Asset pricing with costly short sales*

Julien Hugonnier[†] Rodolfo Prieto[‡]

January 10, 2024

Forthcoming: *Management Science*

Abstract

We study a dynamic general equilibrium model with costly-to-short stocks and heterogeneous beliefs. Costly short sales drive a wedge between the valuation of assets that promise identical cash flows but are subject to different trading arrangements. In particular, we show that the price of an asset is given by the risk-adjusted present value of its future cash flows which include both dividends and an endogenous *lending yield* that we characterize explicitly. This valuation formula implies that stocks with low and high shorting costs should offer similar risk-return tradeoff once returns are appropriately adjusted for lending revenues and thus sheds light on recent empirical findings about the explanatory power of shorting costs in the cross-section of returns.

Keywords: Shorting fees; Securities lending; Heterogeneous beliefs; Dynamic equilibrium.

JEL Codes: D51, D52, G11, G12.

*A preliminary version was circulated under the alternative title "Costly short sales and nonlinear asset pricing". We thank Adem Atmaz (discussant), Snehal Banerjee, Pierre Collin Dufresne, Kent Daniel, Jérôme Detemple, Itamar Dreschler (discussant), Bernard Dumas, Philip Dybvig, Sergei Glebkin, Ahmed Guecioueur, Maarten Meeuwis (discussant), Johannes Muhle-Karbe, Joël Peress, Stephen Ross, Raman Uppal, Junyuan Zou and audiences at Boston University, ESSEC, INSEAD, Universidad de los Andes, the 9th AMaMeF Conference 2019, the 2020 NAW meeting of the Econometric Society, the 2020 SFI Research Days, 2022 EUROFIDAI-ESSEC Paris December Finance Meeting, and EFA 2020 for constructive comments.

[†]EPFL and CEPR, Switzerland. E-mail: julien.hugonnier@epfl.ch

[‡]INSEAD, France. E-mail: rodolfo.prieto@insead.edu

1 Introduction

Securities lending and borrowing is a critical function that makes financial markets more efficient through improved liquidity and price discovery. In the U.S. alone, short selling accounts for more than a quarter of trading volume in the stock market and the value of securities on loans has recently surpassed \$1.4trillion (Gensler, 2021). Historically, the primary suppliers of shares to loan have mostly been investment firms, pension funds, insurance companies, passive funds, and exchange traded funds (ETFs). However, the practice of securities lending is now extending to non institutional investors because it can be a significant revenue source that often offsets management fees and transaction costs.[1] For example, Kashyap, Kovrijnykh, Li and Pavlova (2020) report that securities lending contributed 5% of the total revenues of both BlackRock and State Street in 2017 while the data provider DataLend (2022) reports that the global revenues of security lenders have been growing steadily in the last decade to reach a level in excess of \$9billion in 2021. As a last indication of the current importance of securities lending, we note that the U.S. Securities and Exchange Commission Rule 10c–1, which is currently under review, will soon create a new reporting and disclosure regime for all participants in the securities lending market (Gensler, 2021).

Despite the crucial importance of short selling and the extensive literature on the effects of short sales constraints on asset returns,[2] there are only few studies that explicitly analyze the role of securities lending in the price formation process.[3] In particular, there is currently no commonly accepted theoretical model for the joint endogenous determination of asset returns and lending fees. We contribute to bridging this gap by developing a tractable dynamic general equilibrium model of asset prices and lending fees with a focus on the *return-augmenting* effect of securities lending. Specifically, we consider a continuous-time Lucas economy populated by two groups of investors who have logarithmic utility and heterogenous dogmatic beliefs about the growth rate of the

[1]See, e.g., Table 3 in iShares Report on Securities Lending. In addition, a number of important broker-dealers have recently started lending programs that allow retail customers to earn incremental income. See, for example, the Fully Paid Lending program of Fidelity and the similar programs put in place by TDAmeritrade/Charles Schwab, BNY Mellon, and E–Trade among others.

[2]See among others Seneca (1967), Miller (1977), Harrison and Kreps (1978), Figlewski (1981), Diether, Malloy and Scherbina (2002), Jones and Lamont (2002), Mitchell, Pulvino and Stafford (2002) Scheinkman and Xiong (2003), Ofek, Richardson and Whitelaw (2004), and Atmaz and Basak (2019) for key contributions, and either Reed (2013) or Jiang, Habib and Hasan (2020) for comprehensive surveys.

[3]The short list of such studies includes Duffie (1996), D'Avolio (2002), Duffie, Gârleanu and Pedersen (2002), Cohen, Diether and Malloy (2007), Hanson and Sunderam (2014), Drechsler and Dreschler (2018), Nutz and Scheinkman (2020) and, since the release of our first draft, the work of Atmaz, Basak and Ruan (2021), Gârleanu, Panageas and Zheng (2021), and Chen, Kaniel and Opp (2022) which we briefly review below.

economy.[4] The financial market includes a riskless asset, and two long-lived risky assets that each represent a claim to a constant fraction of the aggregate dividend. The first risky asset (asset 1) can be shorted at a cost that is to be determined in equilibrium, while the second risky asset (asset 2) cannot be shorted.[5] The risky assets in our model have proportional dividends and thus are Siamese twins (see e.g., Froot and Dabora (1999) and De Jong, Rosenthal and Van Dijk (2009)). This assumption allows us to study the effect of costly short sales on ex-ante identical assets *within* a single general equilibrium model where all markets clear *rather than across* different models, while accounting for the fact that the shares of a given stock are often not all available for lending because of a variety of reasons that include the existence of different share classes (Mei, Scheinkman and Xiong, 2009), dual or cross-listings as in the case of ADRs (Blau, Van Ness and Warr, 2012), or the occurrence of an equity carve-out (Lamont and Thaler, 2003).

To establish a short position an investor must borrow the required shares from another investor who holds a long position. In line with the empirical evidence (D'Avolio, 2002, Baklanova, Copeland and McCaughrin, 2015, Gensler, 2021, Chen et al., 2022) we assume that the lending market is intermediated by lending agents. In our model, investors wanting to go short over the next instant are randomly matched with one of the lending agents. Each lending agent sets a shorting fee to maximize the flow of shorting revenues taking as given the aggregate short demand schedule of the investors who are matched with her. Once the terms of the short transactions are set, each lending agent borrows the required shares from the custodian bank that holds securities on behalf of long investors, lends them over an infinitesimal time interval, and transfers back the securities and the induced shorting fees to the custodian bank who in turn redistributes them on a value weighted basis to long investors. This approach is, to the best of our knowledge, new to the literature and allows us to easily integrate costly short sales into an otherwise standard dynamic asset pricing model. Our modelling implies that the shortable asset entitles its owners to a convenience yield that can be endogenously determined in equilibrium by matching the aggregate flows of lending fees paid and received by investors.

[4]The assumption of a constant disagreement simplifies the solution of the model by fixing the identity of the optimist and pessimist, and thereby limiting the number of state variables. Moving to a richer environment with a stochastic disagreement leads to a more cumbersome, yet fully explicit, characterization of the equilibrium but does not affect the underlying economics. We develop such an extension in Appendix B.

[5]Since we endogenize not only the asset prices but also the interest rate, the model is better suited to analyze the global securities lending market and market-wide events, such as the effect of widespread shorting during the dot-com era, rather than idiosyncratic episodes like the recent GameStop short-squeeze.

If asset 1 could be shorted at no cost, then the short sale constraint on asset 2 would have no impact and pricing would be linear in the sense that both assets would offer the same price-dividend ratio. By contrast, our analysis shows that costly short sales drive a wedge between the valuation of the two assets and thereby result in *nonlinear* pricing. In particular, the value of asset 1 in our model represents a fraction of the market portfolio that is strictly greater than its share of dividends and which varies endogenously across times and states to reflect the impact of the short selling frictions at play in the model. This contribution provides a novel general equilibrium approach behind the intuition in Cochrane (2002) and Cherkes, Jones and Spatt (2013) according to which the valuation of a shortable asset includes not only the present value of its future dividends but also the present value of future lending revenues. Perhaps of greater interest, this nonlinearity provides a rational explanation for the mispricing observed in certain famous equity carve-outs (Lamont and Thaler, 2003), such as the partial spin-off of Palm by 3Com that we use to quantitatively illustrate the implications of the model.

Our theoretical framework provides a backdrop to the recent empirical findings of Beneish, Lee and Nichols (2015) and Drechsler and Dreschler (2018) who document that stocks with high lending fees exhibit low average excess returns that cannot be explained by standard factor models such as the three- and four-factor models of Fama and French (1993) and Carhart (1997). In particular, Drechsler and Dreschler (2018) argue that these negative excess returns are compensation for the systematic risk borne by the small fraction of investors who account for most of the shorting activity. They refer to this finding as the shorting premium and construct a portfolio risk factor labeled CME (for *c*heap-*m*inus-*e*xpensive to short) that earns the corresponding abnormal return. Our model offers an alternative explanation for these findings that is aligned with the literature (e.g., Fama and French, 2010) that questions the possibility to generate abnormal returns by stock picking, and argues that the estimated alphas may result from return-augmenting activities like securities lending. Specifically, our framework implies that stock returns satisfy a modified CAPM that includes an explicit downward adjustment for lending fees and, therefore, predicts that lending fees should not have any significant impact on the cross-section of return provided that an appropriate correction is applied before running the estimation. This prediction of the model is supported by empirical evidence in our companion paper Hugonnier and Prieto (2023).

Related literature

To capture the fact that locating shares to borrow may be a time consuming process, Duffie et al. (2002) develop a model with a single stock in which search costs and bargaining over fees generate a *deterministic* price process that includes the present value of the lending fees that accrue to holders of long positions. Vayanos and Weill (2008) extend the search model of Duffie et al. (2002) to include two assets with different lending fees and show that the resulting equilibrium can help understand phenomena such as the different pricing of on-the-run and off-the-run Treasury bonds. By contrast, we study an otherwise frictionless stochastic general equilibrium model where positive loan fees arise due the presence of an intermediated securities lending market.

Our paper is naturally related to the large body of theoretical literature that studies the impact of shorting constraints in trading models where agents have heterogenous beliefs. Earlier contributions in this literature, including the seminal papers of Miller (1977) and Harrison and Kreps (1978), feature discrete-time partial equilibrium models with risk-neutral investors in which the combination of heterogenous beliefs with the impossibility of short selling gives rise to speculative episodes where the asset price exceeds its fundamental value to the most optimistic investor. Scheinkman and Xiong (2003) extend the original setting of Harrison and Kreps (1978) to continuous-time and use it to study the occurrence and properties of bubbles. Detemple and Murthy (1997) and, later, Gallmeyer and Hollifield (2008) study a similar problem but in a dynamic general equilibrium setting with risk-averse investors and show that the imposition of a short sale ban may result in either a price increase or a price decrease relative to a frictionless model. More recently, Nutz and Scheinkman (2020) study a continuous-time version of the model of Harrison and Kreps (1978) in which agents can short the asset subject to exogenous quadratic costs but these costs are dissipated and thus do not accrue to holders of long positions. Our paper advances this literature by proposing a tractable way to model the endogenous determination of securities lending fees in a dynamic general equilibrium setting where lending fees are rebated to holders of long positions.

Our focus and contributions differ markedly from Atmaz et al. (2021) who develop a CARA-Normal model with heterogeneous beliefs, two *independent* risky assets and a riskless asset with exogenous return. In their model, stock prices and shorting costs are linear functions of two Gaussian processes that represent dividends and the stochastic disagreement among agents. Therefore, prices and shorting costs can be negative and, perhaps of greater concern, the model does not rule out situations where some agents that are a priori assumed to be long only end up holding short positions without paying the corresponding cost. By contrast, prices and shorting fees are nonnegative in our framework and, rather than independent assets, we consider Siamese twins which allows us to elicit the premium associated with the possibility of shorting an asset.

Perhaps closest to us, Gârleanu et al. (2021) also propose a continuous-time general equilibrium model that features costly short sales and myopic investors with heterogenous beliefs. The main difference with our paper resides in the modelling of the shorting friction. Specifically, Gârleanu et al. (2021) take the shorting cost as an *exogenously* given function of short interest that they interpret as a cost of matching between competitive brokers and dealers. They show that in some regions of the parameter space this modelling produces multiple equilibria that are associated with different solutions for the lending yield and use this feature to analyze situations where fears among short sellers lead to runs. As we discuss in Section 4.1 below, this equilibrium multiplicity is subject to two important caveats. First, it is quite unlikely to arise in practice as it requires a very large shorting cost. Second, and perhaps more importantly, this multiplicity counterfactually implies that the interest rate and market price of risk experience predictable jumps every time the equilibrium switches from one lending yield to another.

By contrast, we construct a model where securities lending market is intermediated by lending agents with some degree of market power and show that this modelling—which is consistent with existing market conditions in the U.S—produces a unique equilibrium that is not nested among the multiple equilibria of Gârleanu et al. (2021) because in our model the *endogenous* shorting cost cannot be expressed as a deterministic function of the *endogenous* short interest. More recently, Chen, Kaniel and Opp (2022) introduce asymmetric information in a partial equilibrium version of our model to evaluate the implications of non-competitive lending fees. They quantify price *wedges* due to the incremental impact that lenders assign to stocks due to the fee income, similar to the effect we capture *endogenously* by using assets 1 and 2 in our model.

Our work is also related to the broad literature on rational models of limits to arbitrage. See Gromb and Vayanos (2010) for a survey. We highlight Basak and Croitoru (2000) who study a dynamic general equilibrium model with a risky asset and a derivative in zero net supply to show that mispricing can arise between two securities that carry the same risk, if agents are subject to portfolio constraints that prevent them from exploiting the induced arbitrage opportunity. Banerjee and Graveline (2014) obtain similar conclusions in a static CARA-Normal model with quasi-redundant assets and costly short sales. By contrast, we study the implications of costly shorting in a dynamic setting where all risky assets are in positive supply so that both expected returns and volatilities are endogenously determined.

Other contributions to the study of lending fees include Duffie (1996), Krishnamurthy (2002), and Blocher, Reed and Van Wesep (2013). More recently, Nezafat and Schroder (2022) study the role of private information in the equity lending market in a static rational expectations model with endogenous loan fees. There is also a growing literature on strategic short selling that studies the role of short selling in the transmission of information about firm fundamentals. For example, Goldstein and Guembel (2008) show that this channel can lead to negative spillovers, Goldstein, Ozdenoren and Yuan (2013) show that it may distort investment decisions, and Brunnermeier and Oehmke (2014) show that it may lead to situations where short sellers can force a vulnerable institution to liquidate assets at fire-sale prices. In the same vein, Brunnermeier and Pedersen (2005) and Carlin, Lobo and Viswanathan (2007) develop predatory trading models where short sellers exploit undercapitalized arbitrageurs.

The remainder of the paper is organized as follows. The model is presented in Section 2. Section 3 provides a detailed account of the equilibrium construction. Section 4 discusses the endogenous determination of the equilibrium shorting costs and their properties in the one- and two-risky assets cases. Section 5 concludes with empirical implications for the cross section of returns. The proofs of all results are provided in Appendix A and an extension of our benchmark model to the case of stochastic disagreement is found in Appendix B

2 The model

2.1 Fundamental uncertainty

We consider a continuous-time economy on an infinite horizon. There is a single non storable good available for consumption at every date $t \geq 0$ and we assume that its supply e_t evolves according to

$$\frac{de_t}{e_t} = \mu dt + \sigma dZ_t,$$

for some exogenously given constants μ and $\sigma > 0$, where the process $(Z_t)_{t\geq 0}$ is a Brownian motion under some reference probability.

2.2 Agents

The economy is populated by two agents that we index by $a \in \{o, p\}$. Agents observe aggregate consumption as well as the prices of traded assets, but do not observe the increments of the Brownian motion and disagree about their perception of the dynamics of the aggregate consumption process. Specifically, we assume that from the point of view of agent a, this process evolves according to

$$\frac{de_t}{e_t} = \mu^{(a)} dt + \sigma dZ_t^{(a)}$$

for some constant $\mu^{(a)}$, where the process $(Z_t^{(a)})_{t\geq 0}$ is a Brownian motion under the subjective probability of agent a. We denote by

$$\Delta = \frac{\delta}{\sigma} \equiv \frac{\mu^{(o)} - \mu^{(p)}}{\sigma}$$

the disagreement per unit of volatility and assume that $\Delta \geq 0$ so that agent o can be interpreted as being an *o*ptimist and agent p as being a *p*essimist. The assumption of a constant disagreement simplifies the solution of the model by fixing the identity of the optimist and pessimist, and thereby limiting the number of state variables. Moving to a richer model with a stochastic disagreement process leads to a more cumbersome, yet fully explicit, characterization of the equilibrium but does not affect the underlying economics, see Appendix B for such an extension.

Finally, we assume that conditional on their beliefs, the two agents have homogenous logarithmic preferences given by

$$E_0^{(a)} \left[\int_0^\infty e^{-\rho t} \log c_t dt \right]$$

for some constant discount rate $\rho > 0$, where $E_t^{(a)}[\cdot]$ denotes an expectation under the agent's subjective probability measure conditional on the observation of the paths of dividends and market prices, up to date $t \geq 0$.

As is well-known, this specification implies that agents have marginal propensity to consume equal to ρ and choose their portfolio to optimize an instantaneous quadratic criterion (see (8) below). In the context of our model, this myopic behavior also implies that–up to a slight reinterpretation and the addition of a linear term in the drift of the endogenous state variable–all the asset pricing results we derive below remain unchanged if instead of two agents with constant beliefs we consider a steady state population of agents subject to idiosyncratic shocks that shift their perceived growth rate back and forth between a low and a high value.

Investors are initially endowed with long portfolios of the two risky assets. Specifically, we assume that initial endowments are represented by vectors $n^{(o)}, n^{(p)} \in [0,1]^2$ with $n^{(o)} = \mathbf{1} - n^{(p)}$ that give the number of units of each asset in the initial portfolio and determine the initial wealth $W_0^{(a)} \equiv \sum_i n_{ai} S_{i0}$ of both investors.

2.3 Traded assets

The financial market consists of three long-lived assets: A locally riskless asset in zero net supply and two risky securities in positive supply of one unit each.

The price of the riskless asset evolves according to

$$dS_{0t} = r_t S_{0t} dt$$

for some interest rate process r_t that is to be determined in equilibrium. On the other hand, we assume that risky asset $i \in \{1,2\}$ is a claim to a fraction $\eta_i \geq 0$ of aggregate consumption, and that its price evolves according to

$$dS_{it} + \eta_i e_t dt = r_t S_{it} dt + S_{it} \sigma_{it} \left(dZ_t^{(a)} + \theta_{it}^{(a)} dt \right), \tag{1}$$

where the volatility coefficient σ_{it} and the perceived market prices of risk

$$\theta_{it}^{(o)} = \Delta + \theta_{it}^{(p)}$$

are to be determined endogenously in equilibrium. To ensure that the market portfolio $M_t \equiv S_{1t} + S_{2t}$ is a claim to the whole aggregate consumption, we naturally assume that the fractions paid by the risky assets are such that $\eta_1 + \eta_2 = 1$.

The risky assets in our model have proportional dividends and thus are Siamese twins[6] (see, e.g., Froot and Dabora (1999) and De Jong et al. (2009)). This assumption allows us to study the effect of costly short sales on ex-ante identical assets *within* a single general equilibrium model where all markets clear *rather than across* different models, while accounting for the fact that the shares of a given stock are often not all available for lending because of a variety of reasons that include the existence of different share

[6]This setup will admit closed form expressions for all economic quantities. A more general model including multiple risky assets with multiple sources of risk is amenable to the same intermediated securities lending market analysis but a full analytical solution of the model is no longer available.

classes (Mei et al., 2009), dual or cross-listings as in the case of ADRs (Blau et al., 2012), or the occurrence of an equity carve-out (Lamont and Thaler, 2003).

2.4 Shorting frictions

Our point of departure from existing equilibrium models with heterogenous beliefs is that the risky assets are subject to different trading arrangements. Specifically, we assume that shares of asset 2 *cannot* be shorted whereas shares of asset 1 can be shorted by incurring a flow cost per dollar of short as long as the position is maintained.

To sell short, one must first borrow the required shares. In line with the evidence reported by D'Avolio (2002), Baklanova et al. (2015), Gensler (2021), and Chen et al. (2022) among others, we assume that the lending market is intermediated by a number $n \leq \infty$ of ex-ante identical lending agents and that securities are held for investors by a custodian bank. At time $t \geq 0$ an investor who wishes to short over the next instant is randomly matched to one of the lending agents. Each lending agent i sets a shorting fee Φ_{it} to maximize her flow of shorting revenues taking as given the aggregate short demand schedule of the investors who are matched with her at time $t \geq 0$. Once the terms of the short transactions are set, each lending agent borrows the required shares from the custodian bank, lends them over an infinitesimal time interval to the investors matched with her, and transfers the induced shorting fees to the custodian bank who in turn redistributes them on a value weighted basis to holders of long positions. See Figure 1 for an illustration of this mechanism in a model with a single intermediary.

We assume that, at any point in time, a given individual investor can only be matched with a single lending agent. As a result, each lending agent enjoys some degree of market power over the group of investors who are matched with her at a given point in time. To simplify the presentation, we focus throughout on the case where intermediaries are benevolent and enjoy full bargaining power over investors in the determination of the shorting fee. As discussed in Remark 1 below, the model is easily extended to the case where intermediaries do not enjoy full bargaining over investors. Allowing the lending agents to retain a fraction of lending revenues may have different effects on the model depending on what lending agents actually do with these revenues. If these revenues evaporate then one would need to adjust the short market clearing condition (5) to account for the fact that only a fraction of the fees is returned to long agents. This change would percolate through the model but the solution method and the qualitative insights would remain the same. If instead the lending agents redistribute the retained fraction to its shareholders, then different cases may arise depending on the identity of

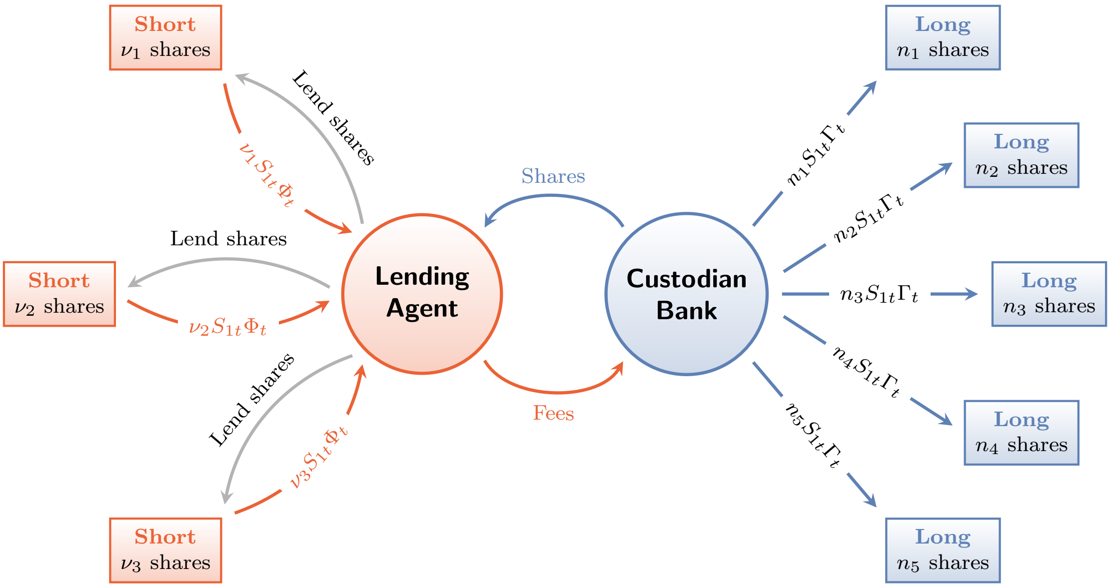

Figure 1: The shorting mechanism. In this example with a single lending agent, the short interest in asset 1 amounts to $\nu = \sum_i \nu_i \geq 0$ shares while the aggregate long position sums up to $\sum_k n_k = 1 + \nu$ shares. As a result, the equilibrium lending yield is $\Gamma_t = \frac{\nu}{1+\nu}\Phi_t$.

these shareholders. If the lending agent is held by optimistic agents and the identities of these agents remain fixed over time then the equilibrium will remain the same. On the other hand, if the intermediary is held by a, possibly time-varying, mix of pessimists and optimists then the dividends of the intermediary would constitute an additional source of income for all investors. This would break the homogeneity of the investors' portfolio and consumption problems and, thereby, make the model untractable.

The fact that at least part of the shorting fees eventually accrue to investors who are long induces a form of interaction between investors. For this interaction to remain competitive, investors have to take as given not only the costs incurred when taking short positions but also the fees that they may receive when they hold shares of the shortable asset. We model this feature by assuming that agents take as given the flow cost of shorting as well as the flow rate of lending fees Γ_t that each dollar invested in asset 1 generates, and determine this rate endogenously in equilibrium by requiring that the flow of lending fees received by long agents equals the flow of costs paid by short agents. To make a clear distinction between the flows paid by short agents and those received or anticipated by long agents, we refer to Φ_{it} as the shorting *cost* charged by lending agent i and to Γ_t as the lending *yield*.

In our model, investors of type o are more optimistic than investors of type p and both have logarithmic utility. Therefore, we know that investors of type p are short whenever investors of type o are and, since all investors cannot be short simultaneously, it follows that shorting activity can only come from the pessimists in equilibrium. Furthermore, the assumption of logarithmic utility implies that the short demand schedule of any investor is proportional to her wealth. Combining these observations shows that, up to a multiplicative factor, all intermediaries face the same optimization problem in equilibrium and it follows that they will all select the same shorting fee $\Phi_{it} \equiv \Phi_t$ at every point in time. In particular, we may assume without loss of generality that the lending market is intermediated by a single agent and thus set $n \equiv 1$ from now on.

2.5 Definition of equilibrium

Combining the above shorting mechanism with the usual self-financing condition shows that the wealth of agent a evolves according to

$$dW_t^{(a)} = \left(r_t W_t^{(a)} - c_t + \Lambda\left(\pi_{1t}; \Phi_t, \Gamma_t\right)\right) dt + \sum_i \pi_{it}\sigma_{it}\left(dZ_t^{(a)} + \theta_{it}^{(a)} dt\right), \tag{2}$$

where $c_t \geq 0$ represents her consumption rate, $\pi_t \in \mathbb{R} \times \mathbb{R}_+$ denote the amounts she invests in the risky assets, and the nonlinear term

$$\Lambda(m; \Phi_t, \Gamma_t) \equiv m^+ \Gamma_t - m^- \Phi_t \tag{3}$$

captures the flow rate of lending fees that she pays or receives.

As is standard, we require agents to maintain strictly positive wealth at all times. Therefore, the optimization problem solved by agent a is

$$\sup_{(c,\pi)} E^{(a)}\left[\int_0^\infty e^{-\rho t} \log c_t dt\right] \text{ subject to (2), (3), and } W_t^{(a)} > 0.$$

Whenever they exist, we denote by $(c_t^{(a)}(\Phi, \Gamma), \pi_t^{(a)}(\Phi, \Gamma))$ the optimal consumption and optimal portfolio of agent a, taking as given the traded asset prices and the pair of flow rates (Φ, Γ) that characterize the short market.

Definition 1. *An equilibrium is a price process* (S_{0t}, S_{1t}, S_{2t})*, a shorting cost* Φ_t*, and a lending yield* Γ_t *such that*

$$\Phi_t \in \underset{x}{\operatorname{argmax}} \left\{ \sum_a \pi_{1t}^{(a)}(x, \Gamma)^- x \right\}, \tag{4}$$

and all markets clear:

$$\begin{aligned} &\sum_a c_t^{(a)}(\Phi, \Gamma) = e_t, && \text{(Consumption)}, \\ &\sum_a \pi_{it}^{(a)}(\Phi, \Gamma) = S_{it}, && \text{(Risky asset } i \in \{1, 2\}), \\ &\sum_a \Lambda\left(\pi_{1t}^{(a)}(\Phi, \Gamma); \Phi_t, \Gamma_t\right) = 0, && \text{(Lending market)}, \end{aligned} \tag{5}$$

where the mapping Λ *is defined in* (3).

The above definition is similar to the classical definition of an equilibrium by Radner (1972) but includes two additional conditions to accommodate the presence of costly short sales in a dynamic general equilibrium setting.

The first condition (4) is an optimality condition that results from our modelling of the security lending market and which requires that the shorting cost maximizes shorting revenues taking as given market prices, the lending yield anticipated by agents, and their short demand schedules $(a, x) \mapsto \pi_t^{(a)}(x, \Gamma)^-$. The second condition (5) requires that the flow of fees received by holders of long positions must equal the flow of lending fees paid by short agents. We treat this condition as a market clearing condition as it matches the flows exchanged between agents, but one could equally view this requirement as a rational expectations condition which ensures that the lending fees anticipated by agents are indeed realized along the equilibrium path.

3 Equilibrium

In this section, we sequentially build an equilibrium for our economy with shorting costs. To facilitate the analysis, we focus throughout on the characterization of an equilibrium in which the *asset volatilities are strictly positive* at all times.

3.1 Individual optimality

The absence of arbitrage requires the net Sharpe ratios offered by the different assets to be such that it is not possible to generate a locally riskless return by combining assets. In our model, this constraint can be intuitively expressed as

$$\max\left\{\theta_{2t}^{(a)}, \theta_{1t}^{(a)} + \gamma_t\right\} - \left(\theta_{1t}^{(a)} + \phi_t\right) \leq 0, \tag{6}$$

where $\gamma_t \equiv \Gamma_t/\sigma_{1t}$ and $\phi_t \equiv \Phi_t/\sigma_{1t}$ denote the lending yield and the shorting cost per unit of volatility. The interpretation of this inequality is clear. Indeed, it simply requires that the largest expected excess return that can be achieved by going in either risky assets (first term) and short in asset 1 (second term) is negative. In addition to this no-arbitrage requirement, in equilibrium we must have that

$$\theta_{2t}^{(a)} = \theta_{1t}^{(a)} + \gamma_t, \tag{7}$$

for otherwise one of the assets would dominate the other and markets would not clear. This equality shows that, in our model, an agent wanting to go long is indifferent between the risky assets once fees are taken into account and will lead to some indeterminacy in the characterization of optimal portfolios: see Proposition 1 below. Importantly, given (7) the no-arbitrage condition (6) boils down to $\gamma_t \leq \phi_t$ which simply requires that borrowing asset 1 to hold it does not generate riskless profits.

The assumption of logarithmic utility implies that, under the above conditions, the optimal consumption rate of agent a is given by

$$c_t^{(a)} = \rho W_t^{(a)}$$

and that the fractions of her wealth $p_{it}^{(a)} = \pi_{it}^{(a)}/W_t^{(a)}$ that she optimally invests in the risky assets solve the mean-variance problem

$$\max_{p\in\mathbb{R}\times\mathbb{R}_+}\left\{\Lambda\left(p_1;\Phi_t,\Gamma_t\right) + \sum_i p_i\sigma_{it}\theta_{it}^{(a)} - \frac{1}{2}\left(p_1\sigma_{1t} + p_2\sigma_{2t}\right)^2\right\}. \tag{8}$$

The following proposition derives the solution to this problem and summarizes the optimal trading behavior of agents, taking as given market prices and the rates (Φ, Γ) that characterize the short market.

Proposition 1. *Assume that condition (7) holds and let $\phi_t \geq \gamma_t$. Then the optimal portfolio of agent a satisfies*

$$p_{1t}^{(a)}(\Phi,\Gamma)\sigma_{1t} = \mathbb{1}_{\left\{\theta_{2t}^{(a)}\geq 0\right\}} x_t + \mathbb{1}_{\left\{\theta_{1t}^{(a)}+\phi_t\leq 0\right\}}\left(\theta_{1t}^{(a)}+\phi_t\right), \tag{9a}$$

$$p_{2t}^{(a)}(\Phi,\Gamma)\sigma_{2t} = \mathbb{1}_{\left\{\theta_{2t}^{(a)}\geq 0\right\}}\left(\theta_{2t}^{(a)}-x_t\right), \tag{9b}$$

where x_t is an arbitrary process such that $0 \leq x_t \leq \theta_{2t}^{(a)}$.

The optimal trading strategy in Proposition 1 admits an intuitive interpretation. If the net Sharpe ratio $\theta_{2t}^{(a)} = \gamma_t + \theta_{1t}^{(a)}$ that agent a associates with long positions in the risky assets is positive, then agent a naturally goes long at the optimum but there is a degree of freedom in the determination of her optimal portfolio because any $p \in \mathbb{R}_+^2$ that delivers the efficient risk exposure

$$\sum_i p_i\sigma_{it} = \operatorname*{argmax}_{x\in\mathbb{R}}\left\{x\theta_{2t}^{(a)} - \frac{1}{2}x^2\right\} = \theta_{2t}^{(a)}$$

is optimal. On the other hand, if $\theta_{2t}^{(a)} \leq 0$ then the agent clearly does not want to go long in either risky asset. Whether she goes short in asset 1 depends on the sign of the Sharpe ratio $-(\theta_{1t}^{(a)} + \phi_t)$ that she associates with a short position in asset 1. If this quantity is positive, then she shorts asset 1 to achieve the efficient risk exposure

$$p_{1t}\sigma_{1t} = \operatorname*{argmax}_{x\in\mathbb{R}}\left\{x\left(\theta_{1t}^{(a)}+\phi_t\right) - \frac{1}{2}x^2\right\} = \theta_{1t}^{(a)} + \phi_t,$$

and otherwise she invests only in the riskless asset.

3.2 Equilibrium shorting cost

Proposition 1 and the discussion preceding it show that the total flow of lending fees induced by a shorting cost process Φ_t is well-defined only if $\phi_t = \Phi_t/\sigma_{1t} \geq \gamma_t$ in which case it is explicitly given by

$$\sum_a \pi_t^{(a)}(\Phi,\Gamma)^-\Phi_t = \sum_a \phi_t\left(\theta_{1t}^{(a)}+\phi_t\right)^- W_t^{(a)}. \tag{10}$$

In our model, agent o is more optimistic than agent p and both have logarithmic utility. Therefore, we know that agent p is short whenever agent o is short and, because agents cannot be short simultaneously in equilibrium, it follows that any shorting activity must come from the pessimist alone. In particular, we must have

$$\theta_{2t}^{(o)} = \theta_{1t}^{(o)} + \gamma_t > 0, \tag{11}$$

so that the optimist is long at all times. In combination with (7), this implies that the Sharpe ratios perceived by the optimist are such that

$$\forall y \geq \gamma_t : \left(\theta_{1t}^{(o)} + y\right)^- = \left(\theta_{2t}^{(o)} - \gamma_t + y\right)^- = 0.$$

As a result, the sum in (10) reduces to the contribution of the pessimist and it follows that the equilibrium shorting cost satisfies

$$\phi_t \in \underset{y \geq \gamma_t}{\operatorname{argmax}} \left\{ y \left(\theta_{1t}^{(p)} + y\right)^- \right\} = \max \left\{ \gamma_t, -\frac{1}{2}\theta_{1t}^{(p)} \right\} + \mathbb{1}_{\left\{\theta_{1t}^{(p)} \geq 0\right\}} \mathbb{R}_+. \tag{12}$$

The above expression shows that when $\theta_{1t}^{(p)} \geq 0$, the shorting cost is undetermined because in such states neither agent wants to go short irrespective of the cost set by the lending agent. To facilitate the presentation, we from now select the smallest element of the above set of maximizers, i.e., we let

$$\phi_t = \max \left\{ \gamma_t, -\frac{1}{2}\theta_{1t}^{(p)} \right\}. \tag{13}$$

This selection is without loss of generality and simply amounts to setting the flow cost to zero on the endogenous set of states $\mathcal{L} \equiv \{(\omega, t) : \gamma_t = 0\}$, where the shorting market is *inactive* in equilibrium.

Substituting the optimal portfolios of Proposition 1 into the market clearing conditions shows that the equilibrium lending yield and shorting cost are related by

$$\gamma_t \sigma_{1t} S_{1t} = (\phi_t - \gamma_t) \left(\theta_{1t}^{(p)} + \phi_t\right)^- W_t^{(p)}. \tag{14}$$

Since the diffusion of asset 1 and the wealth of the pessimist are both strictly positive in equilibrium, this identity combined with (7) and (13) implies that (see the Appendix A for a detailed argument)

$$\mathcal{L} = \left\{ (\omega, t) : \theta_{2t}^{(p)} \geq 0 \right\}, \tag{15}$$

and substituting back into (13) shows that

$$\phi_t = -\frac{1}{2}\theta_{1t}^{(p)}\mathbb{1}_{\{\mathcal{S}\}}, \tag{16}$$

where

$$\mathcal{S} \equiv (\Omega \times \mathbb{R}_+) \setminus \mathcal{L} = \left\{(\omega, t) : \theta_{2t}^{(p)} < 0\right\} \tag{17}$$

denotes the set of states where the shorting market is *active*. These expressions provide key information about the equilibrium properties of the shorting market. Specifically, the equivalent set identities (15) and (17) show that the market is active in equilibrium exactly in states where $\theta_{2t}^{(p)} = \theta_{1t}^{(p)} + \gamma_t < 0$ so that the pessimist perceives a *strictly negative net* Sharpe ratio on long positions in either risky asset. Equation (16) shows that, in those states, the lending agent sets the flow cost ϕ_t to one half of the Sharpe ratio $-\theta_{1t}^{(p)}$ that the pessimist could have obtained *absent frictions*. In response, the pessimist scales down her short demand by a half relative to the frictionless case and, as a result, the flow of lending fees received by the optimist, i.e.

$$\left(\pi_{1t}^{(p)}\right)^- \Phi_t = \frac{1}{4}\left(\theta_{1t}^{(p)}\right)^2 W_t^{(p)}\mathbb{1}_{\{\mathcal{S}\}},$$

amounts to a fourth of her optimal frictionless excess return.

Remark 1 (Nash bargaining)**.** Since the lending agent upholds only the interests of asset owners, this fraction constitute an upper bound on the share that alternative price setting mechanisms may attribute to the optimist. In particular, if the cost was determined by Nash bargaining between the lending agent and the pessimist(s) then (16) would be replaced by $-\frac{b}{2}\theta_{1t}^{(p)}\mathbb{1}_{\{\mathcal{S}\}}$, where $b \in [0, 1]$ represents the bargaining power of the lending agent. In response, the pessimist would now scale down her optimal short demand by a factor $1 - \frac{b}{2}$ and, as a result, the optimist would capture a fraction $\frac{b}{2}(1 - \frac{b}{2})$ of her optimal frictionless excess return. We focus throughout the paper on the polar case where $b = 1$ because it leads to simpler algebraic expressions for equilibrium outcomes, but the structure of the equilibrium and the qualitative properties of the model, including the characterization of the trading regions, remain the same when $b < 1$.

3.3 State variables and trading regions

Combining the equilibrium restriction (7) with the expression of the equilibrium shorting cost in (12) shows that

$$\left\{(\omega,t):\theta^{(p)}_{2t}<0\text{ and }\theta^{(p)}_{1t}+\phi_t\geq 0\right\}=\emptyset, \tag{18}$$

see Appendix A for details. Therefore, it follows from Proposition 1 that the pessimist is strictly long in either or both risky assets in the interior of the set $\mathcal{L}$ where $\theta^{(p)}_{2t}>0$, is fully invested in the riskless asset on the boundary where

$$\partial\mathcal{L}\equiv\left\{(\omega,t):\theta^{(p)}_{2t}=0\right\},$$

and is strictly short in asset 1 on the set $\mathcal{S}$ where $\theta^{(p)}_{2t}<0$. As we now show, the scale invariance of logarithmic preferences allows us to characterize these sets and the resulting pricing of risk/time in terms of a single endogenous state variable

$$s_t\equiv c^{(o)}_t/e_t\in[0,1]$$

that tracks the consumption share of the optimist. To construct the equilibrium evolution of this state variable, we take as reference the subjective probability of the optimist. This choice is without loss in generality.

Since the marginal propensity to consume of both agents is ρ, it follows from market clearing that the price of the market portfolio is

$$M_t=\sum_i S_{it}=\sum_a W^{(a)}_t=\frac{1}{\rho}\sum_a c^{(a)}_t=\frac{e_t}{\rho},$$

and that the endogenous state variable can be expressed as

$$s_t=\frac{W^{(o)}_t}{M_t}=1-\frac{W^{(p)}_t}{M_t}. \tag{19}$$

On the other hand, combining (7), (11), and (14) with the results of Proposition 1 shows that the wealth of the agents evolve according to

$$\frac{dW^{(o)}_t}{W^{(o)}_t}=(r_t-\rho)\,dt+\theta^{(o)}_{2t}\left(dZ^{(o)}_t+\theta^{(o)}_{2t}dt\right),$$

and

$$\frac{dW_t^{(p)}}{W_t^{(p)}} = (r_t - \rho)\, dt + \mathbb{1}_{\{\mathcal{L}\}} \left(\theta_{2t}^{(o)} - \Delta\right) \left(dZ_t^{(o)} + \theta_{2t}^{(o)} dt\right) + \mathbb{1}_{\{\mathcal{S}\}} \left(\theta_{2t}^{(o)} - \Delta - \gamma_t + \phi_t\right) \left(dZ_t^{(o)} + \left(\theta_{2t}^{(o)} - \gamma_t + \phi_t\right) dt\right),$$

subject to (16) and

$$\gamma_t \sigma_{1t} S_{1t} = \mathbb{1}_{\{\mathcal{S}\}}(1 - s_t)\,(\gamma_t - \phi_t) \left(\theta_{2t}^{(o)} - \Delta - \gamma_t + \phi_t\right) M_t. \tag{20}$$

Next, applying Itô's lemma to the second equality in (19) and matching terms, pins down the equilibrium interest rate r_t and the equilibrium market price of risk $\theta_{2t}^{(o)}$ as functions of the state variable and the lending yield:

$$\theta_{2t}^{(o)} = \theta^*(s_t) \, - \mathbb{1}_{\{\mathcal{S}\}} \frac{(1 - s_t)\,(s_t \Delta - \sigma - \gamma_t)}{1 + s_t} \tag{21a}$$

and

$$r_t = r^*(s_t) \, + \mathbb{1}_{\{\mathcal{S}\}} \frac{s_t(1 - s_t)\,(\sigma + \gamma_t + \Delta)\,(s_t \Delta - \sigma - \gamma_t)}{(1 + s_t)^2}, \tag{21b}$$

where

$$\theta^*(s_t) \equiv \sigma + (1 - s_t)\,\Delta,$$
$$r^*(s_t) \equiv \rho - \sigma^2 + \mu^{(o)} s_t + \mu^{(p)}(1 - s_t),$$

denote the market price of risk and the interest rate that would prevail in an otherwise identical economy where either or both of the risky assets can be freely shorted (see e.g., Detemple and Murthy (1997)).

On the shorting region, we have that $\phi_t = -\frac{1}{2}\theta_{1t}^{(p)} > \gamma_t$ from (16) and combining this inequality with (7) and (21a) shows that

$$\sigma + \gamma_t < s_t \Delta.$$

Therefore, it follows from (21) that the presence of costly short selling increases the interest rate and decreases the market price of risk relative to the frictionless case. To understand this result, observe that costly short sales trigger a reduction in the short demand of the pessimist which in turn implies that the optimist's equilibrium portfolio does not have to be as leveraged as in the frictionless case and, thus, explains the changes in the interest rate and market price of risk. The fact that costly short sales result in an increased riskless rate stands in sharp contrast to the effect of other types of frictions such as limited participation (Basak and Cuoco, 1998), borrowing constraints (Kogan, Makarov and Uppal, 2007), and liquidity constraints (Detemple and Serrat, 2003) which all result in a lower interest rate. To understand this difference note that the latter frictions require investors *not* subject to the friction to take highly leveraged positions that they would not have chosen in a frictionless setting. On the contrary, costly short sales imply that pessimistic investors short less than they would have absent costs. This limits how levered optimistic investors have to be to clear market and ultimately results in a higher interest rate. A similar but stronger effect also occurs when short sale are prohibited (Gallmeyer and Hollifield, 2008).

Substituting (17) into (21a) shows that the net Sharpe ratio perceived by the pessimist on long risky asset positions satisfies

$$\theta_{2t}^{(p)} = \mathbb{1}_{\left\{\theta_{2t}^{(p)} \geq 0\right\}} \left(\sigma - s_t \Delta\right) + \mathbb{1}_{\left\{\theta_{2t}^{(p)} < 0\right\}} \frac{\gamma_t \left(1 - s_t\right) + 2\left(\sigma - s_t \Delta\right)}{1 + s_t}. \tag{22}$$

This implies that $\mathcal{L}$ of (15) is contained in the set of states where the consumption share of the optimist lies below the threshold

$$s^* = s^*(\Delta) \equiv \min\left\{1, \frac{\sigma}{\Delta}\right\}$$

and, because the second term on the right hand side of (22) is nonnegative at $s_t = s^*$, we conclude that the trading regions are explicitly given by

$$\mathcal{L} = \left\{(\omega, t) : \theta_{2t}^{(p)} \geq 0\right\} = \left\{(\omega, t) : 0 < s_t \leq s^*\right\}, \tag{23a}$$

$$\mathcal{S} = (\Omega \times \mathbb{R}_+) \setminus \mathcal{L} = \left\{(\omega, t) : \theta_{2t}^{(p)} < 0\right\} = \left\{(\omega, t) : s^* < s_t < 1\right\}. \tag{23b}$$

These expressions show that shorting occurs only in states where the optimists represent a large enough share of the economy. This is intuitive. Indeed, when optimists are very few, prices mostly reflect the opinion of the pessimists and shorting is not necessary. On the contrary, when a large fraction of agents are optimists, equilibrium prices reflect more closely the opinion of the optimists and shorting becomes necessary for the pessimists to express their perception of the risky assets as being overpriced.

As illustrated in Figure 2, the characterization of the trading regions in (23) shows that two mutually exclusive types of equilibria may arise in our model. The first type occurs if

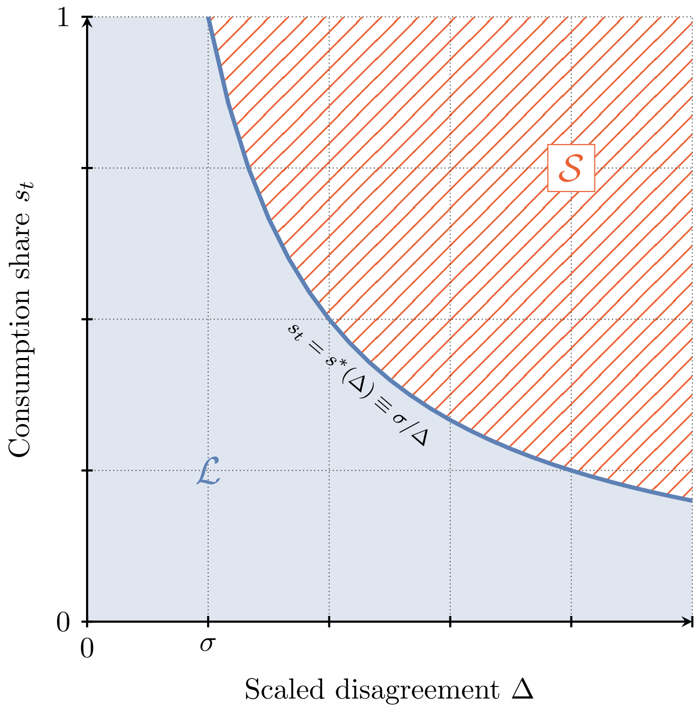

Figure 2: Equilibrium trading regions. The figure illustrates the shape of the equilibrium trading regions and allows us to determine the equilibrium configuration that occurs for each level of disagreement among agents.

the disagreement among agents is so small that $\Delta \leq \sigma$. In that case, both agents are long in the risky asset throughout the state space, and the existence of the shorting market is irrelevant so that the equilibrium is the same as in an otherwise identical frictionless economy with heterogenous beliefs (see Detemple and Murthy (1997)). The second type of equilibrium occurs when the disagreement among agents is such that $\Delta > \sigma$. In that case, the equilibrium includes two non empty trading regions: up to the locus of points $s_t = \sigma/\Delta$, both agents are long and the shorting market is inactive, while strictly above that locus agent 2 holds a short position in asset 1 and the shorting costs she incurs generate a strictly positive flow of lending revenues for the optimist.

Applying Itô's lemma to both sides of the first equality in (19) and matching terms finally shows that the state variable evolves according to

$$\frac{ds_t}{s_t(1-s_t)} = m\left(s_t, \gamma_t\right) dt + v\left(s_t, \gamma_t\right) dZ_t^{(o)},$$

with the functions defined by

$$v\left(s_t,\gamma_t\right)=v\left(s_t,\gamma_t;\Delta\right)\equiv\Delta-\mathbb{1}_{\{s_t>s^*(\Delta)\}}\frac{s_t\Delta-(\gamma_t+\sigma)}{1+s_t}, \tag{24a}$$

$$\begin{aligned}m\left(s_t,\gamma_t\right)=m\left(s_t,\gamma;\Delta\right)\equiv&\left(1-s_t\right)v\left(s_t,\gamma_t\right)^2\\&+\mathbb{1}_{\{s_t>s^*(\Delta)\}}\frac{(s_t\Delta-(\gamma_t+\sigma))((\gamma_t+\Delta)s_t-\sigma)}{(1+s_t)^2}.\end{aligned} \tag{24b}$$

Importantly, the drift and the diffusion of the endogenous state variable are equal to zero at both $s_t=0$ and $s_t=1$. This implies that 0 and 1 are absorbing boundaries for the consumption share process of the optimist, and will allow us to easily derive boundary conditions for equilibrium prices in the next section.

3.4 Price representation

Having characterized the instantaneous pricing of risk and time, we now turn to the pricing of long lived assets. To this end, let

$$\xi^{(o)}_{t,u}\equiv\frac{e^{\rho t}c^{(o)}_t}{e^{\rho u}c^{(o)}_u}=e^{-\rho(u-t)}\frac{s_te_t}{s_ue_u}$$

denote the normalized marginal utility of the reference agent.

Proposition 2. *In equilibrium*

$$S_{1t}=E^{(o)}_t\left[\int_t^\infty\xi^{(o)}_{t,u}\left(e_{1u}+S_{1u}\Gamma_u\right)du\right], \tag{25}$$

$$S_{2t}=E^{(o)}_t\left[\int_t^\infty\xi^{(o)}_{t,u}e_{2u}du\right], \tag{26}$$

where $e_{it}=\eta_ie_t$ denotes the dividend rate of asset $i=\{1,2\}$.

The above proposition shows that there are no rational bubbles in our model. Indeed, the equilibrium prices of the two assets are given by the risk-adjusted present value of the cash flows that they deliver to holders of long positions. The novelty is that, in our model, the cash flows of risky asset 1 include an *endogenous* component $S_{1t}\Gamma_t$ that accounts for the lending fees generated by each share of the asset along the equilibrium path.

This endogenous cash flow component is strictly positive over a set of positive measure if and only if the disagreement $\Delta>\sigma$ so that the shorting region is non empty. In that case, the equilibrium price-dividend ratio of asset 2

$$\mathrm{PD}_{2t} \equiv \frac{S_{2t}}{e_{2t}} = E_t^{(o)} \left[\int_t^{\infty} e^{-\rho(u-t)} \left(\frac{s_t}{s_u} \right) du \right]$$

is strictly lower than that of asset 1

$$\mathrm{PD}_{1t} \equiv \frac{S_{1t}}{e_{1t}} = E_t^{(o)} \left[\int_t^{\infty} e^{-\rho(u-t)} \frac{s_t}{s_u} \left(1 + \mathrm{PD}_{1u}\Gamma_u\right) du \right],$$

and the premium

$$\mathrm{PD}_{1t} - \mathrm{PD}_{2t} = E_t^{(o)} \left[\int_t^{\infty} e^{-\rho(u-t)} \frac{s_t}{s_u} \mathrm{PD}_{1u}\Gamma_u du \right] = \frac{1}{\eta_1} \left(\frac{1}{\rho} - \mathrm{PD}_{2t} \right) > 0, \tag{27}$$

where the second equality follows from the fact that the PD ratio of the market is $1/\rho$, gives the risk-adjusted present value of the stream of holding benefits that accrue to owners of asset 1 in the form of lending fees.

The above inequality implies that, in the presence of costly short sales, the equilibrium pricing rule is *nonlinear* as the risky assets have different price-dividend ratios despite the fact that they are Siamese twins. Specifically, since $\mathrm{PD}_{2t} < 1/\rho$ from (27), we have that the share of asset 2 in the market portfolio

$$\frac{S_{2t}}{M_t} = \rho\eta_2 \mathrm{PD}_{2t} < \eta_2$$

is strictly lower than the share of aggregate dividends that it pays out, while the share of asset 1 in the market portfolio

$$\frac{S_{1t}}{M_t} = \rho\eta_1 \mathrm{PD}_{1t} = 1 - \rho\eta_2 \mathrm{PD}_{2t} > \eta_1$$

exceeds its share of dividends. This nonlinearity is entirely driven by the presence of costly short sales and provides a rational explanation for the apparent mispricing observed in the period following certain corporate restructurings.

For example, Lamont and Thaler (2003) report that after the spin-off by 3Com of 5% of its subsidiary Palm, the extrapolation of the value of the traded Palm shares resulted in an implied valuation that exceeded the market capitalization of the subsidiary 3Com. The key to understand this phenomenon is the observation that at the time of this apparent mispricing, the costs associated with shorting Palm were very high because only the 5% of freely traded Palm shares could be lent to investors wanting to establish a short position. In our model, the $\eta_1 = 5\%$ of freely traded Palm shares are akin to asset 1 so that their price in (25) should include a sizable lending fee component in (27), while the remaining Palm shares held by 3Com are part of asset 2 whose equilibrium price in (26) only reflects the present value of future dividends. We quantitatively illustrate this feature of the model in Section 4.2.

Remark 2. The strict inequality (27) holds irrespective of whether the shorting market is currently active or not. Indeed, we show in Appendix A that the equilibrium evolution of s_t on the long region implies

$$P^{(o)}\left[\left\{\sup_{u\geq t} s_u \in \mathcal{S}\right\}\middle| 0 < s_t \leq s^*\right] = 1, \tag{28}$$

so that the optimist can be certain that, starting from any point in $\mathcal{L}$, her consumption share will eventually enter the open region $\mathcal{S}$ where the trading of the pessimist generates strictly positive lending fees.

4 Equilibrium prices and lending yield

To complete the construction of the equilibrium, it now remains to compute the lending yield and the risky asset prices. To facilitate the presentation, we start with the simpler case of a single risky asset where the solution is in closed-form before turning to the more challenging case of two risky assets. We then calibrate the model to briefly discuss the 3Com/Palm spin-off puzzle.

4.1 One risky asset

When the weight $\eta_1 = 1$, the single risky asset $S_{1t} = M_t$ is the market portfolio and its volatility equals that of the aggregate dividend. Substituting these quantities into (16) and (20) and using (21) shows that in equilibrium

$$\Phi_t = \mathbb{1}_{\{s_t>s^*\}}\frac{(\delta+\Gamma_t)s_t-\sigma^2}{1+s_t}, \tag{29}$$

$$\Gamma_t = \mathbb{1}_{\{s_t>s^*\}}\frac{(1-s_t)(s_t\delta-\Gamma_t-\sigma^2)((\delta+\Gamma_t)s_t-\sigma^2)}{\sigma^2(1+s_t)^2}, \tag{30}$$

where the constant

$$\delta \equiv \sigma\Delta = \mu^{(o)} - \mu^{(p)} \geq 0$$

denotes the unscaled difference in beliefs between the two agents. Solving this system delivers the following result.

Proposition 3. *With a single risky asset, the equilibrium shorting cost and the equilibrium lending yield are given by*

$$\Phi_t = \mathbb{1}_{\{s_t > s^*\}} \frac{s_t(1-s_t)\delta - 2\sigma^2 + \sqrt{s_t^2(1-s_t)^2\delta^2 + 4\sigma^4 s_t}}{2(1-s_t)}, \tag{31}$$

and

$$\Gamma_t = \mathbb{1}_{\{s_t > s^*\}} \frac{-s_t((1-s_t)^2\delta + 4\sigma^2) + (1+s_t)\sqrt{s_t^2(1-s_t)^2\delta^2 + 4\sigma^4 s_t}}{2s_t(1-s_t)}, \tag{32}$$

and both are increasing and convex in δ.

The positive relation between the shorting cost and the difference in beliefs is intuitive. Indeed, an increase in δ implies that agent p becomes more pessimistic than agent o in relative terms and thus triggers an upward shift in her short demand schedule which in turn leads to an increase of the shorting cost. To understand the comparative statics of the lending yield, note that due to market clearing we have

$$\Gamma_t = \frac{\Phi_t \Upsilon_t}{1+\Upsilon_t}, \tag{33}$$

where the *utilization* ratio $\Upsilon_t \equiv \pi_{1t}^{(p)^-}/S_{1t}$ tracks the fraction of shortable shares that are on loan. This measure of short interest is affected by changes in δ both directly through the perceived risk premia and indirectly through the equilibrium shorting cost. However, combining Proposition 1 and (16) reveals that in equilibrium we have

$$\Upsilon_t = \left(\frac{1-s_t}{\sigma}\right)\left(\theta_{1t}^{(p)} + \phi_t\right)^- = \left(\frac{1-s_t}{\sigma^2}\right)\Phi_t, \tag{34}$$

which implies that the comparative statics of Υ_t and thus of Γ_t are the same as those of the equilibrium shorting cost Φ_t. In particular, since the shorting cost is increasing in δ, this identity shows that, in equilibrium, there exists a positive relation between short interest and the divergence in beliefs. This implication of the model is consistent with extensive empirical evidence. In particular, it is well documented that there exists a positive relation between short interest and the dispersion of analysts' earning forecasts taken as a proxy for heterogenous beliefs (see e.g., D'Avolio (2002), Duffie et al. (2002)).

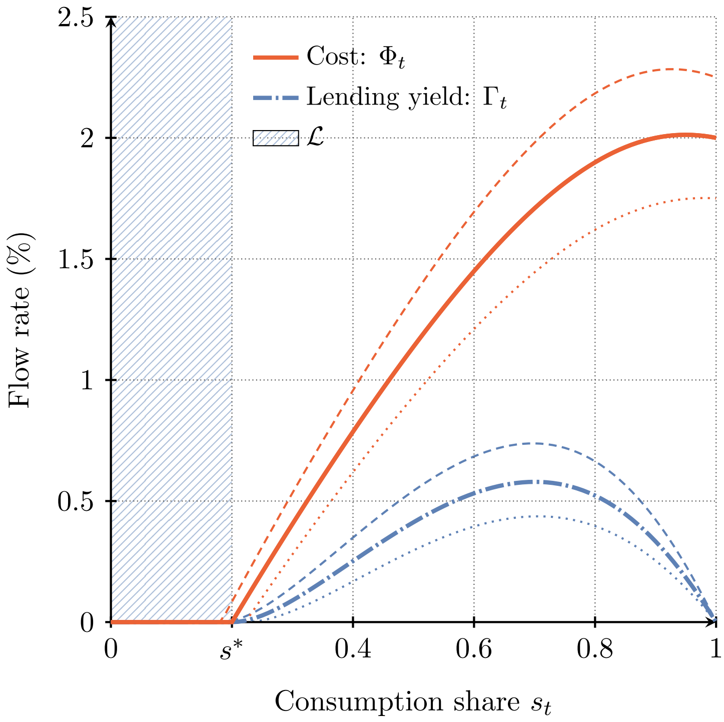

Figure 3: Equilibrium shorting cost and lending yield. The solid and dash-dotted lines represent the equilibrium shorting cost and lending yield as functions of the consumption share of the optimist in a single asset model with $\sigma = 0.1$ and $\delta = 0.05$. The dashed|dotted lines represent the impact of a 10% increase|decrease in the divergence of beliefs δ.

To illustrate the magnitude of the shorting cost and its dependence on the wealth distribution, Figure 3 plots Φ_t and Γ_t as function of the consumption share of the optimist in a model with $\sigma = 10\%$ and $\delta = 5\%$ which are both reasonable if one interprets e_t as modelling aggregate dividends (David, 2008, Belo, Collin-Dufresne and Goldstein, 2015). Given these values, the figure illustrates that the implied lending fee varies between 0 and 2% depending on the state of the economy. This interval of variation is of the right magnitude for most U.S. stocks. Indeed, the empirical analysis in our companion paper Hugonnier and Prieto (2023) shows that over the period 2004-14 the average annual cost of shorting a stock varied between 0.07% and 0.59% in deciles 1 to 9, and jumped to 6.93% in decile 10. See e.g., Beneish et al. (2015), Muravyev, Pearson and Pollet (2018), Drechsler and Dreschler (2018) for similar estimates. The figure shows that the shorting cost starts from zero at the lower end of the shorting region, increases until it reaches a maximum and tappers off to a limit that is explicitly given by

$$\lim_{s_t \to 1} \Phi_t = (1 - s^*) \frac{\delta}{2} = (\Delta - \sigma) \frac{\sigma}{2}$$

as a result of (31). To understand this limit, recall that as $s_t \to 1$ the model converges to one where only the optimist is present. As a result, the market price of risk perceived by the pessimist must converge to its frictionless counterpart $\theta^*(1) - \Delta = \Delta(s^* - 1)$ and the expression for the limiting cost now follows from (16).

The bottom curves of the figure show that the lending yield is a bell-shaped function of s_t that starts from zero at the lower end of the shorting region and comes back to zero as the wealth share of the optimist approaches one. The apparent discrepancy between the limiting behavior of the shorting cost and the lending yield as $s_t \to 1$ can be traced back to the economic nature of these objects. Indeed, Φ_t represents a price that can be meaningfully understood in the limit as the cost for a short position of infinitesimal size (see e.g., Davis (1998), Hugonnier and Kramkov (2004), Hugonnier, Kramkov and Schachermayer (2005)) whereas Γ_t is a flow rate that can be strictly positive only in states where the pessimist holds a non infinitesimal fraction of aggregate wealth.

As usual with logarithmic preferences, the price $S_{1t} = e_t/\rho$ of the single risky asset is unaffected by the presence of frictions. However, it is important to recall that, in our model, this price comprises two parts. Indeed, it follows from Proposition 2 that

$$S_{1t} = E_t^{(o)} \left[\int_t^\infty \xi_{t,u}^{(o)} e_u du \right] + E_t^{(o)} \left[\int_t^\infty \xi_{t,u}^{(o)} S_{1u} \Gamma_u du \right],$$

where the first term

$$E_t^{(o)} \left[\int_t^\infty \xi_{t,u}^{(o)} e_u du \right] = E_t^{(o)} \left[\int_t^\infty e^{-\rho(u-t)} \left(\frac{s_t}{s_u} \right) du \right]$$

gives the (risk-adjusted) present value of futures dividends, i.e., the *fundamental value* of the asset, and the second captures the present value of the flows of lending fees associated with ownership of the asset. As we discuss below in the two risky assets case, the fact that the lending yield is a deterministic function of the endogenous state variable implies that both components can be computed from the solution to a boundary value problem for a nonlinear differential equation, see (39) and (40).

Remark 3 (Exogenously fixed shorting cost)**.** Gârleanu et al. (2021) consider a setting that is very close to ours but assume that the shorting cost is given by an *exogenous* function of short interest which they take to be constant in the baseline version of their model. Working under this alternative assumption it can be shown that

– If $\Delta \leq \sigma$ then agents are always long in the *unique* equilibrium.

– If $\sigma < \Delta \leq \sigma + \phi$ where $\phi \equiv \Phi/\sigma$ denotes the constant shorting cost per unit of volatility then the *unique* equilibrium includes two regions: The region $\mathcal{L}$ over which $s_t < \sigma\Delta$ where both agents hold long positions and the complementary region $\mathcal{O}$ in which the pessimist would short absent costs but does not because the exogenous shorting cost is too high.

– If $\Delta > \sigma + \phi$ then all equilibria include a third region $\mathcal{S}$ over which the pessimist holds a short position in the stock.

In the latter case, the same steps as in Section 3.3 show that the equilibrium interest rate and market price of risk are given by

$$\begin{aligned} r_t = r^*(s_t) &+ \mathbb{1}_{\{\mathcal{O}\}} (1 - s_t) \, \sigma \, (\Delta - \sigma/s_t) \\ &+ \mathbb{1}_{\{\mathcal{S}\}} (1 - s_t) \, s_t \, (\phi - \gamma_t) \, (\Delta - \phi + \gamma_t) \end{aligned} \tag{35}$$

$$\theta_t^{(o)} = \theta^*(s_t) - \mathbb{1}_{\{\mathcal{O}\}} (1 - s_t) \, (\Delta - \sigma/s_t) - \mathbb{1}_{\{\mathcal{S}\}} (1 - s_t) \, (\phi - \gamma_t) \tag{36}$$

where the scaled lending yield $\gamma_t \equiv \Gamma_t/\sigma$ is a solution to

$$\gamma_t \sigma = (1 - s_t) \, (\gamma_t - \phi) \, (\sigma - s_t(\gamma_t + \Delta - \phi)) \tag{37}$$

and the short region $\mathcal{S}$ is determined by the interval of consumption shares s_t for which this equation admits a strictly positive solution.

In our model the endogeneity of the shorting cost ϕ_t guarantees that (37) admits exactly one strictly positive solution and that this solution is equal to zero on the boundary of the shorting region. As a result, the equilibrium is unique and all equilibrium quantities are continuous functions of the endogenous state variable s_t. This is no longer the case when ϕ is exogenously fixed as in Gârleanu et al. (2021). Indeed, a direct calculation shows that, over the region of the parameter space where

$$4\sigma < \phi \quad \text{and} \quad \Delta \in \frac{1}{2} \left[3\phi - \sqrt{\phi \, (\phi - 4\sigma)}, 3\phi + \sqrt{\phi \, (\phi - 4\sigma)} \right],$$

equation (37) admits two strictly positive solutions when s_t lies in an interval $\mathcal{M}$ to the left of the point $\hat{s} \equiv \sigma/(\Delta - \phi)$. As a result, there are infinitely many equilibria that are each associated with a different selection of the lending yield at times where the state variable lies in the interval $\mathcal{M}$.

This multiplicity is interesting from the theoretical point of view, but it is subject to two important caveats that severely limit its practical relevance. First, multiplicity obtains only if the shorting cost is so high that $\phi > 4\sigma$ which is very unlikely to occur. For example, in the ten year sample that we use in Hugonnier and Prieto (2023) this condition is satisfied in 66 of $308,618$ (0.021%) return/month observations associated to 54 different stocks. Second, and perhaps more importantly, the multiplicity of the equilibrium implies that the interest rate and the market price of risk in (35)–(36) experience predictable jumps at every point in time where the equilibrium switches from one branch of the lending yield solution to the other.

4.2 Two risky assets

Consider now the model with two risky assets. Equation (27) shows that in order to compute the equilibrium asset prices it is sufficient to compute the price-dividend ratio of asset 2 or, equivalently, its market share

$$w_t \equiv \frac{S_{2t}}{M_t} = \rho E_t^{(o)}\left[\int_t^\infty \xi_{t,u}^{(o)} e_{2u} du\right] = \rho\eta_2 E_t^{(o)}\left[\int_t^\infty e^{-\rho(u-t)}\left(\frac{s_t}{s_u}\right) du\right].$$

This expression makes it clear that w_t and thus the asset prices

$$(S_{1t}, S_{2t}) = (1 - w_t, w_t)\, M_t$$

depend on an expectation over the future path of the endogenous state variable. On the other hand, since

$$\sigma_{1t} S_{1t} = ((1 - w_t)\sigma - \mathrm{diff}_t(w))\, M_t, \tag{38}$$

it follows from (24) that the drift and diffusion of s_t on the shorting region depend on w_t and its diffusion coefficient

$$\mathrm{diff}_t(w) = \frac{1}{dt} d\left\langle w_t, Z_t^{(o)}\right\rangle$$

through the lending market clearing condition (20). Therefore, the triple $(s_t, w_t, \mathrm{diff}_t(w))$ is the solution to a Forward Backward Stochastic Differential Equation over an infinite horizon (FBSDE, see Ma and Yong (1999) for a thorough introduction).

Since the evolution of the process s_t is fully determined by $(s_t, w_t, \mathrm{diff}_t(w))$, it is natural to look for *Markovian equilibria* in which $w_t = w(s_t)$ for some sufficiently regular bounded function such that

$$w(0) = w(1) = \eta_2, \tag{39}$$

where the equalities follow from the fact that the endogenous state variable is absorbed at the endpoints of the unit interval. Furthermore, Itô's lemma and (38) show that for such a solution we have

$$\mathrm{diff}_t(w) = s_t\,(1 - s_t)\,v\,(s_t, \gamma_t)\,w'(s_t)$$

and therefore

$$\sigma_{1t} S_{1t} = \left((1 - w(s_t))\,\sigma - s_t\,(1 - s_t)\,v\,(s_t, \gamma_t)\,w'(s_t)\right) M_t,$$

where the function $v(s_t, \gamma_t)$ is defined by (24a). Substituting into the short market clearing condition (20), gives a quadratic equation

$$\frac{\gamma_t \sigma}{1 - s_t} = \frac{\gamma_t\,(\gamma_t + \Delta + \sigma)\,w'(s_t)}{(1 - w(s_t))\,(1 + s_t)} + \frac{(s_t \Delta - \gamma_t - \sigma)\,((\gamma_t + \Delta) s_t - \sigma)}{(1 - w(s_t))\,(1 + s_t)^2}$$

that implicitly determines the lending fee

$$\gamma_t = \gamma\,(s_t, w(s_t), w'(s_t))$$

as a function of s_t, $w(s_t)$, and $w'(s_t)$ for all $s_t > s^*$. Substituting this function into (24) then shows that the endogenous state variable evolves according to the *autonomous* stochastic differential equation defined by

$$ds_t = \overline{m}[w](s_t)dt + \overline{v}[w](s_t)dZ_t^{(o)},$$

with the deterministic functions

$$(\overline{m}[w](s), \overline{v}[w](s)) \equiv s\,(1 - s)\,(m, v)\,(s, \gamma\,(s, w(s), w'(s)))\,.$$

This implies that s_t is a Markov diffusion and, since the process

$$e^{-\rho t}\frac{w(s_t)}{s_t} + \rho\eta_2 \int_0^t e^{-\rho u}\frac{du}{s_u} = \rho\eta_2 E_t^{(o)}\left[\int_0^\infty e^{-\rho u}\frac{du}{s_u}\right]$$

is by construction a martingale, it follows that the market weight is a piecewise twice continuously differentiable solution to

$$\rho\left(\frac{w(s)}{s}\right)=\overline{m}[w](s)\left(\frac{w(s)}{s}\right)'+\frac{1}{2}\left(\overline{v}[w](s)\right)^2\left(\frac{w(s)}{s}\right)''+\frac{\rho\eta_2}{s}, \tag{40}$$

subject to the boundary condition (39).

This nonlinear boundary value problem is too complex to admit an explicit solution. We therefore resort to numerical methods to illustrate the quantitative implications of the model. As a first step, we start by observing that on the long region $[0,s^*]$ the differential equation simplifies to

$$\rho w(s)=\rho\eta_2+\frac{1}{2}s^2(1-s)^2\Delta^2 w''(s).$$

A direct calculation shows that, for any given $\varepsilon\in(0,\eta_2)$, the unique solution to this equation with $w(0)=\eta_2$ and $w(s^*)=\varepsilon$ is explicitly given by

$$w(s;\varepsilon)=\eta_2+(\varepsilon-\eta_2)\left(\frac{s}{s^*}\right)^{\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{8\rho}{\Delta^2}}}\left(\frac{1-s}{1-s^*}\right)^{\frac{1}{2}-\frac{1}{2}\sqrt{1+\frac{8\rho}{\Delta^2}}}.$$

Relying on this solution over the long region, we now combine a traditional shooting approach with a collocation method (see, e.g., Miranda and Fackler (2004) and Dangl and Wirl (2004)) to construct a global solution as follows: For each value of ε we implement a Chebyshev collocation to numerically solve equation (40) on the shorting interval $[s^*,1]$ subject to the *initial* conditions

$$0=w(s^*)-w(s^*;\varepsilon)=w'(s^*)-w'(s^*;\varepsilon),$$

and then numerically vary the value of the free constant until the solution satisfies the terminal boundary condition $w(1)=\eta_2$ required by (39).

To illustrate the quantitative implications of the model in the two asset case, we fix the underlying parameters $(\sigma,\delta,\rho)=(10,5,1)\%$ and set $e_t=1$ so that the equilibrium value of the market portfolio is normalized to 100. In the left panel of Figure 4, we plot the equilibrium price of asset 1

$$S_{1t}=M_t-S_{2t}=(1-w(s_t))\,M_t$$

and the present value of its future dividends, i.e., its fundamental value,

$$f_1(s_t)M_t \equiv E_t^{(o)}\left[\int_t^\infty \xi_{t,u}^{(o)} e_{1u} du\right] = \frac{\eta_1 w(s_t) M_t}{1-\eta_1}$$

when $\eta_1 = 50\%$, so that half of the asset supply is available for shorting. In a frictionless environment, these functions would be constant and equal to $\eta_1 M_t = 50$ because, absent shorting costs, asset 1 only entitles its owner to a constant share of dividends. As shown by the figure, this is no longer the case in the presence of a shorting friction. Indeed, since $s^* = \sigma^2/\delta = 0.2 < 1$, we have $\mathcal{S} \neq \emptyset$ and it follows that asset 1 entitles its owners to strictly more than its share of dividends. Since the value of the market is fixed, this implies that the equilibrium price of asset 2, given by the risk-adjusted present value of its dividends, must account for less than $1 - \eta_1 = 50\%$ of the market, and it follows that the price of asset 1 must strictly exceed its frictionless value $\eta_1 M_t = 50$, which in turn must exceed the fundamental value of the asset.

The difference between the market value of the asset and the risk-adjusted present value of its dividends, that is

$$\ell(s_t)M_t \equiv E_t^{(o)}\left[\int_t^\infty \xi_{t,u}^{(o)} S_{1u}\Gamma_u du\right] = \left(1 - \frac{w(s_t)}{1-\eta_1}\right) M_t,$$

represents the risk-adjusted present value of the lending fees associated with ownership of asset 1. The figure shows that this difference is bell-shaped as function of the state variable and can amount to as much as 10% of the market portfolio when half of the asset supply can be shorted. To illustrate the impact of the supply parameter η_1 on the lending component, we plot in the right panel the relative contribution

$$\frac{\ell(s_t)}{1-w(s_t)} = \frac{1-\eta_1-w(s_t)}{(1-\eta_1)(1-w(s_t))}$$

of this component to the price of asset 1 for different values of η_1 ranging from 1% to 100%. As shown by the figure, this contribution is also single-peaked as a function of the state variable and monotone decreasing in η_1. The latter property is intuitive. Indeed, as η_1 decreases, the dividend component of the asset cash flows naturally decreases but the lending fees component remains essentially unchanged because the demand for short positions is not directly affected by the supply parameter η_1, and it follows that lending fees must account for a larger share of the equilibrium asset price.

As discussed in Section 3.4, the nonlinearity of the equilibrium pricing rule can help explain apparent mispricing episodes such as the partial spin-off of Palm by 3Com. To

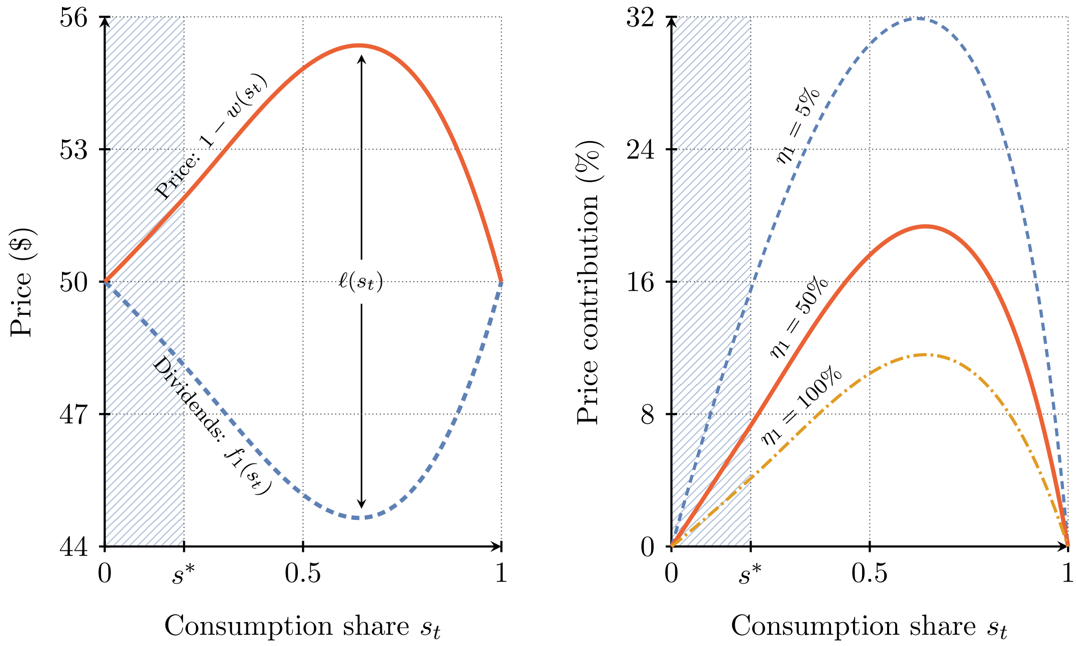

Figure 4: Decomposition of the equilibrium price. The left panel plots the price of the asset 1 (solid line) and the risk-adjusted present value of its dividends (dashed line) when half of the supply is available for shorting. The right panel plots the present value of lending fees as a fraction of the asset price for different values of η_1. In both panels the parameters of the model are set to $\sigma = 10\%$, $\delta = 5\%$, $\rho = 1\%$, $e_t = 1$, and the hatched region indicates the interval over which no shorting activity takes place in equilibrium.

illustrate this point, we identify asset 1 with the 5% of shortable Palm shares and asset 2 with the remaining shares held by 3Com. From the figure, we read that the equilibrium price of the block of shortable Palm shares evaluated at the point s^* is given by $(1 - w(s^*))M_t = 5.87$ and includes 15.5% of lending fees. Extrapolating this price to the remaining Palm shares values asset 2 at $(0.95/0.05)(1-w(s^*))M_t = 111.41$ which exceeds the value $M_t = 100$ of the conglomerate and represents a premium of

$$\left(\frac{0.95}{0.05}\right)\left(\frac{1 - w(s^*)}{w(s^*)}\right) - 1 = 18.35\%$$

relative to the equilibrium price $w(s^*)M_t = 94.14$ of asset 2. Note that these figures are conservative because they are evaluated at the point s^* that signals the entry into the shorting region. If instead we used as reference the point $\operatorname{argmax}(1 - w(s)) \approx 0.62$, then the price of the block of shortable Palm shares would include 32% of lending fees and the relative premium on asset 2 would increase to 46.84%.

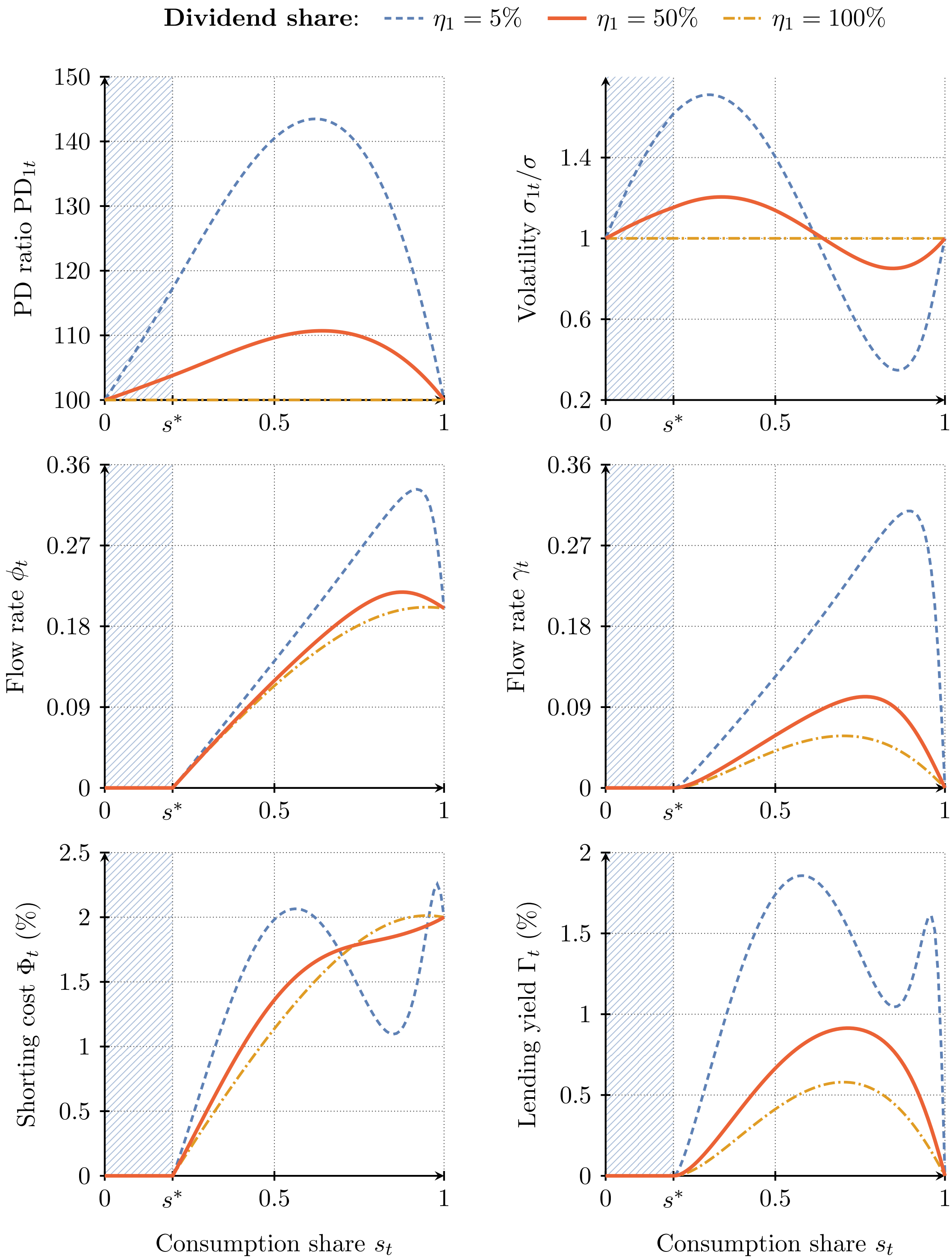

Figure 5: Equilibrium with two risky assets. This figure plots the PD ratio and the volatility of the shortable asset (1st row), the shorting cost and the lending yield per unit of volatility (2nd row), and the shorting cost and the lending yield (3rd row) as functions of the consumption share of the optimist for different values of η_1 in a model with $\sigma = 10\%$, $\delta = 5\%$, $\rho = 1\%$ and $e_t = 1$. In each panel the hatched region indicates the interval of states over which no shorting activity takes place in equilibrium.

Turning to the lending market, Figure 5 plots the price-dividend ratio of the shortable asset PD_{1t} and its volatility σ_{1t}, as well as the shorting cost and lending fees both per unit of volatility (ϕ_t, γ_t) and unscaled (Φ_t, Γ_t). The middle panels show that, when expressed in units of risk, the shorting cost and the lending yield are decreasing in the dividend share η_1 and otherwise behave similarly as in the one asset case of Section 4.1 which here corresponds to the dash-dotted lines. The former feature can be understood as follows: As η_1 decreases, asset 1 becomes more scarce. Therefore, the share of total lending fees that each share entitles to also increases. This tends to push the asset price up and the market price of risk down which in turn implies that the intermediary can charge a higher cost to optimists per unit of volatility as shown by the middle panels of the figure. As the fraction η_1 approaches zero the economy barely includes any shortable stocks but the figure shows that shorting cost and lending yield per unit of volatility continue to increase. This indicates that the shorting market remains active despite the vanishing supply and suggests that the equilibrium is not continuous at $\eta_1 = 0$ because at that point our one asset model coincides with that of Gallmeyer and Hollifield (2008) who analyze the equilibrium effects of a short sale ban.

The bottom panels show that these intuitive properties no longer hold when the shorting cost and the lending yield are expressed as flow rates per dollar of short. This change can be traced back to the oscillatory behavior of the asset volatility in the top right panel, which in turn is implied by the behavior of the PD ratio in the top left panel of the figure. Indeed, since the PD ratio is hump-shaped and the diffusion of the endogenous state variable $s_t(1-s_t)\overline{v}[w](s_t) \geq 0$ vanishes at the endpoints of the state space, we have that the excess volatility

$$\sigma_{1t} - \sigma = \frac{s_t(1-s_t)\overline{v}[w](s_t)(-w'(s_t))}{1-w(s_t)}$$

is positive (negative) over the interval where the PD ratio $(1-w(s_t))/(\eta_1\rho)$ is increasing (decreasing) and equal to zero at $s_t = 0$, $s_t = 1$, and at the point where the PD ratio reaches its maximum. The top right and bottom panels of the figure show that the amplitude of the volatility oscillation is decreasing in the dividend share η_1 and gets gradually transferred to the shorting costs and the lending yield as the shortable asset becomes scarce.

5 Conclusion

We study a dynamic equilibrium model with costly short sales and heterogeneous beliefs. The closed-form solution to the model reveals how costly short sales drive a wedge between the valuation of assets that promise identical cash flows but are subject to different trading arrangements. In particular, we show that the price of an asset is given by the risk-adjusted present value of future cash flows provided that these are augmented to include not only dividends but also an endogenous *lending yield*.

This asset pricing formula implies that, after adjusting for lending revenues, returns satisfy a standard intertemporal capital asset pricing model and sheds light on recent findings about the explanatory power of shorting costs in the cross-section of stock returns. Specifically, it follows from (1), (5), and (7) that, in equilibrium, the expected excess returns on the two risky assets can be expressed as

$$\begin{aligned}\frac{1}{dt}E_t^{(o)}\left[\frac{dS_{it}+\eta_i e_t dt}{S_{it}}\right]-r_t &= \sigma_{it}\theta_{2t}^{(o)}-\mathbf{1}_{\{i=1\}}\Gamma_t \\ &= \sigma_{it}\theta_{2t}^{(o)}-\mathbf{1}_{\{i=1\}}\left(\frac{\Upsilon_t}{1+\Upsilon_t}\right)\Phi_t,\end{aligned}$$

where Φ_t is the shorting cost, Γ_t is the lending yield, and Υ_t captures the fraction of the available inventory that is on loan. This shows that, within our simple framework, the two assets offer the same risk-return tradeoff once the lending revenues of the shortable asset are accounted for. Importantly, the same logic would apply in an extension of the model with multiple shortable stocks paying out imperfectly correlated dividends. Developing such an extension would be very interesting as it would allow to study the cross-sectional effects of costly short sales. Unfortunately, the resulting model would have to be solved numerically because, with more than one shortable asset, the endogenous partition of the state space into disjoint trading regions can no longer be guessed a priori. We leave this challenging extension of the model for future research.

References

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

Atmaz, A., Basak, S., Ruan, F., 2021. Dynamic equilibrium with costly short-selling and lending market. SSRN working paper .

Baklanova, V., Copeland, A.M., McCaughrin, R., 2015. Reference guide to us repo and securities lending markets. FRB of New York Staff Report .

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

Basak, S., Croitoru, B., 2000. Equilibrium mispricing in a capital market with portfolio constraints. Review of Financial Studies 13, 715–748.

Basak, S., Cuoco, D., 1998. An equilibrium model with restricted stock market participation. Review of Financial Studies 11, 309–341.

Belo, F., Collin-Dufresne, P., Goldstein, R.S., 2015. Dividend dynamics and the term structure of dividend strips. Journal of Finance 70, 1115–1160.

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

Blau, B.M., Van Ness, R.A., Warr, R.S., 2012. Short selling of ADRs and foreign market short-sale constraints. Journal of Banking and Finance 36, 886–897.

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

Brunnermeier, M.K., Oehmke, M., 2014. Predatory short selling. Review of Finance 18, 2153–2195.

Brunnermeier, M.K., Pedersen, L.H., 2005. Predatory trading. Journal of Finance 60, 1825–1863.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Carlin, B.I., Lobo, M.S., Viswanathan, S., 2007. Episodic liquidity crises: Cooperative and predatory trading. Journal of Finance 62, 2235–2274.

Chen, S., Kaniel, R., Opp, C.C., 2022. Market power in the securities lending market. Working paper, University of Rochester. Available at SSRN 4100699 .

Cherkes, M., Jones, C.M., Spatt, C.S., 2013. A solution to the Palm-3Com spin-off puzzles. Columbia Business School Research Paper .

Cochrane, J.H., 2002. Stocks as money: convenience yield and the tech-stock bubble. Technical Report. National Bureau of Economic Research.

Cohen, L., Diether, K., Malloy, C., 2007. Supply and demand shifts in the shorting market. Journal of Finance 62, 2061–2096.

Dangl, T., Wirl, F., 2004. Investment under uncertainty: calculating the value function when the Bellman equation cannot be solved analytically. Journal of Economic Dynamics and Control 28, 1437–1460.

DataLend, 2022. Press release: Securities lending markets up 21% in 2021, generating $9.28billion in revenue. URL: www.prnewswire.com/news-releases/(...)301453745.html. accessed on 2022-01-04.

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

Davis, M., 1998. *Option Pricing in Incomplete Markets* in Mathematics of Derivative Securities. Cambridge University Press. Publications of the Newton Institute, pp. 216–226.

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

De Jong, A., Rosenthal, L., Van Dijk, M.A., 2009. The risk and return of arbitrage in dual-listed companies. Review of Finance 13, 495–520.

Detemple, J., Murthy, S., 1997. Equilibrium asset prices and non-arbitrage under portfolio constraints. Review of Financial Studies 10, 1133–1174.

Detemple, J., Serrat, A., 2003. Dynamic equilibrium with liquidity constraints. Review of Financial Studies 16, 597–629.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. Journal of Finance 57, 2113–2141.

Drechsler, I., Dreschler, Q.F., 2018. The shorting premium and asset pricing anomalies. Working paper, NBER and Wharton UPenn .

Duffie, D., 1996. Special repo rates. Journal of Finance 51, 493–526.

Duffie, D., Gârleanu, N., Pedersen, L.H., 2002. Securities lending, shorting, and pricing. Journal of Financial Economics , 307–339.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. Journal of Finance 65, 1915–1947.

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

Froot, K.A., Dabora, E.M., 1999. How are stock prices affected by the location of trade? Journal of Financial Economics 53, 189–216.

Gallmeyer, M., Hollifield, B., 2008. An examination of heterogeneous beliefs with a short sale constraint in a dynamic economy. Review of Finance 12, 323–364.

Gârleanu, N.B., Panageas, S., Zheng, G.X., 2021. A Long and a Short Leg Make For a Wobbly Equilibrium. Technical Report. National Bureau of Economic Research.

Gensler, G., 2021. Proposed Updates to Securities Lending Market. Technical Report. U.S. Securities and Exchange Commission. URL: https://www.sec.gov/news/statement/gensler-securities-lending-market-20211118.

Goldstein, I., Guembel, A., 2008. Manipulation and the allocational role of prices. Review of Economic Studies 75, 133–164.

Goldstein, I., Ozdenoren, E., Yuan, K., 2013. Trading frenzies and their impact on real investment. Journal of Financial Economics 109, 566–582.

Gromb, D., Vayanos, D., 2010. Limits of arbitrage. Annual Review of Financial Economics 2, 251–275.

Hajek, B., 1985. Mean stochastic comparison of diffusions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 68, 315–329.

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

Harrison, J., Kreps, D., 1978. Speculative investor behavior in a stock market with heterogeneous expectations. Quarterly Journal of Economics , 323–336.

Hugonnier, J., Kramkov, D., 2004. Optimal investment with random endowments in incomplete markets. Annals of Applied Probability 14, 845 – 864.

Hugonnier, J., Kramkov, D., Schachermayer, W., 2005. On utility-based pricing of contingent claims in incomplete markets. Mathematical Finance 15, 203–212.

Hugonnier, J., Prieto, R., 2023. Asset pricing with costly short sales: empirical exercise .

Jiang, H., Habib, A., Hasan, M., 2020. Short selling: A review of the literature and implications for future research. European Accounting Review , 1–31.

Jones, C., Lamont, O., 2002. Short-sale constraints and stock returns. Journal of Financial Economics 66.

Karatzas, I., Shreve, S., 1988. Brownian motion and stochastic calculus. second ed., Springer Verlag.

Kashyap, A., Kovrijnykh, N., Li, J., Pavlova, A., 2020. Is There Too Much Benchmarking in Asset Management? Technical Report. University of Chicago Working Paper.

Kogan, L., Makarov, I., Uppal, R., 2007. The equity risk premium and the riskfree rate in an economy with borrowing constraints. Mathematics and Financial Economics 1, 1–19.

Krishnamurthy, A., 2002. The bond/old-bond spread. Journal of Financial Economics 66, 463–506.

Lamont, O.A., Thaler, R.H., 2003. Can the market add and subtract? Mispricing in tech stock carve-outs. Journal of Political Economy 111, 227–268.

Ma, J., Yong, J., 1999. Forward-backward stochastic differential equations and their applications. Lecture Notes in Mathematics, Springer.

Mei, J., Scheinkman, J.A., Xiong, W., 2009. Speculative trading and stock prices: Evidence from Chinese A B share premia. Annals of Economics and Finance 10, 225–255.

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

Miranda, M.J., Fackler, P.L., 2004. Applied Computational Economics and Finance. The MIT Press.

Mitchell, M., Pulvino, T., Stafford, E., 2002. Limited arbitrage in equity markets. Journal of Finance 57, 551–584.

Muravyev, D., Pearson, N.D., Pollet, J.M., 2018. Is there a risk premium in the stock lending market? evidence from equity options. Working paper, Boston College .

Nezafat, M., Schroder, M., 2022. Private information, securities lending, and asset prices. Review of Financial Studies 35, 1009–1063.

Nutz, M., Scheinkman, J.A., 2020. Shorting in speculative markets. Journal of Finance 75, 995–1036.

Ofek, E., Richardson, M., Whitelaw, R., 2004. Limited arbitrage and short-sales restrictions: evidence from the options markets. Journal of Financial Economics 74, 305–342.

Radner, R., 1972. Existence of equilibrium of plans, prices and prices expectations in a sequence of markets. Econometrica 40, 289–303.

Reed, A.V., 2013. Short selling. Annual Review of Financial Economics 5.
Scheinkman, J.A., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy 111, 1183–1220.
Seneca, J.J., 1967. Short interest: bearish or bullish? Journal of Finance 22, 67–70.
Vayanos, D., Weill, P.O., 2008. A search-based theory of the on-the-run phenomenon. Journal of Finance 63, 1361–1398.

A Proofs

Proof of Proposition 1. The solution follows from a direct application of the Karush, Kuhn, and Tucker conditions to (8) subject to (6) and (7). ■

Proof of Proposition 2. Let $\xi_t = \xi_t^{(o)}$ with

$$-\frac{d\xi_t}{\xi_t} = r_t dt + \theta_{2t}^{(o)} dZ_t^{(o)}$$

denote the marginal utility of the optimist. By construction, we have that

$$N_{it} = \xi_t S_{it} + \int_0^t \xi_u \left(e_{iu} + \mathbb{1}_{\{i=1\}} S_{1u} \Gamma_u \right) du$$

are local martingales under $P^{(o)}$ and it follows from Lemma 1 below that these processes are martingales over any finite horizon. In particular, we have that

$$\xi_t^{(o)} S_{it} = E_t^{(o)} \left[\xi_T S_{iT} + \int_t^T \xi_u \left(e_{iu} + \mathbb{1}_{\{i=1\}} \Gamma_u S_{1u} \right) du \right]$$

for all finite $T < \infty$ and therefore

$$\xi_t^{(o)} S_{it} = \lim_{T\to\infty} E_t^{(o)} \left[\xi_T S_{iT} \right] + E_t^{(o)} \left[\int_t^\infty \xi_u \left(e_{iu} + \mathbb{1}_{\{i=1\}} \Gamma_u S_{1u} \right) du \right]$$

by monotone convergence, since the terms below the integral are all nonnegative. To complete the proof, it remains to show that the limit is zero. Let $\lambda_t = 1/s_t - 1$. As shown in the proof of Lemma 1 below, we have that

$$\xi_T S_{iT} \le \xi_T M_T = e^{-\rho T} M_0 \left(\frac{s_0}{s_T} \right) = e^{-\rho T} M_0 \left(\frac{1+\lambda_T}{1+\lambda_0} \right) \le e^{-\rho T} M_0 \left(\frac{1+\Lambda_T}{1+\lambda_0} \right)$$

for some $P^{(o)}-$martingale Λ_t with initial value λ_0 and therefore

$$\lim_{T\to\infty} E_t^{(o)} \left[\xi_T S_{iT} \right] \le \lim_{T\to\infty} \frac{e^{-\rho T} M_0}{1+\lambda_0} \left(1 + E_t^{(o)} \left[\Lambda_T \right] \right) = \lim_{T\to\infty} e^{-\rho T} M_0 = 0,$$

where the last equality uses the fact that $\rho > 0$. Since $\xi_T S_{iT} \geq 0$, this in turn implies that the limit is zero and the proof is complete. ■

Lemma 1. *The process N_{it} is a $P^{(o)}-$martingale on $[0, T]$ for any $T < \infty$.*

Proof. By construction, we have that

$$0 \leq N_{it} \leq N_t \equiv N_{1t} + N_{2t} = \xi_t M_t + \int_0^t \xi_u \left(e_u + S_{1u}\Gamma_u\right) du,$$

and it is thus sufficient to show that the process N_t is a martingale under $P^{(o)}$ over the finite time interval $[0, T]$. Since $S_{it}\sigma_{it} \geq 0$ we have that

$$S_{it}\sigma_{it} \leq \sum_{j=1}^{2} S_{jt}\sigma_{jt} = M_t \sigma.$$

On the other hand, using (30) and the fact that $\gamma_t \leq \phi_t$ shows that we have

$$\gamma_t \leq \phi_t = \mathbb{1}_{\{s_t > s^*\}} \frac{s_t \Delta - (\gamma_t + \sigma)}{1 + s_t} \leq \Delta.$$

Combining this inequality with the definition of ξ_t, we deduce that there are strictly positive constants such that

$$\begin{aligned} |N_t| &\leq \xi_t M_t + \int_0^t \xi_u M_u \left(\rho + \gamma_u \sigma\right) du \\ &\leq \xi_t M_t + \int_0^t \xi_u M_u \left(\rho + \Delta\sigma\right) du \leq C + C' \sup_{u \in [0,T]} \lambda_u \end{aligned} \tag{41}$$

for all $t \in [0, T]$, where $\lambda_t \equiv 1/s_t - 1$. Using Itô's lemma and the dynamics of the consumption share process in (24) shows that

$$d\lambda_t = \lambda_t \left(g(s_t, \gamma_t) dZ_t^{(o)} - f(s_t, \gamma_t) dt\right)$$

for some functions $f, g : [0, 1] \times [0, \phi] \to \mathbb{R}$ such that $f(s, \gamma) \geq 0$ and $|g(s, \gamma)| \leq \Delta$. Therefore, Novikov's condition implies that

$$\Lambda_t \equiv e^{\int_0^t f(s_u, \gamma_u) du} \lambda_t = \lambda_0 \exp\left(-\int_0^t g(s_u, \gamma_u) dZ_u^{(o)} - \frac{1}{2}\int_0^t |g(s_u, \gamma_u)|^2 du\right)$$

is a $P^{(o)}-$martingale on any finite time interval and it thus follows from Doob's maximal inequality that for any $q > 1$ we have:

$$E^{(o)}\left[\sup_{u \in [0,T]} \lambda_u^q\right] \leq E^{(o)}\left[\sup_{u \in [0,T]} \Lambda_u^q\right] \leq \frac{q}{q - q} E^{(o)}\left[\Lambda_T^q\right].$$

Now, since $|g(s,\gamma)| \leq \Delta$ and the function x^q is convex for any $q > 1$, it follows from the mean comparison results of Hajek (1985, e.g., Theorem 1.3) that

$$E^{(o)}\left[\Lambda_T^q\right] \leq \Lambda_0^q E^{(o)}\left[e^{q\Delta Z_T^{(o)} - \frac{1}{2}q\Delta^2 T}\right] = e^{\frac{1}{2}q(q-1)\Delta^2 T}\lambda_0^q.$$

This implies that the right hand side of (41) is $P^{(o)}-$integrable and the required result finally follows from the dominated convergence theorem. ■

Proof of equation (15). If $\theta_{2t}^{(p)} \geq 0$ then it follows from $\gamma_t \leq \phi_t$ and (7) that we have

$$\theta_{1t}^{(p)} + \phi_t \geq \theta_{1t}^{(p)} + \gamma_t = \theta_{2t}^{(p)} \geq 0,$$

and therefore $\gamma_t = 0$ due to (14). To establish the converse implication, assume towards a contradiction that we have the lending yield $\gamma_t = 0$ but $\theta_{2t}^{(p)} < 0$. Then it follows from (7), (13) and (14) that we have

$$0 = \phi_t\left(\theta_{1t}^{(p)} + \phi_t\right)^{-} W_t^{(p)} = \frac{1}{4}\left\{\theta_{2t}^{(p)-}\right\}^2 W_t^{(p)}$$

and therefore $\theta_{2t}^{(p)} \geq 0$, since the wealth of the pessimist is strictly positive. ■

Proof of equation (18). If $\theta_{2t}^{(p)} < 0$ then

$$\begin{aligned}\theta_{1t}^{(p)} + \phi_t = \theta_{1t}^{(p)} + \max\left\{\gamma_t, -\tfrac{1}{2}\theta_{1t}^{(p)}\right\} &= \max\left\{\theta_{1t}^{(p)} + \gamma_t, \tfrac{1}{2}\theta_{1t}^{(p)}\right\}\\ &= \max\left\{\theta_{2t}^{(p)}, \tfrac{1}{2}\left(\theta_{2t}^{(p)} - \gamma_t\right)\right\} < 0,\end{aligned}$$

where the first equality follows from (13) and the third follows from (7). ■

Proof of Proposition 3. For $s_t > s^*$ we have that (30) is equivalent to $g_t(\gamma) = 0$ with the quadratic function defined by

$$g_t(\gamma) \equiv (1-s_t)(s_t\Delta - \gamma - \sigma)((\gamma+\Delta)s_t - \sigma) - \gamma\sigma(1+s_t)^2.$$

Since

$$\begin{aligned}g_t(0) &= (1-s_t)\left(s_t\Delta - \sigma\right)^2 > 0,\\ g_t'(0) &= -\sigma(1+s_t)^2 - (1-s_t)^2(s_t\Delta - \sigma) < 0,\\ g_t''(\gamma) &= -s_t(1-s_t) < 0,\end{aligned}$$

and

$$\lim_{\gamma\to\infty} g_t(\gamma) = -\infty,$$

it is clear that (30) admits a unique strictly positive solution. A direct calculation shows that this solution is given by (32) and substituting into (29) gives (31). The comparative statics follow by (31), (33), and (34) by differentiation. We omit the details. ■

Proof of equation (28). First observe that

$$P_t^{(o)}\left[\left\{\sup_{u\geq t} s_u \in \mathcal{S}\right\}\right] = P_t^{(o)}\left[\{\tau^* < \infty\}\right],$$

where the stopping time

$$\tau^* \equiv \inf\{u \geq t : s_t \geq s^*\}$$

denotes the first time at or after $t \geq 0$ that the Itô process s_t finds itself in the shorting region. To obtain the required probability, we will compute

$$g_t \equiv E_t^{(o)}\left[e^{-\lambda\tau^*}\right] = E_t^{(o)}\left[e^{-\lambda\tau^*}\mathbb{1}_{\{\tau^*<\infty\}}\right],$$

and then let $\lambda \downarrow 0$. On the time interval $[t, \tau^*]$, we have from (24) that the consumption share of the optimist evolves according to the autonomous SDE

$$ds_t = s_t(1-s_t)\Delta\left(dZ_t^{(o)} + (1-s_t)\Delta dt\right).$$

Therefore, it follows from well-known results (see, e.g., Karatzas and Shreve (1988, Chapter 5.7.A)) that $g_t = g(s_t)$, where the function $g : [0,1] \to \mathbb{R}$ is the unique bounded function such that

$$\begin{aligned} \lambda g(s) &= s(1-s)^2\Delta^2\left(g'(s) + \frac{1}{2}sg''(s)\right), && 0 \leq s \leq s^*, \\ g(s) &= 1, && s^* \leq s \leq 1. \end{aligned}$$

Solving this differential equation gives

$$g(s_t) = \mathbb{1}_{\{s_t > s^*\}} + \mathbb{1}_{\{s_t \leq s^*\}}\left\{\frac{s^*}{s_t}\left(\frac{1-s_t}{1-s^*}\right)\right\}^{\frac{1}{2}-\frac{1}{2}\sqrt{1+\frac{8\lambda}{\Delta^2}}},$$

and the desired result now follows from the dominated convergence theorem by letting the constant $\lambda \downarrow 0$ in the definition of g_t. ■

B Stochastic disagreement

In this appendix, we discuss the construction of an equilibrium in an extension of the model where the divergence in beliefs is stochastic and time-varying.

Assume that the economy is populated by two agents indexed by $a \in \{1,2\}$ who have different perceptions of the evolution of the aggregate dividend process. Specifically, assume that in the eyes of agent a

$$\frac{de_t}{e_t} = \mu_t^{(a)} dt + \sigma dZ_t^{(a)},$$

for some agent-specific Brownian motions $Z^{(a)}$ and growth rate process $\mu_t^{(a)}$ such that the scaled divergence in beliefs

$$\Delta_t \equiv \frac{1}{\sigma}\left(\mu_t^{(1)} - \mu_t^{(2)}\right)$$

is adapted to the filtration generated by the observation of the aggregate dividend process. As a typical example, one could consider an Ornstein-Uhlenbeck process for the disagreement, i.e., a process of the form

$$d\Delta_t = -\lambda \Delta_t dt + dZ_t^{(o)} = -(1+\lambda)\Delta_t dt + dZ_t^{(p)},$$

for some strictly positive constant λ, but the exact specification of the divergence process is unimportant for the arguments of this appendix. All the other building blocks of the model, i.e., the agents' preferences, the assets they trade, and the shorting mechanism remain the same as in the benchmark model of Section 2.

If the disagreement never changes sign then this model is essentially equivalent to the benchmark model of Section 2 with the identification $[o, p] = [1, 2]$ if the disagreement is always positive and $[o, p] = [2, 1]$ in the opposite case. Now assume that the disagreement is not signed. In this case, the identity of the optimist is a stochastic process that changes back and forth between $o_t = 1$ when the disagreement is positive and $o_t = 2$ when it is negative. As a result, the equilibrium can be constructed by analogy with that of the benchmark model by observing that the consumption share of agent 1 evolves like the consumption share of the optimist in the benchmark model at times where the disagreement is nonnegative, and as the consumption share of the pessimist at times where it is negative. For brevity we only outline the main steps.

Let $s_t \in [0, 1]$ denote the consumption share of agent 1 which we will use as an endogenous state variable. Proceeding along the lines of Sections 3.2 and 3.3 shows that the equilibrium shorting cost and lending yield satisfy

$$-\phi_t = \mathbb{1}_{\{\mathcal{S}^{(1)}\}} \frac{1}{2}\theta_{1t}^{(1)} + \mathbb{1}_{\{\mathcal{S}^{(2)}\}} \frac{1}{2}\theta_{1t}^{(2)} \tag{42}$$

and

$$\gamma_t \sigma_{1t} (1 - w_t) = (\phi_t - \gamma_t)\, \phi_t \left(\mathbb{1}_{\{\mathcal{S}^{(1)}\}} s_t + \mathbb{1}_{\{\mathcal{S}^{(2)}\}} (1 - s_t)\right), \tag{43}$$

where

$$\mathcal{S}^{(2)} = \left\{(\omega, t) : \theta_{2t}^{(2)} < 0 \le \theta_{2t}^{(1)}\right\} = \left\{(\omega, t) : s_t > s^*\left(\Delta_t^+\right)\right\}$$

gives the region of the state space over which agent 2 is short in asset 1 and agent 1 holds long positions in both risky assets, and

$$\mathcal{S}^{(1)} = \left\{(\omega, t) : \theta_{2t}^{(1)} < 0 \leq \theta_{2t}^{(2)}\right\} = \left\{(\omega, t) : 1 - s_t > s^*\left(\Delta_t^-\right)\right\}$$

gives the region over which agent 1 is short in asset 1 and agent 2 holds long positions in both risky assets. This, in turn, implies that the shorting market is endogenously inactive over the region given by

$$\begin{aligned}\mathcal{L} \equiv (\Omega \times \mathbb{R}_+) \setminus \cup_a \mathcal{S}^{(a)} &= \left\{(\omega, t) : \min_a \theta_{2t}^{(a)} \geq 0\right\} \\ &= \left\{(\omega, t) : 1 - s^*\left(\Delta_t^-\right) \leq s_t \leq s^*\left(\Delta_t^+\right)\right\}\end{aligned}$$

and substituting these expressions into (21) and (24) shows that the equilibrium interest rate, the equilibrium market price of risk perceived by agent 1, and the equilibrium evolution of her consumption share are explicitly given by

$$\begin{aligned}\theta_{2t}^{(1)} = \theta^*(s_t) &- \mathbb{1}_{\left\{s_t > s^*(\Delta_t^+)\right\}} \frac{(1 - s_t)\left(s_t \Delta_t^+ - \sigma - \gamma_t\right)}{1 + s_t} \\ &- \mathbb{1}_{\left\{1 - s_t > s^*(\Delta_t^-)\right\}} \frac{s_t\left((1 - s_t)\Delta_t^- - \sigma - \gamma_t\right)}{2 - s_t}, \\ r_t = r^*(s_t) &+ \mathbb{1}_{\left\{s_t > s^*(\Delta_t^+)\right\}} \frac{s_t(1 - s_t)\left(\Delta_t^+ + \sigma + \gamma_t\right)\left(s_t \Delta^+ - \sigma - \gamma_t\right)}{(1 + s_t)^2} \\ &+ \mathbb{1}_{\left\{1 - s_t > s^*(\Delta_t^-)\right\}} \frac{s_t(1 - s_t)\left(\Delta_t^- + \sigma + \gamma_t\right)\left((1 - s_t)\Delta_t^- - \sigma - \gamma_t\right)}{(2 - s_t)^2},\end{aligned}$$

and

$$\begin{aligned}\frac{ds_t}{s_t(1 - s_t)} = m\left(s_t, \gamma_t; \Delta_t^+\right) dt + v\left(s_t, \gamma_t; \Delta_t^+\right) dZ_t^{(1)} \\ - m\left(1 - s_t, \gamma_t; \Delta_t^-\right) dt - v\left(1 - s_t, \gamma_t; \Delta_t^-\right) dZ_t^{(1)},\end{aligned} \tag{44}$$

where the functions $m(s, \gamma; \Delta)$ and $v(s, \gamma; \Delta)$ are defined as in (24a) and (24b). See Figure 6 for an illustration of the equilibrium trading regions.

To complete the construction of the equilibrium, it now remains to solve for the equilibrium lending yield γ_t and to compute the asset prices. In the one asset case, the derivation follows the same steps as in Section 4.1. In particular, we find that the equilibrium shorting cost and lending yield are given by

$$(\Phi_t, \Gamma_t) = (\Phi, \Gamma)\left(s_t, \Delta_t^+\right) + (\Phi, \Gamma)\left(1 - s_t, \Delta_t^-\right),$$

with the functions $\Phi(s, \Delta)$ and $\Gamma(s, \Delta)$ implicitly defined by the right hand sides of (31) and (32). The comparative statics are very similar to those of the benchmark model

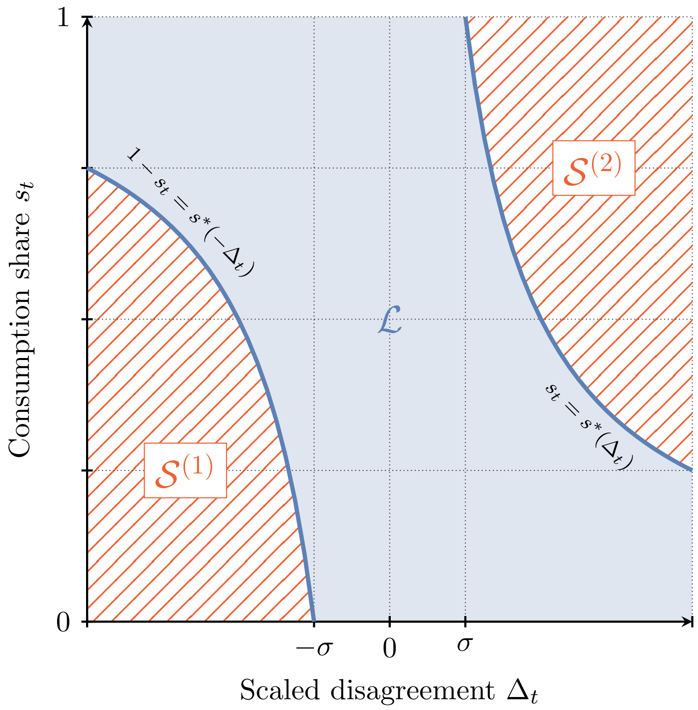

Figure 6: Trading regions with a stochastic disagreement. The figure illustrates the shape of the equilibrium trading regions and allows us to determine the configuration that occurs for each level of disagreement among agents.

with a constant disagreement. In particular, the flow rates (Γ_t, Φ_t) and the equilibrium utilization ratio

$$\Upsilon_t = \frac{1 - s_t}{\sigma^2} \Phi\left(s_t, \Delta_t^+\right) + \frac{s_t}{\sigma^2} \Phi\left(1 - s_t, \Delta_t^-\right)$$

are all convex and u-shaped in the disagreement Δ_t.

The representation of equilibrium prices—or of the fundamental value in the one asset case—is slightly more complex than in the benchmark model because each agent successively participates on both sides of the shorting market. Proposition 1 and the above characterization of the equilibrium trading regions imply that the normalized marginal utility of agent 1 evolves according to

$$-d\xi_t^{(1)}/\xi_t^{(1)} = r_t dt + \left(\theta_{1t}^{(1)} + \mathbb{1}_{\{\mathcal{S}^{(2)}\}}\gamma_t + \mathbb{1}_{\{\mathcal{S}^{(1)}\}}\phi_t\right) dZ_t^{(1)}.$$

Under appropriate integrability assumptions on the disagreement process Δ_t, this expression can be combined with arguments similar to those of the proof of Proposition 2 to show that the equilibrium prices satisfy

$$S_{2t} = E_t^{(1)}\left[\int_t^\infty \xi_{t,u}^{(1)} e_{2u} du\right],$$

and

$$S_{1t} = E_t^{(1)} \left[\int_t^\infty \xi_{t,u}^{(1)} \left(e_{1u} + \mathbb{1}_{\left\{s_u > s^*(\Delta_u^+)\right\}} S_{1u}\Gamma_u + \mathbb{1}_{\left\{1-s_u > s^*(\Delta_u^-)\right\}} S_{1u}\Phi_u \right) du \right]. \tag{45}$$

To understand this expression, note that from the point of view of agent 1 the cash flows that are relevant to the equilibrium valuation of asset 1 depend on which side of the shorting market the agent is. On the set $\mathcal{L}$, the only relevant cash flow is the dividend e_{1t} since the shorting market is inactive. On the set $\mathcal{S}^{(2)}$, the agent is long in asset 1 so that the relevant cash flows are the dividend and the lending yield $S_{1t}\Gamma_t$ associated with each share of the asset, and finally on $\mathcal{S}^{(1)}$, the agent is short so that the relevant cash flows are now given by the dividend and the shorting cost $S_{1t}\Phi_t$ required to maintain a short position. Importantly, if the disagreement process is always positive then the latter region is empty and we recover (17).

To derive a differential equation for the equilibrium price of asset 1, we assume that the scaled disagreement follows an autonomous diffusion process

$$d\Delta_t = \mu(\Delta_t)dt + \Sigma(\Delta_t)dZ_t^{(1)},$$

with values in some set $\mathcal{D} \subset \mathbb{R}$ and then proceed as in Section 4.2 albeit with an additional state variable. Specifically, we look for an equilibrium in which

$$S_{1t} = w\left(s_t, \Delta_t\right) M_t$$

for some sufficiently regular function $w : [0,1] \times \mathcal{D} \to [0,1]$ such that

$$w(0, \Delta) = w(1, \Delta) = \eta_1, \qquad \forall \Delta \in \mathcal{D}. \tag{46}$$

Itô's lemma and (38) show that, in such a Markovian equilibrium, the diffusion coefficient of asset 1 satisfies

$$\begin{aligned} \frac{\text{diff}_t(S_1)}{M_t} = \sigma w\left(s_t, \Delta\right) &+ w'_\Delta(s_t, \Delta_t)\Sigma(\Delta_t) \\ &+ w'_s(s_t, \Delta_t)\left(v(s_t, \gamma_t, \Delta_t^+) - v(1 - s_t, \gamma_t, \Delta_t^-)\right). \end{aligned}$$

Substituting into (42) and (43) then gives a linear-quadratic system that implicitly determines the shorting cost and the lending fee as functions

$$\left(\Phi[w](s_t, \Delta_t), \Gamma[w](s_t, \Delta_t)\right)$$

of s_t, Δ_t, $w(s_t, \Delta_t)$, and the derivatives $(w'_s, w'_\Delta)(s_t, \Delta_t)$. Taking these functions as given, it follows from (44) that the endogenous state variable evolves according to

$$ds_t = \overline{m}[w](s_t, \Delta_t)dt + \overline{v}[w](s_t, \Delta_t)dZ_t^{(1)}$$

for some explicit drift and diffusion functions $(\overline{m}, \overline{v})[w](s, \Delta)$. This, in turn, implies that the pair (s_t, Δ_t) forms a Markov process and, since

$$e^{-\rho t}\frac{w(s_t,\Delta_t)}{s_t} + \int_0^t e^{-\rho u}\Big(\rho\eta_1 + \mathbb{1}_{\{s_u > s^*(\Delta_u^+)\}} w(s_u,\Delta_u)\Gamma[w](s_u,\Delta_u) \\ + \mathbb{1}_{\{1-s_u > s^*(\Delta_u^-)\}} w(s_u,\Delta_u)\Phi[w](s_u,\Delta_u)\Big)\frac{du}{s_u}$$

is a martingale as a result of (45), we deduce that the function $u \equiv w/s$ is a piecewise twice continuously differentiable solution to

$$(\rho - \beta[w](s,\Delta))\, u = \frac{\rho\eta_1}{s} + \mu(\Delta)u'_\Delta + \frac{1}{2}\Sigma(\Delta)^2 u''_{\Delta\Delta} \\ + \overline{m}[w](s,\Delta)u'_s + \overline{v}[w](s,\Delta)\Sigma(\Delta)u''_{s\Delta} + \frac{1}{2}\overline{v}[w](s,\Delta)^2 u''_{s\Delta},$$

subject to the boundary condition (46), where

$$\beta[w](s,\Delta) \equiv \mathbb{1}_{\{s > s^*(\Delta^+)\}}\Gamma[w](s,\Delta) + \mathbb{1}_{\{1-s > s^*(\Delta^-)\}}\Phi[w](s,\Delta)$$

denotes the additional cash flow per dollar of asset value in (45) as a function of the state variables s, Δ taking as given $w(\cdot)$ and $w'(\cdot)$.

A numerical solution to this nonlinear boundary value problem can in principle be constructed using the same collocation approach as in the constant disagreement case of Section 4.2, albeit in two dimensions and subject to the caveat that the differential equation no longer admits an explicit solution on the long region in general. We leave the challenges of this implementation for future research.