Smart Settlement

Mariana Khapko and Marius Zoican*

May 1, 2018

Abstract

Recent regulatory and FinTech initiatives aim to streamline post-trade infrastructures. Does faster settlement benefit markets? We build a model of intermediated trading with flexible settlement and imperfectly competitive securities lending. Faster settlement reduces counterparty risk, but increases borrowing needs. Rigid failure-to-deliver penalties trigger a toxic rat race, as traders aim to lock in low borrowing costs. Excess demand for fast settlement augments lenders' rents. Optimal penalties resemble put options on the security lending market: They protect traders against high settlement costs, but do not eliminate failures-to-deliver. Flexible penalties discipline security lender competition and facilitate faster trade settlement.

Keywords: Market design, trade settlement, security lending, counterparty risk

JEL Codes: D43, D47, G10, G20

---

*Mariana Khapko is affiliated with University of Toronto and Rotman School of Management. Marius Zoican is affiliated with Université Paris-Dauphine. Mariana Khapko can be contacted at mariana.khapko@rotman.utoronto.ca. Marius Zoican can be contacted at marius-andrei.zoican@dauphine.fr. We have greatly benefited from discussion on this research with Matthieu Bouvard, Sabrina Buti, Jean-Edouard Colliard, Gilles Chemla, Olivier Dessaint, Björn Hagströmer, Carole Gresse, Sean Foley, Thierry Foucault, Florian Heider, Andrei Kirilenko, Charles-Albert Lehalle, Katya Malinova, Albert Menkveld, Sophie Moinas, John Nash, Maureen O'Hara, James Thompson, Tomas Thörnqvist, Emiliano Pagnotta, Andreas Park, Christine Parlour, Batchimeg Sambalaibat, Liyan Yang, Darya Yuferova, and Haoxiang Zhu. We are grateful to conference participants at the 2017 NBER Meeting on Competition and the Industrial Organization of Securities Markets, the 2017 SFS Cavalcade, 2017 EFA Meeting, the 13th Central Bank Conference on the Microstructure of Financial Markets, the Swedish House of Finance FinTech Conference, 2017 NFA Meeting, 2017 SFA Meeting, 2nd FIRN Market Microstructure Meeting, Eighth Erasmus Liquidity Conference, Twelfth Early Career Women in Finance Conference, 10th Risk Management Forum in Paris, 2nd Financial Market Infrastructure Conference, IRMC 2017, IFABS 2017, FEBS 2017, Toronto FinTech Conference 2017, as well as to seminar participants at University of Toronto, HKUST, Toulouse School of Economics, HEC Paris, KU Leuven, Norwegian School of Economics, Tilburg University, WHU Vallendar, SAFE Frankfurt, Authorité de Marchés Financiers, Stockholm Business School, ESCP Europe, Luxembourg School of Finance, and Université Paris-Dauphine for insightful comments. Mariana Khapko gratefully acknowledges University of Toronto and the Connaught Fund for a research grant. Marius Zoican gratefully acknowledges Institute Europlace de Finance for a research grant.

Smart Settlement

Abstract

Recent regulatory and FinTech initiatives aim to streamline post-trade infrastructures. Does faster settlement benefit markets? We build a model of intermediated trading with flexible settlement and imperfectly competitive securities lending. Faster settlement reduces counterparty risk, but increases borrowing needs. Rigid failure-to-deliver penalties trigger a toxic rat race, as traders aim to lock in low borrowing costs. Excess demand for fast settlement augments lenders' rents. Optimal penalties resemble put options on the security lending market: They protect traders against high settlement costs, but do not eliminate failures-to-deliver. Flexible penalties discipline security lender competition and facilitate faster trade settlement.

Keywords: Market design, trade settlement, security lending, counterparty risk
JEL Codes: D43, D47, G10, G20

1 Introduction

How fast should trades be settled? A number of recent reforms showcase markets' appetite for faster settlement: In September 2017, the U.S. and Canadian markets migrated from a three-day (T+3) to a two-day (T+2) settlement cycle for equity transactions. Previously, European equity markets had transitioned to T+2 in 2014; Singapore and Japan plan similar reforms in 2018 and 2019, respectively.

Faster settlement reduces counterparty risk. A Boston Consulting Group study (BCG, 2012) commissioned by the Depository Trust & Clearing Corporation (DTCC) estimated a 38% (USD 1 bln) market-wide drop in counterparty risk exposure following U.S. markets' transition from T+3 to T+2. Indeed, a significant number of trades, both exchange-based and OTC, fail to settle on time every day. Reuters estimates that, on an average trading day in 2011, failures-to-deliver on the U.S. equity market reached 4.3% of total traded volume. In 1995, under a five-day settlement cycle, the U.S. equity failure-to-deliver rate amounted to 8.43% (Levitt, 1996). The failure rate is even higher and more volatile for other asset classes: In the week of March 9, 2016, trade failures in U.S. government bonds spiked at $456bn compared to a weekly average of $94bn the year before.[1]

Faster settlement, however, is costly. First, the migration process itself requires a complete overhaul of back-office processes. Second, and more importantly, faster settlement requires broker-dealers to either borrow the securities they need to settle, or to pre-position their trades, and therefore to carry inventory risk. Security lenders typically wield market power, as demonstrated, for example, by recent U.S. lawsuits in 2017 filed by pension funds. Anecdotal evidence suggests that these costs are important: In 2013, the Moscow Stock Exchange transitioned from immediate settlement to T+2, citing security borrowing costs as a key rationale.[2]

Our paper formalizes the trade-off between counterparty risk and security borrowing costs. On the one hand, a long delay between trade and settlement increases exposure to counterparty risk. On the other hand, a short settlement cycle augments expected security borrowing costs. The paper's contribution is (i) to study the impact of settlement cycle length and flexibility on market quality, as well as (ii) to establish the optimal structure of failure-to-deliver penalties.

[1] The sources for this paragraph are the Reuters' 2012 article Persistent trade failure problem gets partial fix, and Financial Times' 2016 article Trade failures of US bonds hit $456bn.

[2] See Reuters, U.S. pension funds sue Goldman, JPMorgan, others over stock lending market (2017); also Bloomberg's 2013 article Moscow Targets March Move to 2 Day Settlement.

Regulators have long envisioned the possibility of flexible trade settlement. In 1996, following the U.S. markets' migration from a five- to a three-day settlement cycle, the Securities and Exchange Commission (SEC) chairman Arthur Levitt emphasized the "staggering interdependence" between financial infrastructures in trade settlement and called for a less rigid system to accommodate the preferences of individual investors (Levitt, 1996). Currently, each trade needs to be processed by a number of different institutions from execution on the exchange until settlement, such as brokerage firms, custodian banks, clearing agencies, and central securities depositories. In a landscape in which traders can update quotes with nanosecond frequency, the state of post-trade financial infrastructure feels dated.

New technologies such as distributed ledgers (e.g., but not limited to, Blockchain) could facilitate the implementation of flexible settlement chains. For instance, in April 2018, the Australian Stock Exchange (ASX) revealed a detailed plan to implement distributed ledger technology (DLT) for securities settlement before 2021. The new system would offer traders the option to settle their trades earlier than T+2.[3] Furthermore, failure-to-deliver contingencies could be accounted for by the use of "smart contracts" (defined by Cong and He, 2018 as self-enforcing digital contracts with automated execution). For example, ASX plans to introduce mandatory automatic security borrowing to rule out failures-to-deliver.

We build a model of interconnected security trading and security lending (repo) markets, with delivery-versus-payment settlement, in which we introduce three frictions.[4] First, buyers and sellers arrive asynchronously at the market (as in Grossman and Miller, 1988) and therefore trading is intermediated. Second, intermediaries can exogenously default: consequently, traders bear counterparty risk. Third, the lending supply is limited and the security lending market is opaque. As in Duffie, Gârleanu, and Pedersen (2002), security lenders can earn rents and borrowing is costly for intermediaries. We distinguish between default and failure-to-deliver: Trades which fail to deliver are eventually settled if the intermediary does not default before locating the asset.

Our main result is that traders' utility is maximized by a contract featuring immediate settlement and a flexible failure-to-deliver penalty. The flexible penalty implements an incentive compatible contract between buyers and intermediaries. That is, an intermediary

[3]See ASX, CHESS Replacement: New Scope and Implementation Plan, April 2018. Other exchanges express a similar interest in DLT and trade settlement flexibility: Fredrik Voss, VP of Blockchain Innovation at Nasdaq, acknowledges DLT would "allow participants to select the pace at which they want to settle, which has been challenging to do in the market today." (`https://goo.gl/mv68il`).

[4]Tobias, Begalle, Copeland, and Martin (2013) point out that securities lending and repo agreements are economically similar, albeit used in the context of different asset classes. In this paper, we use the two terms interchangeably.

only accesses the repo market when the cost of borrowing securities is lower than the marginal counterparty risk it imposes on a buyer through a failure-to-deliver. The optimal penalty is equivalent to a put option written on the repo market price, where the "strike price" increases in counterparty risk and trading urgency. Importantly, the optimal contract does not rule out failures-to-deliver in equilibrium.

An exogenous, large failure-to-deliver penalty ensures that all trades are settled on time, as intermediaries always borrows securities if needed.[5] In this case, immediate settlement is not necessarily optimal. If borrowing costs are large, counterparty risk is low, or both, buyers prefer delayed settlement instead. A longer time-to-settlement allows intermediaries more time to wait for an opposite-side trade, which reduces the likelihood to borrow securities. Intuitively, the optimal settlement time decreases in intermediaries' default risk. If repo markets are perfectly competitive, settlement flexibility allows traders with higher urgency to reduce counterparty risk by settling before others, at a higher price.

If repo markets are not competitive, and if the regulator rules out failures-to-deliver by imposing a large penalty, flexibility leads to a toxic settlement rat race. The intuition is that a larger lending supply strengthens competition between security lenders, and lowers repo borrowing costs for intermediaries. Trades which settle first consume part of the available lending supply, consequently increasing borrowing costs for all future trades to be settled. It follows that each buyer has an incentive to settle before everyone else. Two equilibria can obtain, depending on model parameters: an *immediate settlement* equilibrium and a *delayed settlement* equilibrium, where buyers randomize between a continuum of positive settlement delays and, potentially, also immediate settlement. The settlement rat race is costly as it leads to shorter-than-optimal settlement cycles, excessive security borrowing activity, and economic rents for security lenders.

A flexible failure-to-deliver penalty eliminates the settlement rat race as it stimulates competition between security lenders. In equilibrium, the penalty increases in counterparty risk. For low counterparty risk, intermediaries are given weaker incentives to borrow. Therefore, security lenders compete stronger on prices in order to attract intermediaries. For high counterparty risk, security lenders' rents are larger as they are better able to hold up intermediaries who face higher failure penalties. In equilibrium, trades fail to deliver if counterparty risk is low. Further, the failure-to-deliver rate increases in the security's specialness, that is the opportunity cost to supply an instrument on the security lending

[5]Currently, the Fixed Income Clearing Corporation (FICC) sets a failure-to-deliver penalty at 3% on the settlement value of the trade (minus the Target Fed funds rate in effect the day before the settlement day).

market.

In sum, the paper's message is that a flexible or a short settlement cycle (i.e, T+0), coupled with an option-like penalty for failures-to-deliver, disciplines competition on securities lending markets and therefore improves market quality. Our results support the ongoing reforms around the globe to shorten settlement cycles, and offer insights on how to reduce the costs associated with fast settlement.

2 Related Literature

Our paper contributes to an active discussion on the design of post-trade financial infrastructures. It relates to several branches of literature in the field of market microstructure and financial innovation.

Literature on the securities lending market. D'Avolio (2002) provides a comprehensive overview of the U.S. securities lending market. Large custodian banks are among the most important lenders, acting on behalf of institutional traders such as pension or mutual funds, and especially passive indexeres (i.e., ETFs). Security borrowers are either (i) specialists and market-makers matching buy and sell orders, (ii) derivative traders hedging their positions, or (iii) speculators taking outright short positions. IHS Markit (2016) estimates the global supply of securities in lending programs at USD 15 trillion, generating USD 8.06 billion in lending revenues for the year 2016.

Security lending markets typically exhibit imperfect competition. Chague, De-Losso, Genaro, and Giovannetti (2017) study the Brazilian equity lending market and find that borrowers with higher search costs face higher loan fees. In the same spirit, Huszár and Simon (2018) document that German treasuries lenders behave as oligopolists. Perhaps the most closely related paper to ours, Duffie, Gârleanu, and Pedersen (2002) model the security lending market as a sequential Nash bargaining game, where high search costs allow security lenders to extract economic rents from borrowers. Trade prices are determined on a Walrasian spot market. In contrast, buyers and sellers in our model arrive at the market asynchronously (as in, e.g., Grossman and Miller, 1988), and trading is intermediated. Therefore, a longer settlement delay improves the likelihood of a match on the spot market, and reduces expected security borrowing costs. We solve for the optimal incentive compatible intermediation contract and the corresponding failure-to-deliver penalty.

Beyond imperfect competition, lending fees also depend on the opportunity or transaction costs to supply an instrument on the security lending market. Small-cap, non-index, or illiquid stocks are more difficult to borrow, as are stocks with concentrated ownership (for example, around initial public offerings, as in Geczy, Musto, and Reed, 2002). Duffie (1996) refers to such lending constraints as part of a security's *specialness* and argues that specialness drives up prices. Indeed, Jordan and Jordan (1997) document a liquidity premium for repo specialness for on-the-run U.S. Treasuries. Further, D'Avolio (2002) finds that lending fees can reach "spectacular heights," that is more than 55% in annualized terms.

Several empirical papers show that the supply of securities available for lending correlates with liquidity and failures-to-deliver. Foley-Fisher, Gissler, and Verani (2017) find that the collapse of AIG's security lending programs in 2008 caused a substantial and long-lasting reduction in market liquidity for corporate bonds. Corradin and Maddaloni (2017) document that the 2011 outright purchase program of the Eurosystem for Italian sovereign bonds, which reduced their supply, increased borrowing costs and led to more fail-to-deliver transactions. Finally, Huszár and Simon (2018) find that the withdrawal of German banks from security lending markets around year's end, for financial reporting purposes, reduces the supply of lendable securities and increases lending fees. In our model, failures-to-deliver occur with positive probability even if the optimal contract is implemented, and are more likely if the security is in limited supply.

Literature on counterparty risk. We also contribute to a strand of literature studying counterparty risk. In the aftermath of the 2007-2009 financial crisis, several papers focus the impact of a central clearing counterparty on market quality (Acharya and Bisin, 2014, Acharya, 2009, Pirrong, 2009, Stephens and Thompson, 2017). Loon and Zhong (2014) document that counterparty risk is priced by the market and that the introduction of a CCP in 2009, following the Dodd-Frank Act, led to a decrease in counterparty risk in the CDS market. Duffie and Zhu (2011) discuss economies of scope from having a single CCP across all asset classes. To control systemic risk, Menkveld (2017) proposes a collateral system that internalizes the global effects of highly correlated portfolios across traders, i.e., the "crowding risk." Failures-to-deliver may be caused by miscommunication and operational problems, a failure to receive the assets from an unrelated trade, or short selling (Fleming and Garbade, 2005). Evans, Moussawi, Pagano, and Sedunov (2017) and Fotak, Raman, and Yadav (2014) focus on failures-to-deliver driven by short-selling and find a negative correlation between liquidity and settlement risk.

Central counterparties' main benefit comes through netting existing positions, that is reducing the size of risk exposure. Our paper complements the existing counterparty risk literature by focusing on the *length* of risk exposure.

Literature on financial infrastructure technology. Finally, our paper contributes to a small, yet rapidly growing literature on financial technology in securities markets. We argue that flexible settlement and fail-to-deliver penalties, as implemented for example by smart contracts on a distributed ledger, improve market quality. Malinova and Park (2017) refer to distributed ledgers and smart contracts as "the technology that enables frictionless transfer of value." The regulatory consensus is that distributed ledger technology can eliminate the need for multiple records of a single trade and consequently reduce the number of post-trade intermediaries (ECB, 2016, ESMA, 2016, Brainard, 2016). Investments are already underway: In January 2017, the Depository Trust and Clearing Corporation (DTCC), which processes \$1.5 quadrillion in trades a year, announced a plan to deploy DLT technology for post-trade processing in early 2018.[6]

Benos, Garratt, and Gurrola-Perez (2017) argue that distributed ledgers have the potential to improve efficiency and reduce costs in securities settlement. Koeppl and Chiu (2018) find that large trading volume and strong preferences for fast settlement are necessary conditions for the viability of a public, anonymous Blockchain trade settlement system.[7] We argue that immediate settlement can indeed be optimal, provided there is enough flexibility in failure-to-deliver penalties.

3 Model

3.1 Model primitives

Asset. Consider a continuous-time economy. As in Kyle and Obizhaeva (2017), the game ends at an arbitrarily large but finite T, at which time utilities are evaluated. A single asset pays off a stochastic dividend $v \sim \mathcal{N}(0, \sigma^2)$ at T. The discount rate is normalized to zero.

[6]See Forbes report from January 9, 2017.

[7]A number of papers are also studying the feasibility of trading on a public blockchain: Biais, Bisiere, Bouvard, and Casamatta (2018), Aune, O'Hara, and Slama (2017), and Saleh (2017).

Agents. There are four types of agents in the economy: two identical buyers, two competitive intermediaries, one large seller, and a large number of risk-averse security lenders.[8] Security lenders are typically institutional investors such as pension funds, who we think of as more risk-averse than high-frequency traders or market-makers. Buyers, intermediaries, and sellers are risk-neutral.

Buyers (**B**) can only trade at $t = 0$ and have no endowment of the risky asset. Buyers trade to lock in a private value $\theta > 0$ for the first unit of the risky asset.

Intermediaries (**I**), unlike buyers, are always present at the market. They have no endowment of the risky asset and no private values. Further, intermediaries have access to a risk-free productive technology yielding $R > 1$ at T for every unit invested at $t = 0$, standing in for the intermediaries' shadow cost of capital. They can partially liquidate the investment at any point in time, but cannot reinvest. Intermediaries may exogenously default (e.g., due to shocks in unrelated business activities). The default time is exponentially distributed with rate parameter δ, that is, the probability of default before t is

$$\mathbb{P}\left(\text{Default} < t\right) = 1 - \exp\left(-\delta t\right). \tag{1}$$

The large seller (**S**) has an endowment of at least two units of the risky asset. She arrives at the market at a random time which follows an exponential distribution with rate parameter λ. The large seller arrives at the market before t with probability

$$\mathbb{P}\left(\text{Seller} < t\right) = 1 - \exp\left(-\lambda t\right). \tag{2}$$

Security lenders (**L**) have mean-variance preferences. There are N lenders with risk-aversion coefficient γ and a competitive fringe of lenders with risk-aversion $\Gamma > \gamma$.[9] With probability φ, each lender owns a unit of the risky asset as a perfect hedge for an endowment of an asset that pays off $-v$. Ownership of the risky asset is private information for each lender. A lower φ corresponds to higher *specialness* of the security.

Two assumptions about security lenders are particularly important. First, security lenders have heterogenous costs of supplying the asset. Second, the lendable supply is limited in the following sense: The competitive supply schedule is not perfectly elastic, and therefore

[8]The model could be extended to feature an arbitrary number of buyers and intermediaries without altering the economic trade-offs. The present setup is chosen to ease exposition. In particular, the buyers need not be matched with different intermediaries.

[9]Normalizing $\gamma = 0$, that is assuming some security lenders are risk-neutral, does not change the model results. A positive $\gamma > 0$ allows for an equilibrium contract in which the intermediary never borrows the asset.

intermediaries cannot borrow large quantities without moving the repo price. We choose to model the lending cost heterogeneity through variation in risk aversion. Alternatively, one can model such heterogeneity as variation in search costs. For example, not all prime brokers have access to electronic security lending platforms such as EquiLend. Further, opacity is a realistic feature of security lending markets. In the model, the opacity assumption allows us to back out a non-trivial cross-sectional distribution of repo haircuts. However, the results largely go through also if we assume a transparent repo market.

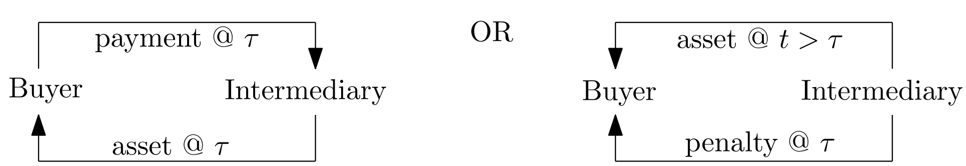
Contract structure. At the start of the game, competitive intermediaries post delivery-versus-payment (DvP) contracts to buy and to sell the asset. A contract to sell the asset to **B** specifies:

1. Time-to-settlement for the trade, $\tau \geq 0$.
2. Payment $p(\tau)$ by **B** to **I** upon settlement at τ.
3. Penalty $z(\tau)$ paid by **I** to **B** at τ upon failure-to-deliver.

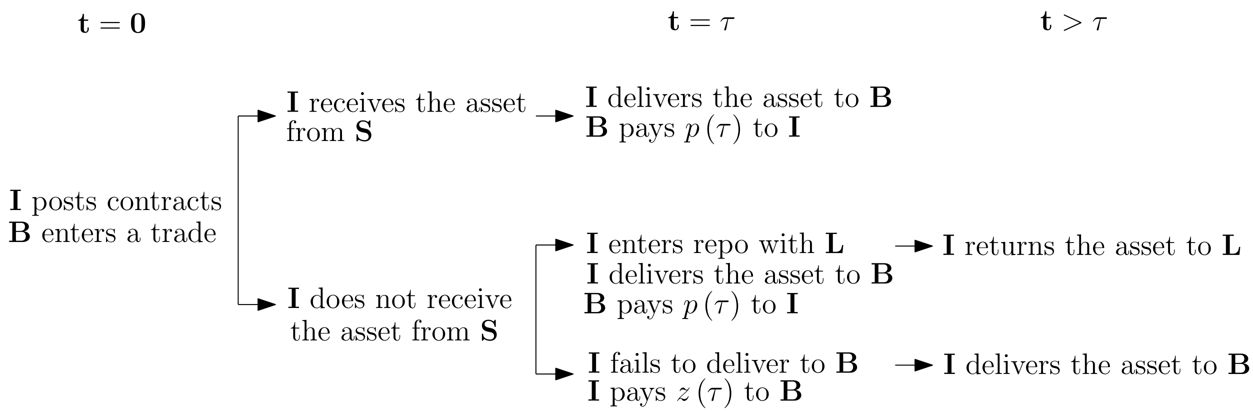
Buy contracts, that is contracts where **I** buys the asset from **S**, can be defined symmetrically. However, since the large seller (i) never defaults, (ii) is risk-neutral and has no private value for the risky asset, and (iii) is originally endowed with at least two units of the risky asset, such buy contracts are trivial. Since intermediaries post offers at $t = 0$, it is optimal for them to set a price of zero since **S** accepts to sell at any non-negative price. Further, the optimal settlement delay is also zero since it minimizes the probability of needing to borrow to settle the sell contract. Finally, since the large seller always settles, any positive penalty ensures there will be no failures-to-deliver.

If the intermediary fails to deliver at τ, her marginal cost of late delivery is zero (i.e., the price paid to **S**). Competitive intermediaries charge a price equal to their marginal cost conditional on non-delivery, that is they receive no payment after τ. They deliver the asset due to, for example, unmodeled reputational concerns.

In what follows, we focus on sell contracts. The structure of the sell contract is illustrated below.



Timing. At $t = 0$, buyers arrive at the market and either choose one of the contracts posted by the intermediaries, or do not trade. When the large seller arrives, she enters a contract with one or more intermediaries: The large seller's contracts settle immediately with a price of zero.



At τ, the intermediary needs to deliver the asset to the buyer. If she does not hold the asset, the intermediary can enter a repurchasing contract with a security lender, to be detailed below.[10] Otherwise, the intermediary fails to settle: **I** pays $z(\tau)$ to the buyer at τ and delivers the asset upon arrival of the seller at $t > \tau$. At $t = T$, payoffs are realized.

Repurchase agreements. The intermediary can enter a repurchasing agreement (repo) with a security lender at τ. In the start leg of the contract, the intermediary **I** pays ℓ to the security lender **L** in exchange for the risky asset. To pay ℓ, the intermediary liquidates part of the risk-free investment, at cost $R\ell$. In the close leg of the contract, **I** delivers the risky asset to **L** and **L** repays ℓ' to **I**. The repo contract is fully collateralized in the following sense: The lender is perfectly hedged against the risk of the intermediary defaulting on the repo contract (i.e., not returning the risky asset). Finally, the repo market is opaque: each lender posts two prices ℓ and ℓ' and is unaware of other lenders' offers.

Let b be the (potentially random) minimum borrowing cost for the intermediary across all security lenders who own the asset, that is

$$b = \min_{i \in \mathbf{L}} R\ell_i - \ell'_i. \tag{3}$$

[10]For tractability, we assume the intermediary borrows the security at the settlement time τ. Alternatively, the intermediary can be allowed to choose the optimal borrowing time before τ. Rational expectations competitive equilibria under the two specifications are isomorphic since buyers perfectly anticipate and incorporate the intermediaries' strategies.

If at τ the intermediary does not own the asset, she borrows if the penalty for late delivery is larger than the borrowing cost minus the trade price received upon timely settlement, that is if the following borrowing condition is true:

$$z(\tau) \geq b - p(\tau). \tag{\textbf{BC}}$$

Moreover, let β_N denote the expected net borrowing cost conditional on there being N lenders with risk aversion γ, in addition to the competitive fringe lenders.

$$\beta_N = \mathbb{E}\left[b \mid N \text{ lenders with risk aversion } \gamma\right]. \tag{4}$$

Equilibrium. We look for symmetric subgame-perfect Nash equilibria in pure and mixed strategies.

> **Definition 1.** An *equilibrium* of the trading game consists of (i) a set of contracts $\{\tau, p(\tau), z(\tau)\}$ posted by each intermediary **I** at $t = 0$ to sell the risky asset; (ii) contract choices of each **B**; (iii) repo contract prices ℓ, ℓ' set by each lender **L**, and (iv) each **I**'s decision at settlement on whether to enter a repo contract and at which price, such that no agent is strictly better off deviating.

3.2 Payoffs

Buyer. First off, consider **B-I** contracts for which the borrowing condition (BC) holds. That is, the intermediary is better off borrowing the asset at τ rather than paying the late-settlement penalty. In this case, the buyer always receives the asset at τ if **I** does not default until τ, regardless of whether **S** arrives at the market or not. The expected utility of the buyer is

$$\mathbb{E}\left[U_{\mathbf{B}} \mid \text{BC true}\right] = e^{-\delta\tau}\left(\theta - p(\tau)\right). \tag{5}$$

Alternatively, consider a **B-I** contract for which the intermediary optimally fails to settle at τ (i.e., borrowing on the repo market is too expensive and BC is false). In this case, the buyer receives the asset at τ if and only if the large seller arrives at the market until then. Otherwise, the buyer may receive the asset at $t > \tau$ if the **S** arrives at the market before **I** defaults. The expected utility of a buyer, if (BC) does not hold, is

$$\mathbb{E}\left[U_{\mathbf{B}} \mid \text{BC false}\right] = e^{-\delta\tau}\left[\left(1 - e^{-\lambda\tau}\right)\left(\theta - p(\tau)\right) + e^{-\lambda\tau}\left(z(\tau) + \frac{\lambda}{\lambda+\delta}\theta\right)\right]. \tag{6}$$

With probability $e^{-\delta\tau}\left(1-e^{-\lambda\tau}\right)$, **S** arrives at the market before τ and **I** does not default: settlement occurs at τ, the buyer receives the asset and pays $p\left(\tau\right)$. With probability $e^{-(\delta+\lambda)\tau}$, **I** pays penalty $z\left(\tau\right)$ to the buyer at τ. In this case, the probability of late settlement, for a large T, can be approximated by

$$\lim_{T\to\infty}\int_{\tau}^{T}\int_{\tau}^{x}\lambda\delta e^{-\lambda y-\delta x}\,\mathrm{d}y\,\mathrm{d}x+\int_{T}^{\infty}\delta e^{-\delta x}\,\mathrm{d}x=e^{-(\lambda+\delta)\tau}\frac{\lambda}{\lambda+\delta}. \tag{7}$$

Intermediary. The expected utility of the intermediary can be written as

$$\mathbb{E}U_{\mathbf{I}}=e^{-\delta T}\mathbb{E}\left[U_{\mathbf{I}}\mid\mathbf{I}\text{ survives}\right], \tag{8}$$

where $\exp\left(-\delta T\right)$ is the probability that **I** does not default until the end of the game. The expected utility of the intermediary, conditional on survival until T, is:

$$\mathbb{E}\left[U_{\mathbf{I}}\mid\mathbf{I}\text{ survives}\right]=\left(1-e^{-\lambda\tau}\right)p\left(\tau\right)+e^{-\lambda\tau}\mathbb{E}\left[\max\left\{p\left(\tau\right)-b,-z\left(\tau\right)\right\}\right]. \tag{9}$$

With probability $1-\exp\left(-\lambda\tau\right)$, **S** arrives at the market until τ. The intermediary buys the asset from the large seller at a price of zero, delivers it at τ to **B** and cashes in the price $p\left(\tau\right)$. With the complementary probability, **S** does not arrive until τ. The intermediary chooses the best of two options: either to pay the penalty $z\left(\tau\right)$, or to enter a repo agreement with **L** at net cost b and deliver at τ against the payment $p\left(\tau\right)$.

Competitive intermediaries earn no excess profits in equilibrium. Since T is an arbitrarily large number, $\exp\left(-\delta T\right)$ is a very small, but *positive* probability. From equation (8), it follows that the competitive condition is equivalent to setting the conditional expected utility in equation (9) to zero.

Lender. The lender **L** receives the repo price ℓ if the intermediary enters a repurchase agreement at τ. If the intermediary does not return the asset, the security lender is exposed to his initial endowment of $-v$ and has expected utility:

$$\mathbb{E}\left[U_{\mathbf{L}}\mid\mathbf{I}\text{ does not return the asset}\right]=\ell-\frac{g}{2}\sigma^{2}\geq0\Rightarrow\ell\geq\frac{g}{2}\sigma^{2},\text{ for }g\in\left\{\gamma,\Gamma\right\}. \tag{10}$$

If the intermediary returns the asset, the lender is perfectly hedged against his initial endowment and has expected utility

$$\mathbb{E}\left[U_{\mathbf{L}} \mid \mathbf{I} \text{ returns the asset}\right] = \ell - \ell'. \tag{11}$$

We restrict our attention to repurchasing agreements that provide a perfect hedge for the lender against the intermediary not returning the asset, it must be that[11]

$$\mathbb{E}\left[U_{\mathbf{L}} \mid \mathbf{I} \text{ does not return the asset}\right] = \mathbb{E}\left[U_{\mathbf{L}} \mid \mathbf{I} \text{ returns the asset}\right] \Rightarrow \ell' = \frac{g}{2}\sigma^2. \tag{12}$$

4 Benchmark equilibrium

In this section, we illustrate how the settlement delay τ drives the key trade-off between counterparty risk and transaction costs. As a point of reference, we consider a variant of the model under two simplifying assumptions.

First, we set $N = 0$ so that intermediaries may only borrow the risky asset from the competitive fringe of security lenders. In Section 5 we allow for $N > 0$, and consequently introduce imperfect competition between security lenders on an opaque repo market. Second, we assume the failure-to-deliver penalty z is set exogenously (e.g., by a regulator) and is large enough such that (**BC**) is always true. This assumption is consistent with, for example, the exchange-mandated automatic borrowing feature proposed by the Australian Stock Exchange. We further assume a large enough private value, such that the buyer has positive expected utility from immediate settlement. This is the case if $\theta > \beta_0$, where β_0 is defined in (4). In Section 6, the buyer and intermediary can contract on the penalty in the contract, allowing for endogenous failures-to-deliver when (**BC**) is false.

Security lenders in the competitive fringe obtain zero expected utility. Therefore, from equation (10), each lender in the competitive fringe offers a repo price $\ell = \frac{\Gamma}{2}\sigma^2$. Consequently, the borrowing cost for the intermediary is β_0, where from equation (4) it follows that

$$\beta_0 = \underbrace{R\ell}_{\text{open leg}} - \underbrace{\ell'}_{\text{close leg}} = \frac{\Gamma}{2}\sigma^2\left(R-1\right). \tag{13}$$

Intuitively, the borrowing cost increases in the lender's risk-aversion, asset volatility, and the opportunity cost of the intermediary (i.e., the safe investment return).

[11] A natural requirement for the repo contract is that the lender is never worse off if the intermediary returns the asset, that is $\ell' \leq \frac{g}{2}\sigma^2$ for $g \in \{\gamma, \Gamma\}$.

From equations (9) and (13), and since the borrowing condition (**BC**) is true, the expected utility of the intermediary is

$$\mathbb{E}\left[U_{\mathbf{I}} \mid \mathbf{I} \text{ survives}\right] = p(\tau) - e^{-\lambda\tau}\frac{\Gamma}{2}\sigma^2 (R-1). \tag{14}$$

Since competitive intermediaries earn zero expected profit, equation (14) pins down the price on the **B-I** contract as a function of the settlement delay τ, that is

$$p(\tau) = \underbrace{e^{-\lambda\tau}}_{\substack{\text{borrowing}\\\text{probability}}} \underbrace{\frac{\Gamma}{2}\sigma^2 (R-1)}_{\substack{\text{borrowing}\\\text{cost}}}. \tag{15}$$

Faster settlement is more expensive: The competitive price decreases in the time to settlement τ. Intermediaries are more likely to need to enter a costly repurchasing agreement if the settlement delay is smaller, as the likelihood of an **S** arrival before τ decreases. The competitive intermediaries transfer any increase in expected costs to buyers through the price $p(\tau)$.

At $t = 0$, each buyer observes the competitive intermediary price schedules (15) and chooses the contract $\{\tau, p(\tau)\}$ that maximizes his expected utility. From equations (5) and (15), **B**'s expected utility is:

$$\mathbb{E}\left[U_{\mathbf{B}}\right] = \underbrace{e^{-\delta\tau}}_{\substack{\text{settlement}\\\text{probability}}} \underbrace{\left[\theta - e^{-\lambda\tau}\frac{\Gamma}{2}\sigma^2 (R-1)\right]}_{\text{payoff upon settlement}}. \tag{16}$$

On the one hand, faster settlement (i.e., lower τ) benefits the buyer as it implies lower counterparty risk, that is, a higher settlement probability. On the other hand, faster settlement is more expensive, as the buyer compensates the intermediary for the higher expected borrowing cost.

Proposition 1 describes the equilibrium of the simplified trading game.

Proposition 1. (Benchmark equilibrium) *If $N = 0$ and the failure-to-deliver penalty is sufficiently large such that* (**BC**) *is always true, then the following strategies form an equilibrium:*

(i) Intermediaries post $\{\tau, p(\tau)\}$ contracts such that: $p(\tau) = e^{-\lambda\tau}\frac{\Gamma}{2}\sigma^2 (R-1)$.

(ii) Buyers choose the contract with time-to-settlement equal to $\tau^\star$, where

$$\tau^\star = \max\left\{0, \frac{1}{\lambda}\log\left[\frac{(\delta+\lambda)}{\theta\delta}\frac{\Gamma}{2}\sigma^2(R-1)\right]\right\}. \tag{17}$$

(iii) Competitive fringe security lenders set repo purchasing price $\ell = \frac{\Gamma}{2}\sigma^2$.

(iv) Intermediaries borrow the asset at $\tau^\star$ if the large seller did not arrive before $\tau^\star$.

From equation (17), immediate settlement is optimal for high enough default risk δ or for low enough borrowing costs, that is if

$$\frac{\delta}{\delta+\lambda}\theta \geq \beta_0 = \frac{\Gamma}{2}\sigma^2(R-1). \tag{18}$$

Corollary 1 provides comparative statics results for the equilibrium time-to-settlement in the simplified game.

Corollary 1. (Benchmark comparative statics) *The optimal time-to-settlement $\tau^\star$ weakly*

(i) increases in the lender's risk aversion (Γ), *asset volatility* (σ), *and the intermediary's opportunity cost* (R), *and*

(ii) decreases in counterparty risk (δ) *and buyer private value* (θ).

Finally, there exists a $\Lambda_0 > 0$ such that $\tau^\star$ increases in λ for $\lambda \leq \Lambda_0$ and $\tau^\star$ decreases in λ for $\lambda > \Lambda_0$.

Buyers choose the time-to-settlement $\tau^\star$ that optimally trades off counterparty risk against the expected repo borrowing cost. A higher default rate for the intermediary leads to faster settlement in equilibrium. Conversely, buyers choose a longer settlement delay if the borrowing cost is larger (either due to asset volatility, **L**'s risk-aversion, or opportunity costs for the intermediary). A longer settlement delay increases the large seller's arrival probability and lowers the likelihood of a costly repo agreement.

From Proposition 1, it immediately follows that the equilibrium price paid by buyers upon settlement at $\tau^\star$ is:

$$p(\tau^\star) = \min\left\{\frac{\delta}{\delta+\lambda}\theta, \frac{\Gamma}{2}\sigma^2(R-1)\right\}. \tag{19}$$

If immediate settlement is optimal, **I** always enters a repo agreement: Therefore, the price is equal to the net borrowing cost. Otherwise, if delayed settlement is optimal, the price paid by **B** at $\tau^\star$ increases in **I**'s default rate and buyers' private value as $\tau^\star$ is lower and **I** is more likely to borrow the risky asset. At the same time, the settlement price decreases in the large seller's arrival rate as borrowing is less likely for higher λ.

The large seller arrival rate λ has a non-linear impact on the equilibrium time-to-settlement. In markets with plenty of potential sellers, that is for very high λ, buyers choose fast settlement as it yields both low counterparty risk and a sufficiently low likelihood of **I** entering a repo agreement (formally, $\lim_{\lambda\to\infty} \tau^\star = 0$). On the other hand, in markets with very few potential sellers, that is for low λ, there are no perfect options: The buyer can either choose immediate settlement and agree to compensate the intermediary for the borrowing cost, or choose a long settlement delay and accept the corresponding counterparty risk. If the borrowing cost is low, the buyers' private value is high, or both, then buyers choose fast settlement in both very thick and very thin markets (high and low λ), and therefore the optimal time-to-settlement first increases and then decreases in the seller's arrival rate.

Figure 1 illustrates the optimal time-to-settlement and corresponding **B**-**I** settlement price as functions of the intermediary's default risk (δ) and the seller's arrival rate (λ).

[insert Figure 1 here]

5 A settlement rat race

In this section, we allow for $N \geq 2$ and therefore introduce imperfect competition between the N security lenders with low risk-aversion $\gamma < \Gamma$. As in Section 3, we assume that (i) buyer's private value is large enough, that is $\theta > \beta_0$, and that (ii) (**BC**) is true and therefore the intermediary always borrows at the settlement time if she does not have the asset. We show that imperfect competition between security lenders, due to the opacity of the repo market, generates a settlement rat race in which buyers choose sub-optimally short settlement delays.

5.1 Competition on the repo market

From equation (10), each of the N security lenders with risk-aversion γ requires a large enough repo price, that is $\ell \geq \frac{\gamma}{2}\sigma^2$, in order to lend the asset to **I**. Since intermediaries have the outside option to borrow from a fringe lender, yielding the net borrowing cost of $\frac{\Gamma}{2}\sigma^2(R-1)$, they only enter a repo contract with one of the N security lenders if

$$R\ell - \frac{\gamma}{2}\sigma^2 \leq \frac{\Gamma}{2}\sigma^2(R-1) \Rightarrow \ell \leq \bar{\ell} \equiv \frac{\Gamma}{2}\sigma^2 - \frac{\Gamma-\gamma}{2R}\sigma^2. \tag{20}$$

It follows that security lenders charge repo prices $\ell \in \left[\frac{\gamma}{2}\sigma^2, \bar{\ell}\right]$. As in Janssen and Rasmusen (2002) or Jovanovic and Menkveld (2017), the opacity of the repo market implies that competition between security lenders does not drive down the repo price to **L**'s reservation value of $\frac{\gamma}{2}\sigma^2$. With conditional probability $(1-\varphi)^{N-1}$, each **L** holding the asset is a "monopolist" who can charge $\bar{\ell}$ to an intermediary. On a transparent market, if two security lenders hold the asset, then they would engage in Bertrand competition and ask for $\ell = \frac{\gamma}{2}\sigma^2$. However, since the repo market is opaque, lenders cannot distinguish whether they are "monopolists" or "Bertrand competitors."

In equilibrium, lenders choose random prices from a distribution $F_N(\cdot)$, such that they are indifferent between all prices in the support of $F_N(\cdot)$. The expected profit of a lender offering to lend securities at a given price ℓ is:

$$\begin{aligned}\mathbb{E}\left[U_{\mathbf{L}}(\ell)\right] &= \varphi \sum_{k=0}^{N-1} \begin{pmatrix} N-1 \\ k \end{pmatrix} \varphi^k (1-\varphi)^{N-1-k} \left[1-F_N(\ell)\right]^k \left(\ell - \frac{\gamma}{2}\sigma^2\right) \\ &= \left(\ell - \frac{\gamma}{2}\sigma^2\right) \varphi \left(1-\varphi F_N(\ell)\right)^{N-1}. \end{aligned} \tag{21}$$

First, with probability φ, the lender owns the asset. The intermediary selects the lender if he offers the lowest price among all **L**. With probability $\begin{pmatrix} N-1 \\ k \end{pmatrix} \varphi^k (1-\varphi)^{N-1-k}$, exactly k out of N competitors also own the asset and a lender offers the lowest price with probability $\left[1-F_N(\ell)\right]^k$. In this case, from equation (10), his profit is $\ell - \frac{\gamma}{2}\sigma^2$. Lemma 1 describes the partial equilibrium distribution for repo prices and establishes that security lenders earn positive rents in expectation.

Lemma 1. (Distribution of repo prices) *If the intermediary requests repo quotes at τ, each lender who owns the asset proposes a price ℓ from the distribution $F_N(\cdot)$, where*

$$F_N(\ell) = \begin{cases} 0, & \text{if } \ell \leq \underline{\ell} \\ \varphi^{-1}\left[1-(1-\varphi)\sqrt[N-1]{\frac{\sigma^2(R-1)(\Gamma-\gamma)}{R(2\ell-\gamma\sigma^2)}}\right], & \text{if } \ell \in \left(\underline{\ell}, \bar{\ell}\right] \\ 1, & \text{if } \ell > \bar{\ell}. \end{cases} \tag{22}$$

and $\underline{\ell} = \frac{\sigma^2}{2}\left[\gamma + (\Gamma - \gamma)\frac{R-1}{R}(1-\varphi)^{N-1}\right]$. *The upper bound of the price distribution,* $\overline{\ell}$*, is defined in equation* (20)*. Further, each lender earns positive expected rents*

$$\mathbb{E}\left[U_{\boldsymbol{L}}\right] = \varphi\left(1-\varphi\right)^{N-1}\frac{\sigma^2}{2}\left(\Gamma - \gamma\right)\frac{R-1}{R} > 0. \tag{23}$$

Let $\Omega_{\mathbf{L}}$ denote the set of lenders who own the asset. The expected net borrowing cost for an intermediary, β_N, is:

$$\beta_N = \left[1-(1-\varphi)^N\right]\left(\mathbb{E}\left[\min_i \ell_i \mid \Omega_{\mathbf{L}} \neq \varnothing\right] \times R - \frac{\gamma}{2}\sigma^2\right) + (1-\varphi)^N\frac{\Gamma}{2}\sigma^2\left(R-1\right). \tag{24}$$

With probability $1-(1-\varphi)^N$, at least one of the N security lenders owns the asset and the intermediaries choose the lowest price offered. With probability $(1-\varphi)^N$, no security lender with risk-aversion γ owns the asset, and the intermediaries need to borrow from the fringe lenders at net cost $\frac{\Gamma}{2}\sigma^2\left(R-1\right)$.

Lemma 2. (Repo borrowing cost) *The expected net borrowing cost* β_N *is*

$$\beta_N = \frac{\sigma^2}{2}\left(R-1\right)\left[\gamma + (\Gamma-\gamma)\left[N\varphi\left(1-\varphi\right)^{N-1} + (1-\varphi)^N\right]\right], \tag{25}$$

which decreases in the number of security lenders N *and increases in the asset specialness, that is, it decreases in* φ*.*

From Lemma 2 it follows that if $N \geq 2$ the expected borrowing cost for the intermediary is lower than in the benchmark equilibrium, that is if **I** always borrows from the fringe lenders. However, $\beta_N > \frac{\gamma}{2}\sigma^2\left(R-1\right)$ since the security lenders obtain positive rents in equilibrium.

We note that on a transparent repo market, both the expected lender's rent and the repo borrowing cost are the same as in equations (23) and (25). However, on a transparent market security lenders quote either their own or the fringe lenders' reservation price. An opaque market is both more realistic and allows for richer predictions about the cross-sectional distribution of repo prices.

Figure 2 illustrates the results in Lemmas 1 and 2, that is the repo price cumulative distribution and expected repo borrowing cost as a function of N and φ.

[insert Figure 2 here]

5.2 Equilibrium

Lemma 2 establishes that repo market rents decrease in the number of security lenders available. Unlike in the benchmark equilibrium, intermediaries' expected borrowing cost depends on the order in which they access the repo market. The first **I** to settle her trade faces expected borrowing cost β_N, as there are at most N security lenders who own the risky asset. Through borrowing, the first intermediary removes part of the available lending supply: the second **I** to settle a trade can borrow the risky asset from at most $N-1$ security lenders (or the fringe) at expected cost $\beta_{N-1} > \beta_N$. Therefore, for intermediaries and buyers, settling the trade before anyone else becomes valuable. In this section, we show that a settlement "rat race" emerges in equilibrium.

From equation (9) and the assumption that intermediaries always borrow rather than fail to deliver, it follows that the expected utility of intermediary j depends not only on the settlement time of her trade, τ_j, but also on that of her competitor, τ_{-j}:

$$\mathbb{E}\left[U_{\mathbf{I}_j} \mid \tau_j, \tau_{-j},\ \mathbf{I}_j \text{ survives}\right] = \begin{cases} p\left(\tau_j\right) - e^{-\lambda \tau_j} \beta_N, & \text{if } \tau_j < \tau_{-j} \\ p\left(\tau_j\right) - e^{-\lambda \tau_j} \beta_{N-1}, & \text{if } \tau_j > \tau_{-j} \\ p\left(\tau_j\right) - e^{-\lambda \tau_j} \frac{\beta_N + \beta_{N-1}}{2}, & \text{if } \tau_j = \tau_{-j}, \end{cases} \tag{26}$$

where if both intermediaries settle at the same time (i.e., $\tau_j = \tau_{-j}$) they are equally likely to approach security lenders first.

We denote by $G(\cdot)$ the symmetric equilibrium cumulative distribution function of the settlement times chosen by the two buyers. Since $G(\cdot)$ could be discontinuous if buyers put positive probability on a particular time-to-settlement, let $\tau^- = \lim_{\varepsilon \downarrow 0} (\tau - \varepsilon)$ such that $G(\tau) - G(\tau^-) = \text{Prob}(\tau_j = \tau)$. Since intermediaries are competitive, it follows that the equilibrium price schedule is

$$p\left(\tau_j\right) = e^{-\lambda \tau_j} \left[G\left(\tau_j^-\right) \beta_{N-1} + \left(G\left(\tau_j\right) - G\left(\tau_j^-\right)\right) \frac{\beta_N + \beta_{N-1}}{2} + \left(1 - G\left(\tau_j\right)\right) \beta_N \right]. \tag{27}$$

The competitive price is equal to the expected borrowing cost on the repo market. With probability $G\left(\tau_j^-\right)$, $\tau_j > \tau_{-j}$ and intermediary j is the second to borrow, that is she borrows at cost β_{N-1}. With probability $1 - G\left(\tau_j\right)$, $\tau_j < \tau_{-j}$ and intermediary j is the first to borrow, that is she borrows at cost β_N. With the complementary probability, $\tau_{-j} = \tau_j$ and each intermediary is equally likely to borrow at β_N or β_{N-1}.

Definition 2. (Benchmark utilities.) Let $V_N(\tau)$ denote the expected utility of each buyer if the repo borrowing cost for competitive intermediaries is fixed at β_N, and let τ_N be the optimal settlement time that maximizes $V_N(\tau)$:

$$\begin{aligned} V_N(\tau) &\equiv \mathbb{E}\left[U_{\mathbf{B}} \mid \text{borrowing cost is } \beta_N\right] = e^{-\delta\tau}\left(\theta - e^{-\lambda\tau}\beta_N\right) \text{ and} \\ \tau_N &\equiv \arg\max_{\tau} V_N(\tau) = \max\left\{0, \frac{1}{\lambda}\log\left[\frac{(\delta+\lambda)}{\theta\delta}\beta_N\right]\right\}. \end{aligned} \tag{28}$$

The benchmark utilities and corresponding optimal settlement delays are equivalent to those in Section 3 for $N > 0$. It immediately follows from Lemma 2 that $V_N(\tau) > V_{N-1}(\tau)$ for all $\tau \geq 0$ and that $\tau_N \leq \tau_{N-1}$, with equality only if $\tau_N = \tau_{N-1} = 0$. The intuition is the same as in the benchmark equilibrium: buyers are overall better off, and would optimally settle faster, when borrowing costs for the intermediary are lower.

From equations (5), (27), and using Definition 2, we can write the expected buyer utility as a linear combination of benchmark utilities, that is

$$\mathbb{E}\left[U_{\mathbf{B}_j}\right] = G\left(\tau_j^-\right)V_{N-1}(\tau) + \left(G(\tau_j) - G\left(\tau_j^-\right)\right)\frac{V_N(\tau) + V_{N-1}(\tau)}{2} + (1 - G(\tau_j))V_N(\tau). \tag{29}$$

Lemma 3. (Buyers' strategy) *If $V_{N-1}(0) + V_N(0) \geq 2V_{N-1}(\tau_{N-1})$, both buyers choose immediate settlement with probability $\pi = 1$.*

If $V_{N-1}(0) + V_N(0) < 2V_{N-1}(\tau_{N-1})$, buyers choose immediate settlement with probability $\pi < 1$, where

$$\pi = \max\left\{0, 2\frac{V_N(0) - V_{N-1}(\tau_{N-1})}{V_N(0) - V_{N-1}(0)}\right\}. \tag{30}$$

With probability $1 - \pi$, each buyer chooses a random settlement time in $[\underline{\tau}, \tau_{N-1}]$ where $\underline{\tau} < \tau_N$ is pinned down by:

$$V_{N-1}(\tau_{N-1}) = (1 - \pi)V_N(\underline{\tau}) + \pi V_{N-1}(\underline{\tau}), \tag{31}$$

such that the buyer's expected utility is $V_{N-1}(\tau_{N-1})$.

From Lemma 3, two types of equilibria emerge depending on the model parameters: an immediate-settlement equilibrium and a delayed-settlement equilibrium. In the immediate-settlement equilibrium, buyers choose the pure strategy $\tau^\star = 0$ and intermediaries always enter repo contracts with security lenders. From equation (27), the price $p(0)$ reflects the average cost of borrowing, that is,

$$p(0) = \frac{1}{2}(\beta_N + \beta_{N-1}). \tag{32}$$

In the delayed-settlement equilibrium, buyers randomize over possible times-to-settlement, using the cumulative distribution function

$$G(\tau) = \begin{cases} \pi, & \text{if } \tau \in [0, \underline{\tau}) \\ \frac{V_N(\tau) - V_{N-1}(\tau_{N-1})}{V_N(\tau) - V_{N-1}(\tau)}, & \text{if } \tau \in [\underline{\tau}, \tau_{N-1}] \\ 1, & \text{if } \tau > \tau_{N-1}. \end{cases} \tag{33}$$

A settlement "rat race" emerges in which each buyer chooses a short settlement delay, so that his trade is more likely to be settled first. Note that, from equation (28), a settlement delay choice $\tau > \tau_{N-1}$ is always sub-optimal for buyers. Therefore, by choosing τ_{N-1} a buyer is sure to be the second to settle his trade. This pins down the equilibrium utility: In any mixed strategy (i.e., delayed settlement) equilibrium, buyers earn $V_{N-1}(\tau_{N-1})$.

Mixed strategies in the delayed-settlement equilibrium do not necessarily have a continuous support. If $V_N(0) > V_{N-1}(\tau_{N-1})$, a buyer can improve his utility if he unilaterally deviates and chooses immediate settlement. In this case, in equilibrium buyers put positive weight $\pi > 0$ on $\tau^\star = 0$. With probability $1 - \pi$, buyers optimally mix over settlement delays in the support $[\underline{\tau}, \tau_{N-1}]$, where $\underline{\tau} > 0$. Otherwise, if $V_N(0) \leq V_{N-1}(\tau_{N-1})$, buyers use a continuous mixed strategy such that no time-to-settlement τ is chosen with positive probability.

The expected utility for buyers, across both the immediate- and delayed-settlement equilibria is therefore

$$\mathbb{E}[U_{\mathbf{B}}] = \max\left\{ \underbrace{V_{N-1}(\tau_{N-1})}_{\substack{\text{delayed-settlement}\\ \text{equilibrium}}}, \underbrace{\frac{1}{2}[V_N(0) + V_{N-1}(0)]}_{\substack{\text{immediate-settlement}\\ \text{equilibrium}}} \right\}. \tag{34}$$

Figure 3 illustrates the buyers' strategy $G(\tau)$ in the delayed-settlement equilibrium for different levels of counterparty risk. Intuitively, higher counterparty risk and lower asset specialness lead to shorter settlement delays in equilibrium.

[insert Figure 3 here]

Proposition 2 formally states the equilibrium of the trading game for $N \geq 2$.

Proposition 2. (Equilibrium, exogenous penalties) *If $N \geq 2$ and the late-delivery penalty z is sufficiently large such that (**BC**) is always true, then the following strategies form an equilibrium:*

(i) Intermediaries post $\{\tau, p(\tau)\}$ contract schedules with $\tau \geq 0$ and $p(\tau)$ as defined in equation (27), where the equilibrium $G(\tau)$ is given in equation (33).

(ii) If $V_{N-1}(0) + V_N(0) \geq 2V_{N-1}(\tau_{N-1})$, buyers choose immediate settlement ($\tau^\star = 0$). Otherwise, buyers choose a random time-to-settlement from the equilibrium cumulative distribution $G(\tau)$ defined in equation (33).

(iii) Securities lenders mix over repo purchasing prices ℓ. The first and second lender to be approached by an intermediary draws from distribution $F_N(\ell)$ and $F_{N-1}(\ell)$, respectively, defined in Lemma 1.

(iv) Intermediaries always borrow the asset at settlement time if the large seller did not arrive before settlement.

Figure 4 illustrates the equilibrium price-settlement delay schedules posted by intermediaries. Everything else equal, a longer time-to-settlement corresponds to a lower probability of costly borrowing, and therefore to a lower price. The price-settlement delay schedule shifts up as specialness increases (i.e., for lower φ), since security lenders are able to extract higher rents and increase the expected borrowing cost for intermediaries. Figure 4 also differentiates between the $\{\tau, p(\tau)\}$ pairs on and off the equilibrium path, that is, the settlement times buyers choose in equilibrium from the competitive schedule.

[insert Figure 4 here]

Corollary 2 states that as the intermediary default rate δ increases, buyers are more likely to choose immediate settlement. There is a smooth transition between delayed- and immediate-settlement equilibria for a positive default rate threshold $\hat{\delta}$, such that buyers choose immediate settlement with probability one if and only if the default rate of the intermediary is higher than the threshold.

Corollary 2. (Immediate settlement probability) *The immediate settlement probability π weakly increases in the default rate δ. Moreover, there exists a unique default rate threshold $\hat{\delta} > 0$ such that for $\delta < \hat{\delta}$ the delayed-settlement equilibrium obtains and for $\delta \geq \hat{\delta}$ the immediate-settlement equilibrium obtains.*

We can measure the cost of the settlement rat race for buyers by comparing their equilibrium expected utility in (34) with a benchmark utility. The rat race emerges either if buyers trade simultaneously at $t = 0$, or they cannot observe each other's choices, or both. Therefore, a natural utility benchmark is to consider the optimal choices buyers would make if they arrive sequentially rather than simultaneously, that is τ_N and τ_{N-1}. Let

$$\mathcal{C} = \underbrace{V_N(\tau_N) + V_{N-1}(\tau_{N-1})}_{\text{benchmark utility}} - \underbrace{\max\{2V_{N-1}(\tau_{N-1}), V_N(0) + V_{N-1}(0)\}}_{\text{equilibrium buyers' utility}} \tag{35}$$

be a measure of the settlement rat race cost for the buyers. Importantly, since borrowing has a real economic cost (intermediary's shadow cost of capital), the settlement rat race cost for the buyers exceeds any additional rents for security lenders due to the higher likelihood of borrowing.

Corollary 3. (Settlement rat race cost.) *The settlement rat race cost $\mathcal{C}$ increases in default risk (δ) and decreases in asset specialness (increases in φ) if the delayed-settlement equilibrium obtains, and decreases in δ and increases in asset specialness (decreases in φ) if the immediate-settlement equilibrium obtains.*

A higher default risk has two effects on the settlement race cost. First, the benchmark optimal settlement time in equation (28) decreases in the default rate δ. Second, from Corollary 2, the settlement race is reinforced as buyers choose faster settlement. For low default risk, the delayed-settlement equilibrium obtains and the second effect dominates, that is, the rat race escalates faster than the corresponding drop in optimal settlement times. Consequently, the cost of the race increases in counterparty risk. For high default risk, the immediate settlement equilibrium obtains. In the immediate-settlement equilibrium, a higher counterparty risk reduces the distance between the optimal time-to-settlement in equation (28) and the equilibrium choice $\tau^{\star} = 0$. In this case, only the first channel is relevant as buyers cannot settle trades any faster. Therefore, in the immediate-settlement equilibrium, the cost of the rat race decreases in counterparty risk δ. A similar reasoning applies when we consider asset specialness (and, implicitly, borrowing costs) rather than counterparty risk.

Corollary 4. (Settlement rat race.) *Let δ_N and δ_{N-1} be the lowest values of the default rate for which immediate settlement is optimal, that is $\tau_N = 0$ and $\tau_{N-1} = 0$ respectively. In equilibrium, $\hat{\delta}$ satisfies*

$$\hat{\delta} \leq \delta_{N-1}, \tag{36}$$

and further $\hat{\delta} \leq \delta_N$ if and only if

$$\frac{1}{2}\left[V_{N-1}(0) + V_N(0)\right] > V_{N-1}\left(\tau_{N-1}(\delta_N)\right). \tag{37}$$

Corollary 4 states that the settlement rat race leads to buyers choosing immediate settlement in equilibrium even for relatively low counterparty risk, for which they would optimally choose delayed settlement in a perfect information setting (where optimal settlement delay choices are τ_N and τ_{N-1}). Therefore, there is excessive demand for immediate settlement in equilibrium, which is not driven by the buyers' desire to reduce counterparty risk, but rather by the desire to obtain a better price through reducing intermediaries' borrowing costs. Figure 5 illustrates the results in Corollaries 3 and 4.

[insert Figure 5 here]

6 Optimal failure-to-deliver penalties

In this Section we allow buyers and intermediaries to also contract on the failure-to-deliver penalty $z(\tau)$. It follows that the borrowing condition (**BC**) is no longer automatically true, and the intermediary only borrows the security at τ if the borrowing cost is low enough, that is if

$$b \leq p(\tau) + z(\tau). \tag{38}$$

Lemma 4. (Optimal borrowing) *Buyers optimally offer incentives for intermediaries to borrow if and only if $b \leq \frac{\delta}{\delta+\lambda}\theta$. Furthermore, in any contract implementing the borrowing decision that maximizes the buyers' expected utility, it is the case that*

$$p(\tau) + z(\tau) = \frac{\delta}{\delta+\lambda}\theta. \tag{39}$$

The result is intuitive. Since intermediaries are competitive, it is the buyers who ultimately bear the borrowing cost. If intermediaries do not borrow at τ, buyers bear additional counterparty risk instead: with probability $\frac{\delta}{\delta+\lambda}$, the intermediary defaults before the large seller arrival. Proposition 3 states the equilibrium in the trading game with endogenous failure-to-deliver penalties.

Proposition 3. (Equilibrium, endogenous penalty.) *If buyers and intermediaries can contract on failure-to-deliver penalties, then the following strategies form an equilibrium:*

(i) Intermediaries offer payment and penalty schedules such that

$$p\left(\tau\right) = e^{-\lambda\tau}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right], \tag{40}$$

$$z\left(\tau\right) = \frac{\delta}{\delta+\lambda}\theta - e^{-\lambda\tau}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right], \tag{41}$$

where the closed form expression for $\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right]$ *is given in equation* (B.54) *in the Appendix.*

(ii) Buyers choose immediate settlement.

(iii) Securities lenders mix over repo purchasing prices ℓ. The first and second lender to be approached by an intermediary draws from a (generalized) distribution $F_N^{\text{gen}}\left(\ell\right)$ and $F_{N-1}^{\text{gen}}\left(\ell\right)$, defined in equation (B.52) *in the Appendix. They charge at most*

$$\overline{\ell}^{\text{gen}} = \frac{1}{R}\left(\min\left\{\frac{\Gamma}{2}\sigma^2\left(R-1\right), \frac{\delta}{\delta+\lambda}\theta\right\} + \frac{\gamma\sigma^2}{2}\right). \tag{42}$$

(iv) Intermediaries only borrow the asset if $b \leq \dfrac{\delta}{\delta+\lambda}\theta$.

First, equilibrium price and penalty schedules have an option-like structure. In particular, the failure-to-delivery penalty resembles a put option on the repo market conditions, with a strike price equal to the additional counterparty risk for the buyer if the intermediary does not borrow, that is $\frac{\delta}{\delta+\lambda}\theta$. We can think of the optimal penalty as insurance against high borrowing costs on the repo market. Figure 6 illustrates that the equilibrium penalty schedule $z\left(\tau\right)$ increases in settlement time and decreases in asset specialness. Intermediaries are optimally penalized more for a settlement failure if the settlement cycle is long. This is intuitive: since the price $p(\tau)$ decreases with the settlement delay, the failure-to-deliver penalty needs to increase to preserve incentives to borrow. Further, for special assets with high borrowing costs, buyers optimally offer intermediaries lower incentives to borrow. Consequently, the failure-to-deliver penalty is lower if the asset supply is more limited.

[insert Figure 6 here]

Second, we note that buyers always choose immediate settlement (i.e., $\tau^\star = 0$) if they can contract on failure-to-deliver penalties. The price $p(\tau)$ in equation (40) is lower for any τ than the additional counterparty risk the buyer bears if the intermediary does not borrow, $\frac{\delta}{\delta+\lambda}\theta$. Therefore, buyers have no incentive to delay settlement in order to avoid borrowing costs.

To further analyze the equilibrium outcomes, we distinguish between three possible scenarios. Depending on the relative magnitude of the borrowing costs and the counterparty risk incurred after the contract settlement time, either (i) $\frac{\delta}{\delta+\lambda}\theta \leq \frac{\gamma}{2}\sigma^2 (R-1)$ and borrowing is never optimal, (ii) $\frac{\delta}{\delta+\lambda}\theta > \frac{\Gamma}{2}\sigma^2 (R-1)$ and borrowing is always optimal, or (iii) $\frac{\delta}{\delta+\lambda}\theta \in \left(\frac{\gamma}{2}\sigma^2 (R-1), \frac{\Gamma}{2}\sigma^2 (R-1)\right]$ and borrowing may be optimal or not.

Pass-through contracts. First, let $\frac{\delta}{\delta+\lambda}\theta \leq \frac{\gamma}{2}\sigma^2 (R-1)$. If counterparty risk is low enough, or costs of borrowing from perfectly competitive lenders are relatively high, it is optimal for **B** that **I** never borrows. Buyers always choose immediate settlement and, consequently, all trades (technically) fail to deliver. A *pass-through contract* emerges, in which no cash changes hands on the equilibrium path: **I** simply delivers the asset to **B** immediately after buying it from the large seller. From Proposition 3, the equilibrium penalty is zero, that is $z(0) = 0$: the intermediary is not penalized for late delivery. While the payment $p(\tau)$ is greater than zero, the intermediary is never able to deliver on time, so the payment is off the equilibrium path.

Always-borrow contracts. Second, let $\frac{\delta}{\delta+\lambda}\theta > \frac{\Gamma}{2}\sigma^2 (R-1)$. If counterparty risk is high enough, or borrowing costs from fringe traders are relatively low, it is optimal for **B** that **I** always borrows. We recover the equilibrium in Section 5, for $\pi = 1$. In addition, the equilibrium contract specifies a penalty schedule $z(\tau)$ as in (41). However, since the intermediary always borrows, paying the late delivery penalty is not on the equilibrium path.

Random fails-to-deliver. If $\frac{\delta}{\delta+\lambda}\theta \in \left(\frac{\gamma}{2}\sigma^2\left(R-1\right), \frac{\Gamma}{2}\sigma^2\left(R-1\right)\right]$, then the buyer would like to offer incentives to the intermediary to borrow from the N security lenders, but not from the fringe lenders. The N security lenders still earn rents, but lower – as the "reservation cost" of the intermediary changes. A trade fails to settle with probability $(1-\varphi)^N$, that is if none of the N security lenders own the asset.

Figure 7 illustrates the failure-to-deliver probability as a function of asset specialness for all three scenarios.

[insert Figure 7 here]

Unlike in Section 5, the repo prices depend on the counterparty risk level. Figure 8 illustrates rents for a security lender, conditional on owning the asset, as a function of counterparty risk.

[insert Figure 8 here]

Repo prices increase in counterparty risk, since the upper bound on the intermediaries' borrowing cost increases in δ. Therefore, for larger counterparty risk, lenders are better able to hold up intermediaries and extract rents from them.

Discussion of zero-penalty contracts. Our analysis in Sections 4 through 6 focuses on either "large enough" or flexible failure-to-deliver penalties. Another possible case would be a zero or "low enough" penalty. From the discussion above, a zero penalty emerges in equilibrium if $\frac{\delta}{\delta+\lambda}\theta \leq \frac{\gamma}{2}\sigma^2\left(R-1\right)$, that is for the pass-through contracts. Otherwise, a zero-penalty contract leads to either rents for intermediaries, sub-optimal security borrowing decisions, or both. If buyers can choose both a price upon successful delivery and a penalty upon failure-to-deliver, they can simultaneously achieve two objectives, that is (i) implement the incentive-compatible repo borrowing choice in Lemma 4 and also (ii) extract the full trading surplus from the intermediary. A zero failure-to-deliver penalty implies buyers have a single tool (i.e., price upon successful delivery) to implement both objectives. Therefore, since buyers cannot penalize intermediaries ex-post, they must increase the price to offer higher borrowing incentives, allowing intermediaries to extract rents. However, since buyers never pay a price higher than the additional counterparty risk $\frac{\delta}{\delta+\lambda}\theta$, they will still optimally choose immediate settlement even if penalties are zero. The rationale is the same as in Proposition 3.

7 Conclusions

This paper provides insights into the consequences of changing the rules for trade settlement on financial markets. The current settlement process, involving several days of delay between trade and settlement, feels at odds with the fast-paced markets of today. Both market participants and financial regulators agree on this. Further, new technologies such as distributed ledgers and smart contracts can offer traders the flexibility to fine-tune settlement cycles on a trade-by-trade or asset-by-asset basis.

We emphasize three insights emerging from our results. First, flexible settlement cycles allow traders to individually balance counterparty risk and borrowing costs on the securities loan market.

Second, high failure-to-deliver penalties are not necessarily a panacea for market quality as they lead to excessive security borrowing. Imperfect competition between security lenders generates a strong incentive to settle trades before everyone else. In an attempt to reduce borrowing costs, traders engage in a settlement rat race. In equilibrium, there is excess demand for fast, or even immediate, settlement. Security lenders earn high rents at the expense of traders, and welfare is reduced.

Finally, we find that the optimal trade contract combines flexible failure-to-deliver penalties with immediate settlement. The equilibrium penalty resembles a put option on the repo market, and it serves a double purpose: First, it provides insurance for traders against high borrowing costs. Second, it improves welfare by fostering competition on the security lending market.

References

Acharya, Viral, and Alberto Bisin, 2014, Counterparty risk externality: Centralized versus over-the-counter markets, *Journal of Economic Theory* 149, 153–182.

Acharya, Viral V., 2009, Centralized clearing for credit derivatives, in *Restoring Financial Stability,* pp. 251–268.

Aune, Rune, Maureen O'Hara, and Ouziel Slama, 2017, Footprints on the blockchain: Information leakage in distributed ledgers, *Working paper.*

BCG, 2012, Shortening the Settlement Cycle, Discussion paper, Boston Consulting Group.

Benos, Evangelos, Rod Garratt, and Pedro Gurrola-Perez, 2017, The economics of Distributed Ledger Technology for securities settlement, *Bank of England Staff Working paper No. 670.*

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, and Catherine Casamatta, 2018, The blockchain folk theorem, *Toulouse School of Economics Working paper No. 17–817.*

Brainard, Lael, 2016, Distributed ledger technology: Implications for payments, clearing, and settlement, *Board of Governors of the Federal Reserve System Speech.*

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

Cong, Lin William, and Zhiguo He, 2018, Blockchain disruption and smart contracts, *Working paper.*

Corradin, Stefano, and Angela Maddaloni, 2017, The importance of being special: Repo markets during the crisis, *ECB Working paper No. 2065.*

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

Duffie, Darrell, 1996, Special repo rates, *The Journal of Finance* 51, 493–526.

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

Duffie, Darrell, and Haoxiang Zhu, 2011, Does a central clearing counterparty reduce counterparty risk?, *Review of Asset Pricing Studies* 1, 74–95.

ECB, 2016, Distributed ledger technology, Discussion paper, European Central Bank.

ESMA, 2016, The distributed ledger technology applied to securities markets, Discussion paper, European Securities and Markets Authority.

Evans, Richard B., Rabih Moussawi, Michael S. Pagano, and John Sedunov, 2017, Etf failures-to-deliver: Naked short-selling or operational shorting?, *Working paper*.

Fleming, Michael J., and Kenneth Garbade, 2005, Explaining settlement fails, *Current Issues in Economics and Finance* 11.

Foley-Fisher, Nathan, Stefan Gissler, and Stephane Verani, 2017, Over-the-counter market liquidity and securities lending, *Working paper*.

Fotak, Veljko, Vikas Raman, and Pradeep K. Yadav, 2014, Fails-to-deliver, short selling, and market quality, *Journal of Financial Economics* 114, 493–516.

Geczy, Christopher C., David K. Musto, and Adam V. Reed, 2002, Stocks are special too: an analysis of the equity lending market, *Journal of Financial Economics* 66, 241 – 269.

Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *The Journal of Finance* 43, 617–633.

Huszár, Zsuzsa R., and Zorka Simon, 2018, The pricing implications of oligopolistic securities lending market: A beneficial owner perspective, *Netspar Working paper* DP 09/2017-032.

IHS Markit, 2016, Securities Lending: Year in Review, Discussion paper, IHS Markit.

Janssen, Maarten, and Eric Rasmusen, 2002, Bertrand competition under uncertainty, *The Journal of Industrial Economics* 50, 11–21.

Jordan, Bradford D., and Susan D. Jordan, 1997, Special repo rates: An empirical analysis, *The Journal of Finance* 52, 2051–2072.

Jovanovic, Boyan, and Albert J. Menkveld, 2017, Dispersion and skewness of bid prices, *Working paper*.

Koeppl, Thorsten, and Jonathan Chiu, 2018, Blockchain-based settlement for asset trading, *Queen's Economics Department Working paper No. 1397.*

Kyle, Albert S., and Anna A. Obizhaeva, 2017, Market microstructure invariance: A dynamic equilibrium model, *Working paper.*

Levitt, Arthur, 1996, Speeding up settlement: The next frontier, *Symposium on Risk Reduction in Payments, Clearance, and Settlement Systems.*

Loon, Yee Cheng, and Zhaodong Ken Zhong, 2014, The impact of central clearing on counterparty risk, liquidity, and trading: Evidence from the credit default swap market, *Journal of Financial Economics* 112, 91–115.

Malinova, Katya, and Andreas Park, 2017, Market design for trading with blockchain technology, *Working paper.*

Menkveld, Albert J, 2017, Crowded positions: An overlooked systemic risk for central clearing parties, *The Review of Asset Pricing Studies* 7, 209–242.

Pirrong, Craig, 2009, The economics of clearing in derivatives markets: netting, asymmetric information, and the sharing of default risks through a central counterparty, *Working paper.*

Saleh, Fahad, 2017, Blockchain without waste: Proof-of-stake, *Working paper.*

Stephens, Eric, and James R. Thompson, 2017, Information asymmetry and risk transfer markets, *Journal of Financial Intermediation* 32, 88 – 99.

Tobias, Adrian, Brian Begalle, Adam Copeland, and Antoine Martin, 2013, *Risk Topography: Systemic Risk and Macro Modeling;* chap. Repo and Securities Lending, pp. 131–148.

A Notation summary

Model parameters and their interpretation.

Parameter	Definition
σ	Asset volatility.
δ	Intermediary default rate.
λ	Large seller arrival rate.
θ	Buyer private value for the asset.
γ, Γ	Risk-aversion of security lenders, with $\gamma < \Gamma$.
N	Number of security lenders with risk aversion γ.
φ	Probability a security lender owns the asset.
R	Opportunity cost of collateral for intermediary.

B Proofs

Proposition 1

Proof. The proof follows immediately from the discussion in the main text. For (i), equation (15) pins down the competitive price for intermediaries.

The optimal settlement time in (ii) solves the first order condition

$$\frac{\partial \mathbb{E}\left[U_{\mathbf{B}}\right]}{\partial \tau} = 0, \tag{B.1}$$

where $\mathbb{E}\left[U_{\mathbf{B}}\right]$ is given in equation (16). It follows that:

$$\frac{\partial \mathbb{E}\left[U_{\mathbf{B}}\right]}{\partial \tau} = e^{-\tau(\delta+\lambda)}\left[\frac{\Gamma}{2}\left(R-1\right)\sigma^2\left(\delta+\lambda\right) - \delta\theta e^{\lambda\tau}\right], \tag{B.2}$$

and therefore $\tau^\star = \max\left\{0, \frac{1}{\lambda}\log\left[\frac{(\delta+\lambda)}{\theta\delta}\frac{\Gamma}{2}\sigma^2\left(R-1\right)\right]\right\}$, that is the expression in equation (17).

The security lenders competitive prices in (iii), that is $\ell = \frac{\Gamma}{2}\sigma^2$, follow from (10) holding with equality. Finally, by assumption, the penalty is high enough such that the intermediary always borrows the asset at $\tau^\star$ if she does not own it. $\square$

Corollary 1

Proof. Let's introduce an auxiliary variable

$$\tilde{\tau} = \frac{1}{\lambda}\log\left[\frac{(\delta+\lambda)}{\theta\delta}\frac{\Gamma}{2}\sigma^2\left(R-1\right)\right] \tag{B.3}$$

so that $\tau^\star = \max\{0, \tilde{\tau}\}$. We take the partial derivatives of $\tilde{\tau}$ with respect to all the parameters and sign them to establish monotonicity, that is

$$\left(\frac{\partial \tilde{\tau}}{\partial \Gamma}, \frac{\partial \tilde{\tau}}{\partial \sigma}, \frac{\partial \tilde{\tau}}{\partial R}, \frac{\partial \tilde{\tau}}{\partial \delta}, \frac{\partial \tilde{\tau}}{\partial \theta}\right) = \Big(\underbrace{\frac{1}{\Gamma\lambda}}_{>0}, \underbrace{\frac{2}{\sigma\lambda}}_{>0}, \underbrace{\frac{1}{\lambda(R-1)}}_{>0}, \underbrace{-\frac{1}{\delta(\delta+\lambda)}}_{<0}, \underbrace{-\frac{1}{\theta\lambda}}_{<0}\Big). \tag{B.4}$$

The partial derivative of $\tilde{\tau}$ with respect to the large seller's arrival rate λ is

$$\frac{\partial \tilde{\tau}}{\partial \lambda} = \frac{1}{\lambda^2}\left[\frac{\lambda}{\delta+\lambda} - \log\left(\frac{\Gamma(R-1)\sigma^2(\delta+\lambda)}{2\delta\theta}\right)\right]. \tag{B.5}$$

Define a function f as

$$f(\lambda, \cdot) = \frac{\lambda}{\delta+\lambda} - \log\left(\frac{\Gamma(R-1)\sigma^2(\delta+\lambda)}{2\delta\theta}\right). \tag{B.6}$$

Since $\frac{1}{\lambda^2}$ is positive, it follows that the sign of $\frac{\partial \tilde{\tau}}{\partial \lambda}$ is the same as the sign of $f(\lambda, \cdot)$. First, we note that $f(\lambda, \cdot)$ decreases in λ since

$$\frac{\partial f(\lambda, \cdot)}{\partial \lambda} = -\frac{\lambda}{(\lambda+\delta)^2} < 0. \tag{B.7}$$

Second, we compute the limits of $f(\lambda, \cdot)$, that is:

$$\lim_{\lambda \to 0} f(\lambda, \cdot) = \log(\theta) - \log\left(\frac{\Gamma}{2}\sigma^2(R-1)\right) \text{ and} \tag{B.8}$$

$$\lim_{\lambda \to \infty} f(\lambda, \cdot) = -\infty. \tag{B.9}$$

Since $\theta > \frac{\Gamma}{2}\sigma^2(R-1)$, $\lim_{\lambda\to 0} f(\lambda, \cdot) > 0$. Then since $f(\lambda, \cdot)$ is monotonous it follows that there exists a unique $\Lambda_0 > 0$ such that $\tilde{\tau}$ increases in λ for $\lambda \leq \Lambda_0$ and $\tilde{\tau}$ decreases in λ for $\lambda > \Lambda_0$. □

Lemma 1

Proof. First, there are no prices ℓ such that a lender offers a given ℓ with positive probability. From Janssen and Rasmusen (2002 pp. 12-13), either $\ell + \varepsilon$ or $\ell - \varepsilon$ with ε sufficiently close to zero would represent a profitable deviation.

In a mixed strategy equilibrium, lenders are indifferent between all prices in the support, that is the following first-order condition is true:

$$\frac{\partial \mathbb{E}\left[U_{\mathbf{L}}(\ell)\right]}{\partial \ell} = 0. \tag{B.10}$$

From (B.10) and equation (21), the cumulative distribution function $F_N(\cdot)$ solves the differential equation:

$$-\frac{1}{2}\varphi\left[1-\varphi F_N(\ell)\right]^{N-2}\left[(N-1)\varphi F_N'(\ell)\left(2\ell-\gamma\sigma^2\right)+2\varphi F_N(\ell)-2\right]=0, \tag{B.11}$$

subject to the boundary condition $F_N\left(\overline{\ell}\right)=1$ where from (20), $\overline{\ell}=\frac{\Gamma}{2}\sigma^2-\frac{\Gamma-\gamma}{2R}\sigma^2$.

The corresponding partial equilibrium mixed strategy is:

$$F_N(\ell)=\begin{cases}0, & \text{if } \ell\leq\underline{\ell}\\ \varphi^{-1}\left[1-(1-\varphi)\sqrt[N-1]{\frac{\sigma^2(R-1)(\Gamma-\gamma)}{R(2\ell-\gamma\sigma^2)}}\right], & \text{if } \ell\in\left(\underline{\ell},\overline{\ell}\right]\\ 1, & \text{if } \ell\geq\overline{\ell}.\end{cases} \tag{B.12}$$

and $\underline{\ell}=\frac{\sigma^2}{2}\left[\gamma+(\Gamma-\gamma)\frac{R-1}{R}(1-\varphi)^{N-1}\right]$ such that $F_N\left(\underline{\ell}\right)=0$.

The lender is indifferent between all prices in the support. If he asks for the maximum price $\overline{\ell}$, $\mathbf{L}$ only makes a profit if no other lender owns the asset. Consequently, the expected utility for $\mathbf{L}$ is:

$$\begin{aligned}\mathbb{E}\left[U_{\mathbf{L}}\right]&=\varphi(1-\varphi)^{N-1}\left(\overline{\ell}-\frac{\gamma}{2}\sigma^2\right)\\ &=\varphi(1-\varphi)^{N-1}\frac{\sigma^2}{2}(\Gamma-\gamma)\frac{R-1}{R}.\end{aligned} \tag{B.13}$$

□

Lemma 2

Proof. The cumulative distribution function of $\min_{i=1,\ldots,k}\ell_i$, if k lenders own the asset, is $1-(1-F_N(\ell))^k$. Since k itself is random, we can write the expectation of $\min_i\ell_i-\frac{\gamma}{2}\sigma^2$ if at least one lender owns the asset (that is, if $\mathbb{1}_{\Omega_{\mathbf{L}}\neq\varnothing}=1$) as:

$$\begin{aligned}&\mathbb{E}\left[\min_i\left(\ell_i-\frac{\gamma}{2}\sigma^2\right)\mathbb{1}_{\Omega_{\mathbf{L}}\neq\varnothing}\right]\\ &=\int_{\underline{\ell}}^{\overline{\ell}}\sum_{k=1}^{N}\binom{N}{k}\varphi^k(1-\varphi)^{N-k}\left(\ell-\frac{\gamma}{2}\sigma^2\right)k\left(1-F_N(\ell)\right)^{k-1}\mathrm{d}F_N(\ell)\\ &=\int_{\underline{\ell}}^{\overline{\ell}}N\varphi\left(1-\varphi F_N(\ell)\right)^{N-1}\left(\ell-\frac{\gamma}{2}\sigma^2\right)\mathrm{d}F_N(\ell)\\ &=N\varphi(1-\varphi)^{N-1}\frac{\sigma^2}{2}(\Gamma-\gamma)\frac{R-1}{R},\end{aligned} \tag{B.14}$$

where the last line follows from the lender being indifferent between all prices in $\left[\underline{\ell},\overline{\ell}\right]$.

It further follows that

$$\mathbb{E}\left[\min_i \ell_i \mid \Omega_{\mathbf{L}} \neq \varnothing\right] = \frac{N\frac{\sigma^2}{2}\left(\Gamma-\gamma\right)\frac{R-1}{R}\varphi\left(1-\varphi\right)^{N-1}}{1-\left(1-\varphi\right)^{N}} + \frac{\gamma}{2}\sigma^2. \tag{B.15}$$

Therefore, after replacing (B.15) in (24) it follows that

$$\beta_N = \frac{\sigma^2}{2}\left(R-1\right)\left[\gamma + \left(\Gamma-\gamma\right)\left[N\varphi\left(1-\varphi\right)^{N-1} + \left(1-\varphi\right)^{N}\right]\right]. \tag{B.16}$$

The partial derivative of β_N with respect to N is

$$\frac{\partial \beta_N}{\partial N} = \underbrace{\frac{\sigma^2}{2}\left(R-1\right)\left(\Gamma-\gamma\right)\left(1-\varphi\right)^{N-1}}_{>0}\left[\varphi + \left(1+\varphi\left(N-1\right)\right)\log\left(1-\varphi\right)\right]. \tag{B.17}$$

We study the sign of $g\left(N\right) \equiv \varphi + \left(1+\varphi\left(N-1\right)\right)\log\left(1-\varphi\right)$. First, note that

$$\frac{\partial g\left(N\right)}{\partial N} = \varphi \log\left(1-\varphi\right) \leq 0, \tag{B.18}$$

since $\varphi \in [0,1]$. Therefore $g\left(N\right)$ weakly decreases in N. Further, $g\left(1\right) = \varphi + \log\left(1-\varphi\right) \leq 0$ for any $\varphi \in [0,1]$ since $\frac{\partial(\varphi+\log(1-\varphi))}{\partial\varphi} = \frac{\varphi}{\varphi-1} < 0$ and $\varphi + \log\left(1-\varphi\right) = 0$ for $\varphi = 0$. It follows that $g\left(N\right) < 0$ for any $N \geq 1$. Consequently, $\frac{\partial \beta_N}{\partial N} < 0$ and the expected borrowing cost β_N decreases in N for $\varphi \in (0,1]$.

The partial derivative of β_N with respect to φ is

$$\frac{\partial \beta_N}{\partial \varphi} = -\underbrace{\frac{\sigma^2}{2}\left(R-1\right)\left(\Gamma-\gamma\right)\left(1-\varphi\right)^{N-2}\varphi N\left(N-1\right)}_{>0} < 0, \tag{B.19}$$

and therefore the expected borrowing cost β_N decreases in φ. □

Lemma 3

Proof.

Sublemma B.1. *Buyers never choose $\tau > \tau_{N-1}$.*

Proof. Since $\tau_{N-1} > \tau_N$, $V_N\left(\tau_{N-1}\right) > V_N\left(\tau\right)$ for all $\tau > \tau_{N-1}$. Therefore, from the definition of τ_{N-1}, for any $\alpha \in [0,1]$,

$$\alpha V_N\left(\tau_{N-1}\right) + \left(1-\alpha\right)V_{N-1}\left(\tau_{N-1}\right) > \alpha V_N\left(\tau\right) + \left(1-\alpha\right)V_{N-1}\left(\tau\right), \forall \tau > \tau_{N-1}. \tag{B.20}$$

It follows that deviations from $\tau > \tau_{N-1}$ to τ_{N-1} are profitable for any probability of being the first to settle the trade. □

Sublemma B.2. *If* $\frac{1}{2}[V_N(0)+V_{N-1}(0)] \geq V_{N-1}(\tau_{N-1})$*, then* $(\tau_j^\star, \tau_{-j}^\star) = (0,0)$ *is a pure strategy equilibrium.*

Proof. If both buyers choose immediate settlement, they are equally likely to settle first or second, and have expected utility $\frac{1}{2}(V_N(0)+V_{N-1}(0))$. If any buyer unilaterally deviates to $\tau > 0$, then he is sure to settle after his competitor and can earn at most $V_{N-1}(\tau_{N-1})$. Therefore, no deviation from $\tau = 0$ is profitable. □

Sublemma B.3. *Buyers never choose a specific* $\tau > 0$ *with positive probability.*

Proof. Assume buyers choose $\tau' > 0$ with positive probability η. With probability η^2, both buyers select τ' and earn $\frac{\eta^2}{2}(V_N(\tau')+V_{N-1}(\tau'))$. A buyer can increase his expected profit by putting zero weight on τ' and choosing instead $\tau' - \varepsilon \geq 0$ with positive probability, for a sufficiently small ε. Such deviation increases the expected profit to $\eta^2 V_N(\tau' - \varepsilon)$. □

Sublemma B.4. *The upper bound of the support for any equilibrium mixed strategy of* $\boldsymbol{B}$*, if such a strategy exists, is* τ_{N-1}*.*

Proof. Assume there exists an equilibrium mixed strategy with cumulative distribution function $G(\tau)$. Further, assume $\overline{\tau} \neq \tau_{N-1}$ is the smallest time-to-settlement such that $G(\overline{\tau}) = 1$, that is the upper bound of the mixed strategy support. From Sublemma B.1 it must be that $\overline{\tau} < \tau_{N-1}$. If a buyer chooses $\overline{\tau}$, he is sure to be the second to settle and obtain $V_{N-1}(\overline{\tau})$. If the buyer instead deviates to τ_{N-1}, since $G(\tau_{N-1}) = 1$, he also is the second to settle but obtains the higher utility level $V_{N-1}(\tau_{N-1}) > V_{N-1}(\overline{\tau})$ (by the definition of τ_{N-1}). Therefore, it must be that $\overline{\tau} = \tau_{N-1}$. □

Corollary B.1. *If there exists an equilibrium mixed strategy for* $\boldsymbol{B}$*, it yields expected utility* $V_{N-1}(\tau_{N-1})$*.*

Proof. From Sublemma B.4, if an equilibrium mixed strategy exists, then τ_{N-1} is in the mixing support and further $G(\tau_{N-1}) = 1$. Therefore, by choosing τ_{N-1} a buyer is guaranteed to be second to settle his trade and earns expected utility $V_{N-1}(\tau_{N-1})$, which is necessarily the expected utility of any other time-to-settlement in the mixed strategy support. □

Sublemma B.5. *If* $\frac{1}{2}[V_{N-1}(0)+V_N(0)] \geq V_{N-1}(\tau_{N-1})$*, then* $(\tau_j^\star, \tau_{-j}^\star) = (0,0)$ *is the unique equilibrium.*

Proof. From Sublemma B.2, $(\tau_j^\star, \tau_{-j}^\star) = (0,0)$ is a pure strategy equilibrium yielding expected utility $\frac{1}{2}[V_{N-1}(0)+V_N(0)]$. Sublemma B.3 further establishes that $(\tau_j^\star, \tau_{-j}^\star) = (0,0)$ is the unique pure strategy equilibrium. It remains to be shown there are no mixed strategy equilibria. If there is one, from Corollary B.1 it yields utility $V_{N-1}(\tau_{N-1})$, which is weakly lower than the utility of the pure strategy equilibrium, that is $\frac{1}{2}[V_{N-1}(0)+V_N(0)]$. Therefore, there can be no mixed strategy equilibrium. □

Sublemma B.5 completes the proof for the first part of Lemma 3. In what follows, we discuss the buyers' equilibrium strategies if $V_{N-1}(\tau_{N-1}) > \frac{1}{2}[V_{N-1}(0)+V_N(0)]$.

Sublemma B.6. *If $V_{N-1}(\tau_{N-1}) > V_N(0)$, then the unique equilibrium is in mixed strategies: buyers randomize between $\underline{\tau}$ and τ_{N-1}, where $\underline{\tau}$ solves*

$$V_N(\underline{\tau}) = V_{N-1}(\tau_{N-1}). \tag{B.21}$$

Proof. First off, if $V_{N-1}(\tau_{N-1}) > V_N(0)$ then since (i) $V_N(\tau_N) > V_{N-1}(\tau_{N-1})$, and (ii) $V_N(\tau)$ increases on $[0, \tau_N]$, it follows that there exists a strictly positive $\underline{\tau} < \tau_N$ such that $V_N(\underline{\tau}) = V_{N-1}(\tau_{N-1})$.

Second, $\underline{\tau}$ must be the lower bound of any equilibrium mixed strategy support. Assume instead $\underline{\tau}^- < \underline{\tau}$ is the lower bound. If the buyer chooses $\underline{\tau}^-$, then he is the first to settle and obtains $V_N(\underline{\tau}^-)$. Since $\underline{\tau} < \tau_N$, it follows that $V_N(\underline{\tau}^-) < V_N(\underline{\tau}) = V_{N-1}(\tau_{N-1})$, which is the expected utility of a mixed strategy equilibrium, from Corollary B.1. Assume now $\underline{\tau}^+ > \underline{\tau}$ is the lower bound. A buyer can profitably deviate by choosing $\underline{\tau} + \varepsilon < \underline{\tau}^+$, then he is the first to settle and obtains $V_N(\underline{\tau} + \varepsilon) > V_N(\underline{\tau}) = V_{N-1}(\tau_{N-1})$.

Third, no buyer can profitably deviate to immediate settlement. Choosing $\tau = 0$ ensures that a buyer is first to settle and gain $V_N(0) < V_{N-1}(\tau_{N-1})$. □

If $\frac{1}{2}[V_{N-1}(0) + V_N(0)] < V_{N-1}(\tau_{N-1}) \leq V_N(0)$ then:

1. Buyers are better off by playing a mixed strategy yielding expected utility $V_{N-1}(\tau_{N-1})$ than by choosing immediate settlement in pure strategies.

2. However, if the competitor plays such a mixed strategy and chooses a random time to settlement between $\underline{\tau} > 0$ and τ_{N-1}, then immediate settlement is a profitable deviation as it guarantees being first to settle the trade and yields expected utility $V_N(0)$.

It follows that buyers choose immediate settlement with some positive probability $\pi < 1$. The expected utility of $\mathbf{B}_j$ from choosing $\tau = 0$ is:

$$\mathbb{E}\left[U_{\mathbf{B}_j}(\tau = 0)\right] = \underbrace{\frac{\pi}{2}[V_{N-1}(0) + V_N(0)]}_{\mathbf{B}_{-j} \text{ chooses } \tau=0} + \underbrace{(1-\pi) V_N(0)}_{\mathbf{B}_{-j} \text{ chooses } \tau>0}. \tag{B.22}$$

The expected utility from choosing immediate settlement needs to be equal, in equilibrium, with the expected utility of any other settlement time choice in the mixed distribution support. Therefore, π is pinned down by

$$\frac{\pi}{2}[V_{N-1}(0) + V_N(0)] + (1-\pi) V_N(0) = V_{N-1}(\tau_{N-1}), \tag{B.23}$$

and therefore

$$\pi = 2\frac{V_N(0) - V_{N-1}(\tau_{N-1})}{V_N(0) - V_{N-1}(0)}. \tag{B.24}$$

If a buyer chooses the lower bound of the mixed strategy distribution, $\underline{\tau}$, then he settles first with probability $1 - \pi$. Therefore, since all points in the mixed strategy support yield the same expected utility, $\underline{\tau}$ solves

$$(1-\pi) V_N(\underline{\tau}) + \pi V_{N-1}(\underline{\tau}) = V_{N-1}(\tau_{N-1}). \tag{B.25}$$

Note that the two mixed strategy cases, that is $\frac{1}{2}[V_{N-1}(0) + V_N(0)] < V_{N-1}(\tau_{N-1}) \leq V_N(0)$ and $V_{N-1}(\tau_{N-1}) > V_N(0)$, can be put together if redefine π as

$$\pi = \max\left\{0, 2\frac{V_N(0) - V_{N-1}(\tau_{N-1})}{V_N(0) - V_{N-1}(0)}\right\}. \tag{B.26}$$

□

Proposition 2

Proof. First, equation (27), evaluated at the equilibrium buyer's strategy $G(\tau)$ given in (33), implies that the intermediaries earn zero expected profit if they rationally anticipate the buyers' and security lenders' equilibrium behaviour. Second, Lemma 3 solves for the equilibrium buyers' strategy given competitive intermediaries and the equilibrium lenders' behaviour. Third, Lemma 1 describes the equilibrium behaviour of security lenders. Finally, we assume intermediaries always borrow the asset at settlement if the large seller did not arrive. □

Corollary 2

Proof. From the envelope theorem and $\tau_N \geq 0$, $V_N(\tau_N)$ decreases in δ since

$$\frac{\partial V_N(\tau_N)}{\partial \delta} = -\tau_N V_N(\tau_N) \leq 0, \forall N \geq 1. \tag{B.27}$$

From Lemma 3, the immediate settlement probability π weakly decreases in $V_{N-1}(\tau_{N-1})$. Since the immediate settlement value functions $V_N(0)$ and $V_{N-1}(0)$ do not depend on δ, it follows that the immediate settlement probability π weakly increases in δ.

From Lemma 3, for $\delta \downarrow 0$, buyers choose an arbitrarily large time-to-settlement and never choose immediate settlement: consequently, $\pi = 0$. For $\delta \to \infty$, both buyers always choose immediate settlement, that is $\pi = 1$. It follows that there exists a $\hat{\delta}$ such that $\pi = 1$ for $\delta > \hat{\delta}$.

□

Corollary 3

Proof. From Lemma 3, buyers choose immediate settlement with probability one if and only if $V_{N-1}(0) + V_N(0) \geq V_{N-1}(\tau_{N-1})$, which therefore corresponds to the case $\delta > \hat{\delta}$. The settlement rat race cost for the buyers is:

$$\mathcal{C} = V_N(\tau_N) + V_{N-1}(\tau_{N-1}) - (V_N(0) + V_{N-1}(0)), \tag{B.28}$$

which decreases in δ and since both $V_N(\tau_N)$ and $V_{N-1}(\tau_{N-1})$ decrease in δ, while $V_N(0)$ and $V_N(0)$ do not depend on δ.

Since Lemma 2 establishes β_N decreases in φ, to show that the cost $\mathcal{C}$ in (B.28) decreases in φ it is enough to show that

$$\frac{\partial\left(V_N(\tau_N)-V_N(0)\right)}{\partial\beta_N}>0, \tag{B.29}$$

for all N and $\tau_N>0$. We compute the partial derivative and obtain

$$\frac{\partial\left(V_N(\tau_N)-V_N(0)\right)}{\partial\beta_N}=1-\left[\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right]^{-\frac{\delta+\lambda}{\lambda}}. \tag{B.30}$$

Since $\tau_N>0$, from Definition 2 it follows that $\frac{\beta_N(\delta+\lambda)}{\delta\theta}>1$. Consequently, since $\delta>0$ and $\lambda>0$, $\left[\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right]^{-\frac{\delta+\lambda}{\lambda}}<1$ which implies the partial derivative in (B.29) is positive.

Conversely, if $\delta<\hat{\delta}$, $V_{N-1}(0)+V_N(0)<V_{N-1}(\tau_{N-1})$ and buyers choose immediate settlement with probability $\pi<1$. The rat race cost is in this case is:

$$\mathcal{C}=V_N(\tau_N)-V_{N-1}(\tau_{N-1}). \tag{B.31}$$

To show the cost increases in δ, it is enough to show that $\frac{\partial^2 V_N(\tau_N)}{\partial N\partial\delta}>0$ or, since β_N decreases in N, that $\frac{\partial^2 V_N(\tau_N)}{\partial\beta_N\partial\delta}<0$. The latter expression can be written as

$$\frac{\partial^2 V_N(\tau_N)}{\partial\beta_N\partial\delta}=\frac{1}{\delta\lambda}\left(\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right)^{-\frac{\delta+\lambda}{\lambda}}\left(\delta\log\left[\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right]-\lambda\right). \tag{B.32}$$

It remains to be shown that $\delta\log\left[\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right]<\lambda$. Since $\theta>\beta_0>\beta_N$, it follows that:

$$\log\left[\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right]<\log\left(1+\frac{\lambda}{\delta}\right)<\frac{\lambda}{\delta}, \tag{B.33}$$

and therefore it follows that $\frac{\partial^2 V_N(\tau_N)}{\partial\beta_N\partial\delta}<0$.

To show that the settlement rat race cost increases in φ, it is enough to show that $\frac{\partial^2 V_N(\tau_N)}{\partial N\partial\varphi}>0$ or, alternatively, that $\frac{\partial^2 V_N(\tau_N)}{\partial\beta_N^2}>0$ since from Lemma 2 β_N decreases in both N and φ. Indeed,

$$\frac{\partial^2 V_N(\tau_N)}{\partial\beta_N^2}=\frac{\delta\theta}{\beta_N^2\lambda}\left(\frac{\beta_N(\delta+\lambda)}{\delta\theta}\right)^{-\frac{\delta}{\lambda}}>0, \tag{B.34}$$

which concludes the proof. $\square$

Corollary 4

Proof. We solve the equations $\tau_N = 0$ and $\tau_{N-1} = 0$ and obtain

$$\delta_N \equiv \frac{\lambda\beta_N}{\theta - \beta_N} \leq \frac{\lambda\beta_{N-1}}{\theta - \beta_{N-1}} \equiv \delta_{N-1}, \tag{B.35}$$

since $\beta_N \leq \beta_{N-1}$ and $\theta > \beta_0 \geq \beta_N$ for any positive N. Clearly, $\hat{\delta} < \delta_{N-1}$. Otherwise, for some $\delta_{N-1} + \epsilon < \hat{\delta}$ (with arbitrarily small ϵ), we have that $\tau_N(\delta_{N-1} + \epsilon) = \tau_{N-1}(\delta_{N-1} + \epsilon)$=0 and consequently:

$$\pi = 2\frac{V_N(0) - V_{N-1}(0)}{V_N(0) - V_{N-1}(0)} = 2 > 1, \tag{B.36}$$

which is a contradiction since $\pi \leq 1$.

For $\hat{\delta} < \delta_N$, the following condition needs to hold

$$\frac{1}{2}\left[V_{N-1}(0) + V_N(0)\right] > V_{N-1}\left(\tau_{N-1}(\delta_N)\right), \tag{B.37}$$

since if $V_{N-1}(\tau_{N-1}(\delta_N)) \geq \frac{1}{2}[V_{N-1}(0) + V_N(0)]$ then from Lemma 3 buyers choose immediate settlement ($\pi = 1$) for δ_N and therefore $\delta_N \geq \hat{\delta}$. □

Lemma 4

Proof. Let $G^{\text{gen}}(\cdot)$ denote the generalized equilibrium time-to-settlement mixing distribution if the penalty $z(\tau)$ is endogenous. Using the same logic as in equations (26) and (27), and acknowledging the intermediary fails to deliver if the penalty is small enough, the expected intermediary utility is:

$$\begin{aligned}\mathbb{E}\left[U_{\mathbf{I}} \mid \mathbf{I}\text{ survives}\right] = {} & \left(1 - e^{-\lambda\tau}\right)p(\tau) \\ & + e^{-\lambda\tau}\left(1 - G^{\text{gen}}(\tau)\right)\mathbb{E}_N\left[\max\left\{p(\tau) - b, -z(\tau)\right\}\right] \\ & + e^{-\lambda\tau}G^{\text{gen}}(\tau^-)\,\mathbb{E}_{N-1}\left[\max\left\{p(\tau) - b, -z(\tau)\right\}\right] \\ & + e^{-\lambda\tau}\left(G^{\text{gen}}(\tau) - G^{\text{gen}}(\tau^-)\right)\frac{1}{2}\Big(\mathbb{E}_N\left[\max\left\{p(\tau) - b, -z(\tau)\right\}\right] \\ & + \mathbb{E}_{N-1}\left[\max\left\{p(\tau) - b, -z(\tau)\right\}\right]\Big),\end{aligned} \tag{B.38}$$

where $\mathbb{E}_N[\cdot]$ denotes expectation conditional on at most N security lenders with risk-aversion γ being available.

We note that the intermediary's borrowing decision depends on the sum between the price and the penalty, and we define $x(\tau) = p(\tau) + z(\tau)$. Let $\alpha_N(\tau) = \mathbb{P}_N(b \leq x(\tau))$, that is the probability that (**BC**) holds conditional on a given N. It follows that the expected utility of the buyer is

$$
\begin{aligned}
\mathbb{E}\left[U_{\mathbf{B}}\right] = & e^{-\delta\tau}\left(1-e^{-\lambda\tau}\right)\left(\theta-p\left(\tau\right)\right) \\
& + e^{-(\delta+\lambda)\tau}\Big\{G^{\text{gen}}\left(\tau^{-}\right)\left[\alpha_{N-1}\left(\tau\right)\left(\theta-p\left(\tau\right)\right)+\left(1-\alpha_{N-1}\left(\tau\right)\right)\left(z\left(\tau\right)+\frac{\lambda}{\lambda+\delta}\theta\right)\right] \\
& + \left(1-G^{\text{gen}}\left(\tau\right)\right)\left[\alpha_{N}\left(\tau\right)\left(\theta-p\left(\tau\right)\right)+\left(1-\alpha_{N}\left(\tau\right)\right)\left(z\left(\tau\right)+\frac{\lambda}{\lambda+\delta}\theta\right)\right] \\
& + \frac{1}{2}\left(G^{\text{gen}}\left(\tau\right)-G^{\text{gen}}\left(\tau^{-}\right)\right)\Big\{\left[\alpha_{N}\left(\tau\right)\left(\theta-p\left(\tau\right)\right)+\left(1-\alpha_{N}\left(\tau\right)\right)\left(z\left(\tau\right)+\frac{\lambda}{\lambda+\delta}\theta\right)\right] \\
& + \left[\alpha_{N-1}\left(\tau\right)\left(\theta-p\left(\tau\right)\right)+\left(1-\alpha_{N-1}\left(\tau\right)\right)\left(z\left(\tau\right)+\frac{\lambda}{\lambda+\delta}\theta\right)\right]\Big\}\Big\}.
\end{aligned}
\tag{B.39}
$$

Combining equations (B.38) and (B.39), we obtain an expression for the buyers' expected utility which is qualitatively similar to equation (29), that is

$$
\begin{aligned}
\mathbb{E}\left[U_{\mathbf{B}}\right] = & G^{\text{gen}}\left(\tau^{-}\right)V_{N-1}^{\text{gen}}\left(\tau\right)+\left(G^{\text{gen}}\left(\tau_{j}\right)-G^{\text{gen}}\left(\tau^{-}\right)\right)\frac{V_{N}^{\text{gen}}\left(\tau\right)+V_{N-1}^{\text{gen}}\left(\tau\right)}{2} \\
& + \left(1-G^{\text{gen}}\left(\tau\right)\right)V_{N}^{\text{gen}}\left(\tau\right),
\end{aligned}
\tag{B.40}
$$

where $V_{N}^{\text{gen}}\left(\cdot\right)$ is a generalization of $V_{N}\left(\cdot\right)$ in (28) as follows:

$$
V_{N}^{\text{gen}}\left(\tau\right)=e^{-\delta\tau}\left[\left(1-e^{-\lambda\tau}\right)\theta+e^{-\lambda\tau}\left(\frac{\lambda}{\lambda+\delta}\theta+\alpha_{N}\left(\tau\right)\left(\frac{\delta}{\lambda+\delta}\theta-\mathbb{E}_{N}\left[b\mid b\leq x\left(\tau\right)\right]\right)\right)\right], \tag{B.41}
$$

for $x\left(\tau\right)=p\left(\tau\right)+z\left(\tau\right)$ and for any $N\geq 1$.

Next, we show that $V_{N}^{\text{gen}}\left(\tau\right)$ reaches its maximum if $x\left(\tau\right)=\frac{\delta}{\lambda+\delta}\theta$, for any N, which is enough to prove the Lemma. Dropping the argument τ for exposition purposes, we want to maximize:

$$
x^{\star}=\arg\max_{x}\mathbb{P}_{N}\left(b\leq x\right)\left\{\frac{\delta}{\lambda+\delta}\theta-\mathbb{E}\left[b\mid b\leq x\right]\right\}. \tag{B.42}
$$

Denote the cumulative distribution of b, conditional on N as $H_{N}\left(\cdot\right):\left[\underline{b}_{N},\overline{b}_{N}\right]\mapsto[0,1]$ and let $h_{N}\left(\cdot\right)$ be the density function. We can therefore rewrite (B.42) as

$$
x^{\star}=\arg\max_{x}\frac{\delta}{\lambda+\delta}\theta H_{N}\left(x\right)-\int_{\underline{b}_{N}}^{x}y\,\mathrm{d}H_{N}\left(y\right). \tag{B.43}
$$

From the first order condition, it follows that $x^{\star}$ solves

$$
\left(\frac{\delta}{\lambda+\delta}\theta-x^{\star}\right)h_{N}\left(x^{\star}\right)=0, \tag{B.44}
$$

and $x^{\star}=\frac{\delta}{\lambda+\delta}\theta$ since $h_{N}\left(x\right)>0$ on the support. Consequently, $p\left(\tau\right)+z\left(\tau\right)=\frac{\delta}{\lambda+\delta}\theta$. □

Proposition 3

Proof. From equation (9) and Lemma 4 we can rewrite the intermediaries' expected utility as:

$$\mathbb{E}\left[U_{\mathbf{I}} \mid \mathbf{I}\text{ survives}\right] = p\left(\tau\right) - e^{-\lambda\tau}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right]. \tag{B.45}$$

Since intermediaries are competitive, they earn zero expected utility. Consequently, the competitive price and penalty schedules become:

$$p\left(\tau\right) = e^{-\lambda\tau}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right] \text{ and} \tag{B.46}$$

$$z\left(\tau\right) = \frac{\delta}{\delta+\lambda}\theta - e^{-\lambda\tau}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right]. \tag{B.47}$$

Further, from (B.46) and (B.47), if the optimal penalty schedule is implemented, the buyers' expected utility in (5) and (6) is the same irrespective of whether the intermediary borrows at settlement or not, that is:

$$\mathbb{E}\left[U_{\mathbf{B}}\right] = e^{-\delta\tau}\left(\theta - p\left(\tau\right)\right). \tag{B.48}$$

From equations (B.46) and (B.48), the optimal settlement time $\tau^\star$ is

$$\tau^\star = \max\left\{0, \frac{1}{\lambda}\log\left(\frac{(\delta+\lambda)}{\theta\delta}\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right]\right)\right\}. \tag{B.49}$$

Since $\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right] \leq \frac{\delta}{\delta+\lambda}\theta$, it follows that $\tau^\star = 0$ and buyers choose immediate settlement in equilibrium.

Securities lenders charge prices ℓ such that an intermediary is not better off by (i) borrowing from a fringe lender, or (ii) failing to settle and paying the penalty. That is, the upper bound on ℓ is pinned down by

$$R\ell - \frac{\gamma}{2}\sigma^2 \leq \min\left\{\frac{\Gamma}{2}\sigma^2\left(R-1\right), \frac{\delta}{\delta+\lambda}\theta\right\}. \tag{B.50}$$

The maximum ℓ security lenders ask for is

$$\overline{\ell}^{\text{gen}} = \frac{\min\left\{\frac{\Gamma}{2}\sigma^2\left(R-1\right), \frac{\delta}{\delta+\lambda}\theta\right\} + \gamma\sigma^2/2}{R}. \tag{B.51}$$

As in Section 5.1, security lenders randomize over repo prices. Following the same steps as in the proof of Lemma 1, we can solve for the (generalized) equilibrium distribution of repo prices,

$$F_N^{\text{gen}}(\ell) = \begin{cases} 0, & \text{if } \ell \leq \underline{\ell}^{\text{gen}} \\ \frac{1}{\varphi} - \frac{1-\varphi}{\varphi} \sqrt[N-1]{\frac{2\min\left\{\frac{\Gamma(R-1)}{2}\sigma^2, \frac{\delta}{\delta+\lambda}\theta\right\} - \sigma^2(R-1)\gamma}{R(2\ell - \gamma\sigma^2)}}, & \text{if } \ell \in \left(\underline{\ell}^{\text{gen}}, \overline{\ell}^{\text{gen}}\right] \\ 1, & \text{if } \ell > \overline{\ell}^{\text{gen}}, \end{cases} \tag{B.52}$$

where

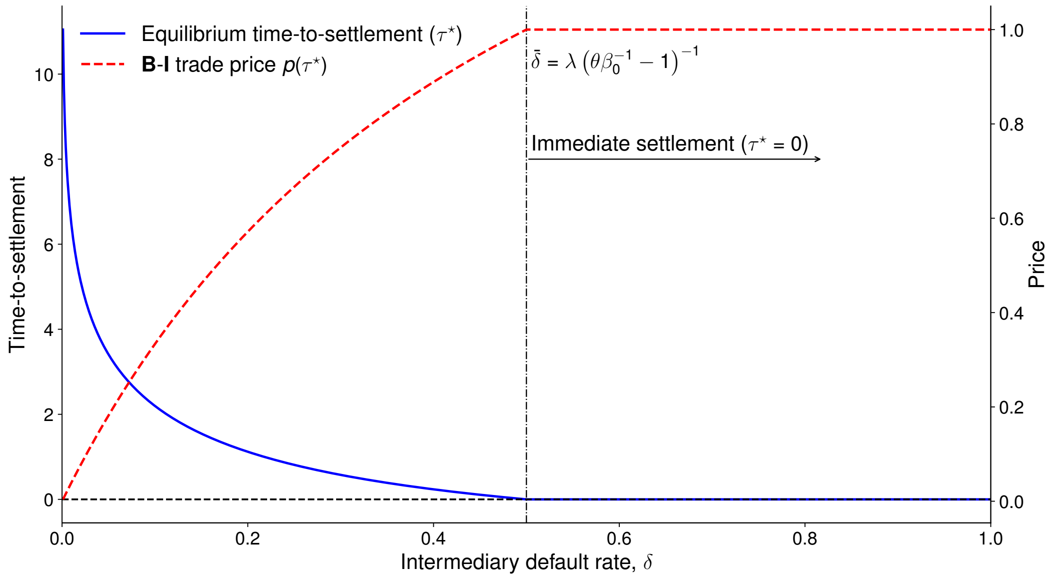
$$\underline{\ell}^{\text{gen}} = \frac{1}{2}\left[\gamma\sigma^2 + (1-\varphi)^{N-1} \frac{2\min\left\{\frac{\Gamma(R-1)}{2}\sigma^2, \frac{\delta}{\delta+\lambda}\theta\right\} - (R-1)\gamma\sigma^2}{R}\right]. \tag{B.53}$$

Finally, we note that we can compute $\mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right]$ in closed form as

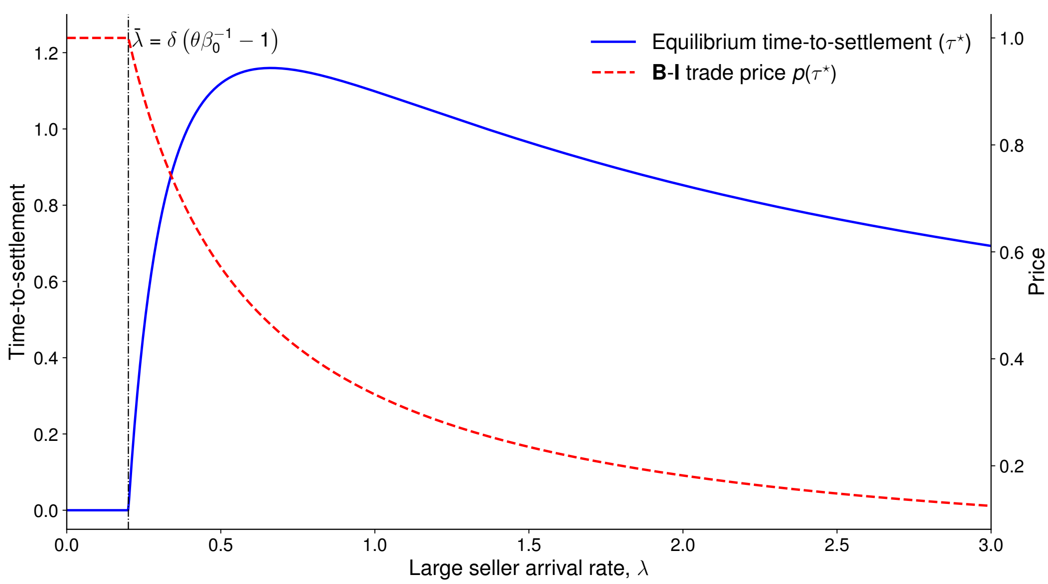
$$\begin{aligned} \mathbb{E}\left[\min\left\{b, \frac{\delta}{\delta+\lambda}\theta\right\}\right] &= N\varphi(1-\varphi)^{N-1}\left[\min\left\{\frac{\Gamma}{2}\sigma^2(R-1), \frac{\delta}{\delta+\lambda}\theta\right\} - \frac{\gamma}{2}\sigma^2(R-1)\right] \\ &+ \frac{\gamma}{2}\sigma^2(R-1)\left(1-(1-\varphi)^N\right) + (1-\varphi)^N \min\left\{\frac{\Gamma}{2}\sigma^2(R-1), \frac{\delta}{\delta+\lambda}\theta\right\}. \end{aligned} \tag{B.54}$$

$\square$

Figure 1: **Benchmark equilibrium time-to-settlement and the corresponding price** This figure illustrates the equilibrium time-to-settlement and the corresponding trade price as a function of the default rate δ (top panel) and the large seller's arrival rate λ (bottom panel). Parameter values: $\sigma = 1$, $\theta = 2$, $\Gamma = 2$, $R = 2$, $\lambda = 0.5$ (top panel), and $\delta = 0.2$ (bottom panel).



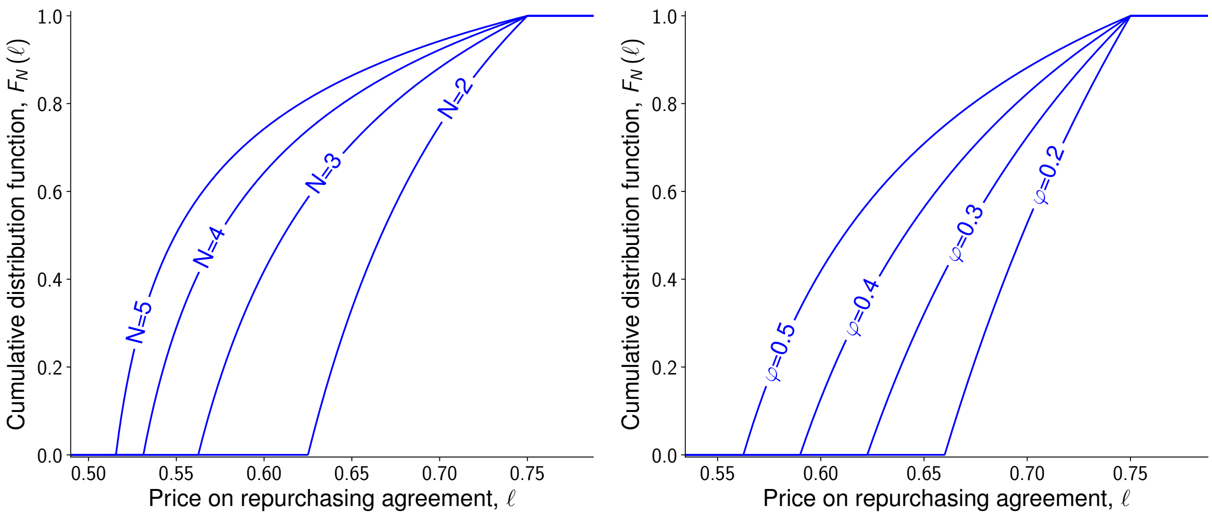
(a) Time-to-settlement, trade price, and counterparty risk.



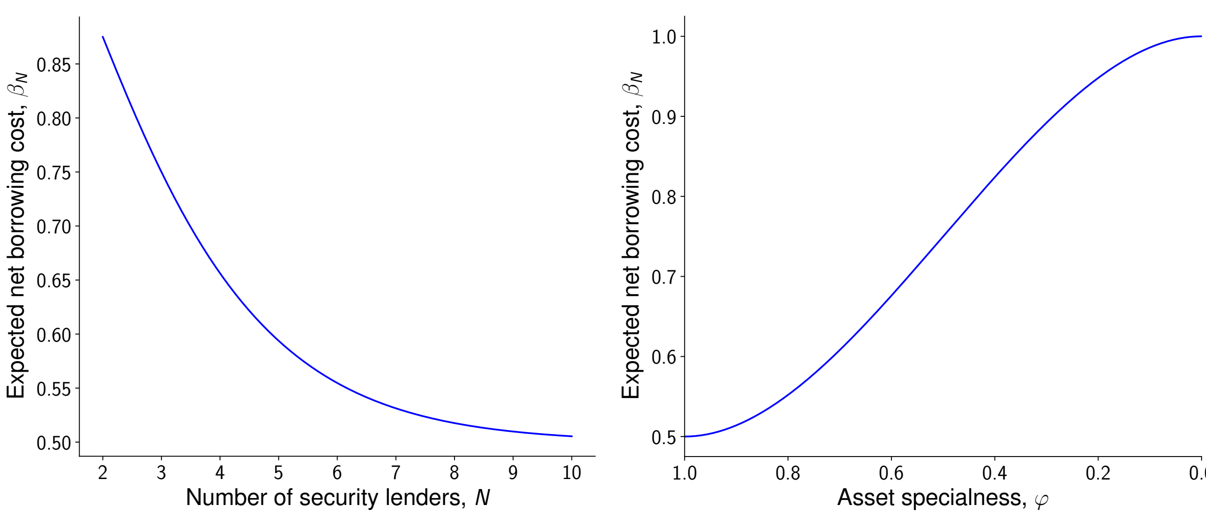
(b) Time-to-settlement, trade price, and seller arrival rate.

Figure 2: **Competition and borrowing costs on the repo market**

Panel (a) illustrates the comparative statics of the distribution of repo prices, $F_N(\ell)$, with respect to the number of lenders, N, and the probability that each lender owns the security, φ, which we refer to as asset specialness. A lower φ corresponds to a higher specialness. Panel (b) illustrates the expected borrowing cost for the intermediary, β_N, as a function of N and φ. Parameter values: $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, and $\varphi = 0.5$.



(a) Cumulative distribution function of repo prices, $F_N(\ell)$.



(b) Expected borrowing cost for the intermediary, β_N.

Figure 3: **Exogenous-penalty equilibrium time-to-settlement distribution**
This figure illustrates the exogenous penalty equilibrium time-to-settlement distribution $G(\tau)$ as a function of the default rate δ. Parameter values: $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, $\varphi = 0.5$ (top panel) and $\delta = 0.2$ (bottom panel).

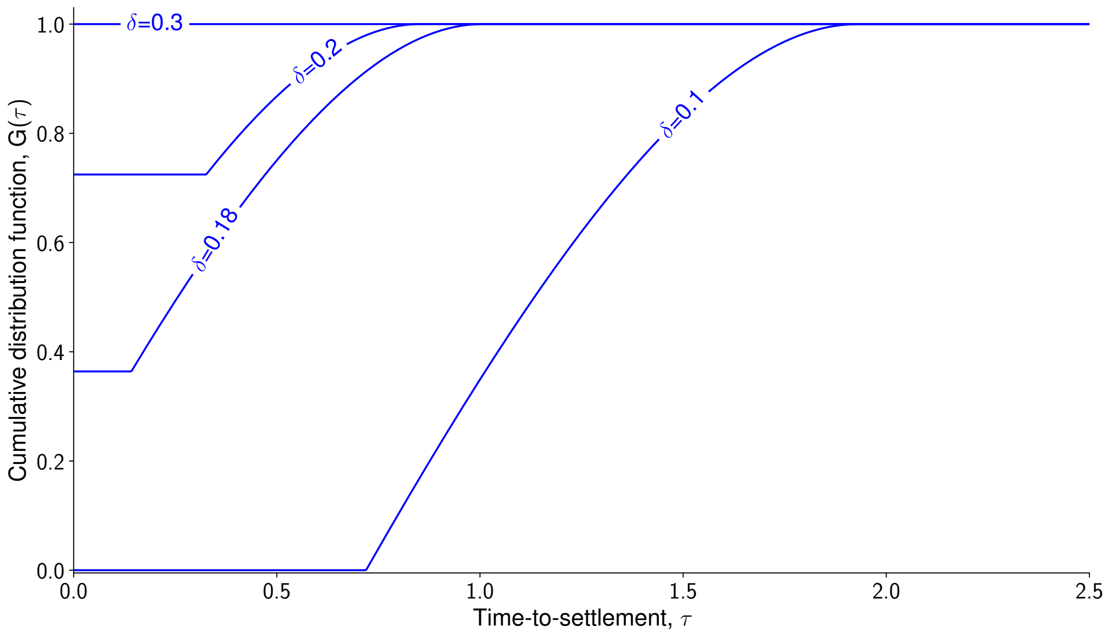


Figure 4: **Price-settlement delay schedules with random borrowing cost**
This figure illustrates the competitive price-settlement delay schedules posted by intermediaries in the exogenous penalty equilibrium, as defined in equation (27) and Proposition 2. We plot the schedule for various levels of φ, which we refer to as specialness of the asset. The red dashed segments correspond to settlement delays buyers choose in a mixed-strategy equilibrium. A red circle at $\tau = 0$ indicates buyers choose immediate settlement with positive probability π in equilibrium; the radius of the circle is proportional to π. Parameter values: $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, and $\delta = 0.2$.

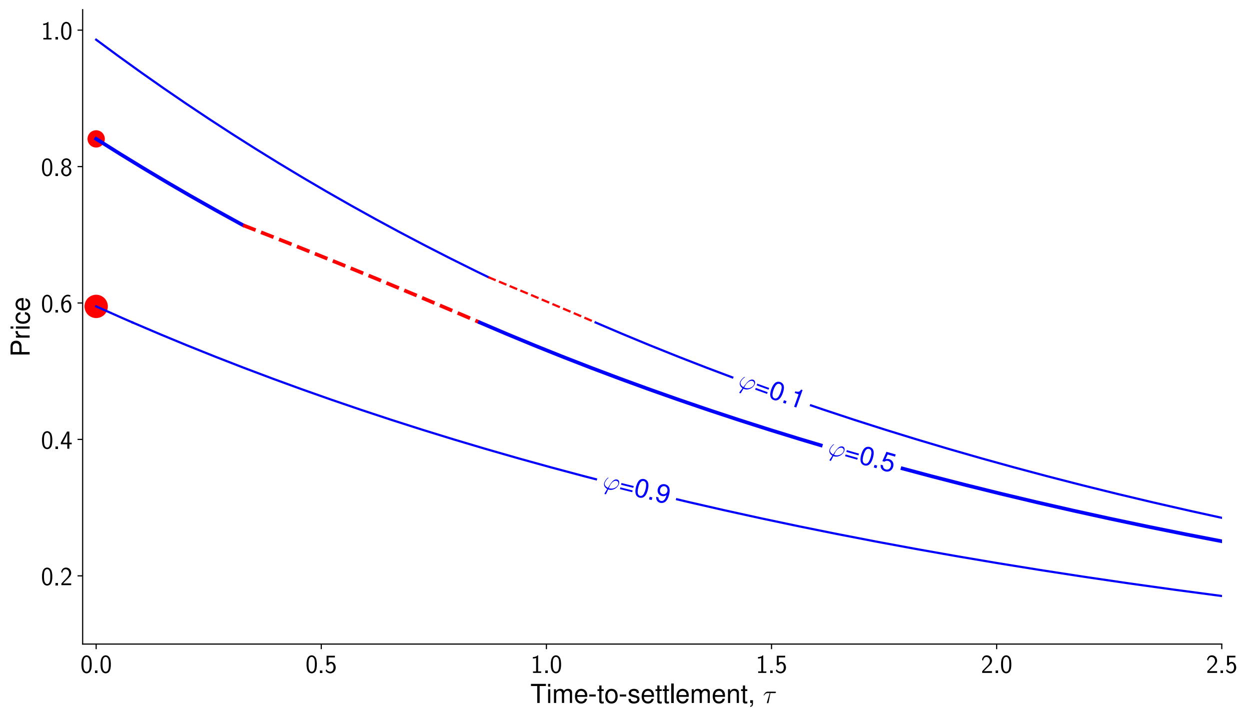


Figure 5: **Cost of the settlement rat race**

This figure illustrates the cost of the settlement rat race for the buyers, $\mathcal{C}$, and the immediate settlement probability, π, as function of the default rate δ. Parameter values: $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, $\varphi = 0.5$.

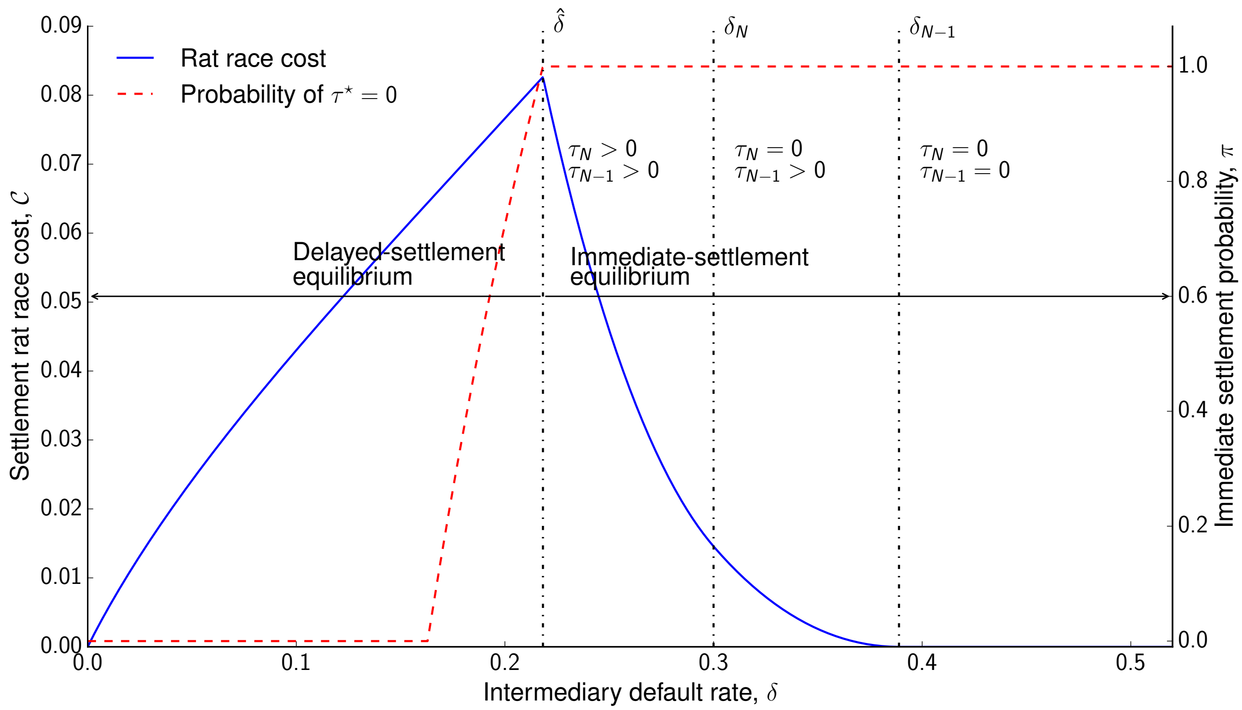


Figure 6: **Failure-to-deliver penalty schedules if intermediary always borrows**
This figure illustrates the equilibrium failure-to-deliver penalty schedules posted by intermediaries, defined in equation (27) and Proposition 3. We plot the schedule for various levels of φ, which we refer to as specialness of the asset. Parameter values: $\theta = 2$, $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, and $\delta = 0.2$.

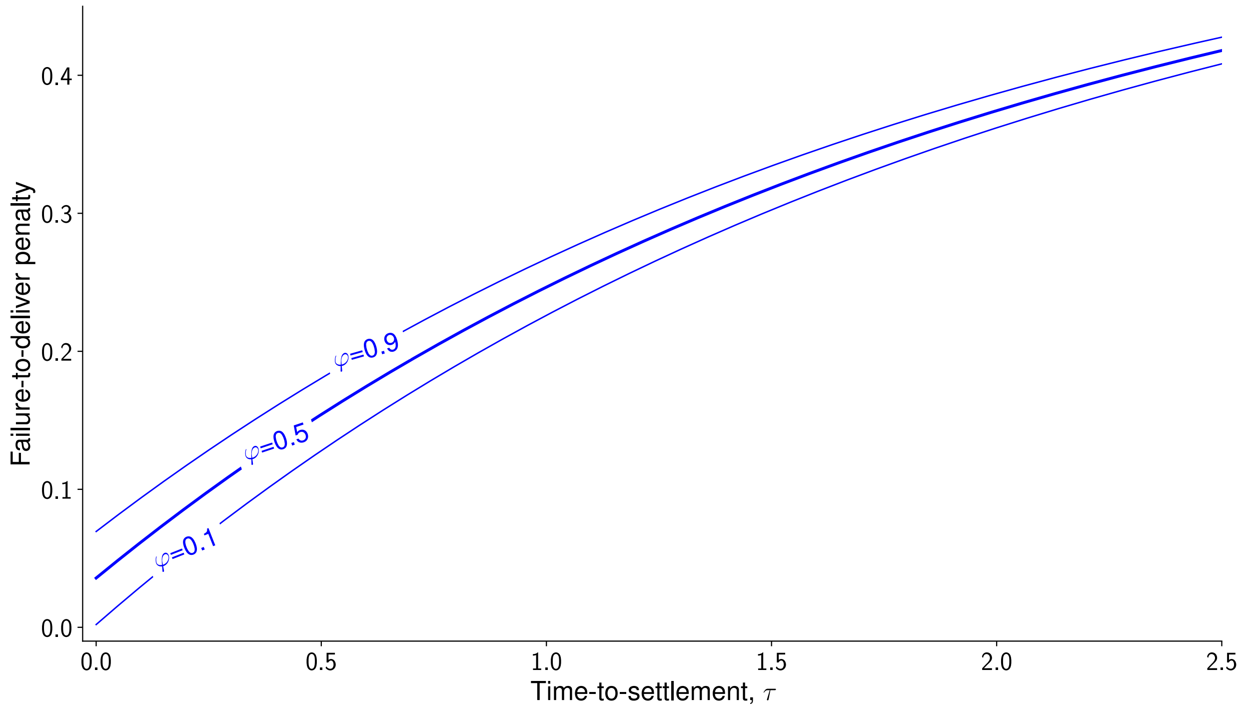


Figure 7: **Failure-to-deliver probability**

This figure illustrates the equilibrium failure-to-deliver probability for a **B-I** trade, as a function of the asset specialness φ. We plot the schedule for various levels of intermediary default rates δ. Parameter values: $\theta = 2$, $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, and $\delta = 0.2$.

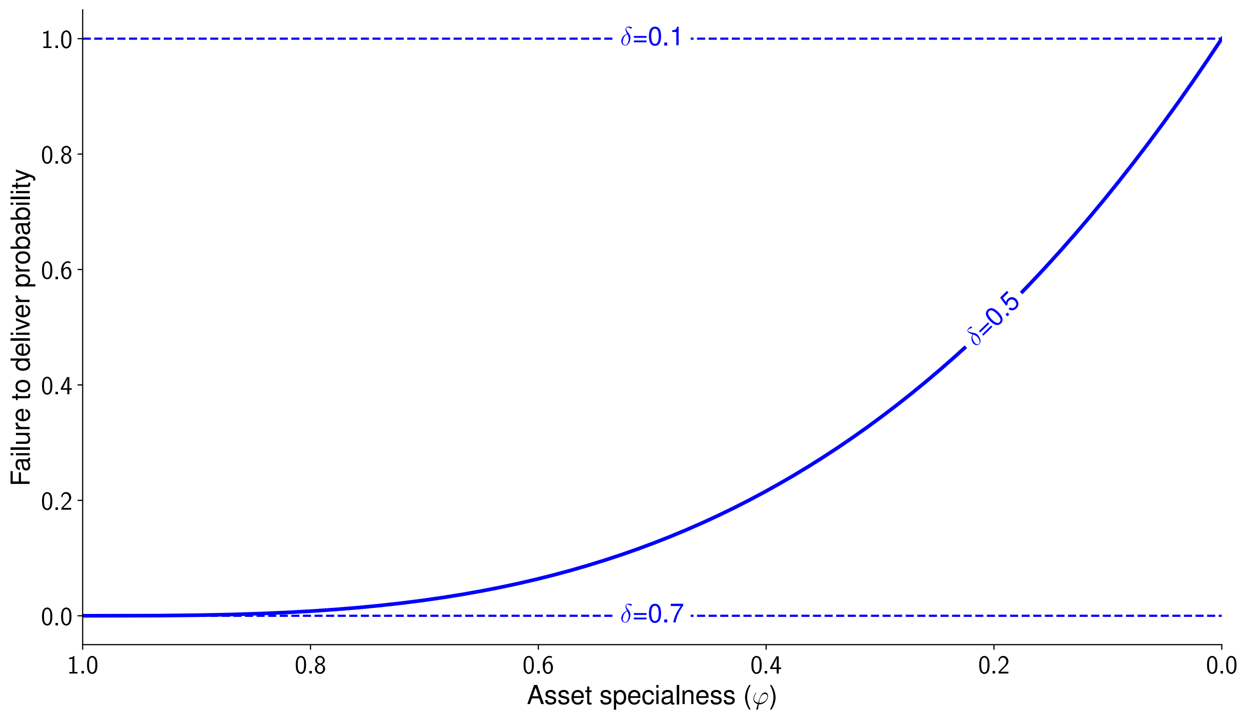


Figure 8: **Expected rents of security lenders**

This figure illustrates the expected utility of a security lender, conditional on owning the asset, as a function of the intermediary default rate δ. We plot the schedule for various levels of φ, which we refer to as specialness of the asset. Parameter values: $\theta = 2$, $\sigma = 1$, $\Gamma = 2$, $\gamma = 1$, $R = 2$, $N = 3$, and $\delta = 0.2$.