = (r-\rho) - r\ln(\gamma r) - rF^i - rG^i\theta_{1t} - rH^i\theta_{1t}^2 - \frac{1}{2\sigma_1^2}\left[\left(\pi_a^i + G^i\tilde{B}_1\upsilon_1^2\right) + \left(2H^i\tilde{B}_1\upsilon_1^2 - \pi_b^i\right)\theta_{1t}\right]^2$$
$$+ \kappa_1\left(\bar{\theta}_1 - \theta_{1t}\right)\left(G^i + 2H^i\theta_{1t}\right) + \frac{1}{2}\upsilon_1^2\left[2H^i + \left(G^i + 2H^i\theta_{1t}\right)^2\right] - \frac{\pi_2^{i^2}}{2\sigma_2^2}.$$

By the method of undetermined coefficients, we must have for $i = \ell, h, s$,

$$0 = (r-\rho) - r\ln(\gamma r) - rF^i - \frac{\left(\pi_a^i + G^i\tilde{B}_1\upsilon_1^2\right)^2}{2\sigma_1^2} + \kappa_1\bar{\theta}_1G^i + \frac{\upsilon_1^2\left(2H^i + G^{i^2}\right)}{2} - \frac{\pi_2^{i^2}}{2\sigma_2^2}, \tag{A.11}$$

$$0 = \kappa_1\left(2\bar{\theta}_1H^i - G^i\right) - rG^i - \frac{1}{\sigma_1^2}\left(\pi_a^i + G^i\tilde{B}_1\upsilon_1^2\right)\left(2H^i\tilde{B}_1\upsilon_1^2 - \pi_b^i\right) + 2\upsilon_1^2G^iH^i, \tag{A.12}$$

$$0 = 2\upsilon_1^2H^{i^2} - (r + 2\kappa_1)H^i - \frac{1}{2\sigma_1^2}\left(2H^i\tilde{B}_1\upsilon_1^2 - \pi_b^i\right)^2. \tag{A.13}$$

To determine the constants $\tilde{A}_1$, $\tilde{B}_1$, a_1, and b_1, and hence verify our conjectured equilibrium shorting fee and the price for stock 1, we first impose the stock market clearing condition, $\lambda\psi_{1t}^\ell + (\frac{1}{2} - \lambda)\psi_{1t}^h + \frac{1}{2}\psi_{1t}^s = Q_1$. Substituting (A.9), and rearranging yields a linear equation

Electronic copy available at: https://ssrn.com/abstract=3516969

of θ_{1t}. By the method of undetermined coefficients, we must also have

$$\left[\left(2\lambda\left(1-\alpha_1\right)H^\ell+\left(1-2\lambda\right)H^h\right)\upsilon_1^2-\left(\frac{1}{2}-\lambda\alpha_1\right)\left(\kappa_1+r\right)\right]r\tilde{B}_1+\left(\frac{1}{2}-\lambda\alpha_1\right)+\lambda\alpha_1\left(1-\alpha_1\right)rb_1=0, \tag{A.14}$$

$$\frac{\left[\lambda(1-\alpha_1)G^\ell+\left(1/2-\lambda\right)G^h\right]\upsilon_1^2\tilde{B}_1+\left(1/2-\lambda\alpha_1\right)\left(\mu_{D_1}/r+\kappa_1\bar{\theta}_1\tilde{B}_1-r\tilde{A}_1\right)+\lambda\alpha_1(1-\alpha_1)\,a_1}{r\gamma\left(\tilde{B}_1^2\upsilon_1^2+\sigma_{D_1}^2/r^2\right)}=Q_1. \tag{A.15}$$

We next impose the stock short-selling and lending market clearing condition for stock 1, $\lambda\alpha_1\psi_{1t}^\ell+\frac{1}{2}\psi_{1t}^s=0$. Substituting (A.9) and again using the method of undetermined coefficients leads to

$$\left[\left(2\lambda\alpha_1H^\ell+H^s\right)\upsilon_1^2-\left(\frac{1}{2}+\lambda\alpha_1\right)\left(\kappa_1+r\right)\right]r\tilde{B}_1-\left(\frac{1}{2}-\lambda\alpha_1\right)+\left(\frac{1}{2}+\lambda\alpha_1^2\right)rb_1=0, \tag{A.16}$$

$$\left(\lambda\alpha_1G^\ell+\frac{1}{2}G^s\right)\upsilon_1^2\tilde{B}_1+\left(\frac{1}{2}+\lambda\alpha_1\right)\left(\frac{\mu_{D_1}}{r}+\kappa_1\bar{\theta}_1\tilde{B}_1-r\tilde{A}_1\right)+\left(\frac{1}{2}+\lambda\alpha_1^2\right)a_1=0. \tag{A.17}$$

We then jointly solve the constants H^ℓ, H^h, H^s, $\tilde{B}_1$, and b_1 from the five equations: (A.13) for $i=\ell,h,s$, (A.14), and (A.16). For this, we first notice (A.13) is quadratic in H^i, which yields two distinct roots, one is positive and the other is negative. We obtain a unique linear equilibrium by imposing the condition that if the disagreement approaches to being constant, i.e., $\upsilon_1\to 0$, we should obtain the economic quantities of the same form for stock 1 and 2. This identification rules out the positive roots of H^i. Consequently, we take the negative root of the quadratic equation (A.13) and obtain for $i=\ell,h,s$,

$$H^i=\frac{-e_1^i-\sqrt{e_1^{i^2}-4e_2^ie_0^i}}{2e_2^i}<0, \tag{A.18}$$

where $e_2^i=2\upsilon_1^2\sigma_{D_1}^2/r^2$, $e_1^i=2\tilde{B}_1\upsilon_1^2\pi_b^i-\left(r+2\kappa_1\right)\sigma_1^2$, and $e_0^i=\pi_b^{i^2}/2$. The constant $\tilde{B}_1$ and b_1 are then jointly solved from the resulting non-linear equation obtained by substituting (A.18) into (A.14) and (A.16), which implies that $\tilde{B}_1$, and b_1 must have the same sign. Again, we impose the condition that if the disagreement approaches to being constant, i.e., $\upsilon_1\to 0$, we should obtain the economic quantities of the same form for stock 1 and 2, in which $\tilde{B}_1$ and b_1 are both positive, so we conclude that is also the case here. Moreover, this is consistent with the special case of our model with $\alpha_1=1$, in which we have a closed-from solution for

Electronic copy available at: https://ssrn.com/abstract=3516969

$\tilde{B}_1 = 1/(r(\kappa_1 + r)) > 0$, which also guarantees $b_1 > 0$.

We next jointly determine the constants G^ℓ, G^h, G^s, $\tilde{A}_1$, and a_1, from the five equations: (A.12) for $i = \ell, h, s$, (A.15), and (A.17). For this, we rearrange (A.12), which yields for $i = \ell, h, s$,

$$G^i = \frac{2\kappa_1\bar{\theta}_1 H^i \sigma_1^2 - \left(2H^i\tilde{B}_1\upsilon_1^2 - \pi_b^i\right)\pi_a^i}{\left(\kappa_1 + r\right)\sigma_1^2 + \left(2H^i\tilde{B}_1\upsilon_1^2 - \pi_b^i\right)\tilde{B}_1\upsilon_1^2 - 2\upsilon_1^2\sigma_1^2 H^i}. \tag{A.19}$$

The constant $\tilde{A}_1$ and a_1 are then backed out from the resulting non-linear equation obtained by substituting (A.19), together with the solved constants, H^ℓ, H^h, H^s, $\tilde{B}_1$, and b_1, into (A.15) and (A.17). Finally, we determine the constant F^i for $i = \ell, h, s$, by substituting (A.18), (A.19) into (A.11) and using the solved constants, which yields

$$F^i = \frac{1}{r}\left[(r - \rho) - r\ln(\gamma r) - \frac{1}{2\sigma_1^2}\left(\pi_a^i + G^i\tilde{B}_1\upsilon_1^2\right)^2 + \kappa_1\bar{\theta}_1 G^i + \frac{1}{2}\upsilon_1^2\left(2H^i + G^{i^2}\right) - \frac{1}{2}\frac{\pi_2^{i^2}}{\sigma_2^2}\right].$$

For stock 2, to determine the constants $\tilde{A}_2$ and a_2 and verify our conjectured equilibrium shorting fee and stock price, we again impose the stock market clearing condition $\lambda\psi_{2t}^\ell + (\frac{1}{2} - \lambda)\psi_{2t}^h + \frac{1}{2}\psi_{2t}^s = Q_2$, and the stock short-selling and lending market clearing condition $\lambda\alpha_2\psi_{2t}^\ell + \frac{1}{2}\psi_{2t}^s = 0$. Substituting (A.10) into the two market clearing conditions and solving yields

$$\tilde{A}_2 = \frac{\mu_{D_2}}{r^2} - \frac{\gamma\sigma_{D_2}^2 Q_2}{r^2} + \left(\frac{1}{2} + \lambda\alpha_2\right)\frac{1}{r}a_2,$$
$$a_2 = \max\left\{0, \frac{1}{r}\frac{\left(1/2 - \lambda\alpha_2\right)\bar{\theta}_2 - \left(1/2 + \lambda\alpha_2\right)\gamma\sigma_{D_2}^2 Q_2}{\left(1/2 + \lambda\alpha_2\right)\left(1/2 - \lambda\alpha_2\right) - \lambda\alpha_2\left(1 - \alpha_2\right)}\right\}.$$

We note that for the parameter restriction (6), stock 2 has a zero shorting fee, $a_2 = 0$, leading to the price, $S_{2t} = D_{2t}/r + \mu_{D_2}/r^2 - \gamma\sigma_{D_2}^2 Q_2/r^2$. As discussed in footnote 11, allowing higher stock 2 disagreement levels such that $\frac{1/2+\lambda\alpha_2}{1/2-\lambda\alpha_2}\gamma\sigma_{D_2}^2 Q_2 < \bar{\theta}_2 \leq 2\gamma\sigma_{D_2}^2 Q_2$ results with a strictly positive shorting fee as determined above, which then can be calibrated to the observed low shorting fee of a typical cheap-to-short stock in the data.

Thus, we have verified our conjectured equilibrium stock prices (A.5) and the shorting fees (A.6) by determining the constants $\tilde{A}_1$, $\tilde{B}_1$, $\tilde{A}_2$, a_1, b_1 and a_2, as well as determining the constants F^i, G^i, and H^i in the indirect utility of each i-type investor (11), $i = \ell, h, s$. Lastly, we use the monotonic relation between the stock 1 shorting fee and the disagreement

Electronic copy available at: https://ssrn.com/abstract=3516969

(7), and rewrite the price of stock 1 in terms of the shorting fee as in (8) by defining $A_1 = \tilde{A}_1 - a_1 \tilde{B}_1/b_1 - \mu_{D_1}/r^2 + \gamma\sigma^2_{D_1} Q_1/r^2$ and $B_1 = \tilde{B}_1/b_1$.

In the costly short-selling economy, the equilibrium stock risk premium is given by $\pi_{nt} = \mu_{nt} - rS_{nt}$, $n = 1, 2$, where μ_{nt} is determined from the drift term in the dynamics of the stock prices (A.5), yielding the stock risk premium as

$$\pi_{1t} = \frac{\mu_{D_1}}{r} + \tilde{B}_1 \kappa_1 \bar{\theta}_1 - r\tilde{A}_1 - (\kappa_1 + r)\, \tilde{B}_1 \theta_{1t},$$
$$\pi_2 = \frac{\mu_{D_2}}{r} - r\tilde{A}_2.$$

For stock 1, using the monotonic relation between the shorting fee and the disagreement (7), together with $A_1 = \tilde{A}_1 - a_1 \tilde{B}_1/b_1 - \mu_{D_1}/r^2 + \gamma\sigma^2_{D_1} Q_1/r^2$ and $B_1 = \tilde{B}_1/b_1$, we obtain (9). For stock 2, substituting $\tilde{A}_2 = \mu_{D_2}/r^2 - \gamma\sigma^2_{D_2} Q_2/r^2$, into π_2 above yields $\pi_2 = \gamma\sigma^2_{D_2} Q_2/r$. The equilibrium volatility of stock $n = 1, 2$, is as determined above and given by $\sigma_1 = \sqrt{\sigma^2_{S_1} + \sigma^2_{\theta_1}} = \sqrt{(\sigma_{D_1}/r)^2 + \tilde{B}_1^2 \upsilon_1^2}$ and $\sigma_2 = \sigma_{S_2} = \sigma_{D_2}/r$, which along with $B_1 = \tilde{B}_1/b_1$ yields (10).

In the benchmark economy with costless short-selling, we solve the equilibrium price for stock n, $n = 1, 2$, again by conjecturing a linear form, $\bar{S}_{nt} = \bar{A}_n + \bar{B}_n \theta_{nt} + D_{nt}/r$, and taking i-type investor's indirect utility $\bar{J}^i(\cdot)$ as in (11). Following the similar steps to those in our model with costly short-selling above, taking $a_1 = b_1 = a_2 = 0$, and substituting (A.9) and (A.10) into the same stock market clearing conditions, immediately yields $\bar{A}_n = \mu_{D_n}/r^2 - \gamma\sigma^2_{D_n} Q_n/r^2$ and $\bar{B}_n = 0$, verifying our conjecture. The stock n mean return and volatility are thus given by $\bar{\mu}_{nt} = \mu_{D_n}/r + D_{nt}$ and $\bar{\sigma}_n = \sigma_{D_n}/r$, respectively, and hence the stock risk premium is $\bar{\pi}_n = \gamma\sigma^2_{D_n} Q_n/r$. The constants $\bar{F}^i$, $\bar{G}^i$, and $\bar{H}^i$ in the indirect utility function are then solved through a similar system of equations of (A.11)–(A.13) by substituting $\bar{A}_n$, $\bar{B}_n$ and setting $a_1 = b_1 = a_2 = 0$. Finally, by substituting $\bar{A}_n$, $\bar{B}_n$, $\bar{F}^i$, $\bar{G}^i$, and $\bar{H}^i$ into (A.8), (A.9), and (A.10), we obtain each i-type investor's optimal consumption $\bar{c}^i$ and portfolio strategy $\bar{\psi}^i_n$, respectively. Specifically, we have $\bar{\psi}^i_{nt} = Q_n + \theta_{nt}/\gamma\sigma^2_{D_n}$ for $i = \ell, h$, and $\bar{\psi}^i_{nt} = Q_n - \theta_{nt}/\gamma\sigma^2_{D_n}$ for $i = s$.

Property (i), which states that the shorting fee is increasing in disagreement follows from the fact that $b_1 > 0$ as discussed above. Property (ii), which states that the stock price is increasing, while its risk premium is decreasing in shorting fee, follows from the fact that $B_1 > 0$. Property (iii), which states that the stock volatility is higher than that in the benchmark economy with costless short-selling is immediate from (10). □

Electronic copy available at: https://ssrn.com/abstract=3516969

Proof of Proposition 2. In the costly short-selling economy, the equilibrium short interest for stock $n = 1, 2$, is given by $\mathcal{SI}_{nt} = -\frac{1}{2}\psi^s_{nt}/Q_n$. For stock 1, substituting the short-sellers' optimal portfolio strategy (A.9) leads to

$$\mathcal{SI}_{1t} = -\frac{1}{2r}\frac{\left(\mu_{D_1}/r + \tilde{B}_1\kappa_1\bar{\theta}_1 - r\tilde{A}_1 + a_1\right) + G^s\tilde{B}_1\upsilon_1^2}{\gamma\sigma_1^2 Q_1} + \frac{1}{2r}\frac{\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r}{\gamma\sigma_1^2 Q_1}\theta_{1t}, \tag{A.20}$$

which after using the monotonic relation between the shorting fee and disagreement (7), along with $A_1 = \tilde{A}_1 - a_1\tilde{B}_1/b_1 - \mu_{D_1}/r^2 + \gamma\sigma^2_{D_1}Q_1/r^2$ and $B_1 = \tilde{B}_1/b_1$, and rearranging becomes (13), with the constant

$$K_1 = \frac{1}{2} + \frac{1}{2}\frac{1}{\gamma\sigma^2_{D_1}Q_1}\frac{a_1}{b_1} - \frac{1}{2r}\frac{\frac{1}{r}\frac{a_1}{b_1} - rA_1 + \frac{1}{r}\gamma\sigma^2_{D_1}Q_1 + \left[\kappa_1\left(a_1 + b_1\bar{\theta}_1\right) + b_1\upsilon_1^2 G^s - 2a_1\upsilon_1^2 H^s\right]B_1}{\gamma\sigma_1^2 Q_1}.$$

The slope coefficient in the predictive regression (12) is given by

$$\beta_{\mathcal{SI}_1} = \frac{\mathrm{Cov}\left[\mathcal{SI}_{1t}, S_{1(t+\tau)} - S_{1t}\right]}{\mathrm{Var}\left[\mathcal{SI}_{1t}\right]}, \tag{A.21}$$

where the covariance between the short interest and stock price changes, derived from (A.5) and (A.20), is

$$\mathrm{Cov}\left[\mathcal{SI}_{1t}, S_{1(t+\tau)} - S_{1t}\right] = -\tilde{B}_1\frac{\upsilon_1^2}{4\kappa_1 r}\frac{\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r}{\gamma\sigma_1^2 Q_1}\left(1 - e^{-\kappa_1\tau}\right), \tag{A.22}$$

and the variance of short interest in (A.20), using $\mathrm{Var}\left[\theta_{1t}\right] = \upsilon_1^2/(2\kappa_1)$, is

$$\mathrm{Var}\left[\mathcal{SI}_{1t}\right] = \frac{\upsilon_1^2}{8\kappa_1 r^2}\left(\frac{\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r}{\gamma\sigma_1^2 Q_1}\right)^2, \tag{A.23}$$

where the constants H^s and b_1 are defined as in the proof of Proposition 1. Substituting (A.22) and (A.23) into (A.21) and using $B_1 = \tilde{B}_1/b_1$ yields (14).

In the costless short-selling benchmark economy, the equilibrium short interest of stock $n = 1, 2$, is obtained using $\bar{\psi}^s_{nt} = Q_n - \theta_{nt}/\gamma\sigma^2_{D_n}$as derived in Proposition 1. We obtain $\overline{\mathcal{SI}}_{nt} = -1/2 + \theta_{nt}/(2\gamma\sigma^2_{D_n}Q_n)$. The slope coefficient in the predictive regression (12) is zero because the stock price does not depend on disagreement.

Electronic copy available at: https://ssrn.com/abstract=3516969

Property (i), which states that the short interest is increasing in shorting fee, follows from the positivity of the partial derivative of short interest with respect to ϕ_{1t}, or, equivalently θ_{1t}, which is given by $\partial \mathcal{SI}_{1t}/\partial\theta_{1t} = \left[\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r\right]/\left(2r\gamma\sigma_1^2 Q_1\right)$ using (A.20). To show $\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r > 0$, we first substitute (A.16) to this inequality and get a necessary and sufficient condition $2H^\ell\upsilon_1^2\tilde{B}_1 - (\kappa_1 + r)\tilde{B}_1 + 1/r + \alpha_1 b_1 > 0$. We then substitute the equilibrium H^ℓ from (A.18) into the left-hand-side of this condition and rearrange to obtain $\kappa_1(r+\kappa_1)\tilde{B}_1^2 + r(\alpha_1 b_1 + 1/r)\tilde{B}_1 - (\alpha_1 b_1 + 1/r)^2 > 0$, which always holds since $0 < \tilde{B}_1 < 1/(r(r+\kappa_1)) + (1-\alpha_1)\lambda\alpha_1 b_1/((r+\kappa_1)(1/2-\lambda\alpha_1))$ as implied by (A.14) with $H^\ell < 0$ and $H^h < 0$. Property (ii) that a higher short interest predicts lower future stock returns follows from the fact that $\beta_{\mathcal{SI}_1}$ is negative. This is because except for the negative sign, all the terms in (14) are positive including the denominator term $\left(\kappa_1 + r - 2H^s\upsilon_1^2\right)\tilde{B}_1 - b_1 + 1/r$, as shown above. □

Proof of Proposition 3. In the costly short-selling economy, the equilibrium short-selling risk for stock n, $n = 1, 2$, is given by $V_{\phi_n} = \lim_{t\to\infty} \mathrm{Var}\left[\phi_{nt}\right]$. For stock 1, substituting the equilibrium shorting fee (7) and using the steady state variance, $\mathrm{Var}\left[\theta_{1t}\right] = \upsilon_1^2/(2\kappa_1)$, yields (15). For stock 2, we immediately obtain $V_{\phi_2} = 0$. The equilibrium short-selling activity for stock n, $n = 1, 2$, is defined as the volatility of short interest changes, $\sigma_{\mathcal{SI}_n} = \sqrt{\mathrm{Var}_t\left[d\mathcal{SI}_{nt}\right]/dt}$. For stock 1, taking the derivative of (A.20), substituting (7), and using the disagreement dynamics (2) and $B_1 = \tilde{B}_1/b_1$ gives (16). For stock 2, we immediately obtain $\sigma_{\mathcal{SI}_2} = 0$.

In the costless short-selling benchmark economy, for stock n, $n = 1, 2$, the shorting fee and thus its variance is zero. The volatility of short interest is calculated using $\overline{\mathcal{SI}}_{nt} = -1/2 + \theta_{nt}/(2\gamma\sigma_{D_n}^2 Q_n)$, which yields $\bar{\sigma}_{\mathcal{SI}_1} = \upsilon_1/(2\gamma\sigma_{D_1}^2 Q_1)$ and $\bar{\sigma}_{\mathcal{SI}_2} = 0$. □

Proof of Proposition 4. In the costly short-selling economy, given the cost of entry ξ, the lenders' optimal size λ^* solves (17). Substituting the indirect utility (11) at $t = 0$ into (17) and rearranging gives (18). □

Appendix B: Parameter Values in Figures

In this Appendix, we discuss the parameter values and their choices employed in our Figures. Table 1 summarizes the parameter values used. We note that the behavior of the equilibrium quantities depicted in our Figures is typical and does not vary much with alternative plausible values of parameters.

Electronic copy available at: https://ssrn.com/abstract=3516969

We start by setting the interest rate to $r = 0.025$ and the time discount factor to $\rho = 0.015$, as often used in other works with similar settings (e.g., Barberis, Greenwood, Jin, and Shleifer (2015)). We also set the absolute risk aversion coefficient to $\gamma = 1$. For ease of comparison, we assume the same parameter values for each stock dividend dynamics (1) and choose the mean of dividend changes, μ_{D_n}, and the volatility of dividend changes, σ_{D_n}, so that they simply match the aggregate dividend growth rate mean and volatility in the data, 0.018 and 0.115, respectively (Bansal and Yaron (2004)). Next, for stock n, $n = 1, 2$, we choose the initial dividend level D_{n0} and the supply of stock Q_n such that the stock prices in our costly short-selling economy are around 100. To do this, we first set the initial dividend to $D_{n0} = 3.83$ such that the initial price-dividend ratio S_{n0}/D_{n0} matches the average price-dividend ratio of the stock market in the data, which is reported by Bansal and Yaron (2004) to be 26, after rounding. Given the level of initial dividend we then set the supply of stock to $Q_n = 5$ for stock price to be around 100.

For the costly-to-short stock 1 partial lending α_1, the fraction of lenders' long position that is actually lent to short-sellers, we use the average ratio of total stock amount actually lent out to total amount available to borrow for stocks on special (i.e., stocks with shorting fees higher than 1%) as a proxy, which is reported by Aggarwal, Saffi, and Sturgess (2015) to be 43%. We set the lenders' size in our model so that the lending supply (fraction of outstanding shares held by lenders) in our model matches the corresponding quantity in the data (Engelberg, Reed, and Ringgenberg (2018) report this to be 18%), which yields $\lambda = 0.078$. We note that we endogenize this quantity in Section 5 and our analysis there yields the lenders' optimal size to be quite close to this value. In our analysis of Section 5, we also set each investors' initial wealth to $W_0 = 5000$ and the cost of entry to $\xi = 100$ (2% of initial wealth).

For the second moment parameters of the disagreement process of costly-to-short stock 1, we use the summary statistics provided by Yu (2011), who report the standard deviation and the one-month autocorrelation of disagreement as 0.0038 and 0.93, respectively. Matching these quantities to the corresponding ones in our model gives $0.0038 = \upsilon_1/\sqrt{2\kappa_1}$, and $0.93 = \exp(-\kappa_1/12)$, which yields the volatility of the disagreement as $\upsilon_1 = 0.0058$, and speed of mean reversion as $\kappa_1 = 0.87$. We set the mean of the disagreement changes such that in our main model, the probability of the costly-to-short stock 1 shorting fee being positive is 99.9%. With the equilibrium shorting fee (7) as in Proposition 1, this probability is given by $\lim_{t\to\infty} \mathbb{P}(\phi_{1t} > 0) = 1 - \Phi\left(-\left(\bar{\theta}_1 + a_1/b_1\right)/\sqrt{\upsilon_1^2/2\kappa_1}\right) = 99.9\%$, which yields the mean of

Electronic copy available at: https://ssrn.com/abstract=3516969

Parameter	Symbol	Value
Stock n, $n = 1, 2$, initial dividend	D_{n0}	3.83
Stock n, $n = 1, 2$, mean of the dividend changes	μ_{D_n}	0.018
Stock n, $n = 1, 2$, volatility of the dividend changes	σ_{D_n}	0.115
Stock n, $n = 1, 2$, supply of stock shares	Q_n	5
Stock n, $n = 1, 2$, long-run mean of the disagreement	$\bar{\theta}_n$	0.0895
Stock 1 volatility of the disagreement changes	υ_1	0.0058
Stock 1 speed of mean reversion of the disagreement	κ_1	0.87
Stock 1 partial lending	α_1	0.43
Population mass of lenders	λ	0.078
Interest rate	r	0.025
Absolute risk aversion coefficient	γ	1
Time discount factor	ρ	0.015
Cost of entry	ξ	100

Table 1: **Parameter values.** This table reports the parameter values used in our Figures.

the disagreement changes as $\bar{\theta}_1 = 0.0895$, which is assumed to be the same for the stock 2 constant disagreement level, $\bar{\theta}_2 = 0.0895$. This procedure yields the parameter values in Table 1.

We further note that with these parameter values, the probability of the the costly-to-short stock 1 disagreement θ_{1t} and the short interest $\mathcal{SI}_{1t}$ being positive in the steady-state become $\lim_{t\to\infty} \mathbb{P}\left(\theta_{1t} > 0\right) = 1 - \Phi\left(-\bar{\theta}_1/\sqrt{\upsilon_1^2/2\kappa_1}\right) = 1 - \Phi\left(-20.35\right) \simeq 1$, and $\lim_{t\to\infty} \mathbb{P}\left(\mathcal{SI}_{1t} > 0\right) = 1 - \Phi\left(-44.04\right) \simeq 1$, as reported in Remark 1 and footnote 15, respectively.

Electronic copy available at: https://ssrn.com/abstract=3516969

References

Aggarwal, Reena, Pedro A. C. Saffi, and Jason Sturgess, 2015, The role of institutional investors in voting: Evidence from the securities lending market, *Journal of Finance* 70, 2309–2346.

Andrei, Daniel, Bruce Carlin, and Michael Hasler, 2019, Asset pricing with disagreement and uncertainty about the length of business cycles, *Management Science* 65, 2900–2923.

Asquith, Paul, Parag A. Pathak, and Jay R. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Atmaz, Adem, and Suleyman Basak, 2018, Belief dispersion in the stock market, *Journal of Finance* 73, 1225–1279.

Atmaz, Adem, and Suleyman Basak, 2019, Option prices and costly short-selling, *Journal of Financial Economics* 134, 1–28.

Banerjee, Snehal, 2011, Learning from prices and the dispersion in beliefs, *Review of Financial Studies* 24, 3025–3068.

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

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-CAPM: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1–24.

Basak, Suleyman, 2000, A model of dynamic equilibrium asset pricing with heterogeneous beliefs and extraneous risk, *Journal of Economic Dynamics and Control* 24, 63–95.

Basak, Suleyman, 2005, Asset pricing with heterogeneous beliefs, *Journal of Banking & Finance* 29, 2849–2881.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Beneish, M. D., C. M. C. Lee, and D. C. Nichols, 2015, In short supply: Short-sellers and stock returns, *Journal of Accounting and Economics* 60, 33–57.

Bhamra, Harjoat S., and Raman Uppal, 2014, Asset prices with heterogeneity in preferences and beliefs, *Review of Financial Studies* 27, 519–580.

Blocher, Jesse, Adam V. Reed, and Edward D. Van Wesep, 2013, Connecting two markets: An equilibrium framework for shorts, longs, and stock loans, *Journal of Financial Economics* 108, 302–322.

Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan, 2010, The good news in short interest, *Journal of Financial Economics* 96, 80–97.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2008, Which shorts are informed?, *Journal of Finance* 63, 491–527.

Brunnermeier, Markus K., Alp Simsek, and Wei Xiong, 2014, A welfare criterion for models with distorted beliefs, *Quarterly Journal of Economics* 129, 1753–1797.

Electronic copy available at: https://ssrn.com/abstract=3516969

Buss, Adrian, and Bernard Dumas, 2019, The dynamic properties of financial-market equilibrium with trading fees, *Journal of Finance* 74, 795–844.

Buss, Adrian, Bernard Dumas, Raman Uppal, and Grigory Vilkov, 2016, The intended and unintended consequences of financial-market regulations: A general-equilibrium analysis, *Journal of Monetary Economics* 81, 25–43.

Chabakauri, Georgy, 2015, Asset pricing with heterogeneous preferences, beliefs, and portfolio constraints, *Journal of Monetary Economics* 75, 21–34.

Chague, Fernando, Rodrigo De-Losso, Alan De Genaro, and Bruno Giovannetti, 2017, Well-connected short-sellers pay lower loan fees: A market-wide analysis, *Journal of Financial Economics* 123, 646–670.

Cohen, Lauren, Karl B. Diether, and Christopher J. Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 62, 2061–2096.

Cvitanić, Jakša, and Semyon Malamud, 2011, Price impact and portfolio impact, *Journal of Financial Economics* 100, 201–225.

Daniel, Kent, Alexander Klos, and Simon Rottke, 2018, Overconfidence, information diffusion, and mispricing persistence, Working paper, Columbia University.

David, Alexander, 2008, Heterogeneous beliefs, speculation, and the equity premium, *Journal of Finance* 63, 41–83.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Desai, Hemang, K. Ramesh, S. Ramu Thiagarajan, and Bala V. Balachandran, 2002, An investigation of the informational role of short interest in the Nasdaq market, *Journal of Finance* 57, 2263–2287.

Detemple, Jérôme B., and Shashidhar Murthy, 1994, Intertemporal asset pricing with heterogeneous beliefs, *Journal of Economic Theory* 62, 294–320.

Detemple, Jérôme B., and Shashidhar Murthy, 1997, Equilibrium asset prices and no-arbitrage with portfolio constraints, *Review of Financial Studies* 10, 1133–1174.

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner, 2009, Short-sale strategies and return predictability, *Review of Financial Studies* 22, 575–607.

Drechsler, Itamar, and Qingyi Freda Drechsler, 2016, The shorting premium and asset pricing anomalies, Working Paper, University of Pennsylvania.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics* 66, 307–339.

Dumas, Bernard, Alexander Kurshev, and Raman Uppal, 2009, Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility, *Journal of Finance* 64, 579–629.

Duong, Truong X., Zsuzsa R. Huszár, Ruth S. K. Tan, and Weina Zhang, 2017, The information value of stock lending fees: Are lenders price takers?, *Review of Finance* 21, 2353–2377.

Electronic copy available at: https://ssrn.com/abstract=3516969

Ehling, Paul, Alessandro Graniero, and Christian Heyerdahl-Larsen, 2017, Asset prices and portfolio choice with learning from experience, *Review of Economic Studies* 85, 1752–1780.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2018, Short-selling risk, *Journal of Finance* 73, 755–786.

Evgeniou, Theodoros, Julien Hugonnier, and Rodolfo Prieto, 2019, Costly short sales and nonlinear asset pricing, Working paper, INSEAD.

Figlewski, Stephen, 1981, The informational effects of restrictions on short sales: Some empirical evidence, *Journal of Financial and Quantitative Analysis* 16, 463–476.

Gallmeyer, Michael, and Burton Hollifield, 2008, An examination of heterogeneous beliefs with a short-sale constraint in a dynamic economy, *Review of Finance* 12, 323–364.

Hanson, Samuel G., and Adi Sunderam, 2014, The growth and limits of arbitrage: Evidence from short interest, *Review of Financial Studies* 27, 1238–1286.

Harrison, Michael J., and David M. Kreps, 1978, Speculative investor behavior in a stock market with heterogeneous expectations, *Quarterly Journal of Economics* 92, 323–336.

Hong, Harrison, and Jeremy C. Stein, 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies* 16, 487–525.

Johnson, Timothy C., 2004, Forecast dispersion and the cross section of expected returns, *Journal of Finance* 59, 1957–1978.

Jones, Charles M., and Owen A. Lamont, 2002, Short-sale constraints and stock returns, *Journal of Financial Economics* 66, 207–239.

Kaplan, Steven N., Tobias J. Moskowitz, and Berk A. Sensoy, 2013, The effects of stock lending on security prices: An experiment, *Journal of Finance* 68, 1891–1936.

Kogan, Leonid, Stephen A. Ross, Jiang Wang, and Mark M. Westerfield, 2006, The price impact and survival of irrational traders, *Journal of Finance* 61, 195–229.

Kolasinski, Adam C., Adam V. Reed, and Matthew C. Ringgenberg, 2013, A multiple lender approach to understanding supply and search in the equity lending market, *Journal of Finance* 68, 559–595.

Liptser, Robert S., and Albert N. Shiryaev, 2001, *Statistics of Random Processes* (Springer).

Longstaff, Francis A., and Jiang Wang, 2012, Asset pricing and the credit market, *Review of Financial Studies* 25, 3169–3215.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Nezafat, Mahdi, Mark Schroder, and Qinghai Wang, 2017, Short-sale constraints, information acquisition, and asset prices, *Journal of Economic Theory* 172, 273–312.

Nutz, Marcel, and José A. Scheinkman, 2020, Shorting in speculative markets, *Journal of*

Electronic copy available at: https://ssrn.com/abstract=3516969

Finance 75, 995–1036.

Ofek, Eli, Matthew Richardson, and Robert F. Whitelaw, 2004, Limited arbitrage and short sales restrictions: Evidence from the options markets, *Journal of Financial Economics* 74, 305–342.

Prado, Melissa Porras, 2015, Future lending income and security value, *Journal of Financial and Quantitative Analysis* 50, 869–902.

Prado, Melissa Porras, Pedro A. C. Saffi, and Jason Sturgess, 2016, Ownership structure, limits to arbitrage, and stock returns: Evidence from equity lending markets, *Review of Financial Studies* 29, 3211–3244.

Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 20, 0–1.

Reed, Adam V., 2013, Short selling, *Annual Review of Financial Economics* 5, 245–258.

Saffi, Pedro A. C., and Kari Sigurdsson, 2010, Price efficiency and short selling, *Review of Financial Studies* 24, 821–852.

Scheinkman, José A., and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1220.

Senchack, A. J., and Laura T. Starks, 1993, Short-sale restrictions and market reaction to short-interest announcements, *Journal of Financial and Quantitative Analysis* 28, 177–194.

Seneca, Joseph J., 1967, Short interest: Bearish or bullish?, *Journal of Finance* 22, 67–70.

Xiong, Wei, and Hongjun Yan, 2010, Heterogeneous expectations and bond markets, *Review of Financial Studies* 23, 1433–1466.

Yan, Hongjun, 2008, Natural selection in financial markets: Does it work?, *Management Science* 54, 1935–1950.

Yu, Jialin, 2011, Disagreement and return predictability of stock portfolios, *Journal of Financial Economics* 99, 162–183.

Zapatero, Fernando, 1998, Effects of financial innovations on market volatility when beliefs are heterogeneous, *Journal of Economic Dynamics and Control* 22, 597–626.

Electronic copy available at: https://ssrn.com/abstract=3516969