Essays on Corporate Finance and Financial Markets

By

Jiaheng Yu

B.A. Economics and Finance
B.S. Pure and Applied Mathematics
Tsinghua University, 2018

S.M. Management Research
Massachusetts Institute of Technology, 2021

SUBMITTED TO THE DEPARTMENT OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Jiaheng Yu
MIT Sloan School of Management
April 15, 2023

Certified by: Hui Chen
Professor of MIT Sloan School of Management, Thesis Supervisor

Certified by: David Thesmar
Professor of MIT Sloan School of Management, Thesis Supervisor

Accepted by: Eric So
Sloan Distinguished Professor of Financial Economics
Professor, Accounting and Finance
Faculty Chair, MIT Sloan PhD Program

Essays on Corporate Finance and Financial Markets

by

Jiaheng Yu

Submitted to the Department of Management
on April 15, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management

Abstract

This thesis consists of three chapters.

Chapter 1 studies the informational role of trade credit and the accounts receivable financing market. I hand collect new data on the contracts of accounts receivable based loans and trade credit terms. I find that sellers experiencing payment delays are primarily financed through accounts receivable based loans. These loans are 2-4% per year more expensive than buyers' borrowing rates and require a 20% average haircut on invoice value. Seller moral hazard that leads to bad-quality products is a determinant of payment delays, and although difficult to observe in existing data, can be uncovered from terms of accounts receivable based loans. Lenders help improve the quality of sellers: sellers who successfully receive credit experience a 5% decline in receivable days and have higher sales and longer relationships with buyers. I propose and structurally estimate a trade credit model that incorporates accounts receivable financing. In the model, the buyer trades off the financial cost and the incentive effect of trade credit and learns from the lender's loan decisions. I show through counterfactual analyses that regulatory limits on payment delays increase the presence of bad products and lower output, while subsidizing accounts receivable financing may increase output at relatively low expense.

In Chapter 2, joint work with Rodney Garratt and Haoxiang Zhu, we study the design of Central Bank Digital Currencies. Banks of different sizes respond differently to interest on reserves (IOR) policy. For low IOR rates, large banks are non-responsive to IOR rate changes, leading to weak pass-through of IOR rate changes to deposit rates. In these circumstances, a central bank digital currency (CBDC) may be used to provide competitive pressure to drive up deposit rates and improve monetary policy transmission. We explore the implications of two design features: interest rate and convenience value. Increasing the CBDC interest rate past a point where it becomes a binding floor, increases deposit rates but leads to greater inequality of market shares in both deposit and lending markets and can reduce the responsiveness of deposit rates to changes in the IOR rate. In contrast, increasing convenience, from sufficiently high levels, increases deposit rates, causes market shares to converge and can increase the responsiveness of deposit rates to changes in the IOR rate.

In Chapter 3, joint work with Jingxiong Hu, we study the effect of "guaranteed close" on the informativeness of market close prices. Passive investment strategies that trade at market close have incurred high transaction fees charged by the primary exchanges. Investment banks undercut the exchanges by executing client orders at close prices set on the exchanges yet charging lower fees. While providing liquidity, banks trade on the order flow information. Using a quasi-experimental shock – an NYSE close auction fee cut – we find that banks' trading activities improve the informativeness of close prices and reduce the cost of passive investment strategies. To explain this finding, we propose a model where dual trading improves price discovery. A bank contributes to price discovery by trading on the informativeness of the orders it receives relative to the market. The implications of our model apply generally to scenarios with multiple trading venues where venue operators trade on order flow data.

Thesis Supervisor: Hui Chen
Title: Nomura Professor of Finance, Professor of Finance

Thesis Supervisor: David Thesmar
Title: Franco Modigliani Professor of Financial Economics, Professor of Finance

Acknowledgments

I am deeply indebted to my advisors Hui Chen, David Thesmar and Emil Verner. Their brilliant guidance is the catalyst of this thesis. Hui is a patient listener to my early-stage research ideas and never fails to provide concrete and constructive feedback along a project's development. I am especially thankful for him being the source of support from the beginning of my PhD study. This thesis would not exist without his continued encouragement. David has been organizing a bi-weekly reading group on corporate finance for many years. I started to join that group in 2020, which for me was a precious venue of academic exchange when everyone worked remotely during the pandemic. Chapter 1 of this thesis was motivated by discussions in that reading group. David also expertly guided the direction of my research with his wisdom and masterful vision of the literature. Emil spent the time reading into my slides and drafts in several iterations and offered advice page by page. To my advisors, I am grateful for their generosity with their time. To me, they are exemplars of great scholars. They are open-minded, allowing me to explore new topics instead of categorizing myself into existing agendas. They also pushed my work in rigor, sharpness and clarity, and make me firmly believe that a good paper is one that is closer to the truth. I aspire to meet the high standard they set for all financial economists.

Beyond my thesis advisors, I am grateful to MIT Sloan's finance group faculty and the finance PhD program. Adrien Verdelhan's dedication to the PhD program is remarkable. I also appreciate his advice and encouragement during my job search. Among others, Eben Lazarus, Jonathan Parker and Antoinette Schoar in particular were extremely generous with their time. For helpful comments, I am also grateful to Dan Greenwald, Tong Liu, Debbie Lucas, Christopher Palmer, and Daniel Paravisini. My special thanks go to Haoxiang Zhu. My passion for finance research is partly due to its practical use and policy implications, and Haoxiang's work further inspired me on this angle. I am fortunate to have had the opportunity to learn from him by having him as a coauthor and a mentor. Outside of MIT, I would like to thank Carola

Frydman, Ravi Jagannathan and John Mondragon, who mentored me and saw me as a serious researcher before my PhD study.

I benefited tremendously from friends and colleagues. Yixin Chen, Fangzhou Lu, Yupeng Wang, and Jian Sun gave me lots of advice as friends, and made my PhD much more enjoyable. My peers Maya Bidanda, Joanne Im, Jonathan Jensen and J.R. Scott are great classmates and officemates. Yura Olshanskiy introduced me to empirical studies on market microstructure. I would also like to thank my coauthors and friends Jingxiong Hu at Kellogg, and Shumiao Ouyang at Princeton.

Finally, may this thesis be dedicated to my parents and my wife. My father Fusheng is the closest-to-me examplar of an assiduous scholar, and my mother Chunying gives me unconditional love and encourages me to pursue the better self. Thousands of miles away from Cambridge, my wife Vivian Su provided much understanding, support and love during the long period that was needed to bring this thesis to fruition.

Contents

List of Figures

List of Tables

Chapter 1

Getting the Banks on Board: Accounts Receivable Financing in the US

Firms typically sell goods to customers on credit rather than requiring immediate cash payment. As of 2021, accounts receivable is the second largest financial asset on the aggregate balance sheet of nonfinancial business in the United States, totaling $4.5 trillion (U.S. Flow of Funds Account 2021). Contributing to this volume, it is particularly concerning when large investment-grade buyers with easy access to capital markets borrow from their smaller weaker suppliers. And numerous policies have been adopted to force large buyers to pay faster, for example, in Chile (Breza and Liberman, 2017), China, Colombia, France (Barrot, 2016), and US (Barrot and Nanda, 2020).[1]

Should a government regulate payment delays? Existing studies show that shortening payment delays alleviates the financial constraints of sellers (Murfin and Njoroge,

[1]Colombia: SMEs need to be paid within 45 days starting from 2022. See The Law on Fair Payment, Law No. 2024 of 2020 at https://bu.com.co/en/insights/noticias/law-fair-payment-terms-issued. China: state-owned companies need to pay SMEs within 30 days starting from September 2020. See Order No. 728 of the State Council of the People's Republic of China issued in July 2020 at http://www.lawinfochina.com/display.aspx?lib=law&id=33251. US: Federal Quickpay reform of 2011 and Supplierpay reform in 2014. Chile and France: please see the cited papers.

2015; Barrot, 2016; Barrot and Nanda, 2020). However, payment delays may be optimally chosen by buyers to address the information asymmetry problems inherent in trade relationships. For example, previous work has considered seller moral hazard that affects product quality as a determinant of payment delays (Smith, 1987; Long, Malitz, and Ravid, 1993; Lee and Stowe, 1993; Kim and Shin, 2012). Meanwhile, market solutions for sellers to finance the accounts receivable do exist. The costliness of these financing options, and their role in addressing the information asymmetry problems, are all relevant for this policy discussion.

This paper studies the informational role of payment delays and the accounts receivable financing market. To be precise, I quantify the role of payment delays in addressing seller moral hazard and the extent to which lenders providing accounts receivable financing screen out bad-quality sellers. Both tasks are challenging as the latent actions that affect product quality are difficult to observe. My strategy is to use new data on the accounts receivable financing market, which actively prices the risks of sellers producing bad products. To implement the strategy, I also construct two other new datasets: accounts receivable due from individual customers and contractual payment delays in supply agreements.

I proceed in three steps. First, I hand collect contracts of accounts receivable based loans entered into by all US public companies from 2000 to 2020, and document important new facts based on this new dataset. I focus on two major forms of loans: accounts receivable based credit lines and factoring agreements (henceforth, A/R financing).[2] Second, I show reduced form evidence that seller moral hazard is a determinant of payment delays, and that lenders of A/R financing add value to the trade relationships by screening and monitoring the sellers. Third, I propose and structurally estimate a trade credit model that explicitly incorporates A/R financing, and use the estimated model to analyze the effects of counterfactual policies.

The following is a new set of facts based on my new dataset on A/R financing.

[2] Both use borrowers' accounts receivable as the borrowing base and the collateral. See Lemmon et al. (2014) for accounts receivable securitization, another important form of credit based on accounts receivable that large non-financial firms often use. Having cross-checked with their data, I find that borrowers in my sample, which are smaller firms, rarely use accounts receivable securitization.

First, A/R financing is widely used. During 2000–2020, 24.9% of US public firms with average assets between 250 million and 5 billion USD have borrowed from A/R financing. So did 41.2% of US public firms with negative median net trade credit days (payable days less receivable days) and at least one rated customers. In more than 60% of cases, committed credit by the lender exceeds 70% of the face value of the borrower's total accounts receivable. Second, A/R financing is an important form of credit for small, young firms that are short of other types of collateral. 77% of the borrowers are unrated and another 20% are rated at non-investment grade. The borrowers and non-borrowers of similar size seldomly use other debt instruments, and when they do, the interest rates on other debt instruments are higher than A/R financing. These facts are consistent with Luck and Santos (2022), who find that 40% of all credit lines of small firms are secured by accounts receivable and inventory, and accounts receivable is as valuable a form of collateral as real estate.[3] A/R financing is also an important funding source during the financial crises.[4]

Third, there is a 2-4% per year average spread between a seller's A/R financing interest rate and its customers' average unsecured borrowing rate. This comparison is enabled by my hand-collected data on the composition of accounts receivable due from individual customers.[5] Ex-ante, it is not obvious this spread should be positive, since the lender's recourse against the seller in A/R financing may hedge the customer's payment risk. Meanwhile, the lenders (henceforth, A/R lenders) do not lend against the full invoice value, that is, the full collateral value, and the median haircut is 20%. The spread and the haircut could be the grounds for regulating payment delays: by asking a large buyer to pay any seller quicker, the total financing cost of the seller and the buyer can be reduced, if everything else can be kept equal. But if this is true, it is puzzling why Coase theorem does not apply in the first place: by paying a seller

[3] Luck and Santos (2022) find that using accounts receivable as collateral can lower a firm's credit spread by 23 bps, while using real estate as collateral can lower the credit spread by 21 bps, after controlling for other loan characteristics.

[4] The outstanding credit and new originations of A/R financing loans did not dry up in financial crises. This is consistent with Chodorow-Reich et al. (2022) and Greenwald, Krainer, and Paul (2021), who show that credit lines provided by large banks help firms weather the crises, and is to some extent contrary to the view in Costello (2020).

[5] Freeman (2020) collects the same accounts receivable data to study determinants of trade credit.

quicker and asking for a lower product price, a large buyer can extract surplus from the lowered total financing cost. In resolving the puzzle, the literature has considered that payment delays can incentivize sellers to deliver high quality products. By withholding payment, buyers have extra time to verify moral hazard actions taken by the sellers that may be unobservable at delivery. Murfin and Njoroge (2013), Breza and Liberman (2017) and Gofman and Wu (2022) provide evidence consistent with this seller moral hazard channel.

My data on the composition of accounts receivable allows me to strengthen the evidence on the seller moral hazard channel. I show that a buyer pays its seller slower if the seller is from an industry with higher product quality risk. Following Murfin and Njoroge (2013), I use warranty claims filed against firms as a proxy of firms' product quality risk, since warranty claims represent the ex-post charges to firms caused by defective products, repairs and returns. The advantage of my analysis is that I control for buyer fixed effects and the sellers in my data reside in quite diverse industries so that product quality risk can be better measured.

As another piece of evidence, I find that the haircut of A/R financing contracts (adjusted by the interest rate) is close to be linearly increasing in payment delay. My strategy to uncover the *magnitude* of seller moral hazard utilizes the slope of this relationship. If payment delays are used to address seller moral hazard, then the longer is the delay, the more likely a buyer can discover bad quality products and issue more returns, which causes higher dilution to the invoice value. Anticipating this, A/R lenders should use higher haircut and interest rate to cover the dilution.[6] In other words, the slope of this relationship between haircut (adjusted by interest rate) and payment delay maps to the incentive role of payment delay: the speed at which a buyer detects and returns bad products after delivery. This strategy is appealing because returned goods are not directly observable from standard accounting data.[7]

[6] Gateway Commercial Finance, a factoring company, illustrates that "[when] defining your factoring advance rate [and interest rate], a factoring company will analyze your business dilution rate ... Dilution is the difference between the gross value of your invoices and the payment that is actually collected from your customers. The difference can be represented by a number of factors such as returned goods, bad debt write-offs, discounts offered, etc." See https://gatewaycfs.com/invoice-factoring/cash-advances/.

[7] Sales disclosed in a firm's balance sheets already excluded trade discounts and returned sales.

Besides informing us of seller moral hazard, I ask, does A/R financing add value directly to trade relationships? My paper highlights the screening and monitoring role of A/R lenders. It is not new that lenders produce information and monitor the borrowers (Diamond, 1984; Chava and Roberts, 2008; Sufi, 2009).[8] A/R financing is unique in that a lender's loan payoff is closely related to product quality hence the lender has skin-in-the-game partially aligned with buyers. Also, payment delays naturally get these lenders "on board" the trade relationship, and their screening and monitoring role is important for our understanding of the benefit and cost of payment delays. And my hypothesis is that lenders help a buyer work with the good sellers on top of the incentive mechanisms the buyer has imposed.

To provide evidence of the screening and monitoring role of A/R lenders, I examine the impact of obtaining A/R financing on sellers. I find that a seller's receivable days decline by 5% (3 days) after it first borrows from A/R financing. The effect persists for more than 5 years. It is estimated from an event study analysis, accounting for a rich set of observables and fixed effects, and has no preexisting trends. I interpret this result as: getting a loan from lenders improves the likelihood that the seller can deliver good products, after which the buyer only needs to use weaker incentives and becomes more eager to satisfy the seller's liquidity need, hence pays quicker. Supporting this view, the decline in receivable days is smaller for sellers that provide more homogeneous goods like gas, oil, and mines, for which the quality of goods can be easily verified. The decline is also smaller for older first-time borrowers, and is larger when lenders are more experienced. Also supporting this view, sales and SG&A expenses of borrowers increase by 5.7% and 4.1% after the loan is granted, and compared to a matched sample of sellers that are non-borrowers, borrowers have 0.68 years (29%) longer relationships with customers. Alternative explanations are difficult to explain these results. First, the decline in receivable days is not due to the pledged accounts receivable being moved off balance sheet.[9] Second, it is unlikely

[8]For monitoring, A/R lenders use covenant-based monitoring extensively like other credit lines lenders. They also conduct site visits and appraise the value of accounts receivable very frequently, usually monthly or even weekly (Office of the Comptroller of the Currency, 2000). See Gustafson, Ivanov, and Meisenzahl (2021) for evidence of lenders' active monitoring of borrowers.

[9]Borrowers can move accounts receivable off balance sheet only with non-recourse factoring, which

that borrowers in need of liquidity simultaneously make buyers pay quicker. Third, the results are not due to changes in the composition of a seller's customer base.

My empirical evidence also rejects two counterarguments for A/R financing. One counterargument is that by making sure sellers receive cash upfront, A/R financing unwinds the incentive effects of payment delays in addressing seller moral hazard. To restore the incentives, buyers may need to delay payments even longer, otherwise they have to work with other sellers that are ex-ante more transparent. Another counterargument suggests that after sellers secure A/R financing, buyers have more "excuses" to squeeze the sellers and further delay payments. Both counterarguments point out the potential harm of A/R financing to trade relationships, but are at odds with my evidence.

To answer the policy question of whether a government should regulate payment delays, I build the first trade credit model that explicitly considers A/R financing. In the model, A/R financing terms are related to seller moral hazard and payment delay. And the optimal payment delay is again determined by seller moral hazard and the A/R financing terms. What drives my model is that longer payment delay makes moral hazard actions costlier for the seller,[10] and that the lender has independent information on the quality of the seller. Seller moral hazard can be interpreted as poor quality of management, financial misconduct, under-investment in equipment and materials, and substitution of effort between multiple product lines, etc., which I lump together under the name "misbehavior". These misbehaviors are hard to be fully known by any single buyer. A key trade-off determines the optimal payment delay: longer payment delay induces less misbehavior but increases the seller's financing cost and consequently the total production cost.

The model generates predictions that find support in the data. It predicts that more profitable buyers, including those with higher product market power, use longer payment delays, because they lose more when the seller's product is inferior. This is related to the notion that high bargaining power buyers pay slower (Fabbri and

I have excluded from the analysis. Other types of A/R financing have recourse against borrowers, and are treated as loans for accounting purpose.

[10]The micro-foundation of this assumption can be found in Kim and Shin (2012).

Klapper, 2008; Klapper, Laeven, and Rajan, 2011). It also predicts that industries producing more opaque products use more trade credit because moral hazard is more severe, as documented by Giannetti, Burkart, and Ellingsen (2011) and this paper. Lastly, it predicts that if a seller receives A/R financing, its receivable days will decline and its relationship length with buyers is longer as documented in this paper, because the lender's information helps buyers work with better sellers.

I move on to structurally estimate the parameters of this model. To make sure the model applies to payment delays of US public firms, I hand collect the contractual payment terms in all bilateral supply agreements filed in SEC filings from 2000 to 2020.[11] The main identification challenges in estimating the model are that a seller's borrowing decision is endogenous, and that the loan terms are affected by unmodeled variables that may correlate with payment delays.[12] I treat the selection bias caused by endogenous borrowing decision as an omitted variable according to Heckman (1979), that may correlate with payment delays. This allows me to tackle both identification challenges by using instrumental variables for payment delays. As the first instrumental variable, I use liquidity shocks to buyers, measured as the fraction of debt that needs to be refinanced in the subsequent year, following Almeida et al. (2011). The second instrumental variable I use is lagged industry-level payment delays. The two instrumental variables affect bilateral payment delays but are largely uncorrelated with omitted variables that affect loan terms.

Using the estimated model, I conduct two counterfactual analyses. First, I consider forcing a universal reduction in payment delays to 30 days. Sellers on average have lower default probability and lower financing cost. However, the presence of bad products increases by 12.6% and total production declines by 26.3%. Second, I consider subsidizing A/R lenders. It can increase a firm's annual production quantity by \$2.6 million at the government's (or outside investor's) expense of \$1.57 million. What makes the government subsidy a positive NPV project is that it enables the

[11] Costello (2013, 2019) collect the same data, for a shorter time period, and study different economic questions.

[12] Borrowing decisions may be related to growth opportunities and the continuation value of selling to customers. The omitted variables may include, for example, macro-economic shocks.

buyers to use longer payment delays to prevent bad quality products, which in turn boosts the buyers' production. This is a relevant exercise, as in many countries, accounts receivable exceeds 20% of corporate assets yet the A/R financing market is not as developed as in the US.

Related Literature This paper contributes to the existing literature in at least four ways. To start, it offers the first large-scale empirical study on accounts receivable management and financing. As far as I know, the latest empirical study on this topic for US individual firms is Mian and Smith (1992), who use a 1982 survey of 600 firms. Klapper (2006) and Bakker, Udell, and Klapper (2004) use country-level aggregate data to study factoring. There is some international evidence using data from particular factoring programs, such as in Tunca and Zhu (2018). Most recently, Greenwald, Krainer, and Paul (2021), Chodorow-Reich et al. (2022), and Luck and Santos (2022) use proprietary Fed Y14 data to study the credit supply and pricing of a rich set of loan contracts. Although their data covers accounts receivable based credit lines, it only covers loans extended by large banks and does not contain all the loan terms details needed in my study. They also study very different economic questions.

My paper offers the first trade credit model structurally estimated by the related data. It uncovers seller moral hazard and highlights the role of the A/R financing market. In the trade credit literature, besides the moral hazard theory, trade credit from a small seller to a large buyer can be explained by the seller's incentive to price discriminate buyers while circumventing antitrust regulations (Giannetti, Serrano-Velarde, and Tarantino, 2021); and by savings in tax expenses (Brick and Fung, 1984; Desai, Foley, and Hines Jr, 2016). There is also a large literature that examines trade credit provision best seen as from large sellers to small buyers, and the central puzzle it answers is why small buyers are not financed directly by banks: see among others, Petersen and Rajan (1997); Biais and Gollier (1997); Frank and Maksimovic (1998); Wilner (2000); Burkart and Ellingsen (2004); Cunat (2007).[13] I am unable to

[13]While all these theories have merit and can be *partially* applied to the small seller – large buyer setting, some of the critiques of Burkart and Ellingsen (2004) apply: "Price discrimination

account for all the theories and all the variations of the trade credit data. Rather, I focus on the informational role of trade credit that is crucial for the small seller – large buyer setting, and I argue that the existence of a large A/R financing market further rationalizes trade credit usage. As one empirical contribution, I combine pieces of publicly available data on trade credit and A/R financing for all US public firms. Given the difficulty in obtaining seller-buyer level data, recent studies on trade credit are usually based on foreign data (Fabbri and Klapper, 2008; Giannetti, Serrano-Velarde, and Tarantino, 2021) and proprietary data from platform companies that may be restricted in certain firm characteristics.[14] I also complement recent studies on how a seller's access to debt affects trade credit.[15]

My estimated structural model can inform policies. Empirical studies have found rich consequences when sellers are paid quicker. Murfin and Njoroge (2015), Barrot (2016), Beaumont and Lenoir (2019) and Barrot and Nanda (2020) find that quicker payments alleviate the financial constraints of small sellers, while Breza and Liberman (2017) find that payment delays are optimally chosen by firms, and when the government forces quicker payments, buyers reduce the purchases from small sellers and conduct vertical integrations. My paper reconciles these findings and can evaluate the aggregate consequences of counterfactual policies. When my model is estimated for individual industries, it could help policymakers consider industry-specific regulations.

Finally, this paper further the inquiry into the financial frictions in production networks. Liu (2019) rationalizes government subsidy to upstream sectors by market imperfections including financial frictions. Bigio and La'O (2020) cite financial frictions

theories cannot account for trade credit in competitive markets; the collateral liquidation theory cannot account for trade credit in service industries; product quality theories cannot account for trade credit in homogeneous goods industries...".

[14]For example, Klapper, Laeven, and Rajan (2011) use data from PrimeRevenue with 56 buyers, and Costello (2020) uses data from Credit2B that lacked detailed information on buyers. Retail industries appear to be more represented in datasets from these platform companies.

[15]Giannetti, Serrano-Velarde, and Tarantino (2021) use the reforms of anti-recharacterization laws in Italy, which increased banks' willingness to accept a firm's accounts receivable as collateral, and show that sellers grant more trade credit after the reforms. Billett, Freeman, and Gao (2021) find the opposite in US: higher debt capacity (also due to anti-recharacterization law change) of the sellers decrease trade credit. Neither paper looks at A/R financing loans, that prove to be crucial for small sellers burdened by accounts receivable.

as a contributor to sectoral distortions. Measures of financial frictions in production networks, however, need to account for the informational role of financial intermediaries. While a few recent papers show clear evidence that monetary and liquidity shocks pass through the production networks (Giannetti and Saidi, 2019; Costello, 2020; Alfaro, García-Santana, and Moral-Benito, 2021), it remains unknown to what extent the pass-through is attenuated by deep pocketed A/R lenders (Kiyotaki and Moore, 1997) .

The rest of the paper proceeds from empirics to the theory. The next section introduces the institutional details on A/R financing and the data used in this paper. In Section 1.2, I document facts related to the prices, risks and wide use of A/R financing. In Section 1.3, I show evidence that seller moral hazard is one of the determinants of payment delays. In Section 1.4, I present evidence that A/R lenders improve the quality of the sellers. The empirical sections motivate my model in Section 1.5, which details the method to uncover seller moral hazard using A/R financing data. Estimation of the model and counterfactual analyses are discussed in Section 1.6. Section 1.7 concludes.

1.1 Institutional Details and Data

1.1.1 Institutional details

This paper focuses on accounts receivable based credit lines and factoring agreements (A/R financing), the two major forms of credit backed by accounts receivable. Figure 1-1 below shows the timeline of how A/R financing works with a real example in my data. The seller first makes the sales and delivers products to the buyer. Usually upon or within a few days after the delivery, an invoice is generated. The accounts receivable witnessed by the invoice can then be pledged to a lender as collateral. The lender calculates a borrowing base that specifies the maximum amount that the seller can borrow, and the borrowing base is a fraction of the face value of accounts receivable. When the seller receives the buyer's payment after some delay, it pays the lender the principal and interest of the borrowed funds, and keeps whatever is remaining. In some cases, the buyer's payment goes directly to a lockbox account set up by the lender rather than going to the seller. The lender takes the principal and interest and remits the remaining funds to the seller. These loan contracts are all in the form of credit lines. They usually last for 2-3 years (Office of the Comptroller of the Currency, 2000). In our data, the most common contracts have 1 year duration with 1 year automatic renewal. Many new contracts are simply extensions of old contracts, with some minor changes in loan terms. Borrowers typically use A/R financing repeatedly for many years until they find cheaper financing options. Some borrowers use A/R financing for more than 15 years.

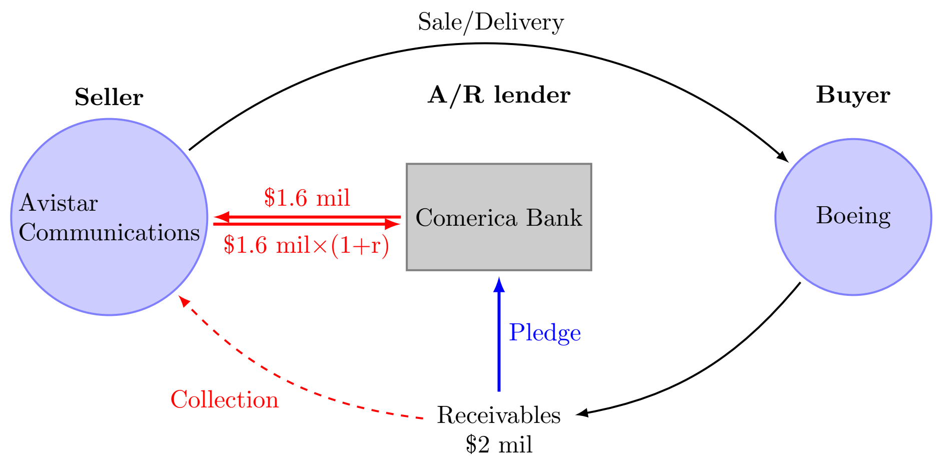


Figure 1-1: Example of A/R financing

What differentiates accounts receivable based credit lines and factoring agreements is the recourse status. Traditionally, factoring agreements are one-point sale of some specific invoices to lenders. My data suggests that factoring agreements have modernized to take the form of credit lines. That is, the seller can borrow and repay as time goes, as long as the outstanding amount does not exceed credit limit, and factoring lenders also take in all eligible invoices as collateral. By recourse status, there are two types of factoring agreements: recourse factoring and non-recourse factoring. Having a recourse means lenders can ask the seller to pay for the debt if the buyer's payment is not enough to cover the principal and interest. The majority of factoring agreements in my data have recourse, and non-recourse factoring appears to be used only by firms in trouble (close to bankruptcy) or firms that have difficulty in collecting customer payment, or firms that are very small and young.[16] As a result, non-recourse factoring is much more expensive, and the median interest rate is about 1.5% per month. All of the accounts receivable based credit lines have recourse against the seller, just as regular collateralized loans. This means recourse factoring agreements are no different from accounts receivable based credit lines. Instead, they just reflect different name conventions.[17] Non-recourse factoring constitutes a sale of accounts receivable, and allows firms to move accounts receivable off balance sheet. My data shows that very few lenders offer non-recourse factoring, among which CIT Group is the single major lender.[18] Given that non-recourse factoring is rare and quite special, I leave it out from this paper's main analysis.

The most important terms of A/R financing contracts are the advance rate, credit limit and interest rate. The loan in Figure 1-1 is made in 2002 between Avistar Communications and Comerica Bank. Its advance rate is 80% (haircut = 20%), credit limit is $4.5 million, and interest rate is prime rate + 0.25% per year. Advance rate specifies the borrowing base, which means that the lender is willing to advance funds to the seller up to 80% of the face value of the eligible accounts receivable. The median and average advance rate in my data is 80%, and can vary from 60% to 90%. The face value of eligible accounts receivable is reappraised weekly or monthly, and certified

[16] For example, a borrower says that "factoring was necessitated by the cash flow problems created by our customer's inabililty to make payments on his work with us in the typical time frame we contracturally provide for payments." See https://www.sec.gov/Archives/edgar/data/0000854171/000100009603000185/renegade123102.txt. When a firm grows, it usually switches from factoring to accounts receivable based credit lines. For example, Diversified Corporate Resources, Inc. (OTC Bulletin Board: HIRD) started to borrow from accounts receivable based credit lines in 2006 to move away from the higher cost of factoring receivables. See https://www.sec.gov/Archives/edgar/data/0000779226/000110465906023126/a06-8626_1ex99d1.htm.

[17] Other aspects like covenants, loan terms, etc. are also similar between them. Firms explicitly describe in their filings that the accounts receivable based credit lines they enter are effectively factoring contracts (with recourse). For example, "The Company plans to file an application to enter into a revolving line of credit ("Credit Agreement") with an independent bank.... Under the terms of the proposed Credit Agreement, which will be a factoring arrangement...". See https://www.sec.gov/Archives/edgar/data/0001112987/000111650201501472/intercallnet-10qsb.txt.

[18] This is also evidenced by the practitioners: "these days, very few factors offer accounts receivable financing solutions that are truly on a non-recourse, notification basis ... [So called "non-recourse" factoring agreements] products may, on their surface, state that they are purchase agreements with a seller and buyer of receivables. On closer inspection, however, such agreements typically contain terms that are viewed by courts as incompatible with a true sale." See https://crisismanagementupdate.com/the-risks-of-factoring-agreements. CIT Group, Inc. went bankrupt in 2009.

by borrowing base certifications(Office of the Comptroller of the Currency, 2000). The certification could be issued by the lender or sometimes an auditing company the lender hires to conduct field inspections (Gustafson, Ivanov, and Meisenzahl, 2021). The lender may also require month-end accounts receivable aging reports for the amounts on the certificate. As in other credit lines, sellers can borrow and repay as long as the outstanding debt does not exceed the credit limit ($4.5 million in the example). Interest rate is almost always set equal to a floating benchmark rate plus a margin, where the benchmark is usually the lender's prime rate or one-month LIBOR.

What determine these loan terms? Both buyer default and return of goods can reduce the payment eventually collected from the buyer. In other words, they reduce the collateral value of accounts receivable. Also, since these loans have recourse against sellers, seller default risk matters. Also relevant are a few other factors, for example, the level of concentration of the customer base (due to diversification benefit), and the amount of credit being used (due to fixed cost of setting up the loans) (Office of the Comptroller of the Currency, 2000).

The determinant of loan terms that this paper focuses on is the return of goods.[19] This arises from disputes and returns of below quality products, or untimely delivery in cases where invoice is generated before delivery. This is indeed an important concern when lenders set loan terms. For example, Gateway Commercial Finance says that "[industries] that supply products generally receive lower advance [rate] than industries that simply provide service. There are fewer opportunities for a dispute to arise in case of provision of services, for example once delivered to a client a service cannot be returned or found to be defective."

The lender takes in all eligible accounts receivable of the borrower from all customers as collateral. The reason to take in all accounts receivable is to prevent the borrower from cherry-picking the riskier accounts receivable to pledge. The lender will register perfected first-priority UCC-1 (uniform commercial code) liens on all the

[19] I focus on sellers that work with large creditworthy buyers that observationally never default to exclude the buyer default risk. Later, I discuss how seller default risk and other factors like order discounts are isolated when estimating seller moral hazard from loan terms. The idea is that these factors do not interfere with the sensitivity of loan terms to payment delays.

accounts receivable.[20] When the seller defaults, the pledged accounts receivable is regarded as "cash collateral". That means it is very difficult for the seller to divert the accounts receivable in bankruptcy. Chapter 11 Bankruptcy Code does not permit the use of cash collateral by the borrower unless the secured lender consents or the court authorizes the use of it after notice and a hearing. The court may authorize such use only if the secured party is adequately protected against loss (Weintraub and Resnick, 1982).

Sometimes the borrowers also post additional collateral, like inventory, equipment, real estate and other assets, in securing the A/R financing loan. In my data, about 40% of the contracts mention the inclusion of inventory as collateral, but there are usually separate and much lower credit limit on the lending backed by inventory compared to that backed by accounts receivable. The advance rate on inventory is also much lower.[21] Office of the Comptroller of the Currency (2000) suggests that "lenders who finance distressed companies or acquisitions are more likely to include the less liquid collateral when calculating the borrowing base. When they do, the agreement generally limits such collateral to a small percentage of the borrowing base." And OCC instructs the lenders that "as much as possible, an ARIF [A/R financing] loan should be supported by working assets".

What types of accounts receivable are eligible? The definition of eligible receivables is generally specified in the loan contract and the most common requirement is on the invoice's age. For example, most contracts specify that the eligible ones are invoices issued within 90 days ago or 120 days ago.[22] Besides, receivables commonly designated as ineligible are those that are unbilled, delinquent for a period or owed by an insolvent borrower, as well as receivables that exceed concentration limits, or due

[20] A uniform commercial code (UCC) filing is a notice registered by a lender when a loan is taken out against a single asset or a group of assets. A UCC filing creates a lien against the collateral a borrower pledges for a business loan.

[21] For a small random sample of A/R financing contracts, I collected the advance rate against inventory and find the median is around 40%.

[22] For example, the loan agreement of Monster Worldwide says: "Borrowings under the Agreement are based on 90% of eligible accounts receivable, which are amounts billed under 120 days old and amounts to be billed as defined in the Agreement." See Form 8-K filed on October 9, 2003 by Monster Worldwide, Inc at https://www.sec.gov/Archives/edgar/data/0001020416/000104746903033033/a2119977z8-k.htm.

from foreign customers or government. The amount of eligible accounts receivable will be stated on the borrowing certificate, and is one way that the lender insulates itself against buyer default risk.

Finally, the A/R financing market appears to be relatively competitive, with 188 banks and 440 non-banks lending to US public firms in the past 20 years. The majority of A/R financing loans have single lenders. That means the screening and monitoring motive of lenders is stronger as these single lenders have more skin-in-the-game. Young small firms without strong cash flows use A/R financing to finance themselves, by borrowing against accounts receivable from creditworthy customers. In my data, 27% of the borrowers have negative EBITDA when they receive A/R financing funds, and for 16% of borrowers the debt/EBITDA ratio is higher than 6. Without having the accounts receivable as a liquid primary source of payment and collateral, these loans would be considered very risky and incur scrutiny from bank regulators.[23]

1.1.2 A/R financing data

Source and coverage I collect data on 5569 A/R financing contracts, namely, accounts receivable based credit lines and factoring agreements (recourse and non-recourse), entered into by 2965 non-financial non-utility Compustat firms.[24] These contracts include loan originations and subsequent amendments that introduce changes to at least one of the three key loan terms: advance rate, interest rate, and credit limit. The data source is SEC filings of all US firms. Regulation S-K requires firms to file material contracts, including loan and credit agreements, as exhibits to the SEC filings. Loan agreements are not necessarily filed as exhibits, but in most cases are simply described in main text of the filings. I search for keywords "eligible accounts receivable", "eligible receivable", and keywords such as "factoring agreement/arrangement" from all 8-K, 10-K, 10-Q, and S-1 filings over the past 20 years. A/R financing is

[23] Interagency Guidance on Leveraged Lending and OCC suggest that borrower's leverage in excess of 6 times of the Total Debt/EBITDA ratio raises concerns for most industries. See Chernenko, Erel, and Prilmeier (2022) for details.

[24] I exclude firms in the financial industry (SIC from 6000 to 6999) and the utility industry (SIC from 4900 to 4999).

identified by reading through the search results and looking for the most significant feature: the loan's borrowing base is pinned by accounts receivable. I trace the loan usage as much as possible for individual firms that ever borrowed from A/R financing. More details on the data collection can be found in the Appendix A.1.

DealScan also covers a small subset of the A/R financing loans. One drawback of DealScan data is the lack of coverage on the advance rate.[25] I cross check the DealScan data with the hand collected data. To match DealScan data to Compustat, I use Chava and Roberts (2008)'s "Refinitiv LPC DealScan–S&P Global Market Intelligence Compustat match linktable", which forms the basis of most syndicated loan research. The construction of this link file has been an ongoing process since 2002, and the latest version of it contains matches through the end of 2017. The matches are also very similar with the more recent linktable generated by advanced algorithms (Cohen et al., 2021) . For A/R financing loans in DealScan that cannot be matched to Compustat using the linktable, I performed a hand match.

Table 1.1 reports that only 18% of the A/R financing loans are covered by DealScan. This is consistent with Chernenko, Erel, and Prilmeier (2022) who find that, comparing with Capital IQ loan data (also based on SEC filings) on 750 middle-market public firms, DealScan covers only about half of the bank loans and almost none of the direct non-bank loans. Non-banks account for 47% of contracts and 43% of total committed credit in the A/R financing market, and for 21% of the contracts borrowers did not disclose the lenders' names. Besides, non-recourse factoring agreements account for only 2% of all the A/R financing contracts.

Table 1.1 also shows that smaller firms are more likely to use A/R financing according to my data, in terms of both the number of borrowers and the number of contracts. 76.8% of the borrowers are unrated, while 20.4% of the borrowers are rated below investment grade. The borrowers' customers are generally large firms. For 47% of the borrowers, their largest customers are rated above investment grade. For 84.2% of the borrowers, their firm sizes are smaller than 5% of their own customers' average

[25] DealScan provides a borrowing base dataset, but A/R financing loans' advance rates are rarely available in that dataset.

size.

Panel A of Figure 1-3 shows the industry breakdown of the borrowers. 51.8% of the borrowers are in manufacturing industries and 23.5% are in services industries. Panel B shows the trend of committed credit in A/R financing over the past 20 years. The new originations and outstanding credit did not dry up during the financial crisis, showing that A/R financing is a relatively reliable funding source for firms. Panel A of Table 1.2 shows summary statistics for A/R financing loan terms. The mean and median advance rate are 80%. The mean and median interest rate are 5.9% and 5.5% per year. For more than 50% of firms, credit limit covers more than 88% of total accounts receivable, while the average coverage ratio as a fraction of total accounts receivable is 1.86.

Since my data is based on firms' SEC disclosures, the data coverage is better for smaller borrowers, and for the loan contracts entered into by firms when they are young. As a borrower grows larger, it may gradually treat the A/R financing loans as immaterial thus does not disclose them. Another reason why we miss some later-stage loan contracts is that firms may simply refer to a similar loan earlier used, yet does not mention the keywords I use in searching through the filings. Such selection bias in the data is however less concerning for my analysis, because my analysis focuses on smaller borrowers and the first times they borrow.

Comparison with the Y14 data The proprietary Fed Y-14 data used in Chodorow-Reich et al. (2022); Luck and Santos (2022) and Greenwald, Krainer, and Paul (2021) also covers A/R financing loans. The data comes from regulatory report of bank holding companies with more than $100 billion in total consolidated assets. It contains loan-level information, and separates loans by collateral types including: real estate, fixed assets, accounts receivable & inventory, cash, other specified assets, blanket lien, and unsecured.

Compared to the Y14 data, my hand collected data has several advantages. First, my data covers loans from 188 banks and 440 non-banks, while the Y14 data only 33 large banks. This means loans to many small borrowers are missing in the Y14 data.

Also, the Y14 data only covers loans with a commitment above \$1 million, while in my data 10% of the A/R financing loans have less than \$1 million commitment. Second, the majority of banks start to report data after 2012Q3 in the Y14 data, while my data goes back reliably to 2000 (and earlier). This helps me identify the screening and monitoring role of A/R lenders that may be more salient for young small firms. Third, I have data on more loan terms details such as advance rate, that help me uncover seller moral hazard.

Table A.2 compares my data and the Y14 data. It compares the aggregate outstanding committed credit of A/R financing as of December 2019 in my data and that according to the Y14 data.[26] For smaller firms, committed credit in my data is equal to 40% in Y14, and for large firms the fraction is smaller. This comparison appears to show my data covers a small fraction of all the A/R financing loans. However, I interpret this comparison with much caution. My data only considers non-financial and non-utility firms, while the Y14 data includes the other firms. My data also adopts a stricter definition of A/R financing – accounts receivable has to be the determinant of borrowing base rather than just a part of the collateral. Besides, my data is based on disclosure, so covers loan originations better than the outstanding loans at particular times.

1.1.3 Sales relationship and trade credit data

Sales relationship According to the Statement of Financial Accounting Standard (SFAS) No.14, a firm is required by SEC to disclose major customers that occupy more than 10% of its sales in the 10-K filings. The disclosed relationship and the cumulative amount of sales to individual customers at fiscal year ends are collected by the Compustat historical customer segment file, which I use for sales relationship between firms. This file lacks standardized names and identifiers of the disclosed customers, and many authors exerted separate efforts to manually create identifiers of the customers. I use a linktable that recently became available, which contains the customers' firm identifiers (i.e., Compustat gvkeys). The linktable is constructed

[26] I do not have access to the Y14 data. This comparison is based on summary statistics of the Y14 data in Chodorow-Reich et al. (2022).

by WRDS, by matching customers' names with historical CRSP and Compustat companies through a fuzzy name-matching algorithm and is verified manually and periodically. The matches are further calibrated and complemented by publicly available data and data contributed by researchers (Cen et al., 2017; Cohen and Frazzini, 2008).

Trade credit Following Freeman (2020), I collect data on a firm's year-end balance of accounts receivables due from individual customers from SEC 10-K filings. FASB No.105 effective in 1990 requires firms to disclose information regarding concentrations of credit risk from individual or aggregate counterparties. This leads many firms to report accounts receivable balances due from major customers. My data collection is restricted to sellers in the Compustat historical customer segment file. For these sellers, we know who are their customers from the Compustat file. Also, I require both sellers and buyers to be Compustat non-financial non-utility firms. I read through all the 10-Ks during 2000–2020 of the qualified sellers to collect this accounts receivable data. More details on this data collection process are described in the Appendix A.1.

This results in a seller-buyer pair by year panel dataset on the trade credit used between each pair of seller and buyer. Panel B of Table 1.2 shows summary statistics of this data. I am able to have 5537 observations for 758 sellers and 672 buyers during 2000–2020. This is comparable to the sample size obtained by Freeman (2020). About 70% of the sellers in the sample are in manufacturing industries, while 20% are in services industries. About 40% of the buyers are in manufacturing industries, and 20% are in retail industries. The average receivable days from individual buyers, defined as accounts receivable from a buyer/sales to a buyer × 360 days, equals 68.6 days.

Payment terms data To obtain contractual payment delays in calendar days, I collect the payment terms, i.e., trade credit terms, between seller and buyer firms from 2137 bilateral supply agreements. The data collection approach follows and extends that in Costello (2013, 2019). Among all SEC filings (10-K, 10-Q, 8-K and

S-1) between 2000 and 2020, I search for exhibits with "supply" or "procurement" in the title and "buyer" and "supplier" or "seller" in the first few paragraphs to locate the supply agreements. Then I read through the verified supply agreements to find and collect the payment terms. More details on the data collection can be found in the Appendix A.1.

Table A.3 in the Appendix tabulates the payment terms commonly used in each industry for the past 20 years, in the same spirit of Ng, Smith, and Smith (1999) who use survey data on 950 sellers. The two-part terms like "2/10 net 30" that give early discount is very rare. Rather, late payment interest is sometimes used and the common late interest terms are prime rate + a margin (usually 2–3%) , and 1.5–2% per month. "Net 30" is the most common terms, which means full payment is due 30 days after the invoice date; after that date the buyer is in default. There are considerable variations across industries. Panel B of Table 1.2 reports that the average contractual payment delay is 32.5 days.

1.1.4 Additional data

Other debt instruments To investigate what other debt instruments A/R financing borrowers and similar-sized non-borrowers use, I collect data on these other debt instruments from Capital IQ, following Chernenko, Erel, and Prilmeier (2022). This appears to be the most comprehensive data on debt instruments. According to Chernenko, Erel, and Prilmeier (2022), DealScan is missing half of the bank loans and all of the non-bank loans for medium-market Compustat firms. The other debt instruments include term loans, debentures, capital leases, and so on. The Debt Structure page of Capital IQ gives the name and the type of debt instruments a firm uses. Capital IQ collects this data from debt related agreements in firm's SEC filings and other sources. Data on the loan terms details and dates of borrowing are only available for a subset of the debt usages.

Warranty claims Following Murfin and Njoroge (2013), I use the warranty claims filed against firms as a proxy for product quality risk. Manufacturer warranty law, governed by the UCC, the Magnuson-Moss Warranty Act, and the FTC rules, disciplines a seller's express or implied guarantees of a product's quality and reliability. If the product malfunctions or does not work as promised, the consumer or purchaser may, under the warranty, be able to return the item for a refund, receive a replacement, or have the item repaired for free. Written warranty is not required by law in the US,[27] and firms may follow industry conventions and local regulations to provide written warranty.

If a firm provides warranty, accounting rules require it to disclose the amount of warranty claims in 10-K and 10-Q filings. FASB Interpretation No.45 in 2003 requires all public companies to publish details about all their guarantees including warranties. ASC 460 in 2009 further requires a warranty provider to explain the methodology that led them to determine the amount of warranty liabilities they expect.[28] WarrantyWeek newsletter collects data on warranty claims from individual firm's filings from 2003 to 2021, and is the source of my data. I have data on a firm's annual expense on warranty claims for a total of 1038 firms from 2003 to 2021. I compute the warranty claims as a percentage of past sales, and aggregate to 4-digit SIC code level to form an industry-level measure of product quality risk. Panel D of Table 1.2 shows summary statistics of product quality risk measure. On average, a firm's expense on warranty claims is about 0.7% of its sales.

Product market price elasticity of demand To get a measure of buyer bargaining power, and to characterize the production functions, I use a measure of the product market price elasticity of demand. To do so, I follow Alfaro et al. (2019) and assume buyers have CES production functions. Then price elasticity of demand is measured by the elasticity of substitution of products, which is estimated from US import data. Following Alfaro et al. (2019), I estimate the elasticity at HS10 product code level using the method of Feenstra (1994) and Broda and Weinstein (2006) and the US import data from Peter Schott's website (Schott, 2008), then aggregate to the 4-digit SIC code industry level. As a robustness check, I also estimate the elasticity,

[27]See FTC's Businessperson's Guide to Federal Warranty Law at https://www.ftc.gov/business-guidance/resources/businesspersons-guide-federal-warranty-law

[28]For details, see https://www.warrantyweek.com/archive/ww20181220.html.

using the improved hybrid method of Soderbery (2015). Demand elasticity is only available for manufacturing industries given the scope of US import data. Panel C of Table 1.2 shows summary statistics of the demand elasticities.

1.2 Prices, Risks and the Wide Use of A/R Financing

1.2.1 How expensive is A/R financing?

To answer the question in this title, I compare a firm's A/R financing interest rate with the average unsecured borrowing rate of its buyers. To measure a firm's buyers' borrowing rates, I first use the interest rates of unsecured credit lines in DealScan, and then supplement the interest rates with the commercial paper rates published by Fed when buyers have short-term credit ratings. I include the fees to both the A/R financing rate and credit lines interest rate, and the fees are comparable between the two.

Then I weight the unsecured borrowing rates of the buyers by the accounts receivable balance due from these buyers, and calculate the average. Since accounts receivable balance due from individual buyers are not always available, I use two approaches. The first one is to directly weight the borrowing rates by sales to individual buyers, since sales is the most important predictor of accounts receivable. The second one is to impute accounts receivable due from individual buyers based on sales to these buyers and a rich set of observables, and then use the imputed accounts receivable balance to weight the buyers' borrowing rates. Table A.1 reports the imputation model I use.

Figure 1-4 shows the result of this comparison when the buyers' borrowing rates are weighted by sales. Panel A of Figure A-1 shows the result of the comparison when buyers' borrowing rates are weighted by imputed accounts receivable balance. We see that the results under the two approaches are quantitatively similar. Both figures show there is a 2-4% per year average spread between a seller's A/R financing interest rate and its customers' average unsecured borrowing rate. Without seller moral hazard that dilutes invoice value through returned goods, this spread would be striking: since the risk of an A/R financing loan would have been smaller than a buyer's unsecured loan given that the recourse against seller that provides insurance for the lender. This spread also means the 2-4% per year expense paid to the financial system could have been saved if a buyer pays sellers quicker.

For robustness, Figure A-1(b) in the appendix shows the spread between a seller's A/R financing interest rate and its customers' average secured borrowing rate. To do this comparison, one needs some caution as most of the large firms' credit lines are unsecured,[29] and secured loans are often used when borrowers are riskier.[30] This selection bias distorts the interest rate of secured borrowings upward. I use the estimates from Luck and Santos (2022) to address the selection bias. Depending on the specifications, the spread between A/R financing interest rate and its customers' average secured borrowing rate is about 3-4% per year. Finally, this interest rate spread is increasing in seller's default risk. Panel B of Figure 1-4 shows that when the seller's credit rating is lower, the spread is higher.

1.2.2 The wide use of A/R financing and the determinants of use

My data suggests that A/R financing is widely used. Table 1.3A splits the Compustat non-financial non-utility firms into deciles by size. To measure trade credit usage, I use net trade credit borrowing days (payable days less receivable days). We see that firms in the smallest size decile are the largest trade credit borrowers, with the median firm paying its suppliers 97.2 days later than it is paid. Firms in the largest decile are also net trade credit borrowers, with the median firm paying its supplier 7.5 days later than it is paid. Firms who are burdened by accounts receivable are

[29] Luck and Santos (2022) find that 73% of the large firms' credit lines are unsecured. Chodorow-Reich et al. (2022) find that up to 70% of credit lines to firms with more than $5 billion in assets are unsecured.

[30] This selection bias is well documented. For example, Luck and Santos (2022) illustrate that "the result from our pooled model is in line with evidence from earlier studies that riskier borrowers are more likely to pledge collateral".

those in the middle of the size distribution. The median firms in the middle deciles are paid 3.2 to 6.7 days later than they pay their suppliers. They are also the most likely to borrow from A/R financing. Around 29–32% of these firms have borrowed from A/R financing at least once. The wide use of A/R financing is consistent with Luck and Santos (2022), who find that in the Y14 data, 47% and 22% of credit lines for private and public firms are secured by accounts receivable and inventory.

A seller is more likely to use A/R financing when its customers are creditworthy, since accounts receivable from creditworthy customers is a better collateral. In Table 1.3B, I restrict the sample to sellers with at least one rated buyers. Then I split firms into deciles by their trade credit borrowing days. I find that 17% – 26% of firms have outstanding A/R financing loans when they are net trade credit lenders, and 38% – 53% of net trade lenders have used A/R financing at least once during 2000 – 2020. Table 1.3 also shows that A/R financing is not solely used by small firms or firms that are net trade credit lenders. Some large firms and net trade credit borrowers also use A/R financing.

Besides trade credit usage and creditworthiness of customers, what are the other determinants of A/R financing? In Table A.4, I run logit and linear probability regressions to study the determinants of A/R financing borrowing. I find that firms that are smaller, have lower cash balance, lower market to-book value, higher debt to EBITDA ratio, yet are more profitable, are more likely to be A/R financing borrowers. I have controlled for industry $\times$ year fixed effects in these regressions. The result indicate that A/R financing borrowers tend be small, financially weak firms, yet are growing or have growth potential.

1.2.3 The pecking order

What are the capital structures of A/R financing borrowers? I study whether A/R financing borrowers and non-borrowers of similar firm size use other debt instruments. I select a 10% random subsample from A/R financing borrowers and match them to a set of non-borrowers by size and industry. Then I collect the debt instruments each firm in the subsample use from Capital IQ.

There seems to be a pecking order in the choice of debt instruments for these firms. Table 1.4 shows that the majority of A/R financing borrowers do not use other debt instruments. Among the other debt instruments, capital lease appears to be the most common instrument. 26.4% of the borrowers have also used capital lease. Similar results apply for the matched non-borrowers: 72% of them have no recorded debt instruments, while 23.9%, 12.3% and 10.3% of them have used capital lease, revolving credit and term loans.

A rough comparison of the interest rates shows that these other debt instruments are more expensive. The median interest rate on the other debt instruments is 4.4%, while the median interest rate on A/R financing is 3.6%. This is consistent with Luck and Santos (2022), who find that accounts receivable are as valuable a collateral as real estate. It reduces the loan spread by 21 bps, compared to 23 bps for real estate. This result comes from regressions that control for the a rich set of characteristics including maturity, bank fixed effects, borrower fixed effects, and time fixed effects.

Of course, I apply a caution when interpreting these results: Capital IQ may not cover all the debt instruments that firms use. As supporting evidence on the pecking order, the median leverage of matched non-borrowers 10.8%, lower than median firm in Compustat, and lower than the borrowers whose median leverage is 18.6%. This suggests that non-borrowers do not substitute for other debt instruments, rather, they appear to be not able to secure credit from A/R financing.

1.2.4 Loan pricing in originations and renegotiations

The way A/R financing loan is priced in originations and renegotiations impacts my model's assumptions. One feature of A/R financing is that renegotiations of loan terms are frequent. Changes in credit limit are the most frequent, with a median spell of 1.53 years (mean is 2.38 years). The median spell between interest rate changes is 1.80 years (mean is 2.55 years). And the median spell between advance rate changes is 2.19 years (mean is 3.21 years). On the one hand, the frequent renegotiations make A/R financing terms flexible. On the other hand, the informational content of loan term changes might be small.

The changes in credit limits are often stated by borrowers to be due to seasonal sales and sales expansions. Indeed, in Table A.5, I show that credit limit is increasing in a firm's accounts receivable balance, and is decreasing in cash flow (EBITDA). So it appears that lenders adjust credit limits to accommodate borrower's liquidity needs, and what really binds the loan amount is the borrowing case: accounts receivable. This makes me not to explicitly model credit limit.

Most of the variation in advance rate and interest rate is at the firm level. Table A.5 shows that in the cross section, advance rate and interest rate are related to borrower cash flow and collectibility of accounts receivable. But for an existing borrower, changes in these observable characteristics do not appear to lead to loan term changes. Unobservable characteristics like the aging and composition of accounts receivable might lead to these loan term changes.

Some of the loan term changes are due to breach of covenants. Since the majority of A/R financing loans have single lenders, the cost of renegotiation is low. The frequent renegotiations are consistent with Garleanu and Zwiebel (2009), who study the optimal allocation of control rights in the context of asymmetric information about the magnitude of potential asset substitution. They show that when renegotiation costs are low, this information asymmetry results in optimal covenant that are tight and thus frequently violated, but also frequently relaxed upon violation. Taken together all these, the informational content of loan term changes might be small, thus is not tested or modeled.

1.3 Seller Moral Hazard and Payment Delays

In this section, I provide evidence that seller moral hazard affects payment delays. To start, I check whether supply agreements allow buyers to return bad products to sellers after delivery, that is, after when the invoice is usually generated and accounts receivable is formed. A substantial fraction of supply agreements in my data have such provisions. In Appendix A.1, I provide an example where the buyer can return bad products to the seller within 45 days after delivery. Although the length of the inspection period does not always precisely equal to the payment delay, they seem to be positively correlated.

I use warranty claims filed against firms as a proxy of firms' product quality risk, since warranty claims represent the ex-post charges to firms caused by defective products, repairs and returns. One caution here is that firms may endogenously choose whether to provide warranty and the strength of warranty protection, and it is unclear whether firms with higher or lower product quality risk are more likely to provide warranty. I mitigate this potential selection bias by aggregating the warranty claims data to industry level, under the assumption that firms within an industry are likely to obey industry conventions in terms of warranty provision. Since product quality risk is a persistent feature, I take the average of warranty claims over 2003–2021 and construct a cross-sectional measure of product quality risk: warranty claims as a percentage of past sales for 160 4-digit SIC industries.

I show that as a buyer purchases from multiple sellers, it pays the seller from an industry with higher product quality risk more slowly. Figure 1-6a is a binscatter plot that shows the positive correlation between payment delays and product quality risk measured by warranty claims. I only control for buyer firm fixed effects and year fixed effects. One advantage of my analysis is that my data on the composition of accounts receivable of firms allows me to observe payment delays at the seller-buyer-year level, so it is possible to control for buyer fixed effects. This is not available in existing work (Klapper, Laeven, and Rajan, 2011; Giannetti, Burkart, and Ellingsen, 2011; Murfin and Njoroge, 2013). Another advantage of my analysis is that sellers in my data reside in very diverse industries and are most represented in manufacturing industries. This makes it possible to distinguish sellers based on the industry measure of warranty claims.

Table 1.5 further strengthens this result. First, I run regressions at the seller level, and show that sellers with higher product quality risk are paid more slowly. Such regressions apply to a much larger sample of firms, as receivable days at seller level is readily available based on regular balance sheet data. In the regressions, I control for seller 2-digit SIC industry × year fixed effects, and seller characteristics including size, leverage and cash flow. The result shows that when warranty claims as a percentage of sales increase by 1%, the seller is paid 7 days (about 12%) more slowly. Second, I run regressions at the seller-buyer pair level, in the same spirit of Figure 1-6a. I again find a positive correlation between product quality risk and payment delay, after I control for buyer × year fixed effects and seller characteristics including size, leverage and cash flow. These results suggest that seller moral hazard affects payment delays.

There are, however, at least two reasons why warranty claims are not perfect measures of the returns of bad product. First, warranty claims are usually filed by consumers rather than buyer firms, and the warranty period can last from a few months to many years. Buyer firms do not need to return bad products via warranty claims. Rather, supply agreements specify a typically short period of time right after delivery for the buyer to return bad products, as the example in Appendix A.1. To the extent that warranty claims reflect product quality risk, it is correlated with buyer returns. However, we cannot use warranty claims to quantify the magnitude of seller moral hazard that impact payment delays. Second, warranty claims data is only available for certain manufacturing firms, and only a small sample of firms have warranty claims data.

These reasons make the approach of using A/R financing loan terms to back out seller moral hazard much more appealing. First, the return of bad products directly affects loan payoffs, and are reflected in the loan terms. In particular, the adjusted haircut, that is, nominal haircut deducting interest payment, can be interpreted as the buffer set aside by lenders to absorb the dilution of invoice value caused by return of bad products. And I show in Figure 1-6b that the adjusted haircut is indeed approximately linearly increasing in payment delays: the longer is payment delay, the more bad products a buyer can discover and return to the seller, and anticipating this, the lender sets higher adjusted haircut. This relationship serves the basis of how I back out the magnitude of seller moral hazard, which I detail in the next section. Second, A/R financing data covers more firms and are not restricted to manufacturing industries. Third, to make sure our analysis is relevant for policies, we should also focus on the sample of firms that borrow from A/R financing. The trade relationships that might benefit from a reduction in payment delays are exactly those that include A/R financing borrowers, who are in liquidity needs and have to borrow at a higher cost compared to their customers.

1.4 The Screening and Monitoring Role of Lenders

In this section, I provide evidence on the screening and monitoring role of A/R financing lenders, and show that they are valuable for the buyers. Since lenders have skin-in-the-game in a seller's product's quality, their screening and monitoring may deter the seller's moral hazard actions. Also, lenders may screen out sellers with high default risk and restrain a seller's default risk after it enters into the borrowing relationship. Lenders' impact on both seller moral hazard and seller default risk are beneficial for buyers who value high-quality and stable supply of products, and to some extent moral hazard is correlated with the default risk. Due to this correlation, I am unable to distinguish between these two margins on which lenders' screening and monitoring are targeted. Nonetheless, the purpose is to show that by getting the lenders involved in the trade relationships, the likelihood of sellers producing high quality products can be improved.

1.4.1 Empirical evidence

What happens to firms before and after A/R financing? I adopt the following event-study specification:

$$Y_{it} = \sum_{k=-5,k\neq-1}^{5} \alpha_k \text{AR Financing}_{i,\{t-t_0=k\}} + \text{Controls}_{it} + \text{Firm FEs} + \text{Year FEs} + \epsilon_{it}, \tag{1.1}$$

where the dependent variable Y_{it} is either receivable days (accounts receivable/sales $\times$ 360), log sales, payable days (accounts payable/purchases$\times$ 360), or log SG&A expense of firm i in year t. AR Financing is a dummy that takes the value of one when a firm has ever used A/R financing. t_0 is the time when a firm borrows from A/R financing for the first time. Control variables include size, leverage, cash balance (quick ratio), cash flow (EBITDA/total assets), and market to book ratio (Tobin's Q). I restrict the sample to 5 years before and after a firm borrows from A/R financing for the first time.

The regression can also be interpreted as a dynamic difference-in-differences design, with the caution that borrowing from A/R financing is endogenous. Also, I restrict the sample to borrowers with large customers: those rated at investment grade, in the top 20% of Compustat firm's size distribution, or are 20 times larger than seller, which makes sure the customers never default and have easy access to financing.[31] Since matching borrowers with similar non-borrowers requires quite many empirical considerations, I restrict the estimation sample to be only the borrowers. Unreported results suggest the conclusions are robust to including matched non-borrowers in the estimation sample.

Figure 1-5 plots the estimated coefficients α_k for these outcome variables respectively. Consistent with the screening and monitoring role of A/R lenders, receivable days decline after a firm borrows from A/R financing for the first time. The decline is about 3 days (5%) on average, and persists for at least 5 years. There is no significant preexisting trend in receivable days. The firm's sales together with SG&A expense also increases, suggesting A/R financing is used by firms to expand sales.[32] Finally, there is no evidence that borrowers pay their own suppliers faster after receiving A/R financing. This implies A/R financing improves the net trade credit days for borrowers, and suggests the borrowing decision is not driven by borrower's own suppliers. These results are also reported in Table 1.6.

I exploit cross-sectional variations to find supporting evidence for the screening and monitoring channel. The results are reported in Table 1.6. First, I find that borrowers in agricultural and mining industries experience a smaller decline in receivable days after the first time they borrow from A/R financing. These firms provide clearly

[31]I restrict the sample to borrowers with large customers for a cleaner interpretation. I should mention that this restriction only lowers the power of my test since it leads to a smaller sample. The effect is actually stronger in the full sample.

[32]Although firms increase SG&A spending after A/R financing, I find no evidence that Capex increases after A/R financing. This suggests the credit is mainly used to support short-term sales-related working capital needs, rather than to fund long-term investment.

homogeneous goods that are easy to check the quality at delivery, so have less seller moral hazard problem that affects the payment delay. Hence, lenders' screening and monitoring role does not add much value. It is not cherry-picking to consider only borrowers in agricultural and mining industries to generate this result. The difficulty here is that it is not obvious lenders can add more value for firms in which industry, so a comparison across all industries is not much useful. There are 116 agricultural and mining industries borrowers in this analysis, and the sample size grants the test some power.

Second, the value added by A/R lenders should decline with a firm's age when it first borrows from A/R financing. This is a reputation effect: uncertainty of a firm's product quality declines over time. And it implies the decline in receivable days would be smaller for the older firms, which is exactly what I find. Third, lenders that are more experienced in providing A/R financing loans should have more expertise in identifying good sellers and are more effective monitors. Having them on board the trade relationship better resolves the seller's product quality risk. Thus, sellers that borrow from the more experienced lenders should have a larger decline in receivable days. I measure lender experience by the total number of borrowers to which a lender provides loans. This hypothesis is also verified, as shown in Table 1.6.

Suppose two firms both reach out to A/R lenders for financing, while one receives the credit and the other doesn't. What are the causes and consequences of the two different experiences? The following analysis further highlights the screening and monitoring role of A/R lenders. Specifically, I compare A/R financing borrowers with matched non-borrowers that have similar financing needs. I show that the matched non-borrowers failed to receive A/R financing, not because the quality of their customer base (collateral value of accounts receivable) is lower. Yet, it is more likely that their own quality is lower: non-borrowers have lower sales and lower relationship length with customers.

To measure financing needs, I run logit regression to generate a propensity score of borrowing, based on a rich set of observables that capture firm's financing need. Table A.4 reports the results of the logit regressions. I adopt the specification in

Column (2) of the table to compute the propensity score of borrowing. Specifically, the observable variables I use include net trade credit, size, cash balance, cash flow, Q, leverage, debt/EBITDA ratio as well as industry $\times$ year fixed effects.

Based on this propensity score measure, I construct a matched sample of non-borrowers. For each borrower, I find two non-borrowers in the same industry, that are closest to the borrower in terms of the propensity score of borrowing, as well as age, size, receivable days, and payable days. I match the samples on ages, and further restrict borrowers and non-borrowers to young firms (on average younger than 10 years old during 2000-2020), in order to minimize lender's selection based on reputable sellers (free riding on buyer's information). Size, receivable days and payable days are important determinants of a firm's decision to borrow from A/R financing, so I match on them explicitly rather than through the propensity score. These are also information commonly known to lenders and buyers that I control for, to identify the value added by the lenders.

Table 1.7 shows the results of this comparison. The outcome variables include the quality of firm's customer base and measures of a firm's performance. First, non-borrowers have similar amount of uncollectible accounts receivable than borrowers. They also have similar concentration of customer bases, measured by the Herfindahl–Hirschman Index (HHI) of sales to individual customers. The borrowers and non-borrowers also have customers of similar size. This suggest that the reason non-borrowers in this sample were not able to get a loan is not because of customer's higher default risk or concentration or size, but more likely to be due to their own poor quality. Second, I find that borrowers on average have 0.68 years longer relationship with customers than non-borrowers. They also have about 35% higher sales than non-borrowers. If borrowers are higher quality sellers, then naturally buyers have longer relationship with them and purchase more from them. These are consistent with the screening and monitoring role of A/R financing.

1.4.2 Alternative explanations

I consider several alternative explanations for the finding that receivable days decline after a seller first borrows from A/R financing. First, the decline in receivable days is not due to the pledged accounts receivable being moved off balance sheet. Firms can sell receivables off-balance sheet to an unconsolidated SPV through securitization, or to a factor through non-recourse factoring. Having cross-checked with Lemmon et al. (2014)'s data on firms that have used securitization, A/R financing borrowers in my sample rarely use securitization. I have also excluded non-recourse factoring from the analysis.

The second alternative explanation is that when a seller needs liquidity, it borrows from A/R lender and simultaneously asks buyers to accelerate payment. However, I find no evidence that the receivable days of a firm would decline in response to a decline in cash flow. Quite opposite to the view that a buyer would accommodate a seller's liquidity need, Ersahin, Giannetti, and Huang (2022) in fact find that firms in operational difficulties caused by natural disasters provide more trade credit to their buyers. They interpret this finding as: a seller hit by operation difficulty needs to preserve the relationship with the buyers by transferring surplus to them via more trade credit, otherwise the buyers would switch to purchase from other sellers in healthier situations. This alternative explanation is also hard to explain why older firms have smaller decline in receivable days, or why firms borrowed from experienced lenders experience larger decline in receivable days, or the difference in relationship length with customers between borrowers and non-borrowers.

The third alternative explanation is that the borrowing decision and the decline in receivable days are both driven by composition changes of buyers. For example, when a seller expands sales to small buyers, it may be able to bargain with the small buyers to have quicker payment, and receivable days as a measure of average payment delay will decrease. And meanwhile, it is the expansion in sales that leads the seller to borrow from A/R financing. I reject this alternative explanation by directly showing that the composition of accounts receivable or sales of the borrower does not change.

I use the following specification:

$$Y_{ijt} = \sum_{k=-5,k\neq-1}^{5} \alpha_k \text{AR Financing}_{i,\{t-t_0=k\}} + \beta_k \text{AR Financing}_{i,\{t-t_0=k\}} \times \text{Large Cus}_j + \text{Controls}_{ijt} + \text{Firm FEs} + \text{Year FEs} + \epsilon_{ijt}, \tag{1.2}$$

where Y_{ijt} is either the fraction of seller i's accounts receivable due from buyer j in year t, or the fraction of seller i's sales to buyer j in year t. Large Cus_j is a dummy that takes the value of one if buyer j's size is above the median of seller i's buyers. Panel (e) and (f) in Figure 1-5 plot the regression coefficient β_k. Panel B of Table 1.6 report the regression results. Indeed, the composition of accounts receivable or sales of a firm does not change after it first borrows from A/R financing.

1.4.3 Discussion

My empirical evidence rejects two counterarguments for A/R financing that point out its harm to trade relationships. One counterargument is that by making sure sellers receive cash upfront, A/R financing unwinds the incentive effect of payment delays in addressing seller moral hazard. To restore the incentives, buyers may need to delay payments even longer, otherwise they have to work with other firms that are ex-ante more transparent. Both predictions are inconsistent with my evidence. What is neglected by this counterargument is that seller moral hazard also affects the collateral value of invoice, so it may not be in the lender's interest to destroy the incentives of payment delays. The fact that A/R financing loans has large haircut on the invoice value, leaves room for the seller's eventual payoff to be contingent on product quality. Another counter argument suggests that after sellers secure A/R financing, buyers have more "excuses" to squeeze the sellers and further delay payments. This is also at odds with my evidence. Overall, my evidence points to the benefit of A/R financing to trade relationships.

My results are to some extent consistent with Billett, Freeman, and Gao (2021). They find that when policy shocks make accounts receivable securitization easier for lenders, which presumably makes it cheaper for a seller to pledge accounts receivable to get a loan, the seller's receivable days decline. However, they do not use A/R financing data so it is unclear whether the policy shocks actually lowered the cost of A/R financing, since many A/R lenders carry the accounts receivable on their balance sheet rather than sell them to the securitization market. Their interpretation of the decline in receivable days is also different: easier access to debt increases the bargaining power of sellers, such that sellers can better negotiate with the buyers to have quicker payments. However, market power alone does not rationalize payment delays. Unless the product's price is regulated,[33] it is always optimal for a buyer to extract surplus by asking for a lower product price rather than by delaying payment. My interpretation of the decline in receivable days instead relies on the screening and monitoring role of lenders.

1.5 A Model of A/R Financing and Trade Credit

My model is related to Biais and Gollier (1997) but key differences are as follows. Their paper features the "large seller, small buyer" setting, where the buyer has no cash and needs to borrow to purchase from the seller, and the question is whether the bank or the seller should lend to the buyer. The seller and the bank have independent information on the *buyer*, and the seller's information is communicated to the bank through trade credit decision. My model features the "small seller, large buyer" setting, where the buyer has abundant cash. The buyer and the bank have independent information on the *seller*. Trade credit terms and loan terms are jointly determined, and the bank's loan decision communicates its private information to the buyer.

1.5.1 Agents

I consider three types of agents: sellers, buyers and banks. They are risk neutral and rational. Banks are competitive and earn zero expected profits. Buyers and sellers can both apply for loans from the bank. I normalize the interest rate of buyers

[33]Even if direct price discrimination is regulated by anti-trust laws, there are many ways firms can circumvent such regulations, e.g., by customizing minor changes in product specifications for different customers. In fact, loyalty, discount and hidden fees are often used by sellers to favor customers.

to 0. The cost of funds for the sellers' banks is g, which is equal to interest rate at which the banks can borrow. g may be interpreted as cost of information production and regulation constraints for A/R lenders. Buyers face inelastic demand of their products, and the demand elasticity is ρ. Credit granting (or rationing) and loan terms for the seller are endogenous. The loan terms include advance rate α and interest rate r.

The seller produces goods at a unit cost and delivers Q units of goods to one buyer in every period. To operate the seller needs to invest I unit of input in its production process during each period.[34] I is drawn from a distribution with *c.d.f.* $F(\cdot)$. If the investment is not fully undertaken, the seller goes bankrupt. The seller receives payment pQ from the buyer with a delay D after delivery.[35] She has no cash at the beginning of the period, so I must be financed. Her only collateral is accounts receivable, which she pledges at the bank and receives a loan. She needs to repay the bank when the buyer makes the payment, that is, when the accounts receivable collateral dissolves. In addition to doing business with the buyer, the seller is assumed to have another exogenous source of income, which, at the time of buyer's payment, generates cash flow C with probability δ, or 0 with probability $1 - \delta$. Should the buyer's payment made after the delay not suffice to repay the bank, the seller would have to use this cash flow C to repay the bank. That is, bank can use the recourse against the seller's cash flow C when necessary. The seller has reservation utility η every period.

The buyer produces final goods at zero additional cost after the seller's delivery. The final goods generates a random cash flow eventually for the buyer. This cash flow depends on the seller's misbehavior in the production process. When the seller's misbehavior is at level b (equivalently, b can be thought of as the reduction of effort), there are two states: product quality is bad with probability $\phi(b)$, and good with probability $1-\phi(b)$. The private benefit to the seller is $\psi(b)$. Bad product probability

[34] This investment need I could represent, for example, payment for worker wages.

[35] Some supply agreements in my sample uses late payment interest rate to punish buyers that pay later than the contracted dates. However, I do not model late payment interest, for tractability. In my model, contractual payment delay is optimal for buyers so they have no incentive to pay even later.

$\phi(b)$ and private benefit $\psi(b)$ are both increasing in misbehavior b. [36]

Both the bank and the buyer have private information on the severity of seller's moral hazard problem, in the sense as I describe in Section 5.3. The buyer can return some of the bad-quality products and get refund from the seller. The amount of refund is related to payment delay, which I discuss in Section 5.4. We are interested in solving for the payment delay, sales, relationship length, and loan terms in equilibrium.

1.5.2 Timeline of the model

Figure 1-2 shows the model's timeline. Time is discrete. Let $t = 0, 1, 2, ..., T, ...$ denote the time of delivery. Before the delivery, seller and buyer first enter into a supply agreement that contracts on payment delay, price and quantity. Then bank offers the loan contract to the seller after observing the supply agreement. Upon signing both the supply agreement and the loan contract, buyer and bank observe signals about seller's quality, as we discuss in Section 5.3. After contracts are signed, the seller decides whether to misbehave. After the delivery of product, accounts receivable is generated as evidenced by the invoice. Then the investment need I is realized, and the seller pledges the accounts receivable to the bank to finance this investment need. If the investment need is not fully met, the seller immediately goes bankrupt. Otherwise, the seller continues operation.

After detecting bad quality products and returning them to the seller for a refund, the buyer pays the contracted invoice amount, either to the seller if seller is not bankrupt or to the bank otherwise. The buyer's payments are thus made at $t = D, 1+D, 2+D, ...$ Before the next delivery, the buyer can choose to end the relationship with the seller and find a new seller. This is consistent with the observation that supply agreements in my data can often be ended with just appropriate early notice.[37] I assume supply agreements and loan contracts can be rewritten every period. This is

[36]The state of bad quality product corresponds to the case when buyer's product becomes obsolescent due to sellers not exerting effort in Kim and Shin (2012). This also reflects the idea that "net terms are expected when buyers are unable to distinguish high-quality from low-quality sellers. The period of time over which payment may be delayed will be a function of the time required to verify seller performance." (Smith, 1987; Long, Malitz, and Ravid, 1993)

[37]See Appendix A.1 for features of the supply agreements in my data that are consistent with this assumption.

mainly to simplify the model, such that I do not need to solve commitment problems and other modeling difficulties introduced by dynamic contracts. The assumption is also realistic to a great extent. A/R financing contracts are frequently renegotiated with changes in loan terms, as we discussed in Section 1.2. Buyers and sellers often communicate about contract terms like production quantities,[38] and both are able to end the supply agreement with much flexibility. A buyer can also always choose to pay more quickly than the contractual payment delay if he finds it optimal. However, the caveat is that when the duration of contracts is optimally chosen, it may be correlated with payment delays and loan terms. Right now, I do not have the data to control for contracts' duration.

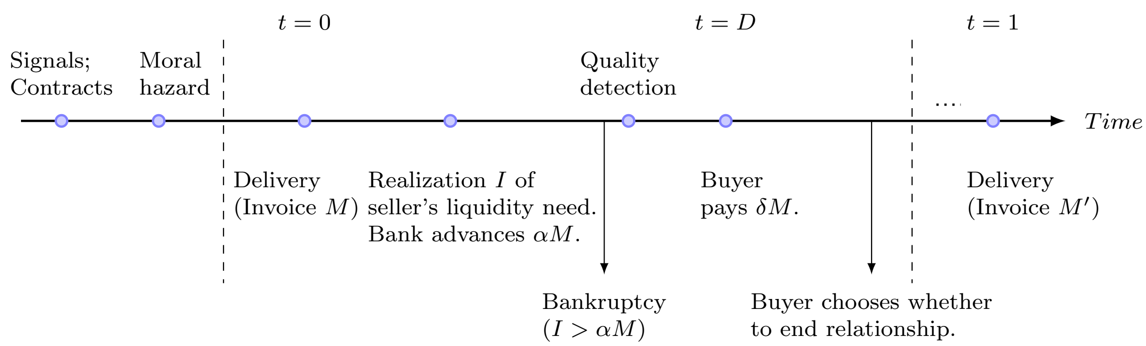


Figure 1-2: Model Timeline

1.5.3 Information structure

There are two types of sellers. The good type has no moral hazard problem, in the sense that it always commits to produce good products, and bad product occurs at probability 0. The bad type has a moral hazard problem. The seller knows its type, while the bank and the buyer do not have perfect information on it. The proportion of good types is h, while the proportion of bad types is $1 - h$.

The bank and the buyer have private information about the seller's type. If the seller is bad then the signal of the buyer, denoted θ, takes the value 1 (good) with probability σ_c and 0 (bad) with probability $1-\sigma_c$. If the seller is good then the signal of the buyer always takes the value 1. Hence, conditional on the signal of the buyer

[38] Again, see Appendix A.1 for features of the supply agreements in my data that are consistent with this assumption.

being good, the probability that the seller is good is

$$Pr(good|\theta = 1) = \frac{h}{h + (1-h)\sigma_c}, \tag{1.3}$$

while conditional on the signal of the buyer being bad, the probability of the seller being bad is 1. This is analogous to the monotone likelihood property in contract theory studies.

The larger σ_c is, the larger is the probability that the bad seller can receive a good signal (error rate is higher), hence the lower the precision of the buyer's information. That only the bad buyers can be misclassified as good can of course be thought of as, a normalization. Qualitatively identical results would be obtained if the good buyers could also be misclassified, as long as the seller's signal would be informative. In that case, we solve for separating equilibrium where both bank and buyer adopt cut-off strategies, sellers whose perceived quality are bad enough are screened out.

The information structure of the bank is similar to that of the buyer. If the seller is bad, the signal of the bank, denoted γ, takes the value 1 (good) with probability σ_b and 0 (bad) with probability $1 - \sigma_b$. Conditional on the signal of the bank being good, the probability that the seller is good is

$$Pr(good|\gamma = 1) = \frac{h}{h + (1-h)\sigma_b}, \tag{1.4}$$

while conditional on the signal of the bank being bad, the probability of the seller being bad is 1. Although I use the narrative that the bank has signals on the seller's type, to the extent that these signals help the buyer work with better sellers, they can also be interpreted as a reduced form way to model the bank's monitoring role.

A seller's type is fixed.[39] The bank and the buyer receive signals every period before they sign contracts. As in Biais and Gollier (1997), I assume the bank and

[39] A simple extension is to let the seller's type be redrawn every couple of years. I discuss this assumption in the model estimation section.

the buyer receive signals and sign contracts simultaneously. The assumption that the bank and the seller make simultaneous offers, contingent on each other's offers, attempts to capture in a simple way the following stylized facts. On the one hand, when providing loans, banks in practice do examine in detail the amount and composition of accounts receivable of the sellers. On the other hand, since the seller rationally expects the investment need, if the bank refused to provide financing, the seller would not undertake the supply agreement.

Finally, I assume that conditional on the seller's type, the signals of the buyer and the bank are independent. Conditional on the seller's type, the probability of a low-quality product and investment need are independent, and are independent to the signals.

1.5.4 Trade credit

Instead of paying the seller at delivery, the buyer pays with a delay D after delivery. Late payment adds to the financing cost and bankruptcy risk of the seller, but can be used to incentivize low misbehavior. The buyer can return detected low-quality products worth of νD to the seller every period. That is, the seller receives pQ in the good product state, and $pQ - \nu D$ in the bad product state. This lets the seller have some skin-in-the-game in product quality, and is a reduced form way to model the incentive role of payment delays in attacking moral hazard. Indeed, this assumption is equivalent to saying bad states are not fully observable at any time during each period. Otherwise, payment can be fully contingent on the states, rather than be governed by payment delay. So ν can be interpreted as a measure of the speed of revelation of the states. [40]

Buyer operates in an industry producing constant elasticity of substitution (CES) varieties for consumers, so it faces a demand curve

$$q = A^{1/(1-\rho)} p_c^{-1/(1-\rho)},$$

[40] This assumption is a reduced form exhibition of Kim and Shin (2012). In Kim and Shin (2012), the longer the payment delay, the closer the payment is to discovery of bad product, and the stronger is the incentive in inducing seller's high effort. Other models of moral hazard have only one period, so the strength of incentive is not explicitly related to length of payment delay.

where $\rho = \frac{\sigma-1}{\sigma} \in (0,1)$, and σ is the CES parameter. When ρ is higher, the demand curve is more elastic, meaning the buyer has less market power. A is a demand shifter. For any quantity Q, the value of sales is $Q^\rho A$.

Sellers are competitive, and their expected profit after deducting interest expenses should be higher than the reservation utility η. For a bad seller (with moral hazard), this means

$$F(\alpha pQ)\Big[pQ - \phi(b)\nu D - \bar{I}e^{rD}\Big]/e^{\beta D} - Q + \psi(b) \geq \eta \tag{1.5}$$

where β is the discount rate of the seller. αpQ is the maximum amount the bank is willing to advance, so $F(\alpha pQ)$ is the probability of satisfying the investment need and continuing operation, and $\bar{I}$ is the expected investment need conditional on $I \leq \alpha pQ$. I assume when the seller goes bankrupt, its eventual payoff is zero. The first term on the left hand side is thus the expected buyer payment deducting loan repayment, the second term is the production cost the seller needs to pay regardless of bankruptcy, and the third term is private benefit. For simplicity, I assume $\beta = r$, that is, the seller's discount rate is equal to borrowing rate. I also assume $i = I/pQ$ follows a distribution with *c.d.f.* denoted $f()$. This makes sure price and quantity do not affect seller's default probability, hence makes the model more tractable. This also enables the model to match data by normalizing a firm's investment needs by its sales (firm size). Other parametric assumptions of the model are summarized later.

The reservation price of bad seller's product can be solved

$$p = \frac{[\eta + Q - \psi(b)]e^{rD}}{f(\alpha)Q} + \frac{\bar{I}e^{rD}}{Q} + \frac{\phi(b)\nu D}{Q}. \tag{1.6}$$

The reservation price of good seller's product can be solved

$$p = \frac{(\eta + Q)e^{rD}}{f(\alpha)Q} + \frac{\bar{I}e^{rD}}{Q}. \tag{1.7}$$

The following assumption makes sure the private benefit of misbehavior makes bad seller's reservation price lower than the good seller. So it is not possible to separate bad seller from good seller by offering a low price. Offering a lower price will first drive out the good sellers.

Assumption 1.1 $\psi(b) > \phi(b)\nu D$ *(private benefit is large).*

The interpretation of the assumption is that the private benefit from misbehavior is larger than the punishment caused by refund. This would necessarily hold in equilibrium for a profit maximizing seller, otherwise misbehavior b should always be 0.

Since buyer cannot distinguish a good seller from a bad seller, it has to offer a uniform price to all the sellers. Because p does not affect the incentive, buyer lets p equal the reservation price of the good seller. So Equation (1.7) is the expression of seller's price. As it shows, the product price is increasing in payment delay D. This is consistent with Amberg, Jacobson, and von Schedvin (2021), who find that trade credit incurs a price premium.[41] The two terms in Equation (1.7) correspond to reasons why this is the case: as payment is delayed longer, first there is a general discounting effect on reservation utility and production cost, second, the financing cost of investment is higher. The product price is decreasing in the quantity ordered Q. This is consistent with common observations that sellers give discount to large purchases. And this is a different property from the price discrimination role of trade credit studied in Giannetti, Serrano-Velarde, and Tarantino (2021), where a seller charges higher price to larger buyers. The price formula also captures the cost of seller default: when the continuation probability $f(\alpha)$ is lower, default probability is higher, and the input price for the buyer is higher.

Consider the situation where the buyer uses its own signal (θ) to decide whether to maintain relationship with the seller. We consider the following equilibrium. It is not willing to purchase from a seller whose signal is bad, because in this case the seller is known to be bad type. And when the signal is good, the buyer will maintain

[41] Amberg, Jacobson, and von Schedvin (2021) find empirical evidence that product prices in transactions involving trade credit include a trade credit price premium, the size of which is determined by the contracted loan maturity and an implicit interest rate, which, in turn, is a function of the selling firm's liquidity costs and the buying firm's default risk.

relationship with the seller.[42]

Some of the sellers with a good signal are of bad type hence have moral hazard problem. Suppose the buyer wants to incentivize the bad seller not to misbehave. The payment delay needs to make sure the seller has enough skin-in-the-game in the product quality. Denote $M = pQ$ as the invoice amount.

The incentive compatible (IC) constraint of the bad seller to induce misbehavior level b is in the following.

$$\begin{aligned} b = \arg\max_{\hat{b}} \; & [1-\phi(\hat{b})]M(1-\alpha e^{rD}+\delta C/M)+\phi(\hat{b})M(1-\delta)\max[0,1-\alpha e^{rD}-\nu D/M] \\ & +\phi(\hat{b})M\delta\max[0,1+C/M-\alpha e^{rD}-\nu D/M]+\psi(\hat{b}) \end{aligned} \tag{1.8}$$

The first term is the expected payoff to seller when its product is of good quality. The payoff equals buyer payment deducting loan payments and adding cash flow from seller's other activities. The second and third term are the payoffs to seller when its product is of bad quality and cash flow from other activities is zero and C, respectively. The fourth term is private benefit from misbehavior. When seller goes bankrupt, its firm value is always 0, so the probability of bankruptcy due to not being able to finance the investment need does not show up in the IC constraint.

The FOC of the bad seller is

$$\begin{cases} \phi'(b)\nu D = \psi'(b) & \text{if } 1-\alpha e^{rD} \geq \nu D/M \\ \phi'(b)[\delta\nu D+(1-\delta)M(1-\alpha e^{rD})] = \psi'(b) & \text{if } 1-\alpha e^{rD} \in [\nu D/M - C/M, \nu D/M) \\ \phi'(b)(1-\alpha e^{rD})M = \psi'(b) & \text{if } 1-\alpha e^{rD} < \nu D/M - C/M. \end{cases} \tag{1.9}$$

From these conditions, we can see there are three cases depending on the advance rate and interest rate of the bank. Henceforth, let me call $1-\alpha e^{rD}$ the adjusted haircut of the A/R financing loan. This adjusted haircut is deeply connected with

[42] In a variation of my model where good sellers can also receive a bad signal, bank will ration credit and buyer will end relationship when the perceived seller quality is bad enough.

the loan's influence on the incentives of payment delays. In the first case, the haircut is very high, and bank keeps enough buffer off collateral value to absorb dilution caused by payment delays. Hence, the dilution does not suffice to push the seller to limited liability protection, nor does the bank need to use the recourse against seller's cash flow from other activities. In the second case, the haircut does not fully cover collateral value dilution, and seller needs to use its cash flow from other activities to repay the bank. When this additional cash flow is zero, the seller is in limited liability, and the incentive effect of payment delays is dampened. In the third case, the haircut is very low, and even the seller's cash flow from other activities is not enough to cover the dilution. This is the case A/R financing gives the seller too much cash flow from the project upfront, and destroys all the incentive of payment delay. The buyer can then only rely on the bank to provide incentive. I only consider the first two cases, since my empirical evidence in Section 1.4 is at odds with the third case.

Now we see that payment delays can incentivize the bad seller to not misbehave (see Equation (1.9)), but the downside is that it increases the financing cost and the "discounting cost" of both good seller and bad seller (see Equation (1.7)). It is obvious that when the fraction of good sellers is 0, that is, $\Pr(good) = 0$, the moral hazard problem is very severe. And the buyer has to incentivize the seller to not misbehave since misbehavior strongly reduces buyer's profit. When the fraction of good seller is 1, that is, $\Pr(good) = 1$, there is no moral hazard problem to solve. In the intermediate cases, the optimal payment delay D depends on the fraction of good sellers.

1.5.5 Bank credit rationing

Consider the situation where the bank uses its own signal (γ) to decide whether to lend to the seller. We consider the following equilibrium. It is not willing to lend when its signal is bad. If the bank has observed a good signal (γ=1), it lends to the seller at advance rate α and interest rate r. For simplicity I do not model credit limit, although it is an important A/R financing term. In more complicated models as in Holmström and Tirole (1998, 2000), the credit limit is related to the borrower's liquidity needs.

We now solve for the advance rate and interest rate that maximizes bank profit. Suppose seller borrows αM. The bank's profit is

$$
\begin{aligned}
\Pi =& \Pr(good)\alpha e^{rD} M + \Pr(bad)[1-\phi(b)]\alpha e^{rD} M \\
&+ \Pr(bad)\phi(b)(1-\delta)\min[1-\nu D/M, \alpha e^{rD}]M \\
&+ \Pr(bad)\phi(b)\delta\min[1+C/M-\nu D/M, \alpha e^{rD}]M - \alpha e^{gD} M. \qquad (1.10)
\end{aligned}
$$

The first two terms are bank's loan payoff from good seller and bad seller when product quality is good. The third and fourth term are bank's loan payoff from bad seller when product quality is bad. In the third term, seller has cash flow from other activities while in the fourth term it does not. Finally the last term is the borrowing cost of the bank itself.

Lemma 1.1 *(Loan terms) The bank's optimal advance rate and interest rate satisfy*

$$1-\alpha e^{rD} = \nu D/M - C/M \qquad (1.11)$$

$$e^{rD} = \frac{1-\nu D/M + C/M}{1-\nu D/M + C/M - Pr(bad)\phi(b)(1-\delta)C/M} e^{gD} \qquad (1.12)$$

All the proofs of the model are in Appendix A.3. Intuitively, to maximize profit, bank sets adjusted haircut to cover the dilution of invoice value caused by return of bad quality products, and bank fully exploits the value of recourse against seller. The interest rate is then given by bank's break-even condition. As we can see, the interest rate is higher than bank's cost of fund, as the bank is not able to receive full repayment when product quality is bad and seller does not have additional cash flow from other activities. The cash flow C from other activities and its associated probability δ can be interpreted as measures of seller default risk. When δ is lower or C is lower, the seller is "riskier", in which case Equation (1.12) tells us A/R financing's interest rate r is higher. This is consistent with the earlier empirical finding that the interest rate spread between A/R financing and buyer's unsecured borrowing rate is increasing in seller's default risk. (See Panel B of Figure 1-4).

In this model, credit rationing is caused by moral hazard as in Holmstrom and Tirole (1997), and is reflected by the fact that advance rate is lower than 1. Another way to model advance rate is by assuming there is adverse selection on quality of collateral.[43] This can be done on the basis of this model by letting the bad seller's customers be more likely to default. By restricting my estimation sample to large creditworthy buyers, it is reasonable to assume that this adverse selection channel is not empirically strong in my sample.

My modeling of bankruptcy cases is drastically simplified. I abstract from bankruptcy costs. Sellers meet their obligations by paying out any cash they have, but there is limited liability. In addition, since bank registered liens on the accounts receivable, it has seniority on buyer's payment over the seller and any other debt holders, and seller does not appropriate any of the buyer's payment.[44]

1.5.6 Equilibrium

In this subsection I analyze a separating Bayesian Nash equilibrium, where the seller obtains bank credit and maintains relationship with buyer if and only if both signals (θ and γ) are positive. I also describe the dynamics of learning over time. In the separating equilibrium, bad sellers with two good signals obtain financing and maintains supply relationship. In that sense, there is not full separation, since bad types with two good signals are pooled with good types. Nonetheless, for simplicity I will refer to the equilibrium I analyze as "separating".

To make the model cleaner and the estimation possible, I make the following parametric assumptions.

Assumption 1.2 *(Parametric assumptions) (1)* $\phi(b) = b^2$, $b \in [0, 1]$. *(2)* $\psi(b) = zbpQ$. *Private benefit is proportional to sales. (3)* $\nu = vpQ$. *Returned goods is*

[43] Indeed, lenders rely on loan-to-value ratio adjustments to manage default risk in other secured lending markets, such as in corporate lending markets (Benmelech, Garmaise, and Moskowitz, 2005; Benmelech and Bergman, 2009) and in derivatives markets (Capponi et al., 2022).

[44] Banks can easily use lockbox account to gain full control of the buyer's payments.

proportional to sales. (4) $i = I/pQ$ *follows c.d.f.* $f(x) = 1 - exp(-x/s)$*. Investment need follows exponential distribution with mean* s*. And I denote* $\bar{i} = \mathbb{E}(i|i < \alpha)$*. (5)* $C = cpQ$*. Seller's cash flow from other activities is proportional to sales.*

Some comments are as follows. The parametric form of $\phi(b)$ guarantees that it is strictly increasing and convex in b. z is a measure of the severity of moral hazard and v is a measure of the effectiveness of payment delays in addressing this moral hazard. Both parameters are important for understanding the policy counterfactual and will be estimated. Private benefits, returned goods, investment need, and cash flow from other activities are all assumed to be proportional to sales. This essentially normalizes these quantities by firm size, and makes the estimation of the model possible.

The following lemma summarizes the conditions on bank's advance rate (α), interest rate (r), the payment delay (D), price (p) paid to the seller, and production quantity (Q) in equilibrium.

Lemma 1.2 *(Equilibrium) In the separating equilibrium where (1) the bank lends to sellers only when it has a good signal and (2) the buyer maintains supply agreement only with sellers for which it has a good signal and to which the bank provides loans, the bank's advance rate is*

$$1 - \alpha e^{rD} = vD - c. \tag{1.13}$$

The interest rate is

$$e^{rD} = \frac{1 - vD + c}{1 - vD + c - Pr(bad)\phi(b)(1-\delta)c} e^{gD}. \tag{1.14}$$

The price of seller's product is given by

$$pQ = \frac{(\eta + Q)e^{rD}}{f(\alpha)(1 - \bar{i}e^{rD})}. \tag{1.15}$$

The payment delay and quantity of goods are given by

$$\rho Q^{\rho-1}A\Big[1-m(D)\Big]-\frac{e^{rD}[1-vDm(D)]}{f(\alpha)(1-\bar{i}e^{rD})}=0 \tag{1.16}$$

$$Q^{\rho}A\frac{d(1-m(D))}{dD}-[1-vDm(D)]\frac{dpQ}{dD}+vpQ\frac{dDm(D)}{dD}=0. \tag{1.17}$$

where $m(D)=Pr(bad)(\frac{z}{2[vD-(1-\delta)c]})^2$ *is the probability of bad quality products.*

The interpretation is as follows. Loan terms are derived in the previous subsection. The payment delay D and quantity of goods Q are given by first order conditions of buyer's profit maximization problem. First, given any payment delay, Equation (1.16) pins down the quantity of production Q. It equates the marginal profit and marginal cost of production. Marginal profit is higher when the probability of bad products $m(D)$ is lower, and is also higher when demand elasticity ρ or demand shock A are higher since they raises the profit of any unit of product. The marginal cost depends on the payment delay, seller borrowing interest rate, probability of default, which all affect the input price. Second, given quantity of production, Equation (1.17) pins down payment delay D. It equates the marginal improvement in product quality caused by the incentive effect and the marginal increase in input cost.

Importantly, Equation (1.13) specifies the relation between loan terms and payment delay: the adjusted haircut is linearly increasing in payment delay, with the slope being the deep parameter v that governs the incentive provided by payment delay. We have seen in Figure 1-6b that the adjusted haircut in the data is indeed approximately linearly increasing in payment delays. I use the slope of this relationship to uncover v.

The following proposition summarizes the dynamics of the fraction of good sellers Pr(*good*) and expected relationship length of firms.

Proposition 1.1 *(Relationship length) Suppose at period* t*, the history of the buyer's and the bank's signals are* $\theta^t=(\theta_0,\theta_1,...,\theta^t)=1$ *and* $\gamma^t=(\gamma_0,\gamma_1,...,\gamma^t)=1$*, the probability that the seller is good is* $h_t=Pr(good|\theta^t=1,\gamma^t=1)=\frac{h}{h+(1-h)(\sigma_b\sigma_c)^t}$. *Let*

$\hat{f} = f(\alpha)[1 - \phi(b)(1-\delta)]$. *The expected relationship length is*

$$E(T|\theta^t = 1, \gamma^t = 1) = h_t \frac{f(\alpha)}{[1-f(\alpha)]^2} + (1-h_t)(1-\sigma_b\sigma_c)\frac{\hat{f}}{[1-\hat{f}\sigma_b\sigma_c]^2}. \quad (1.18)$$

The interpretation is as follows. There are two reasons why a seller's relationship with the buyer ends: first the seller goes bankrupt, and second, either the buyer or the bank receives a bad signal. The expected relationship length is longer when seller's continuation probability in every period $f(\alpha)$ is higher. More interestingly, the relationship length is also longer when the probability of seller being good h_t is higher. This is because conditional on a higher h_t, the probability of bad signals in the future is lower.

In Figure 1-7, I show how equilibrium quantities vary with different levels of Pr(*good*) under a numerical example. We see that payment delay D is decreasing in Pr(*good*), while sales quanity Q and expected relationship length $\mathbb{E}(T)$ are increasing in Pr(*good*). It is also worth noting that in this numerical example, the advance rate is about 80%, matching the average advance rate in data.

The patterns in Figure 1-7 can be proved, and are summarized in the following proposition. Besides, the proposition also summarizes one prediction on the relation between payment delays and demand elasticity that is consistent with my empirical evidence.

Proposition 1.2 *(Equilibrium property) In the equilibrium characterized by Lemma 1.2: (1) As* Pr(*good*) *increases, payment delay D decreases, sales quantity Q and expected relationship length between seller and buyer both increase. (2) Payment delay D is decreasing in demand elasticity ρ.*

A direct corollary of property (1) is that receivable days will decline, and the expected relationship length with customer will increase, after a seller receives AR financing, since the bank's loan decision increases Pr(*good*). This is consistent with the empirical evidence in Section 1.4. As for property (2), it predicts that buyers with higher product market power use longer payment delay. This prediction is related to the notion that high bargaining power buyers pay slower (Fabbri and Klapper, 2008;

Klapper, Laeven, and Rajan, 2011), and does not require the large buyer "misuse" its market power. In fact, this model rationalizes such notion by having rational buyer that considers the costs of trade credit, yet chooses to use longer payment delay to maximize profit.

I present direct evidence consistent with property (2) in Table 1.8. I run regressions of a seller's receivable days on its buyer's demand elasticity. The regressions are run at seller-buyer pair level. Receivable days of a seller from individual buyer again comes from my hand collected data on the composition of accounts receivable. And buyer's demand elasticity ρ is estimated from US import data (see Section 1.1 for details). The result shows that a seller's receivable days from a buyer is indeed decreasing in buyer's demand elasticity, that is, increasing in buyer's product market power. I have controlled for seller $\times$ year fixed effects to sharpen the result: among the buyers of a given seller and in the same year, the buyer with higher product market power pays more slowly than other buyers.

1.5.7 Comparative statics

I now present comparative statics of the model that inform us how parameters of the model can be identified by the data.

Figure 1-8 shows the comparative statics with respect to parameters to be estimated, under a numerical example. We see that the decline in payment delay after a firm borrows from A/R financing as we witnessed in Section 1.4 is related to the precision of the bank's signal. The higher is the bank's signal's noise σ_b, the magnitude of the decline is smaller. After a firm borrows from A/R financing, there should be a increase in expected relationship length, again as we witnessed in Section 1.4. The magnitude of the increase should be smaller when the bank's signal's noise σ_b is higher. When a seller in the data does not borrow from A/R financing because of no liquidity need, or when we look at the period of time before a seller's first A/R financing borrowing, its quality should still be gradually revealed overtime to the buyer, even bank does not provide any information. That is, the payment delay should decrease, and expected relationship length should increase, as the seller-buyer relationship lasts over time. The magnitude of this decrease (increase) is smaller when the customer's signal's noise σ_c is higher. These results are shown in the first two panels of Figure 1-8.

In the last 4 panels of Figure 1-8, we see that payment delay is increasing in private benefit parameter z, $(1 - \alpha e^{rD} + c)/D$ is increasing in the parameter characterizing the incentive role of payment delay v, seller's sales pQ is increasing in its reservation utility η, and expected relationship length is increasing in the fraction of good seller h. These can all be derived from the optimality conditions of the equilibrium. It is worth to comment that in my model the seller's reservation utility η has little effect on payment delay. Instead, the buyer's demand elasticity ρ is closely related to payment delay. This is precisely how "bargaining power" affects payment delays in my model.

The comparative statics shown in Figure 1-8 can be proved, and are summarized in the following proposition.

Proposition 1.3 *(Comparative statics) (1) For a seller that borrows from bank, there is a decline in payment delay ($\Delta D < 0$) after bank grants credit. The decline is smaller when bank's signal noise σ_b is larger. The increase in expected relationship length $\mathbb{E}_{t+1}(T) - \mathbb{E}_t(T)$ is smaller when σ_b is larger. (2) For all sellers, payment delay (expected relationship length) declines (increases) over time. And the decline (increase) is smaller when buyer's signal noise σ_c is larger. (3) Payment delay D is increasing in z. $(1-\alpha e^{rD}+c)/D$ is increasing in v. Seller's sales pQ is increasing in reservation utility η. Expected relationship length $\mathbb{E}(T)$ is increasing in the fraction of good seller h.*

1.6 Estimation and Counterfactuals

1.6.1 Calibrated parameters

I calibrate the parameters of the model that are not intuitively identified by the data: the cost of funds for the seller's banks g, seller's cash flow generated from other activities c which is related to the seller's own default risk hence the value of the recourse, and the distribution parameter of investment need s. The primary purpose of the calibration is to keep the number of estimated parameters reasonable and ensure the estimated parameters are identified in the model. The calibrated parameters are

$$g = 0, c = 0.18, s = 0.4 \tag{1.19}$$

where g is set to be zero, meaning that seller's bank and buyer's bank have the same cost of fund. It may be reasonable to assume the cost of fund of seller's bank (A/R lenders) is higher from buyer's bank (large banks who can use funds from deposit). This is because, investigating the funding source of the A/R lenders, I find that factors and commercial finance companies obtain their funds mainly from invested capital of PE/VC investors or bank borrowings.[45] However, one drawback of my model is its inability to distinguish the value of recourse governed by δ and bank's cost g. I thus shut down the funding cost difference between the banks by setting g to zero, and focus on other parameters that are related to the incentive role of payment delay. In this way, δ can be estimated from A/R financing interest rate.

Cash flow from other activities c is chosen to makes sure my model matches the average level of adjusted haircut in the data. Finally, the average investment need parameter s is calibrated by matching the unconditional expected relationship length with data. In the data, average relationship length between seller and buyer is 2.89 years. This puts restrictions on the parameter s, which in turn governs the probability of seller default in my model. I choose $s = 0.4$ and show later this helps me match the data moments.

1.6.2 Estimation strategy

The identification strategy is to use data on observable variables – advance rate and interest rate of A/R financing contracts, receivable days, relationship length, and seller's revenue and production quantity – to infer properties of unobservable parameters – the incentive role of payment delays v, the private benefits z, probability that seller has additional cash from other activities that governs the value of recourse

[45]In Biais and Gollier (1997), sellers also need to pay an extra cost when borrowing from banks.

δ, the precision of bank's and buyer's information σ_b and σ_c, the fraction of sellers with moral hazard problems h, and seller's reservation utility η. Besides the variables we have seen, I measure revenue by total sales (Compustat item revt), and measure production quantity by cost of goods sold (Compustat item cogs).

I use GMM to estimate model parameters in the sample of A/R financing borrowers. To do so, I construct the following moment conditions from the equilibrium conditions of the model.

$$\begin{aligned} f_1(\alpha_{it}, r_{it}, D_{it})\mathbb{1}_{after} - \epsilon_{it} &= 0 \\ f_2^b(Q_{it}, pQ_{it}, D_{it})\mathbb{1}_{before} + f_2^a(Q_{it}, pQ_{it}, D_{it})\mathbb{1}_{after} - \xi_{it} &= 0 \\ \mathbb{E}(T_{it} \times \mathbb{1}_{after}) - \mathbb{E}(T_{it} \times \mathbb{1}_{before}) - u_i &= 0 \end{aligned} \tag{1.20}$$

Explanations are as follows. I suppressed the i, t subscripts in the two indicator variables $\mathbb{1}_{before}$ and $\mathbb{1}_{after}$. Closely following the empirical section, observations before a firm first borrows from A/R financing is classified as $\mathbb{1}_{before} = 1$, while those within the 5 years after the first borrowing are classified as $\mathbb{1}_{after} = 1$. The 5 years restriction is to avoid the many unmodeled shocks in the later years that could contaminate the estimation, and mapping to the model, this can be interpreted as that seller types are redrawn every 5 years. Separating the observations into two classes allows us to utilize the variation before and after A/R financing.

Table 1.9 summarizes all model parameters and intuitively how each parameter is estimated. The idea is to utilize the comparative statics results in the previous section. The first set of moment conditions $f_1(\cdot)$ are the bank's FOC conditions in Equation (1.13) and (1.14). The second set of moment conditions $f_2(\cdot)$ are the buyer's FOC conditions in Equation (1.16) and (1.17). The third moment condition comes from Equation (1.18) that characterizes the dynamics of expected relationship length. $\epsilon_{it}, \xi_{it}, u_i$ are unmodeled noises, which are unobservable to econometricians, and equal to 0 in expectation. The only difference between f_2^b and f_2^a is the following: I set the investment need of the seller to zero, and $\sigma_b = 1$ for the time before A/R financing

(f_2^b), while investment need and σ_b are estimated for the time after A/R financing (f_2^a). The third moment condition essentially compares the expected relationship length before and after a firm first borrows from A/R financing. The fact that I do not use the variation of relationship length over time for the estimation is due to the measurement errors from the seller-buyer trade relationship data. Also, I emphasize the difficulty in measuring expected relationship length T_{it} before a firm's A/R financing. This is due to that firms typically borrow from A/R financing when they are young, thus the number of seller-buyer observations is small. So I use the relationship length of the matched non-borrower in Section 1.4 as the "counterfactual" measure of T_{it} before a firm's A/R financing. Mapping to the model, I set $\sigma_b = 1$ for the case of $\mathbb{1}_{before} = 1$, assuming that the matched never-borrowers have the same investment need, and can access funds from a source that does not provide screening and monitoring service as the A/R lenders.

Table 1.9 also lists a few other model parameters that are directly measured, which I discuss now. First, buyer's demand elasticity ρ is measured by the approach of Soderbery (2015) using 2000–2020 US import data, at 4-digit SIC code industry level. Second, demand shock A is measured by $A = Rev/COGS^{\rho}$. In the model, buyer's revenue is $Q^{\rho}A$ and cost of goods sold is pQ. Since in reality a firm has many suppliers, it is difficult to find the price of goods from each supplier. Here I use the approximation that p is close to 1, that is, suppliers are close to break-even, to derive the measure for A.

Before the GMM estimation, I perform some pre-treatment to the data, to partial out heterogeneities in the data that are not modeled. First, to completely exclude the risks of customer default, I calculate the invoice value pQ as total A/R minus the uncollectible A/R. That is, the face value of invoice is adjusted by the uncollectible amount. This is because invoices that are estimated to be uncollectible, for example, those that are aged more than 90 days, would not be considered eligible as collateral by lenders. Invoice value and production quantities are then scaled by firm's total asset, as a normalization.

Second, my measure of payment delays is based on receivable days AR_{it}/Rev_{it} at year ends. The advantage of using receivable days over contractual payment delays in supply agreements is that a buyer can always choose to pay quicker or slower than the contractual delay, and receivable days measure how quickly sellers are effectively paid. Also, receivable days are available for all public firms, and provide variations over time that help me estimate the model. However, receivable days is not a measure in calendar days. To convert it into the space of calendar days, I do the following transformation to compute D_{it}:

$$D_{it} = \frac{AR_{it}}{Rev_{it}} \cdot \mathbb{E}_{i \in Industry\, k} \frac{\tilde{D}_i}{AR_{it}/Rev_{it}} \tag{1.21}$$

This is an heuristic approach. We observe the contractual payment delays $\tilde{D}_i$ in calendar days for a subset of the firms. For each industry k, I compute the ratio of contractual payment delays to receivable days. This industry-level ratio is then used to scale receivable days to arrive at D_{it}.

Finally, we also need to know the time between deliveries, because the model assumes the frequency at which bank and buyer receive signals equals the frequency of delivery. Since one seller delivers products to many buyers, and the deliveries to different customers happen at different time, to measure the time between deliveries an empirically daunting task. I assume delivery happens every 30 day. Using different delivery time is not material for the estimation and the conclusions, but just scales parameters differently.

Instrument variables There are reasons to believe that the error terms $\epsilon_{it}, \xi_{it}, u_i$ in Equation (1.20) are correlated with the observable variables. So I use instrument variables to address the endogeneity concerns.

The main identification challenges here are that seller's borrowing decision is endogenous, and that loan terms are affected by unmodeled variables that correlate with payment delays.[46] Since I only estimate the model in the sample of A/R financing borrowers, this leads to selection bias. I treat the selection bias caused by endogenous

[46]Borrowing decisions may be related to growth opportunities and continuation value of selling to customers. The omitted variables may include macro-economic shocks and seller default risks.

borrowing decision as an omitted variable according to Heckman (1979), that may correlate with payment delays D. This allows me to tackle both identification challenges by using instrumental variables for payment delays. As the first instrumental variable, I use liquidity shocks to buyers, measured as the fraction of debt that needs to be refinanced in the subsequent year, following Almeida et al. (2011). The second instrumental variable I use is lagged industry-level average payment delays. The two instrumental variables affect bilateral payment delays but are largely uncorrelated with omitted variables that affect loan terms.

Table A.6 presents the first-stage regressions of the instruments. Liquidity shocks to a buyer are correlated with a seller's receivable days in the cross section, with F-statistics safely above 10. In the time series, liquidity shocks do not correlate much with a seller's receivable days after controlling for seller fixed effects. However, the cross-sectional correlation is enough for my purpose of uncovering seller moral hazard. By restricting the sample to large buyers, I am confident that these liquidity shocks do not increase buyer default risk. Instead, buyers absorb the liquidity shocks by paying the sellers slower. Our second instrument, lagged industry-level average payment delays is strongly correlated with a seller's payment delay both in the cross section and in the time series, with F-statistics much above 10.

1.6.3 Estimation results

Since demand elasticity ρ is only available for manufacturing industries, my estimation is carried out only for firms in manufacturing industries. The sample contains 2169 firms with 18,049 observations. Table 1.10 reports the estimation results. The results suggest, buyer can return $v = 26.5\%$ of the bad-quality products to seller per 30 days. There is a $\delta = 67.5\%$ probability that seller can generate additional cash flows from activities other than the trade with the buyer, which serves as the basis of the value of lender's recourse option. The noisiness (error rate) of bank's signal is $\sigma_b = 0.47$ compared to the noisiness of buyer's signal $\sigma_c = 0.70$. The high precision of bank's signal reflects the large decline in receivable days after a firm's first A/R financing borrowing. The initial fraction of good sellers in the economy is $h = 32.8\%$.

Seller's reservation utility is $\eta = 0.06$, meaning that a seller requires approximately a 6% net profit to be willing to supply products. Among these parameters, the noisiness of bank's signal is σ_b and the initial fraction of good sellers h are not very well identified. This reflects the general difficulty in separating moral hazard from adverse selection in a structural model. However, they are still significant at 95% under standard errors clustered at the firm level.

Under the estimated parameters, the model is able to match a few key data moments. As shown in Table 1.11, my model matches the average payment delay and relationship length in the data particularly well. My model-implied adjusted haircut and interest rate are slightly higher than the data. Given the simplicity of the model and that there are lots of data heterogeneity I have not captured, I conclude that my model matches the data reasonably well.

1.6.4 Counterfactual analyses

What happens when the government cuts payment delays? I consider a counterfactual scenario where the government restricts the payment delays of the average US firm to be 30 days, while the average payment delay in the data is 51.7 days. Table 1.12 shows the effects of this regulatory limit. While the regulation can decrease seller's default probability from 15.6% to 10.3% (a 5.3% decline), the presence of bad quality products increases by 12.6%. This is because under the regulation buyers do not have long enough time to discover and return the bad quality products. This, in turn, incurs more moral hazard actions that lead to bad quality products. As a result of worsening product quality, the buyer's total production will decline by 26.3%. Seller's gross profit will be reduced by 32.4% given the contraction in buyer's purchase. This shows that the well intended policy could actually harm the sellers in terms of reducing their gross profit. A benevolent government needs to trade off the regulation's effect on reducing seller default versus the effect on reducing production and seller profit.

What happens when the government subsidizes A/R lenders? I consider a counterfactual scenario where the government can lower A/R lender's cost of fund by 5%

per year by directly subsidizing the lenders. Table 1.12 shows the effects of this policy. With subsidized A/R financing, payment delays will increase by 2.04 days. This counterfactual increase illustrates the difficulty of identifying the screening and monitoring role of lenders using a reduced-form approach (like a difference-in-differences method): shocks that make the financing cheaper or more available may increase payment delay, and would appear as if seller quality has declined under the framework that seller moral hazard determines payment delay. Despite the increase in payment delay, total production and the seller's gross profit both increase, and the presence of bad products slightly decreases. The seller's default probability increases by 0.1%, which is insignificant. It is worth mentioning that the government can increase a firm's annual production quantity by $2.6 million at an annual expense of $1.57 million. What makes the subsidy a positive NPV project is the fact that when A/R financing is cheaper, the buyer uses longer time to inspect products. That reduces seller moral hazard, and in turn decreases the lender's loss and the A/R financing rate, constituting a feedback effect. A government can effectively achieve this goal by allowing more outside investors to provide the "subsidy", for example, by lifting potential regulations that restrict lenders from conducting A/R financing.

1.7 Conclusion

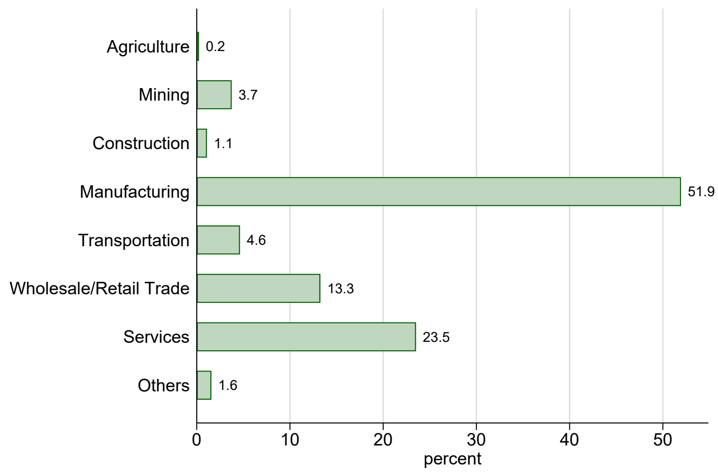
This paper adopts an efficient view of trade credit and A/R financing. Payment delays are optimally chosen by buyers to address seller moral hazard that leads to bad quality products. Sellers burdened by payment delays can in turn seek credit from a large mature A/R financing market when they need financing. Loan terms of A/R financing help us uncover seller moral hazard, because bad quality products dilute the collateral value of invoice. A/R financing adds value directly to trade relationships by screening and monitoring the sellers that borrow, which further rationalizes the use of trade credit. I quantify the informational roles of trade credit and A/R financing to conduct welfare analysis for policies. Some limitations of the paper should be mentioned. First, I do not consider other reasons why a government may want to regulate payment delays: for example, to enhance financial stability (Kiyotaki and Moore, 1997) or curb externalities (Glode and Opp, 2021). Second, my model is estimated by data on US public firms, while moral hazard problems of private firms could be more severe. Third, my model applies better to trade relationships where large buyers value high quality sellers, and instruments other than payment delays are not enough to regulate seller moral hazard.

This paper does not have the means to study the frictions in the A/R financing market. Although these frictions have limited impact on this paper's analysis and conclusions, they make A/R financing unlikely to be the panacea to all sellers that are burdened by accounts receivable. For example, A/R lenders tend to avoid portfolios highly concentrated in a few customers and many lenders carry the loans on their balance sheets which impedes the diversification of risks. There is a lack of flexibility in selecting the invoice to pledge – A/R lenders would only take in all the accounts receivable of a borrower as collateral, to avoid adverse selection on collateral risks. New financial products like reverse factoring and deep-tier factoring, which allow sellers to select the accounts receivable to pledge, provide sellers with such flexibility. Future research may shed more light on the frictions in the A/R financing market to inform policies that aim to enhance the supply chains.

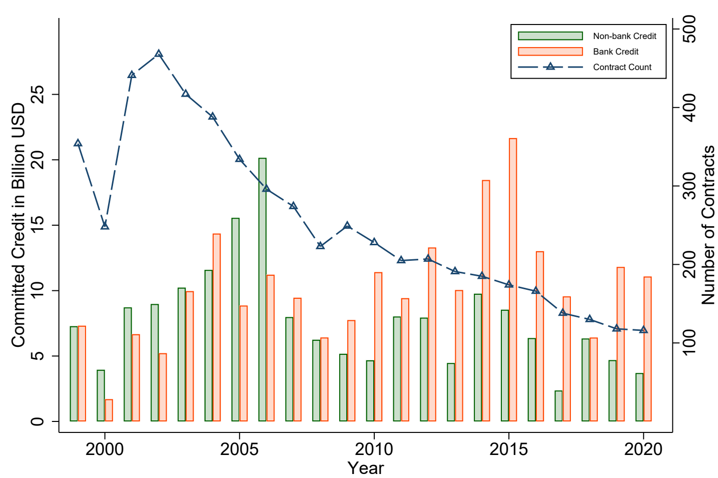
1.8 Figures and Tables

Figure 1-3: Industry Breakdown of A/R Financing Borrowers and Trend of Credit

(a) Industry Breakdown of A/R Financing Borrowers



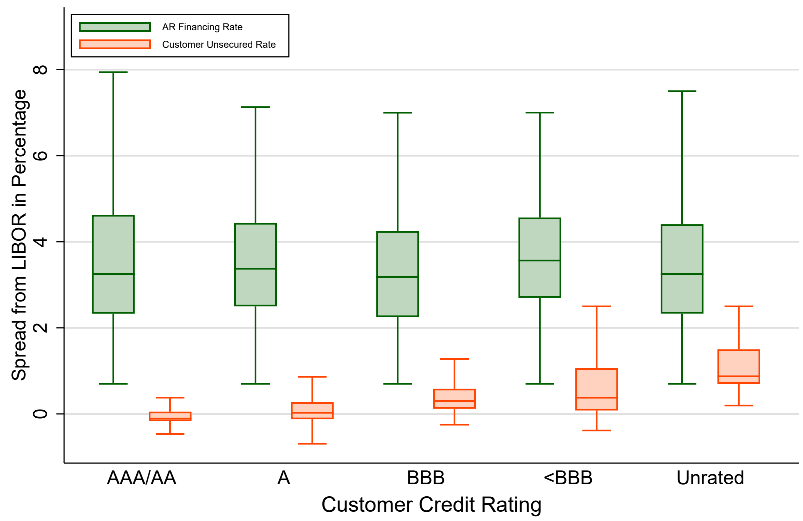
(b) Committed Credit from A/R Financing per Year



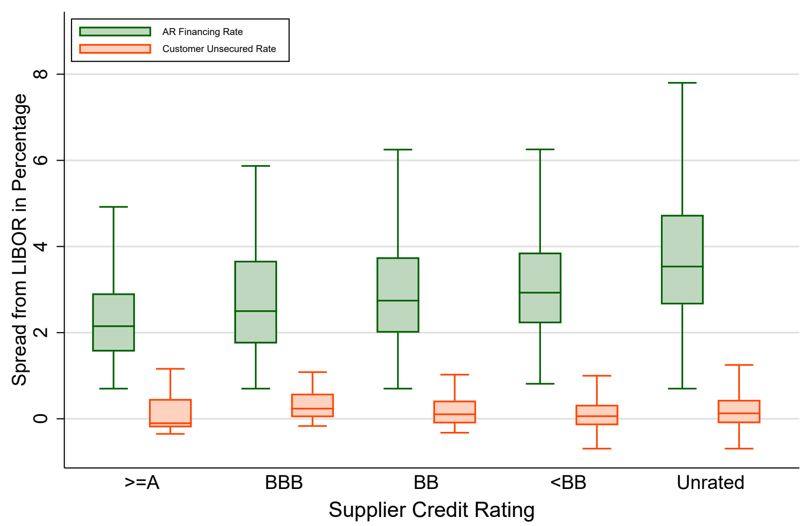
Notes: Panel (a) plots the industry breakdown of all US public non-financial non-utility firms who have used A/R financing during 2000–2020. A/R financing includes accounts receivable based credit lines as well as recourse and non-recourse factoring agreements. Panel (b) plots the number of A/R financing contracts entered each year and the committed credit provided by A/R financing lenders during 2000 − 2020. Number of contracts includes loan originations and amendments to at least one of the three key loan terms: advance rate, interest rate, and credit limit. Committed credit is calculated as the sum of credit limits of all known outstanding A/R financing contracts. I assume a firm has an outstanding A/R financing contract if it is within 2 years after the last loan origination or amendment. This is a conservative measure of total committed credit. Banks are depository institutions identified by Call Reports issuers. All remaining known lenders are classified as non-banks, and loans with unknown lenders are excluded from the calculation of the total committed credit.

Figure 1-4: Interest Rates of A/R Financing and Customer Firms' Unsecured Borrowing Rates

(a) Interest Rate Spreads over LIBOR by Customer Credit Rating



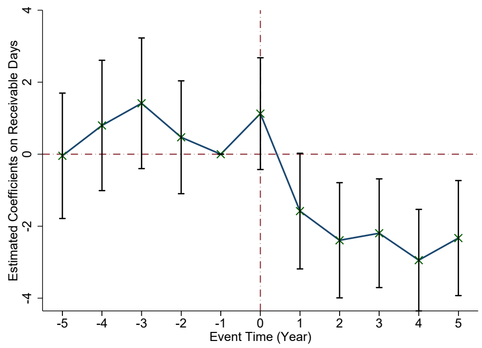
(b) Interest Rate Spreads over LIBOR by Supplier Credit Rating



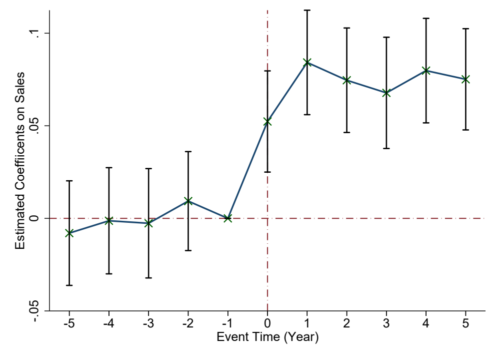
Notes: This figure compares the interest rates a seller pays when using A/R financing versus the interest rates the seller's customers on average pay in unsecured borrowings. Unsecured borrowings include unsecured credit lines and commercial papers. For each seller, I compute its customers' average interest rates of unsecured borrowings, weighted by the sales to each customer. For both A/R financing interest rate and customer average unsecured interest rate, I report the spread over concurrent LIBOR. Panel (a) plots the spreads by the customer firm's credit rating. Panel (b) plots the spreads by the seller firm's credit rating. Both panels are boxplots showing the median (middle horizontal line in boxes), 25% and 75% percentiles (top and bottom of boxes) as well as 25% - 1.5 quartile range and 75% percentile + 1.5 quartile range (two ends of the ticks). Outliers are removed.

Figure 1-5: Dynamics of Firms Before and After A/R Financing

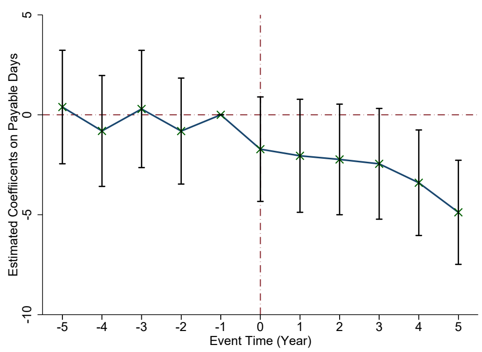
(a) Receivable Days



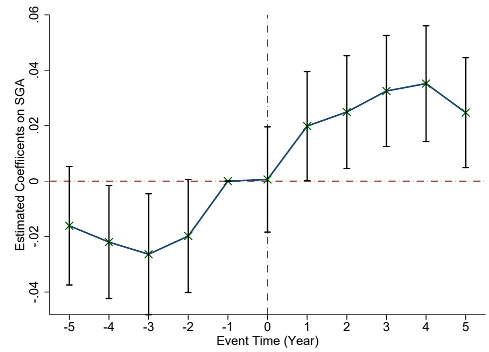
(b) Log of Total Sales



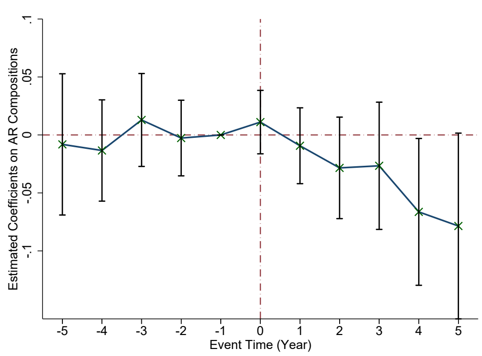
(c) Payable Days



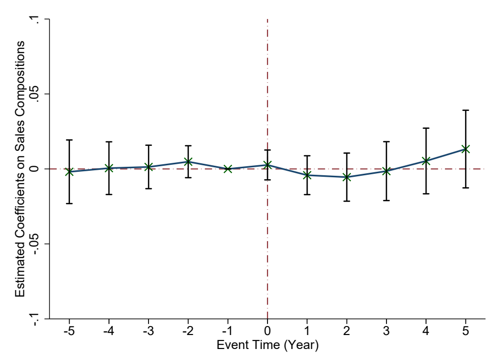
(d) Log of SGA Expense



(e) Composition of Accounts Receivable



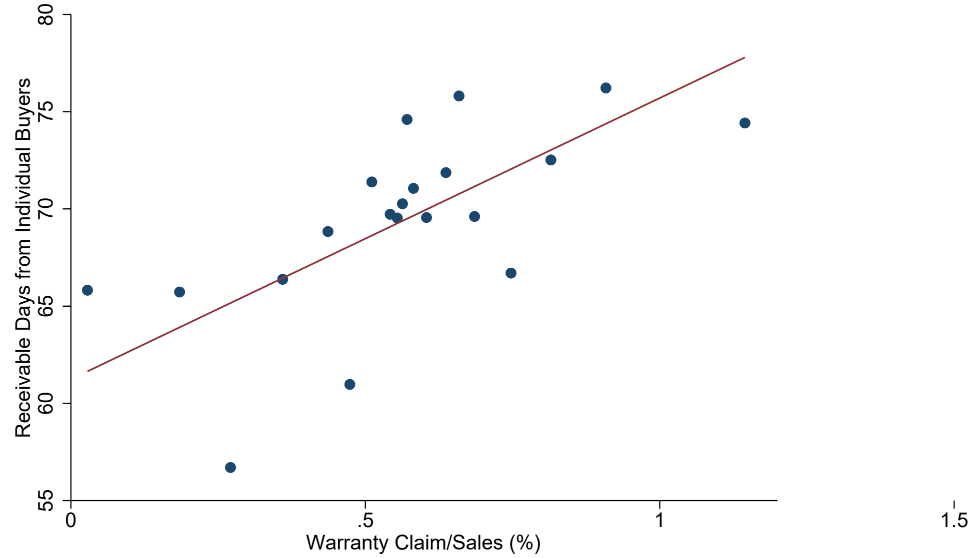
(f) Composition of Sales



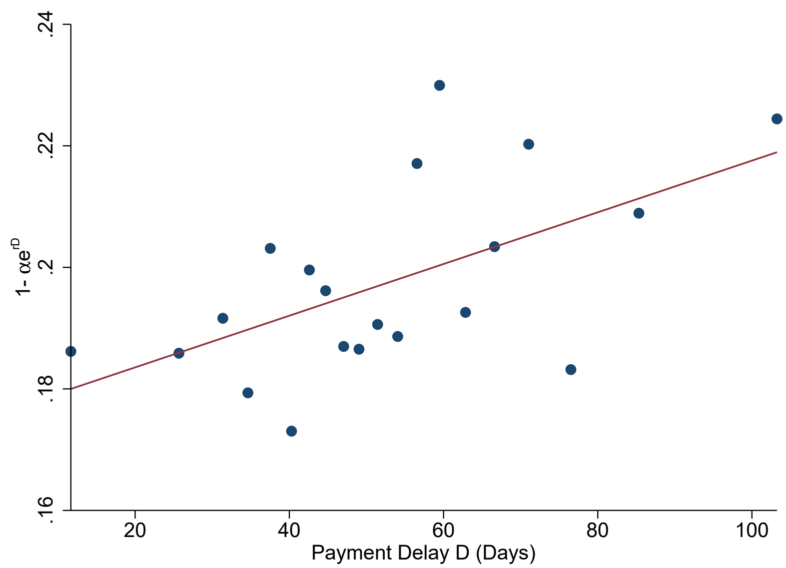
Notes: This figure presents the dynamics of firms before and after a seller borrows from A/R financing for the first time. Panel (a) to (d) plot the OLS coefficients α_k from the following regression: $Y_{it} = \sum_{k=-5,k\neq-1}^{5} \alpha_k \text{AR Financing}_{i,\{t-t_0=k\}} + \text{Controls} + \text{Firm FEs} + \text{Year FEs} + \epsilon_{it}$, where t_0 denotes the first year that firm i borrows from A/R financing. The coefficient with $k = -1$ is excluded as a benchmark category. Receivable days for sellers are calculated as trade accounts receivable (Compustat item RECTR or RECT if RECTR is missing) divided by sales and multiplied by 360. Payable days are calculated as accounts payable (Compustat item AP) divided by purchases (cost of goods sold + change in inventory) and multiplied by 360. Panel (e) and (f) are based on regressions at the seller-buyer pair level. They plot coefficients on the interaction of AR Financing dummy with an indicator of large customer, that is, β_k in regression equation (1.2). In all the panels, I plot 95% confidence intervals. Standard errors in Panel (a) to (d) are clustered at the seller level, while in Panel (e) and (f) are clustered at the seller-buyer pair level.

Figure 1-6: Seller Moral Hazard, Payment Delay, and A/R Financing Terms

(a) Payment delay is increasing in product quality risk

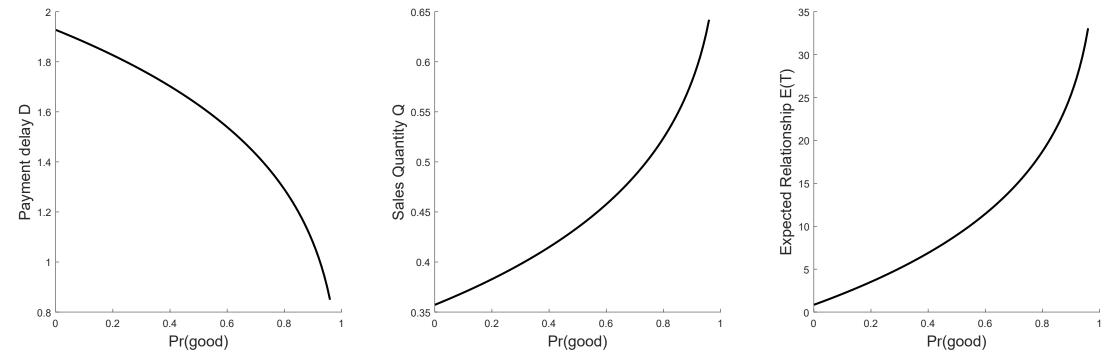


(b) Bank's optimality condition: adjusted haircut is linearly increasing in payment delay



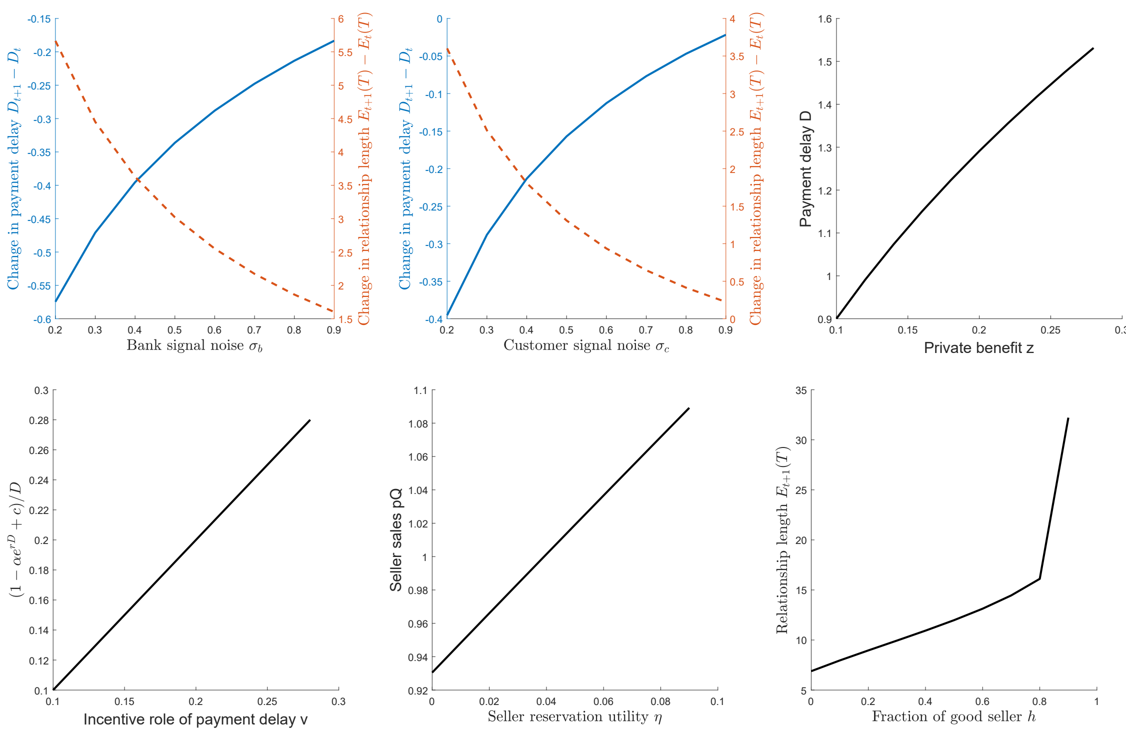
Notes: Panel (a) is a binscatter plot of a seller's receivable days from individual buyers against an industry-level measure of seller's product's risk. Product quality risk is computed as the warranty claims against firms in seller's industry, divided by these firms' sales. I only control for buyer firm fixed effects and year fixed effects. Panel (b) is a binscatter plot of adjusted haircut $1 - \alpha e^{rD}$ against payment delay D in the actual data, a test of the bank's optimality condition. Payment delay D is calculated as the receivable days at the time of borrowing scaled by industry-level ratio of contractual payment delay/receivable days. For the adjusted haircut, α is advance rate and r is interest rate of A/R financing contract. I only control for 2-digit SIC level industry fixed effects, and as in the main empirical analysis, I restrict the sample to borrowers with large customers to exclude the default risk of customers.

Figure 1-7: Equilibrium Quantities when the Fraction of Good Sellers in the Economy Varies



Numerical example: $t = 1$ means 30 days, that is, delivery happens every 30 days. $f(x) = 1 - e^{-x/s}$ where $s = 0.5$. $v = 0.2, z = 0.2, \sigma_b = \sigma_c = 0.5, A = 2, \rho = 0.8, \eta = 0, g = 0.04/12, c = 0.1, \delta = 0.8$. Pr($good$) on the x-axis is the fraction of good sellers in the economy.

Figure 1-8: Comparative Statics with respect to Parameters



Numerical example: $t = 1$ means 30 days, that is, delivery happens every 30 days. I start with following parameters at time t: $f(x) = 1 - e^{-x/s}$ where $s = 0.5$. $v = 0.2, z = 0.2, \sigma_b = \sigma_c = 0.5, A = 2, \rho = 0.8, \eta = 0, g = 0.04/12, c = 0.1, \delta = 0.8$, and $h_t = h = 0.5$. Then I vary one parameter, while fixing the other parameters, and show the model implied quantities at time $t + 1$.

Table 1.1: A/R Financing Data Source, Lender Types, and Borrower Characteristics

A. *Loan types and data source*

Loan Type (Source)	No. Contracts	*Percent %*	Committed Credit ($b)	*Percent %*
AR credit line (hand collected)	4071	73%	322.3	65%
AR credit line (Dealscan)	974	17%	156.5	32%
Recourse factoring (hand collected)	386	7%	13.0	3%
Non-recourse factoring (hand collected)	138	2%	4.8	1%
Total	5569	100%	496.6	100%

B. *Lender types*

Lender Type	No. Contracts	*Percent %*	Committed Credit ($b)	*Percent %*
Bank	1703	31%	175.4	35%
Non-bank	2723	49%	238.1	48%
Lender unknown	1143	21%	83.1	17%

C. *Number of loan contracts and committed credit by firm size*

Firm Size (Assets in $m)	No. contracts	Committed Credit (Mean,$m)	No. firms	Committed Credit (Total, $b)
0-50	1724	8.5	908	11.0
50-250	1680	24.7	907	28.9
250-1000	1154	89.6	622	70.1
1000-5000	751	250.9	401	128.8
5000-	260	755.7	127	110.6
Total	5569		2965	

D. *Percent of borrowers by characteristics*

Borrower ratings	AAA - A	BBB	<BBB	Unrated
	0.3%	2.6%	20.4%	76.8%
Customer ratings	AAA - A	BBB	<BBB	Unrated
	39.3%	7.7%	4.3%	48.6%
Size/customer size	< 5%	5-10%	10-25%	>25%
	84.2%	6.3%	5.1%	4.4%

Notes: In Panel A, I report the loan types and data sources of the A/R financing contracts I collected. For each loan type and data source, I report the number of contracts and total committed credit. Contracts include both originations and amendments to the three loan terms: advance rate, interest rate, and credit limit. I treat multiple tranches within a given loan package as only one contract. To avoid double counting committed credit, if a firm has multiple contracts in the same year, I include only the contract with the largest credit limit. In Panel B, I report the number of loan contracts and the total committed credit written by each type of lenders. In Panel C, I report the number of loan contracts, committed credit by borrower size. In Panel D, I report borrower characteristics. Borrower ratings are borrowers' S&P long term credit ratings. Customer ratings include the borrower's largest customer's S&P long term credit rating. Size/customer size equals a borrower's total asset divided by the median total asset of the borrower's customers. 2965 firms have received A/R financing during 2000 –2020, from 188 banks and 440 non-banks.

Table 1.2: Summary Statistics

A. *A/R financing contracts*	Count	Mean	SD	p10	p50	p90
Committed credit/accounts receivable	4894	1.86	3.56	0.14	0.88	3.41
Advance rate	3146	0.80	0.10	0.75	0.80	0.85
Interest rate (annualized %)	4318	5.90	2.72	2.69	5.50	9.50
Firm age upon first contract	2965	12.95	12.56	1.00	9.00	32.00
B. *Trade credit and sales relationships (seller-buyer-year level)*						
A/R from individual buyers (% of total)	5537	23.28	20.15	6.00	17.00	49.90
Sale to individual buyers (% of total)	4630	23.44	18.72	8.53	17.27	48.00
Receivable days from individual buyers	4636	68.55	66.23	17.59	50.26	132.00
Contractual payment delay (days)	1863	32.46	18.02	10.00	30.00	60.00
C. *Buyer characteristics (industry-year level)*						
Demand elasticity (2000-2020)	1713	0.67	0.19	0.38	0.69	0.92
Demand elasticity (1990-2001)	1719	0.82	0.12	0.64	0.84	0.95
D. *Warranty claims (industry level)*						
Warranty claims/sales (%)	160	0.70	0.50	0.08	0.62	1.51

Notes: This table reports the summary statistics of key variables used in this paper. In Panel A, I report summary statistics on A/R financing contracts. In Panel B, I report summary statistics on trade credit and sales relationships. I only include seller-buyer pairs where both the seller and the buyer are public firms. There are 695 sellers and 527 buyers in the sample where receivable days from individual buyers are available. In Panel C, Demand elasticity (1990-2001) is first constructed at the HS10 product level by the method of Feenstra (1994) and Broda and Weinstein (2006), then aggregated to 4-digit SIC code industry level by total US import trade value of the products. Demand elasticity (2000-2020) is first constructed at HS10 product level using an improved hybrid method of Soderbery (2015), based on 2000–2020 import data from Peter Schott's website (Schott, 2008), then aggregated to 4-digit SIC code industry level. The demand elasticity sample includes 112 manufacturing industries. In Panel D, Warranty claims/sales is 4-digit SIC code industry level average of the warranty claims filed against a firm divided by its sales. The original warranty claims data covers 1038 firms for the time period of 2003 – 2021.

Table 1.3: The Wide Use of A/R financing among US Firms

A. *Full Compustat sample*

Decile (assets)	Smallest	2	3	4	5	6	7	8	9	Largest
Median net trade credit days	96.8	13.6	0.8	-3.0	-5.5	-5.7	-6.7	-3.4	0.3	7.8
(payable days - receivable days)										
Under A/R financing	3%	9%	12%	14%	13%	13%	12%	12%	9%	5%
Have used A/R financing	8%	20%	26%	29%	29%	31%	32%	31%	26%	14%

B. *Firms with customers that have credit ratings*

Decile (trade credit)	Lowest	2	3	4	5	6	7	8	9	Highest
Median net trade credit days	-67.2	-38.3	-25.1	-15.2	-7.3	0.5	8.5	20.0	41.4	137.2
(payable days - receivable days)										
Under A/R financing	17%	20%	21%	22%	26%	20%	19%	18%	15%	12%
Have used A/R financing	38%	42%	43%	47%	53%	48%	44%	38%	32%	26%

Notes: In Panel A, the sample is all Compustat non-financial non-utility firms during 2000–2020. For each year's cross-section, I rank firms by total assets to form deciles, then I report for each decile the median net trade credit days and the fraction of firms that have outstanding A/R financing contract, and the fraction that have ever used A/R financing during 2000–2020. In Panel B, the sample is all Compustat non-financial non-utility firms that have at least one customers with long term credit ratings. I rank firms in each year's cross-section by net trade credit days and form deciles, then I report for each decile the median net trade credit days and the fraction of firms that have outstanding A/R financing contract, and the fraction that have ever used A/R financing during 2000–2020.

Table 1.4: Debt Instruments of A/R Financing Borrowers and Matched Non-Borrowers

Panel A: A/R Financing Borrowers

	Debt Type	N firms	Percent of firms
Other Debt	Capital Lease	101	26.4%
	Corporate Convertible	17	4.5%
	Corporate Debentures	15	3.9%
	Foreign Currency Debenture	1	0.3%
	Other Loans or Borrowings	9	2.4%
	Preferred Security	6	1.6%
	Preferred Stock	1	0.3%
	Revolving Credit	75	19.6%
	Term Loans	55	14.4%

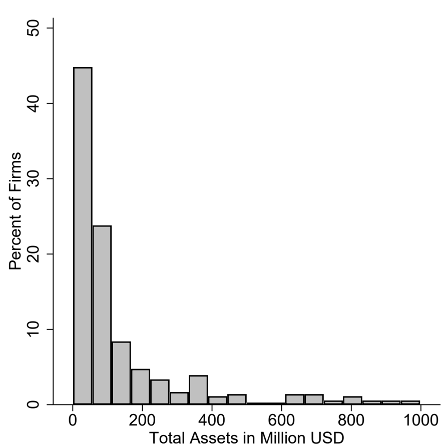
Panel B: Matched Non-A/R Financing Borrowers

	Debt Type	N firms	Percent of firms
No Debt		335	72.2%
Has Debt	Capital Lease	111	23.9%
	Corporate Convertible	16	3.4%
	Corporate Debentures	8	1.7%
	Credit Facility	4	0.9%
	Other Loans or Borrowings	12	2.6%
	Preferred Security	5	1.1%
	Preferred Stock	1	0.2%
	Revolving Credit	48	10.3%
	Term Loans	60	12.9%

Notes: In Panel A, I report the debt instruments used by a random sample of 381 A/R financing borrowers (15% random sample of all borrowers with less than $1 billion total asset). For each borrower, I match 2 non-borrowers in the same 2 digit SIC industry and of similar size and cash balance. This constitutes a sample of 465 matched non-borrowers. In Panel B, I report the debt instruments used by these non-borrowers. The median leverage of A/R financing borrowers is 18.6%, and is 10.8% for non-borrowers.

Figure: Size Distribution of Borrowers and Non-Borrowers in Table 1.4

(a) A/R Financing Borrowers



(b) Matched Non-A/R Financing Borrowers

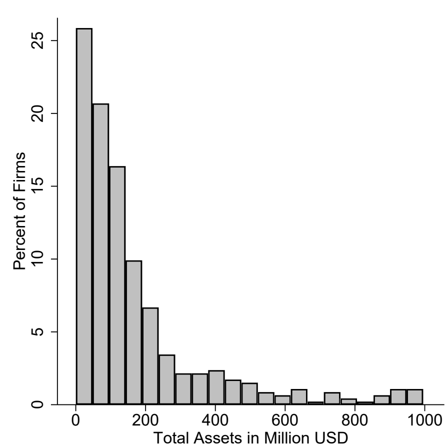


Table 1.5: Payment Delay Is Increasing in Product Quality Risk

Dep Var: **Seller Receivable Days**	Seller-Level		Seller-Buyer Level			
	(1)	(2)	(3)	(4)	(5)	(6)
Warranty Claim/Sales(%)	7.096***	7.165***	12.399**	13.596**	11.163**	11.491*
	(4.640)	(4.455)	(2.411)	(2.074)	(2.131)	(1.741)
log(Total Asset)		1.503***			-0.261	-0.050
		(6.215)			(-0.252)	(-0.034)
log(Seller Age)		-1.545***				
		(-3.283)				
log(Relationship Length)					-3.013*	-1.086
					(-1.696)	(-0.465)
Leverage		-4.375*			-2.762	-4.187
		(-1.716)			(-0.414)	(-0.538)
EBITDA/Total Asset		1.815***			-18.630***	-18.808*
		(6.232)			(-2.745)	(-1.866)
N	67530	63864	2759	1651	2744	1639
Adj. R^2	0.096	0.119	0.190	0.053	0.200	0.057
N Sellers	8030	8030	427	427	425	425
N Buyers			216	216	215	215
Seller Industry × Year FE	Yes	Yes				
Buyer FE			Yes		Yes	
Year FE			Yes		Yes	
Buyer × Year FE				Yes		Yes

Notes: In this table, I regress a seller's receivable days on a measure of product quality risk: Warranty Claim/Sales. I calculate Warranty Claim/Sales by first dividing the warranty claims filed against sellers in each 4-digit SIC industry by sales of these sellers, and then taking the average at 4-digit SIC industry level. In Column (1) and (2), receivable days for sellers are calculated as total trade accounts receivable divided by sales and multiplied by 360. Here, I control for seller's 2-digit SIC code industry times year fixed effects, and standard errors are clustered at the seller's industry level. In Column (3) − (6), receivable days for sellers are calculated as trade accounts receivable from a individual buyer divided by sales to that buyer and multiplied by 360. Standard errors are clustered at the buyer level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: The Screening and Monitoring Role of A/R Financing

Panel A: Evidence of A/R lender's screening and monitoring role

	Receivable Days					Payable Days	log(Sales)	log(SGA)
Post	-2.798***	-2.860***	-5.969***	-1.792*	1.100	-1.131	0.044***	0.024**
	(-3.23)	(-3.26)	(-4.79)	(-1.83)	(0.65)	(-0.81)	(2.63)	(1.97)
Post × Agriculture /Mining Industry		1.303						
		(0.42)						
Post × Age			0.199***					
			(4.88)					
Post × Big Lender				-3.830*				
				(-1.88)				
Post× Lender Experience					-0.898*			
					(-1.67)			
N	16605	16605	15850	16605	11382	15680	16619	16004
Adj. R^2	0.584	0.584	0.576	0.585	0.571	0.559	0.964	0.972
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Robustness check: no composition change in A/R or sales

	Percent of Sale	Percent of AR
Large Customer × Post	-0.003	-0.019
	(-0.31)	(-0.67)
N	6643	978
Adj. R^2	0.696	0.435
Seller×Buyer FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes

Notes: This table reports the difference-in-differences regressions that estimate changes in key outcome variables with respect to A/R financing borrowings. In Panel A, regressions are at the seller-year level. In Panel B, regression is at the seller-buyer-year level. Post is a dummy that takes the value of one if the time is after when the seller first borrows from A/R financing. Agriculture/Mining Industry is a dummy that takes value of one if seller has SIC code below 1499. Age is the seller's age when it first borrows from A/R financing. Big Lender is a dummy that takes the value of one if seller's A/R lender made loans to more than 30 borrowers (the dummy equals one for approx. 25% of lenders). Lender Experience is measured as log of the number of borrowers the seller's A/R lender made loans to. Large Customer (in Panel B) is a dummy that takes the value of one if a buyer is above the median size of a seller's customers. Control variables include size, leverage, cash balance (quick ratio), cash flow (EBITDA/total assets), and market to book ratio (Tobin's Q). I restrict the sample to before and within 5 years after a seller borrows from A/R financing for the first time. Standard errors are clustered at the firm level in Panel A, and at the seller - buyer pair level in Panel B. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: Comparing A/R Financing Borrowers with Non-Borrowers of Similar Financing Needs

	Treat			**Control**			**Diff.**	t-stat
	Mean	Variance	Skewness	Mean	Variance	Skewness		
A. *Variables to match on*								
Age	5.04	3.04	0.31	5.00	3.04	0.31	0.04	0.44
log(Total Assets)	4.56	2.62	0.32	4.60	2.34	0.31	-0.05	-0.52
Receivable Days	67.18	1104.30	2.01	65.59	845.54	1.46	1.59	0.97
Payable Days	84.26	18152.46	6.02	78.22	14497.54	6.70	6.04	0.95
Propensity Score	0.19	0.01	-0.17	0.19	0.01	-0.12	0.00	0.60
B. *What prevent non-borrowers from getting A/R financing?*								
Uncollectible AR/Sales	0.01	0.00	10.53	0.01	0.00	22.48	0.00	-1.20
Customer HHI	0.60	0.02	-0.36	0.61	0.03	-0.22	-0.01	-1.04
log (Customer Total Asset)	9.67	3.16	-0.96	9.51	3.59	-0.91	0.15	1.45
C. Screening/monitoring role of A/R financing								
Relationship Years	2.98	11.82	1.81	2.30	7.99	2.22	0.68^{***}	3.76
log(Sales)	4.53	2.94	0.30	4.18	3.24	0.00	0.35^{***}	3.58
Sales/Assets	1.26	0.60	1.85	0.94	0.41	1.60	0.31^{***}	8.17
N Firms	704			776				

Notes: In this table, I perform a matched sample comparison between A/R financing borrowers and non-borrowers of similar financing needs. I restrict both borrowers and non-borrowers to firms younger than 10 years so that lenders learn little from a seller's past interaction with buyers. For each A/R financing borrower with less than 1 billion total assets, I match 2 non-borrowers in the same industry, and are closest to the borrower in terms of Mahalanobis distance of age, log(total assets), receivable days, payable days and propensity score of borrowing. Propensity Score is calculated based on the logit model in Table A.4, and measures a firm's financing need. Panel A shows the distribution of the variables that the matching is based on. Panel B shows the distribution of variables on customer characteristics that may prevent a firm from getting A/R financing. Panel C shows that sellers who receive A/R financing have longer relationships with customers and have higher sales.

Table 1.8: Buyers with Lower Product Demand Elasticity (Higher Market Power) Pay Sellers Slower

Dep Var: **Seller Receivable Days**	(1)	(2)	(3)	(4)
Buyer ρ (2000-2020)	-45.22***	-49.51**	-84.77***	-61.15**
	(-2.81)	(-2.69)	(-3.35)	(-2.17)
N	1376	1376	499	499
Adj. R^2	0.344	0.348	0.251	0.273
Buyer ρ (1990-2001)	-118.0**	-119.4**	-180.5***	-142.5***
	(-2.15)	(-2.12)	(-3.49)	(-3.01)
N	1376	1376	499	499
Adj. R^2	0.343	0.345	0.257	0.284
N Sellers	222	222	68	68
N Buyers	170	170	104	104
Seller FE	Yes	Yes		
Year FE		Yes		
Seller × Year FE			Yes	Yes
Controls				Yes

Notes: In this table, I regress a seller's receivable days from individual customers on the customers' product market's demand elasticities. The demand elasticity $\rho \in [0, 1]$ is estimated from US import data at HS10 product code level, then aggregated to 4-digit SIC code industry level following Alfaro et al. (2019). ρ (1990-2001) uses the demand elasticity data of Feenstra (1994) and Broda and Weinstein (2006), based on 1990–2001 US import data. ρ (2000-2020) is estimated using an improved hybrid method of Soderbery (2015), based on 2000–2020 US import data from Peter Schott's website (Schott, 2008). Demand elasticity is only available for manufacturing industries given the scope of US import data. Controls include customer's log total assets and credit ratings. Standard errors are double clustered at customer's industry and year level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.9: Summary of the Parameterization of the Model

Panel A. Estimated parameters

Parameter	Description	↑ parameter leads to
v	returned goods per unit of time	↑ $\widehat{haircut}$ /payment delay
z	private benefit of bad seller	↑ payment delay
δ	probability that seller has additional cash flow	↑ spread between A/R financing rate and customer unsecured rate
σ_b	noise of bank's signal	↓ decline of payment delay after A/R financing, ↓ expected relationship length of borrowers
σ_c	noise of customer's signal	↓ decline of payment delay over time
h	fraction of good sellers in economy	↑ expected relationship length
η	seller's reservation utility	↑ seller's revenue

Panel B. Calibrated parameters

Parameter	Description	Value
ρ	customer's elasticity of demand	measured from US import data
A	demand shock	measured from customer's Revenue and COGS

Notes: This table summarizes all the model parameters. Panel A lists the parameters that I estimate using Generalized Method of Moments. It reports how each parameter is identified by the data. Panel B lists the parameters that are directly measured from the data.

Table 1.10: GMM Estimation Results of Model Parameters

Parameter	Explanation	Value	95% CI
v	Returned goods per unit of time	0.265***	[0.253, 0.276]
		(44.65)	
z	Private benefit of seller	0.334***	[0.175, 0.494]
		(4.12)	
δ	Probability that seller has additional cash flow	0.675***	[0.660, 0.689]
		(90.12)	
σ_b	Noisiness of bank's signal	0.472**	[0.069, 0.874]
		(2.30)	
σ_c	Noisiness of buyer's signal	0.703***	[0.355, 1.051]
		(3.96)	
h	Fraction of good type seller	0.328**	[0.362, 0.620]
		(2.20)	
η	Seller reservation utility	0.0592***	[0.0543, 0.0641]
		(23.90)	

Notes: This table reports the parameter estimates for the model. t statistics based on standard errors clustered at the firm level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.11: Fitness of the Model

	Model	Data
Adjusted haircut $1-\alpha e^{rD}$	25.30%	18.60%
Interest rate r	6.50%	5.80%
Payment delay D (days)	50.01	51.70
Relationship length T (years)	2.76	2.89

Notes. To show the fitness of the model, this table reports the mean of variables of interest implied by the model and the mean in the data. To calculate model implied quantities, I first calculate the implied probability of each seller being the good type in each year, given the parameter estimates. Then I compute the equilibrium quantities implied by the model that depend on such probabilities. And finally I take the average over the estimation sample after winsorizing at 2% tails to remove outliers.

Table 1.12: Effects of Counterfactual Policies

	Model Predicted	Counterfactual 1	Counterfactual 2
Payment delay D (days)	50.01	30.00	52.05
Production quantity Q (\$mil)	64.15	47.26	66.79
Seller gross profit $(p-1)Q$ (\$mil)	43.36	29.30	44.12
Prob(bad product)	5.2%	17.9%	5.0%
Prob(seller default)	15.6%	10.3%	15.7%
Government expense (\$mil)	0.00	0.00	1.57

Notes: This table shows the results of two counterfactual analyses. In Counterfactual 1, I restrict the payment delays to be within 30 days. In Counterfactual 2, I introduce government subsidy that lowers the A/R lender's cost of fund by 5% per year. Model predicted numbers are based on the parameter estimates combined with the annual production of the median US public firm, whose annual COGS is 64.15 million USD, taken as the analogy of Q of the model. The numbers under counterfactual columns are also for the median US public firm.

Chapter 2

The Case for Convenience: How CBDC Design Choices Impact Monetary Policy Pass-Through

> *"If all a CBDC did was to substitute for cash – if it bore no interest and came without any of the extra services we get with bank accounts – people would probably still want to keep most of their money in commercial banks."*
> *— Ben Broadbent, Deputy Governor of the Bank of England, in a 2016 speech*

A central bank digital currency (CBDC) "is a digital payment instrument, denominated in the national unit of account, that is a direct liability of the central bank" (BIS, 2020). Over the last few years, interest in CBDC has grown to the point where at present 90 percent of central banks are investigating options for introducing CBDCs (BIS, 2021). As indicated in Broadbent's remarks (in the epigraph), policymakers initially contemplated a CBDC that duplicated features of cash, without adding design characteristics that would make it more likely to compete with money issued by commercial banks – the so called disintermediation problem. However, more recently central banks have taken a broader view, and have been more open to the possibility that CBDCs can help them to fulfill their mandates, either in the present or the future. Central banks are increasingly viewing CBDCs as a way to improve the payment system, promote financial inclusion, enhance monetary policy transmission, and reduce systemic risk (BIS, 2020).

The likelihood that a CBDC will achieve any of the desired central bank objectives depends upon its design features and how they interact. CBDCs can offer both pecuniary benefits, in the form of interest payments, and nonpecuniary benefits. These can include a host of features that enhance the performance of CBDC as a medium of exchange. Examples include the quality of the user interface, processing speed, privacy and access to markets.[1] We lump these possibilities together under the heading "payment convenience."

In this paper, we explore the implications of introducing a CBDC that is interest-bearing and offers payment convenience. All else equal, consumers will prefer to hold a currency that pays higher interest. However, if a currency is easier to use, or accepted at more places, then these non-pecuniary benefits may offset interest payments. Hence both interest rate *and* payment convenience are important choice parameters for a CBDC, as both will determine consumer demand for CBDC and hence its ultimate impact on monetary policy objectives.

We seek to evaluate design choices for CBDC in a model that is descriptive of the current US financial system, which is characterized by large excess reserves and in which the main monetary policy variable is interest on reserves (IOR).[2] Crucially, we also want our analysis to capture heterogeneity in bank size. Bank size matters for two key reasons. First, bank size impacts the cost basis of loans, as it determines

[1]A publicly provided CBDC could be less expensive to use and more widely accessible than existing, privately offered payment methods, and it could offer access to new platforms and services. See, for example, the Bank of England discussion paper on CBDC (Bank of England, 2020) which describes the potential for third-party payment interface providers to provide overlay services on top of CBDC balances. A CBDC could also provide privacy. Policy makers have argued that the central bank is specially positioned to provide privacy in payments because the central bank does not have a profit motive to exploit consumer payment data (Lagarde, 2018).

[2]In the United States, IOR has been paid since October 2008. Since the financial crisis of 2008-09, interest on excess reserves (IOER) has become the Federal Reserve's main policy tool to adjust interest rates. In July 2021, the Federal Reserve renamed IOER to interest on reserve balances (IORB), as required reserves are currently zero. For simplicity, we use the acronym IOR.

the likelihood of retained reserves. Loan issuance involves the creation of deposits. When these deposits are spent by the borrower there is a chance that the recipient will belong to the same bank, in which case there is no associated transfer of reserves to another bank. This likelihood is not negligible for the largest banks in the US economy and hence the impact of retained reserves should not be ignored.[3] Second, large bank deposits may offer a higher convenience value than small bank deposits. For example, a large bank could have a more expansive network of branches and ATMs, a better mobile App, or a wider range of other (unmodeled) services. The higher convenience value of its deposit may allow the large bank to offer a lower deposit rate than the small bank and yet still maintain a larger market share.

In our model, the CBDC is offered through commercial banks. While CBDC balances are the direct liability of the central bank, we envision that commercial banks will act as the central bank's agents to conduct KYC (Know Your Customer) and AML (Anti-Money Laundering). This "tiered" design of CBDC is consistent with the recent pilot of E-Krona in Sweden, the CBDC experiment in China, and the Banking for All Act in the United States.[4]

The heterogeneous-bank model we develop is new. The first part of our analysis seeks to validate the model by demonstrating that it explains aspects of US deposit markets that are not explained by existing, homogeneous-bank models of the US economy with large reserves. In particular, we are able to explain the observed lack

[3]During 2010-2020, based on Call Reports data, the top four largest banks captures 35% of US deposit market. Their deposit market shares are large *and* stable over the last decade, with the averages listed in the following: Bank of America (11%), Chase (10%), Wells Fargo (10%), and Citi (4%). Given the stationarity of deposit shares, and the fact that deposits are created through commercial banks making loans (Bank of England Quarterly Bulletin, 2014), the fraction of a lent dollar that ultimately flows to each bank should approximate the deposit shares, regardless of the number of transactions or transfers. Hence the high deposit market shares of the largest banks have non-negligible impact on their opportunity costs of making loans. In local deposit markets, concentration is also salient. Drechsler, Savov, and Schnabl (2017) measure this concentration by Herfindahl index (HHI), calculated by summing up the squared deposit-market shares of all banks that operate in a given county. Their calculation indicates that around 50% of US counties have an HHI that is higher than 0.3 (at least one bank's deposit share is greater than 30%), and 25% of US counties have an HHI that is higher than 0.5 (at least one bank's deposit share is greater than 50%).

[4]The Act argues that Digital Dollar Wallets should provide a number of auxiliary services including debit cards, online account access, automatic bill-pay and mobile banking. These features (in particular mobile banking which could give access to a variety of platforms that customers of a particular bank might otherwise not have access to) could result in a CBDC with its own convenience value. Similar provisions are outlined in the ECB's digital euro report (2020).

of interest-rate pass-through in US deposit markets. Interest rate pass-through in the US economy is far from complete. Drechsler, Savov, and Schnabl (2017) find that "[f]or every 100 bps increase in the Fed funds rate, the spread between the Fed funds rate and the deposit rate increases by 54 bps." Duffie and Krishnamurthy (2016) document a sizable dispersion of a broad range of money market interest rates, which widened as the Fed raised its interest on reserves. Our analysis shows that the low correlation between movements in policy rates that determine the fed funds rate and movements in deposit rates can be partly attributed to the differential impact these changes have on banks of different sizes.

The intuition behind this result is as follows. The large bank's ability to offer a lower rate than the small bank may place them at a "corner solution" introduced by the fact that deposit rates cannot go below zero. The large bank will set a deposit rate of zero, that is non-responsive to changes in the policy rate, until that rate rises to a level where the zero lower bound on deposit rates is no longer binding. We will show that the zero lower bound on deposit rates binds only when IOR rate is low. Hence for low levels of IOR we expect deposit rates set by a large share of the banking sector to be non-responsive to changes in IOR. For high levels of IOR, the lower bound is not binding and we expect all banks to adjust deposit rates in response to changes in IOR (or the federal funds rate if this becomes the relevant opportunity cost of lending). We demonstrate that both of these model predictions seem to be true empirically.

In light of these observations regarding weak pass-through for low IOR rates, we examine how outcomes in both the deposit market (deposit shares at the different banks and deposit rates) and the lending market (loan volumes and rates) may be impacted by the choice of design characteristics of a CBDC. First, we vary the CBDC interest rate. Increasing the CBDC interest rate while holding the IOR rate and the CBDC convenience value fixed raises the deposit rates of both banks, thus bringing their weighted average closer to the IOR rate. Setting the CBDC interest rate equal to the interest rate on reserves would result in full monetary policy pass-through. However, by forcing both banks to raise interest rates, a higher CBDC interest rate makes it more difficult for the small bank to compete with the large bank by offering higher deposit rate. Thus, a higher CBDC interest rate reduces the market share of the small bank in deposit and lending markets, further widening the large bank-small bank gap.

Second, we vary the CBDC convenience value, holding the IOR rate and CBDC interest rate fixed. Making the CBDC more convenient weakens the market power of the large bank by narrowing the convenience gap between the two banks. For example, by hosting a convenient CBDC, a small community bank partially "catches up" with large global banks in offering payment functionalities. The most immediate implication is that a convenient CBDC results in a lower deposit rate at the small bank, because the small bank does not have to compensate depositors as much for forgoing the large bank's convenience. The deposit rate at the large bank initially remains unchanged, and hence the average deposit rate for the market falls as convenience is increased from zero. However, as convenience rises, a point is eventually reached where the large bank is no longer constrained by the zero lower bound and starts to raise its deposit rate to compete with the small bank. The result is an increase in the average deposit rate. The implication is that for any given level of IOR, pass-through of IOR to deposit rates is reduced for low levels of convenience and increased for high levels of convenience.

Finally, we address the issue of how a given CBDC design impacts the sensitivity of deposit rates to changes in the IOR rate. In the equilibrium where the lower bound on deposit rates is not binding, pass-through is complete regardless of the levels of the CBDC interest rate or convenience value. In equilibria where the lower bound is binding, we establish two results that hold under some distributional assumptions. First, within the constrained equilibrium, where the large bank's deposit rate is non-responsive to changes in the IOR rate, the response of the small bank's deposit rate to increases in the IOR rate decreases as the CBDC interest rate increases and increases as convenience increases. Second, higher levels of the CBDC interest rate increase the range of IOR rates for which the large bank's deposit rate is non-responsive and higher levels of convenience decrease the range. Thus, a positive interest on CBDC necessarily weakens monetary policy transmission from IOR to deposit rates when the IOR rate is low, while increasing convenience necessarily increases monetary policy transmission.

The paper is organized as follows. Section 2.1 provides a literature review. Section 2.2 introduces the model and characterizes the constrained and unconstrained equilibrium. There is a critical level of IOR rate at which the economy transitions from the the constrained to unconstrained equilibrium. Hence, this analysis provides an explanation for weak pass-through of IOR rate to deposit rates at low levels of IOR and strong pass-through at high levels. Section 2.3 separately evaluates the direct impact on deposit rates of increasing the CBDC interest rate s and convenience level v on deposit rates. Section 2.4 examines the sensitivity of deposit rates to changes in the IOR rate under different CBDC designs. Section 2.5 concludes.

2.1 Literature

Our work builds on previous literature that has modelled deposit and lending markets in the current regime of large excess reserves. In Martin, McAndrews, and Skeie (2016), a loan is made if its return exceeds the marginal opportunity cost of reserves, which can be either the federal funds rate or the IOR rate, depending on the regime. Our model differs in that we have multiple banks and hence lent money may return to the same bank as new deposits. Hence, our opportunity cost of lending is lower. Nevertheless, we share the conclusion that the aggregate level of bank reserves does not determine the level of bank lending.

There is now a growing literature that seeks to examine the impact of CBDC on deposit and lending markets. The conclusions vary and depend upon the level of competition, the interest rate on the CBDC, and other features (e.g., liquidity properties of CBDC and reserve requirements). Keister and Sanches (2021) consider a competitive banking environment in which deposit rates are determined jointly by the transactions demand for deposits and the supply of investment projects. If the CBDC serves as a substitute for bank deposits, then its introduction causes deposit rates to rise, and the levels of deposits and bank lending to fall.

In contrast, if banks have market power in the deposit market, the introduction of a CBDC does not disintermediate banks, as banks can prevent consumers from holding the CBDC by matching its interest rate. This lowers their profit margin, but does not lower the level of deposits, and may even increase it. This is true in the model proposed by Andolfatto (2021), where the bank is a monopolist. In that paper, an interest bearing CBDC causes deposit rates to rise and the level of deposits to increase. Likewise, in that paper, banks have monopoly power in the lending market, and, as in Martin, McAndrews, and Skeie (2016), lending is not tied directly to the level of deposits, Hence, a CBDC does not impact the interest rate on bank lending or the level of investment.

Chiu, Davoodalhosseini, Jiang, and Zhu (2019) also consider banks with market power and show that an interest-bearing CBDC can lead to more, fewer or no change in deposits, depending on the level of the CBDC interest rate. In an intermediate range of rates, the CBDC impacts the deposit market in a manner similar to Andolfatto (2021) in that banks offer higher deposit rates and increase deposits. Since, similar to Keister and Sanches (2021), lending is tied to the level of deposits, adding the CBDC results in increased lending.

Our work is closest to Andolfatto (2021). We do not specify the overlapping generations framework that he uses to make money essential. However, like Andolfatto (2021), in our model, reserves are abundant, lending is determined by a performance threshold, and banks have monopoly power in lending market. Hence, lending is determined not by deposit levels, but instead by the opportunity cost of funds. In our model, this opportunity cost is lower than the IOR rate, since we allow for the realistic feature that reserves come back to the lending bank with a probability that depends on the deposit market share. Unlike Andolfatto (2021), and the other works mentioned above, we incorporate two key design aspects of CBDC, interest rate and convenience value, and we examine the combined impact these features have on market outcomes in an environment with heterogeneous banks.

The impact of adding a CBDC can be richer in the presence of other frictions. In a model with real goods and competitive banks, Piazzesi and Schneider (2020) find that the introduction of CBDC is beneficial if all payments are made through deposits and the central bank has a lower cost in offering deposits. However, they also find that the CBDC can be harmful if the payer prefers to use a commercial bank credit line, but the receiver prefers central bank money. Parlour, Rajan, and Walden (2022) argue that a wholesale CBDC that enhances the efficiency of interbank settlement system could exacerbate the asymmetry between banks if the CBDC does not distinguish net-paying and net-receiving banks. Agur, Ari, and Dell'Ariccia (2022) consider an environment where households suffer disutility from using a payment instrument that is not commonly used. They examine trade-offs faced by the central bank in preserving variety in payment instruments and show that the adverse effects of CBDC on financial intermediation are harder to overcome with a non-interest-bearing CBDC.

Fernández-Villaverde, Sanches, Schilling, and Uhlig (2021) extend the analysis of CBDC to a Diamond and Dybvig (1983) environment in which banks are prone to bank runs. In this setting, the fact that the central bank may offer more rigid deposit contracts allows it to prevent runs. Since commercial banks cannot commit to the same contract, the central bank becomes a deposit monopolist. Provided that the central bank does not exploit this monopoly power, the first-best amount of maturity transformation in the economy is still achieved.

Brunnermeier and Niepelt (2019) and Fernández-Villaverde, Sanches, Schilling, and Uhlig (2021) derive conditions under which the addition of a CBDC does not affect equilibrium outcomes. Key to their result is the central bank's active role in providing funding to commercial banks in order to neutralize the CBDC's impact on their deposits.

2.2 Model and Equilibrium

2.2.1 Setup

The economy has a large bank (L) and a small bank (S).[5] There are $X = X_S + X_L$ reserves in the banking system, where X_S denotes the reserve holding of the small bank and X_L denotes the reserve holding of the large bank.[6] For simplicity, the banks start off holding reserves as their only asset, balanced by exactly the same amount of deposits. Following Martin, McAndrews, and Skeie (2016), we assume that the level of reserves X is exogenously determined by the central bank and is assumed to be large. The central bank pays the two commercial banks an exogenously determined interest rate f on their reserve holdings, which is called interest on reserves (IOR). The large and small banks pay depositors endogenously determined deposit rates r_L and r_S, respectively. Thus, if nothing else happens, bank j's total profit would be $X_j(f - r_j)$.

Commercial bank deposits are valuable not just for the interest they pay, but also for the payment services they provide, the benefit of which we refer to as convenience value. The convenience value of deposits in the small bank is normalized to be zero. The convenience value of deposits in the large bank is a random variable $\delta \geq 0$ that has the twice differentiable, cumulative distribution function G. We make the following assumption on G that we impose throughout the paper:

Assumption 2.1 *The function G satisfies $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$ for any $\delta \in [0, f - s + v]$.*

This condition ensures the second order condition of a bank's optimization problem is satisfied. The condition is also used in the comparative statics analysis. Bounding the curvature of G bounds the masses of agents who value large bank's deposits highly and lowly, and ensures that both banks will compete to win additional depositors by

[5] The assumption of two banks is, of course, a simplification. However, the situation may accurately describe the retail depositors' decision making process. Using survey data, Honka, Hortaçsu, and Vitorino (2017) find that US consumers were, on average, aware of only 6.8 banks and considered 2.5 banks when shopping for a new bank account. More than 80% considered fewer than 3 banks when shopping for a new bank account.

[6] We normalize the size of an individual loan to be \$1, so reserves are in units of the standard loan size. For example, if a loan size is \$1 million and the actual reserve is \$1 trillion, then in our model, X is interpreted as $\$1\,trillion/\$1\,million = 10^6$.

raising their deposit rates when f and other parameters, which we introduce in the following paragraphs, change. Each depositor in the economy draws their large-bank convenience value δ independently from the distribution G. This process reflects the idea that enhanced payment services are not valued the same by all depositors.

The central bank offers a "retail" CBDC that is universally available. The CBDC has two features: it pays an interest rate of s to depositors who use it and it provides a per-dollar convenience value $v \geq 0$ to users that is the same across all depositors.[7] The convenience value can be interpreted as a benefit that depositors receive from transacting using central bank money. This benefit can include access to platforms on which CBDC can be spent, aspects of the mobile user interface (app features) or any other account services that are associated with central bank accounts.

We assume that the CBDC is offered via commercial banks, and that money can be transferred seamlessly between a depositor's deposit account at a commercial bank and their CBDC account offered via the commercial bank.[8] Because a depositor can transfer money between her deposit account and her CBDC account at no cost, she can obtain a convenience value in payments that is equal to the maximum of the two options. A depositor at the large bank receives convenience value $\max(\delta, v)$ and a depositor at the small bank receives convenience value $\max(0, v) = v$. The convenience value v acts as a lower bound on the payment convenience obtained by all depositors and thus narrows the gap between payment convenience levels that depositors receive across banks of different sizes.

There is a unit mass of agents, and each potentially plays three roles: entrepreneur (borrower), worker, and depositor. The main heterogeneity among the agents is their

[7]The uniform nature of CBDC convenience value reflects the idea that CBDC should ideally create no discrimination. This is also without loss of generality. If v varies across people, let's say $v = \bar{v} + \tilde{v}$, where $\bar{v}$ is the average convenience value that can be adjusted by the central bank, and $\tilde{v}$ represents the individual deviation from the average, we would only need to let G describe the distribution of $\delta - \tilde{v}$.

[8]The Chinese CBDC experiment pivots around the e-CNY wallet mobile phone app. Embeded in the app are interfaces connecting to deposit accounts at eight authorized commercial banks, as well as AliPay and TenPay. Users can transfer money seamlessly between the deposit accounts and the e-CNY wallet, with just a click, and make payments from the e-CNY wallet. The transfer incurs no fee. The CBDC launched in Nigeria through the eNaira wallet app has the same characteristics. Chiu et al. (2019) also assume that CBDC and deposits are perfect substitutes in terms of payment functions.

convenience value for large bank deposits.

The model has four periods. At $t = 0$, the commercial banks set the deposit rates r_L and r_S. The central bank sets the interest on reserves rate f, the CBDC interest rate s, and the CBDC convenience value v. In the model, f, s, and v are exogenous, and r_L and r_S are endogenous. At the start of the model, a fraction m_L of agents have existing deposits at the large bank and a fraction $m_S = 1 - m_L$ of agents have existing deposits at the small bank. The amount of deposits per capita across agents is identical. This means $m_L = X_L/X$ and $m_S = X_S/X$. Because users can seamlessly transfer money between the CBDC account and the deposit account, a CBDC account offered via a commercial bank effectively provides the same convenience value as the deposit account of that commercial bank. For this reason, the CBDC interest rate s is a lower bound on banks' deposit rates, i.e., $r_L \geq s$ and $r_S \geq s$.

At $t = 1$, the agents act as entrepreneurs and workers. Each agent is endowed with a project, and each project requires \$1 of investment and pays $A > 1$ with probability q_i and zero with probability of $1 - q_i$, where q_i has the distribution function Q and A is a commonly known constant. The expected payoff per dollar invested is thus q_iA. Each agent can only borrow from the bank where she keeps her deposit (the "relationship" bank). The bank prices the loan as a monopolist. If the loan is granted, the entrepreneur pays \$1 to a randomly selected agent from the same population. The selected agent plays the role as a worker and completes the project. The main point of introducing workers is to generate some money flow in the economy.

At $t = 2$, agents play the role as depositors. Workers who receive wages choose where to deposit the wage. The depositor can pick either the large bank or the small bank, and within a bank, the depositor can pick either the bank's own deposit account or the CBDC account. These choices are made after considering the depositor's own convenience value for large bank deposits, the convenience value of the CBDC, and all relevant interest rates.

At $t = 3$, the projects succeed or fail. The banks earn interest on reserves and pays depositors according to their deposit holdings and the deposit rates.

2.2.2 Bank deposit creation

For the purpose of illustration it is convenient to illustrate the deposit creation process by considering a discrete set-up, in which we characterize the bank's decision to make a single loan. The condition on bank lending that we derive will be applicable to the continuum model in which borrowers (i.e., the entrepreneurs) are infinitesimal.

The tables below show the sequence of changes in the large bank's balance sheet in the loan process. The changes in the small bank's balance sheet in the loan process are entirely analogous.

1. Before lending, the large bank starts with X_L reserves. Its balance sheet looks like:

Asset	Liability
Reserves X_L	Deposits X_L

2. If the large bank makes a loan of \$1, it immediately creates deposit of \$1 in the name of the entrepreneur. The balance sheet of the bank becomes:

Asset	Liability
Reserves X_L	Deposits X_L
Loans 1	New Deposits 1

3. Eventually, the entrepreneur will spend her money to pay a worker. The large bank anticipates that, in expectation, a fraction α_S of the \$1 new deposit will be transferred to the small bank, leading to a reduction of reserves by the same amount. The fraction α_L remains in the bank because the worker has an account with the same bank. The bank's balance sheet becomes:

Asset	Liability
Reserves $X_L - \alpha_S$	Deposits X_L
Loans 1	New Deposits α_L

If the large bank makes the \$1 loan to entrepreneur i, and charges interest rate R_i, its total expected profit, by counting all items in the balance sheet, will be

$$\underbrace{(X_L - \alpha_S)f}_{\text{Interest on reserves}} + \underbrace{[q_i(1+R_i) - 1]}_{\text{Gross profit on the loan}} - \underbrace{(X_L + \alpha_L)r_L}_{\text{Cost of deposits}}. \tag{2.1}$$

If the large bank does not make the loan, then its total profit will be

$$X_L(f - r_L). \tag{2.2}$$

The large bank's marginal profit from making the loan, compared to not making it, is

$$\pi_i = \underbrace{q_i(1+R_i) - (1+f)}_{\text{Net profit on the loan}} + \underbrace{\alpha_L(f - r_L)}_{\text{Profit on deposit}}. \tag{2.3}$$

In the expression of π_i, the net profit on the loan reflects the true opportunity cost of capital. Besides the usual profit on the loan, the large bank makes an additional profit equal to $\alpha_L(f - r_L)$. This is because each \$1 lent out stays with the large bank with probability α_L and earns the bank the IOR-deposit spread of $f - r_L$. The corresponding term for the small bank's marginal profit of lending is $\alpha_S(f - r_S)$. In the equilibrium we characterize, it will be the case that $\alpha_L(f - r_L) > \alpha_S(f - r_S)$, i.e., the large bank's convenience value of deposits translates into an advantage in the lending market. Such a feature would not be present if banks were homogeneous.

2.2.3 Equilibrium

We solve the model backward in time.

Deposit market at $t = 2$. A depositor with a large-bank convenience value of δ faces four choices:

	Large bank		Small bank	
	Deposit	CBDC	Deposit	CBDC
Convenience value	$\max(\delta, v)$	$\max(\delta, v)$	v	v
Interest rate	r_L	s	r_S	s

Obviously, the small bank attracts no depositors if $r_S < r_L$. So $r_S \geq r_L$ in equilibrium. For technical simplicity, whenever a depositor is indifferent between two choices, their preference is the small bank, the large bank, and finally the CBDC, in this order.[9]

We will characterize parameter conditions under which $r_S > r_L$. This implies that a depositor with convenience value δ chooses the large bank if and only if

$$\delta > v \text{ and } r_L + \delta > r_S + v \Rightarrow \delta > r_S - r_L + v. \tag{2.4}$$

Therefore, the eventual market shares of the banks in the newly created deposits are

$$\alpha_L = 1 - G(r_S - r_L + v) \tag{2.5}$$

$$\alpha_S = G(r_S - r_L + v). \tag{2.6}$$

Loan market at $t = 1$. In the previous section we derived the marginal profit of a bank from making a loan. While the entrepreneur is infinitesimal here, expression (2.3) still applies.

The monopolist position of each bank in the lending market implies that a bank can make a take-it-or-leave-it offer to the entrepreneur. The bank's optimal interest rate quote would be $R_i = A - 1$ (or just tiny amount below), and the entrepreneur, who has no alternative source of funds, would accept. The lending bank takes the full surplus.

[9]The tie-breaking rule between commercial banks and the CBDC is without loss of generality because a commercial bank can always offer ϵ above s so that depositors strictly prefer commercial bank deposits to the CBDC. The tie-breaking rule also preserves continuity in the fractions of depositors as parameters change to make depositors indifferent.

Hence, the large bank makes the loan if and only if

$$q_i A - (1 + f) + \alpha_L (f - r_L) > 0, \tag{2.7}$$

or

$$q_i > q_L^* = \frac{1 + f - \alpha_L (f - r_L)}{A}. \tag{2.8}$$

Exactly the same calculation for the small bank yields the comparable investment threshold

$$q_S^* = \frac{1 + f - \alpha_S (f - r_S)}{A}. \tag{2.9}$$

Choice of deposit rates at $t = 0$. Again, we start with the large bank. The large bank makes profits in two ways. Because the large bank is a monopolist when lending to its customers, its first source of profit is on the loans, $m_L \int_{q_l^*} (qA - 1 - f) dQ(q)$. The second source of the large bank's profit is on the interest rate spread. The existing deposit in the banking system is $X = X_L + X_S$. As discussed above, the lending process also creates new deposits. The amount of new deposit created by the large bank is $m_L(1 - Q(q_L^*))$, by the normalization that each loan is of \$1. Likewise, the small bank creates new deposit $m_S(1 - Q(q_S^*))$. When the two banks compete for depositors by setting the deposit rates r_L and r_S, we already show above that a fraction $\alpha_L = 1 - G(r_S - r_L + v)$ of total deposits end up with the large bank, enabling the large bank to collect a spread of $f - r_L$ per unit of deposit held.

Adding up the two components, we can write the large bank's total profit as

$$\begin{aligned} \Pi_L &= m_L \int_{q_L^*}^{1} (qA - 1 - f) dQ(q) + [X_L + X_S + m_L(1 - Q(q_L^*)) + m_S(1 - Q(q_S^*))] \alpha_L (f - r_L) \\ &= m_L \int_{q_L^*}^{1} [qA - (1 + f) + \alpha_L (f - r_L)] dQ(q) + [X_L + X_S + m_S(1 - Q(q_S^*))] \alpha_L (f - r_L). \end{aligned} \tag{2.10}$$

Likewise, the small bank's total profit is

$$\Pi_S = m_S \int_{q_S^*}^{1} [qA - (1+f) + \alpha_S(f - r_S)]dQ(q) + [X_L + X_S + m_L(1 - Q(q_L^*))]\alpha_S(f - r_S). \tag{2.11}$$

As discussed before, the CBDC interest rate puts a lower bound on commercial banks' deposit rates, i.e., $r_L \geq s$, and $r_S \geq s$. There are two cases. The first is that $r_L > s$, so that the CBDC interest rate does not constrain the commercial banks' deposit rates. We call the first case the *unconstrained equilibrium.* The second case is that $r_L = s$, i.e., the CBDC interest rate binds the large bank's deposit rate. We call the second case the *constrained equilibrium.*

Unconstrained equilibrium. Assuming that Π_L is strictly quasi-concave in r_L, the sufficient condition for a unique maximum of the function Π_L with respect to r_L is

$$\begin{aligned} \frac{d\Pi_L}{dr_L} &= m_L(1 - Q(q_L^*))\frac{d[\alpha_L(f - r_L)]}{dr_L} - m_L \underbrace{[q_L^* A - (1+f) + \alpha_L(f - r_L)]}_{=0} \frac{dq_L^*}{dr_L} \\ &\quad + [X_L + X_S + m_S(1 - Q(q_S^*))]\frac{d[\alpha_L(f - r_L)]}{dr_L} - m_S \alpha_L (f - r_L) Q'(q_S^*) \frac{dq_S^*}{dr_L} \\ &= \Big[X_L + X_S + m_L(1 - Q(q_L^*)) + m_S(1 - Q(q_S^*))\Big] \cdot \Big[(f - r_L)G'(r_S - r_L + v) - 1 \\ &\quad + G(r_S - r_L + v)\Big] - m_S \alpha_L (f - r_L) Q'(q_S^*) \frac{(f - r_S)G'(r_S - r_L + v)}{A}. \end{aligned} \tag{2.12}$$

Likewise, the first-order condition of the small bank is

$$\begin{aligned} \frac{d\Pi_S}{dr_S} &= \Big[X_L + X_S + m_L(1 - Q(q_L^*)) + m_S(1 - Q(q_S^*))\Big] \cdot \Big[(f - r_S)G'(r_S - r_L + v) \\ &\quad - G(r_S - r_L + v)\Big] - m_L \alpha_S (f - r_S) Q'(q_L^*) \frac{(f - r_L)G'(r_S - r_L + v)}{A}. \end{aligned} \tag{2.13}$$

For simplicity, let $Q(\cdot)$ be the uniform distribution on $[0, 1]$. And further impose a stationarity condition that the market shares of deposits $\{\alpha_j\}$ are identical to the starting market shares $\{m_j\}$. The first-order conditions simplify to

$$0 = \frac{d\Pi_L}{dr_L} = \Big[X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)\Big] \cdot \Big[(f - r_L)G'(r_S - r_L + v) - 1 + G(r_S - r_L + v)\Big] - \frac{1}{A}\alpha_S\alpha_L(f - r_L)(f - r_S)G'(r_S - r_L + v), \quad (2.14)$$

$$0 = \frac{d\Pi_S}{dr_S} = \Big[X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)\Big] \cdot \Big[(f - r_S)G'(r_S - r_L + v) - G(r_S - r_L + v)\Big] - \frac{1}{A}\alpha_L\alpha_S(f - r_L)(f - r_S)G'(r_S - r_L + v). \quad (2.15)$$

From the above conditions we derive

$$(r_S - r_L)G'(r_S - r_L + v) + 2G(r_S - r_L + v) = 1. \quad (2.16)$$

Proposition 2.1 *Suppose that the profit function Π_j is quasi-concave in r_j, $j \in \{L, S\}$ and that $G(v) < 0.5$. Let r_L and r_S solve equations (2.14)–(2.15). If $r_L > s$ and $r_S > s$, then it is an unconstrained equilibrium that the banks set r_L and r_S as their deposit rates. In this equilibrium:*

1. *The large bank sets a lower deposit rate ($r_L < r_S < f$) and has a larger market share ($\alpha_L > \alpha_S$) than the small bank.*

2. *The large bank uses a looser lending standard than the small bank does ($q_L^* < q_S^*$).*

Proofs are in Appendix A.

The condition $G(v) < 0.5$ ensures that the CBDC does not increase the market share of the small bank so much that it fully eliminates the large bank's convenience value advantage in its deposits. Consequently, the small bank still needs to compete by offering a higher deposit rate than the large bank.

Further intuition of the equilibrium may be gained by considering an example. Suppose that $G(\delta) = \delta/\Delta$, where $\delta \in [0, \Delta]$ for a sufficiently large Δ. Then $G'(\cdot) =$

$1/\Delta$. The two first-order conditions reduce to

$$\frac{f - r_L}{\Delta} = 1 - \frac{r_S - r_L + v}{\Delta} + B \tag{2.17}$$

$$\frac{f - r_S}{\Delta} = \frac{r_S - r_L + v}{\Delta} + B \tag{2.18}$$

where

$$B \equiv \frac{\frac{1}{A\Delta}\alpha_L \alpha_S (f - r_L)(f - r_S)}{X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)} > 0. \tag{2.19}$$

As the total reserve X becomes large, B becomes close to zero. So the equilibrium deposit rates of the two banks become approximately $r_L \approx f - \frac{2}{3}\Delta + \frac{1}{3}v$ and $r_S \approx f - \frac{1}{3}\Delta - \frac{1}{3}v$. This shows directly how an increase in convenience reduces the spread between deposit rates.

Constrained equilibrium. The second case of the equilibrium is that the CBDC interest rate s becomes binding for the large bank. Recall the tie-breaking rule that at $r_L = s$, depositors use the large bank.

The small bank's profit function and first-order condition are as before:

$$0 = \frac{d\Pi_S}{dr_S} = [X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)] \cdot [(f - r_S)G'(r_S - s + v) - G(r_S - s + v)] - \frac{1}{A}\alpha_L \alpha_S (f - s)(f - r_S)G'(r_S - s + v). \tag{2.20}$$

By contrast, the large bank's first order condition takes an inequality because the conjectured optimal solution is at the left corner:

$$0 > \left.\frac{d\Pi_L}{dr_L}\right|_{r_L \downarrow s} = [X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)] \cdot [(f - s)G'(r_S - s + v) - 1 + G(r_S - s + v)] - \frac{1}{A}\alpha_S \alpha_L (f - s)(f - r_S)G'(r_S - s + v). \tag{2.21}$$

Proposition 2.2 *Suppose that the profit function Π_j is quasi-concave in r_j, $j \in \{L, S\}$. Suppose that $v < \bar{v}$ for some $\bar{v}$ that may depend on f and s. Let r_S solve equation* (2.20)*. If, at r_S, equation* (2.21) *also holds, then it is a constrained equilibrium that the large bank sets s and the small bank sets r_S as their deposit rates. In* *this equilibrium:*

1. *The large bank sets a lower deposit rate ($s < r_S$) and has a larger market share ($\alpha_L > \alpha_S$) than the small bank.*

2. *The large bank uses a looser lending standard than the small bank does ($q_L^* < q_S^*$).*

The condition that v cannot be too high guarantees that the small bank still wishes to compete by offering a higher deposit rate. This is analogous to the restriction on $G(v)$ in Proposition 2.1.

While the model described in this section is stylized, it potentially explains an important fact about the U.S. deposit market: deposit rates are below and only partially responsive to the key policy rate (Federal Funds rate and IOR) set by the central bank. We discuss in Appendix B how our model, under a certain parameterization and without the laborious calibration to the data, already generates predicted deposit rates that are largely similar to actual U.S. deposit rates from 1986 to 2021. We contend that this conformance provides essential support for the validity of the model's predictions regarding the introduction of a CBDC.

Parlour, Rajan, and Walden (2022) also analyze asymmetries in the banking sector and their consequences. In their model, a bank that is a net payer incurs an additional settlement cost and hence reduces lending, compared to net-receiving bank. In this sense, the net payer bank in their model looks like the small bank in ours. Despite similar predictions on lending, the two models are driven by different mechanisms. In our model, there is no exogenous cost associated with interbank settlement; rather, the main advantage of the large bank in lending is a higher likelihood that a lent dollar stays with the large bank and earns interest on reserves from the central bank. Moreover, the size of the large bank's advantage depends on the interest rate paid on reserves and the CBDC design, including its interest rate and convenience value, as we see in the next section.

Using confidential FedWire transaction data, Li and Li (2021) calculate the volatility of daily net payments as a fraction of daily gross payments for various banks. They find that banks with higher payment volatility pay a higher deposit rate and have lower loan volume growth, controlling for a set of observables. While our model does not have payment volatility, our predictions are consistent with the negative cross-sectional correlation they compute between the deposit rate and lending.

2.3 Impact of CBDC Interest Rate and Convenience Value

In this section, we discuss the consequences of varying the CBDC interest rate s or convenience value v. The results are summarized in Propositions 2.3 and 2.4.

2.3.1 Impact of CBDC interest rate s

When the federal reserve introduced the overnight reverse repo program (ONRRP) as a temporary facility to support its IOR policy, it began by testing the facility by varying the ONRRP rate between 1 basis point and 10 basis points, while holding the IOR rate fixed at 25 basis points. Here we examine how market outcomes change as s varies from a rate of 0 to f, while holding f fixed.

We focus on the case where, given a fixed value of v, f is sufficiently low that the constrained equilibrium applies. This case is most relevant to the current economic environment in the United States. In the unconstrained equilibrium, market outcomes are invariant to the CBDC interest rate s by definition.

Before we provide a formal statement of the comparative statics, it is useful to illustrate the impact of CBDC interest rate changes in an example. The top row of Figure 2-1 plots the behavior in the deposit markets as the CBDC interest rate rises from 0 to $f = 2\%$. The charts are computed numerically using a uniform distribution for G and a zero CBDC convenience value ($v = 0$). As we see in the top left plot, raising the CBDC interest rate increases the deposit rates of both banks as well as the weighted average deposit rates. The top right plot shows the corresponding changes in deposit market shares α_j, $j = L, S$, which are easily computed from (2.5) and (2.6). Since the large bank's deposit rate rises faster than the small bank's, the large bank gains market share from the small bank. Intuitively, the small bank competes with the large bank primarily by offering a higher deposit rate. As s increases, the maximum spread $f - s$ shrinks, limiting the small bank's ability to compete with its interest rate choice. Once the deposit rates are equal at f, the large bank obtains the entire market share of depositors, given the higher convenience value of its deposits.

The bottom row of Figure 2-1 illustrates the impact raising the CBDC interest rate has on the lending market. Raising the CBDC interest rate changes the incentives to make loans via the expected profit on the interest rate spread, $\alpha_j(f - r_j)$. Because, as shown above, both α_S and $f - r_S$ decrease in s, so does $\alpha_S(f - r_S)$. Thus, the small bank's loan quality threshold, $q_S^* = \frac{1+f-\alpha_S(f-r_S)}{A}$, increases in s, and its loan volume, $\alpha_S(1 - q_S^*)$, decreases in s. In this example, the large bank's loan quality threshold, $q_L^* = \frac{1+f-\alpha_L(f-r_L)}{A}$, increases in s, and its loan volume $\alpha_L(1 - q_L^*)$, also increases due to its larger market share. In this example, the total loan volume declines in s.

The following proposition characterizes the impact of CBDC interest rate s on the deposit and lending markets in more general cases.

Proposition 2.3 *For a sufficiently large X, increasing the CBDC interest rate in a constrained equilibrium has the following impact on the deposit and lending markets:*

	As s increases	
	Large	*Small*
Deposit rates r_L and r_S	↑	↑
Deposit market shares α_L and α_S	↑	↓
Weighted average deposit rate	↑	
Loan quality thresholds q_L^* and q_S^*	↑	↑
Loan volume $\alpha_L(1 - Q(q_L^*))$ and $\alpha_S(1 - Q(q_S^*))$	↑ *or* ↓	↓
Total loan volume, i.e., total deposit created	? (↓ *if* $G'' \leq 0$)	

Most of the qualitative aspects illustrated in Figure 2-1 are true generally, and are analytically proven in Proposition 2.3. The exceptions are that, in general, the large bank's loan volume may go up or down in s and when $G'' > 0$, we do not know what will happen to total loan volume.[10]

[10] An example illustrating the ambiguity in large bank loans volumes is seen by setting $G(\delta) =$

2.3.2 Impact of CBDC convenience value v

A convenient CBDC reduces the large bank's convenience advantage and hence has an impact even if its interest rate is zero. We illustrate the impact of a convenient CBDC by considering this polar case in Figure 2-2. The top row shows the outcomes in the deposit market. As v rises, the inconvenience disadvantage of the small bank shrinks. As long as the large bank's deposit rate remains at the floor rate, the small bank can afford to lower its interest rate and still capture a growing market share. Once v gets large enough, the large bank responds by raising its interest rate; however, the small bank can still afford to continue lowering its deposit rate for the same reason that the convenience gap between the two banks continues to shrink. Throughout this process the large bank loses market share and the small bank gains market share, albeit at a slower rate once the large bank is no longer constrained. The overall impact of increasing the CBDC convenience value is the convergence of the deposit rates and market shares for the two banks.[11]

The CBDC convenience value has a nuanced impact on the weighted average deposit rates. In a constrained equilibrium, a higher v results in a lower weighted average deposit rate when the large bank's deposit rate is at the lower bound. That is, a convenient CBDC weakens the transmission of monetary policy to the deposit market through IOR. Once the economy transitions to an unconstrained equilibrium with a sufficiently high v, however, a higher CBDC convenience value increases the average deposit rate, increasing the transmission of monetary policy.

The bottom row of Figure 2-2 shows the outcomes in the lending market. Because the two deposit rates and the deposit market shares get closer to each other as v rises, it is unsurprising that the loan quality thresholds and loan volume of the two banks are also getting closer to each other. In this example, the total loan volume is almost

---

$\delta/0.035, A = 1.05, X = 10, f = 0.02, v = 0$. Then, in the constrained equilibrium, the large bank's loan volume first increases and then decreases with s.

[11] In fact, a modest CBDC convenience value may be enough to fully level the playing field. Under the uniform distribution of large-bank preference δ, when v rises to the point $v = \Delta/2$, depositors with $\delta > \Delta/2$ strictly prefer the large bank, and depositors with $\delta < \Delta/2$ strictly prefer the small bank. That is, the deposit market shares become equal and so do the deposit rates, loan quality thresholds, and loan volume.

invariant to v, and the most salient effect is the reallocation of loans from the large bank to the small one.

Figure 2-3 below further demonstrates the potential impact of CBDC convenience value on loan volume, using a different parametrization of $f = 3\%$ and $s = 1.25\%$. These parameters lead to a constrained equilibrium, with $r_L = s$. In this example, the total lending volume (left axis) is first decreasing in v and then increasing in v. The magnitude of the axes suggests that the more salient action is, again, the shift of lending from the large bank to the small one.

The next proposition summarizes the comparative statics with respect to v.

Proposition 2.4 *For a sufficiently large X, the impact of increasing v is given in the following table:*

As v increases	*Constrained*		*Unconstrained*	
	Large	*Small*	*Large*	*Small*
Deposit rates r_L and r_S	*Flat(=s)*	$\downarrow$	$\uparrow$	$\downarrow$
Deposit market shares α_L and α_S	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
Weighted average deposit rate	$\uparrow$ *or* $\downarrow$ *($\downarrow$ if $G'' \leq 0$)*		$\uparrow$ *or* $\downarrow$ *($\uparrow$ if $G'' \geq 0$)*	
Loan quality thresholds q_L^* and q_S^*	$\uparrow$	$\downarrow$	$\uparrow$	$\downarrow$
Loan volume $\alpha_L(1-q_L^*)$ and $\alpha_S(1-q_S^*)$	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
Total loan volume, i.e., total deposit created	$\uparrow$ *or* $\downarrow$		$\uparrow$ *or* $\downarrow$ *($\downarrow$ if $G'' \geq 0$)*	

As in the previous subsection, most of the comparative static results that apply for uniform G are true more generally and are stated in Proposition 2.4. There are a few exceptions. The impact on the weighted average interest rate, $\alpha_S r_S + \alpha_L r_L$, is ambiguous. An increase in v shifts market share to the small bank and reduces the small bank's deposit rate, but the small bank has a higher deposit rate to start with, so the overall effect can go in either direction. In the constrained equilibrium, a concave G means that relatively more depositors have a weak (but still positive) preference for large bank's deposits, so a higher v quickly eliminates the large bank's advantage. As a result, the small bank can afford to reduce its deposit rate quickly, leading to a lower weighted average $\alpha_S r_S + \alpha_L r_L$.[12] In the unconstrained equilibrium, r_L increases in v. A convex G means that relatively more depositors have a strong (but still positive) preference for large bank's deposits, so the large bank raises r_L aggressively compared to the reduction in r_S, leading to a higher $\alpha_S r_S + \alpha_L r_L$.[13] The ambiguity in total loan volume for the constrained case is illustrated by the numerical example in Figure 2-3 where G is uniform and total loan volume first decreases and then increases in v. Total loan volume unambiguously decreases in v in the unconstrained equilibrium if G is weakly convex. Intuitively, the weighted average deposit rate increases in v, so the IOR-deposit rate spread is compressed, which discourages lending. When G is strictly concave, however, the change in total loan volume resulting from an increase in v can be in either direction.

2.4 CBDC Design and Monetary Policy Pass-through

We now address the issue of how a given CBDC design impacts the sensitivity of deposit rates to changes in the IOR rate f. In the unconstrained equilibrium, deposit rates of both the large and the small bank move one-for-one with the IOR rate f. This means the pass-through of f is perfect when the large bank is not constrained, that is, when the large bank competes with the small bank on the deposit rate margin. In the constrained equilibrium, the large bank's deposit rate is capped at CBDC interest rate s, and only the small bank's deposit rate reacts to changes in f, hence the pass-through of f to the average deposit rate is much weaker.

When the IOR rate f is low, deposit and lending markets are characterized by the constrained equilibrium and when f is high they enter into the unconstrained equilibrium. Let f^* denote the threshold value of the economy transitions from the constrained equilibrium to the unconstrained equilibrium. Since the pass-through

[12] When $A = 1.5, X = 10, f = 0.02, s = 0, G = Gamma(5, 150)$ with mean 1/30, then $0 < G''(r_S - r_L + v) < G'(r_S - r_L + v)/f$, yet weighted average deposit rate first increases then decreases in v, in the constrained equilibrium.

[13] When $A = 1.5, X = 10, f = 0.02, s = 0, G = Gamma(3, 200)$ with mean 3/200, then $-G'(r_S - r_L + v)/f < G''(r_S - r_L + v) < 0$, yet weighted average deposit rate first increases then decreases in v, total loan volume first decreases then increases in v, in the unconstrained equilibrium.

of f is vastly different between the constrained equilibrium and the unconstrained equilibrium, it is important to understand how the CBDC interest rate s and the convenience value v affect the cut-off value of IOR, f^*, that separates the two equilibria. We solve for f^*, from the following FOCs.

$$\begin{aligned}
0 &= [X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)] \cdot [(f^* - r_S)G'(r_S - s + v) - G(r_S - s + v)] \\
&\quad - \frac{1}{A}\alpha_L\alpha_S(f^* - s)(f^* - r_S)G'(r_S - s + v). \\
0 &= [X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)] \cdot [(f^* - s)G'(r_S - s + v) - 1 + G(r_S - s + v)] \\
&\quad - \frac{1}{A}\alpha_S\alpha_L(f^* - s)(f^* - r_S)G'(r_S - s + v). \qquad (2.22)
\end{aligned}$$

The following proposition summarizes the results.

Proposition 2.5 *For a sufficiently large X:*

1. *In the unconstrained equilibrium, r_L and r_S move one-for-one with f;*

2. *In the constrained equilibrium, $\frac{dr_S}{df}$ decreases with s and increases with v if $\frac{G''(\delta)}{G'(\delta)}$ is increasing in δ; and*

3. *f^* increases with s. f^* decreases with v if $G''(\delta) \geq 0$.*

With abundant reserves, the spread between IOR and the deposit rate is the main factor that determines profits for the banks. In the unconstrained equilibrium, the two banks compete primarily for deposit market shares, and the market shares do not vary with f. As a result, r_L and r_S move one-for-one with f.

Part 2 of Proposition 2.5 requires that $G''(\delta)/G'(\delta)$ is an increasing function as a sufficient condition.[14] Such distributions have decreasing probability density functions and large mass at small values. Under these assumptions, in the constrained equilibrium, higher s or lower v decreases the sensitivity of the small bank's deposit rate r_S to changes in f. This occurs under the stated convexity condition because when s is higher, or v is lower, the convenience value for the large bank of the marginal

[14]This condition is satisfied by Gamma distributions with a shape parameter less than 1. The proposition is also true for exponential distributions, where $G''(\delta)/G'(\delta)$ is a non-zero constant.

consumer that is indifferent between choosing the small bank and the large bank takes a lower value where the convexity of G is lower. This is where r_S needs to react less to offset the change introduced by f. Since r_S becomes less sensitive to f as s increases or v decreases, pass-through is decreased.

A higher s increases the cutoff value f^*, which means the Fed needs to set a higher IOR to enter into the high pass-through region. Intuitively, there needs to be a large spread between IOR and s in order to induce the large bank to compete with the small banks via its deposit rate policy. A higher s necessarily increases f^*. A higher v reduces the large bank's competitive advantage, and forces it to compete with the small bank sooner; that is, a higher v decreases the cutoff value f^*. The sufficient condition for this is that G is convex. Convex G means relatively more depositors have a strong preference for the large bank's deposits, so the large bank competes sooner on the deposit rate margin to compensate for the reduction in the convenience advantage.

2.5 Conclusion

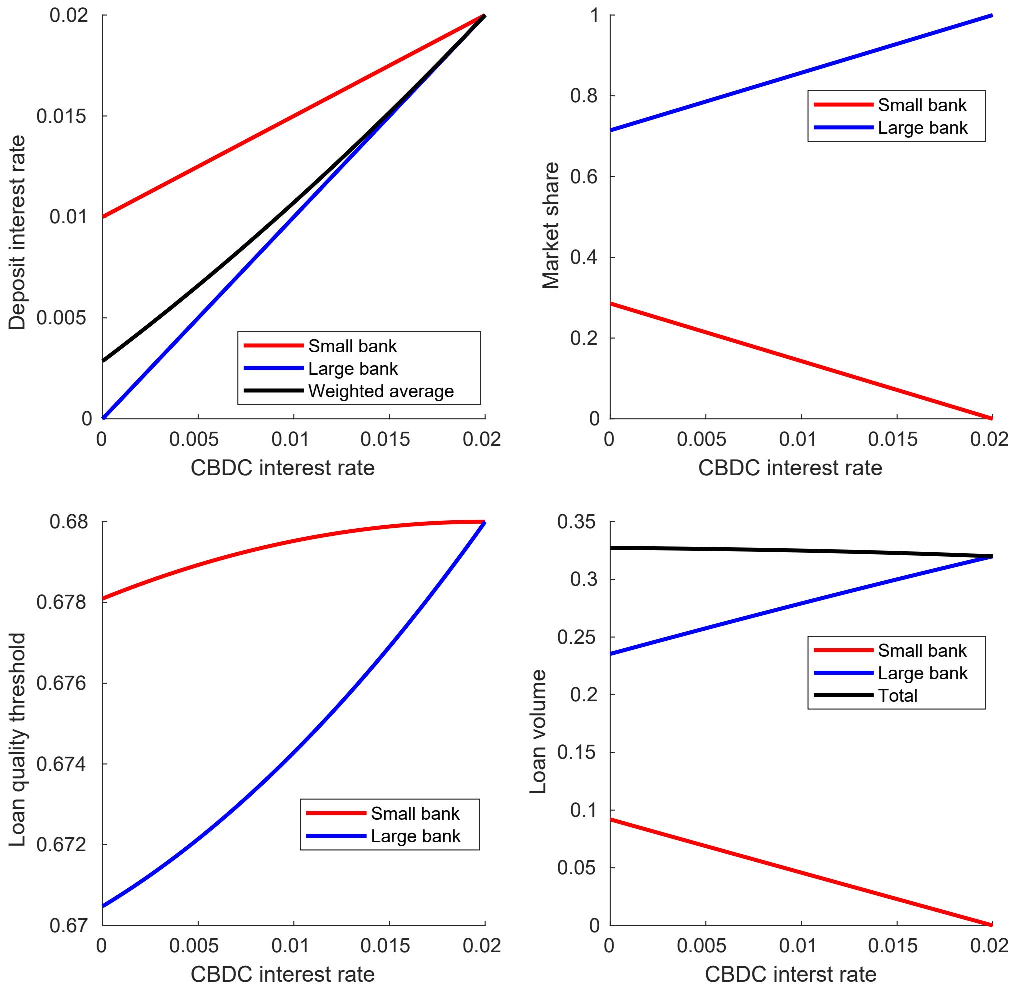
Payment convenience is a crucial aspect of CBDC design that may be more desirable than interest rate policy. A highly convenient CBDC produces sufficient competitive pressure in deposit markets to raise deposit rates for any given level of IOR and increases the responsiveness of deposit rates to IOR rate changes. Convenience also has favorable effects on market composition by leveling the playing field. Interest rate policy is less desirable in the sense that it may weaken the responsiveness of deposit rates to IOR rate changes and it increases the inequality of market shares.

An interesting aspect of our analysis is that the provision of CBDC impacts equilibrium outcomes even though the currency is not held in equilibrium. Hence there is no disintermediation. This is also true in Chiu et al. (2019) and Garratt and Lee (2021), where the option to use CBDC changes the equilibrium outcome even it is not exercised. An exception is Keister and Sanches (2021), where the CBDC has specific liquidity benefits that leads to its use. The idea that a central bank introduces a program to influence market rates by increasing the bargaining power of lenders is not new. Early descriptions of the overnight reverse repurchase agreement facility that the Federal Reserve Bank of New York began testing in September 2013 indicated that "the option to invest in ON RRPs [Overnight Reverse Repurchase Agreement Facility] also would provide bargaining power to investors in their negotiations with borrowers in money markets, so even if actual ON RRP take-up is not very large, such a facility would help provide a floor on short-term interest rates..." (Frost et al., 2015).

The results of our paper could be extended in multiple directions. One possible extension is to add short-term investment vehicles such as money market mutual funds and repurchase agreements that typically pay higher interest rates than bank deposits but cannot be easily used for processing payments. If the CBDC pays a sufficiently high interest rate, it is possible that money would flow out of these short-term investment vehicles into the CBDC, i.e., investors would earn returns from the Fed rather than short-term Treasury Bills. This additional channel is unlikely to affect lending because money market investors do not make loans. Another possibility is to consider heterogeneous CBDC interest rates paid to banks of different sizes, which adds yet another degree of freedom in the central bank's toolkit. In particular, the central bank could use heterogenous CBDC interest rates to fine-tune the competitive positions of large and small banks. These extensions are left for future research.

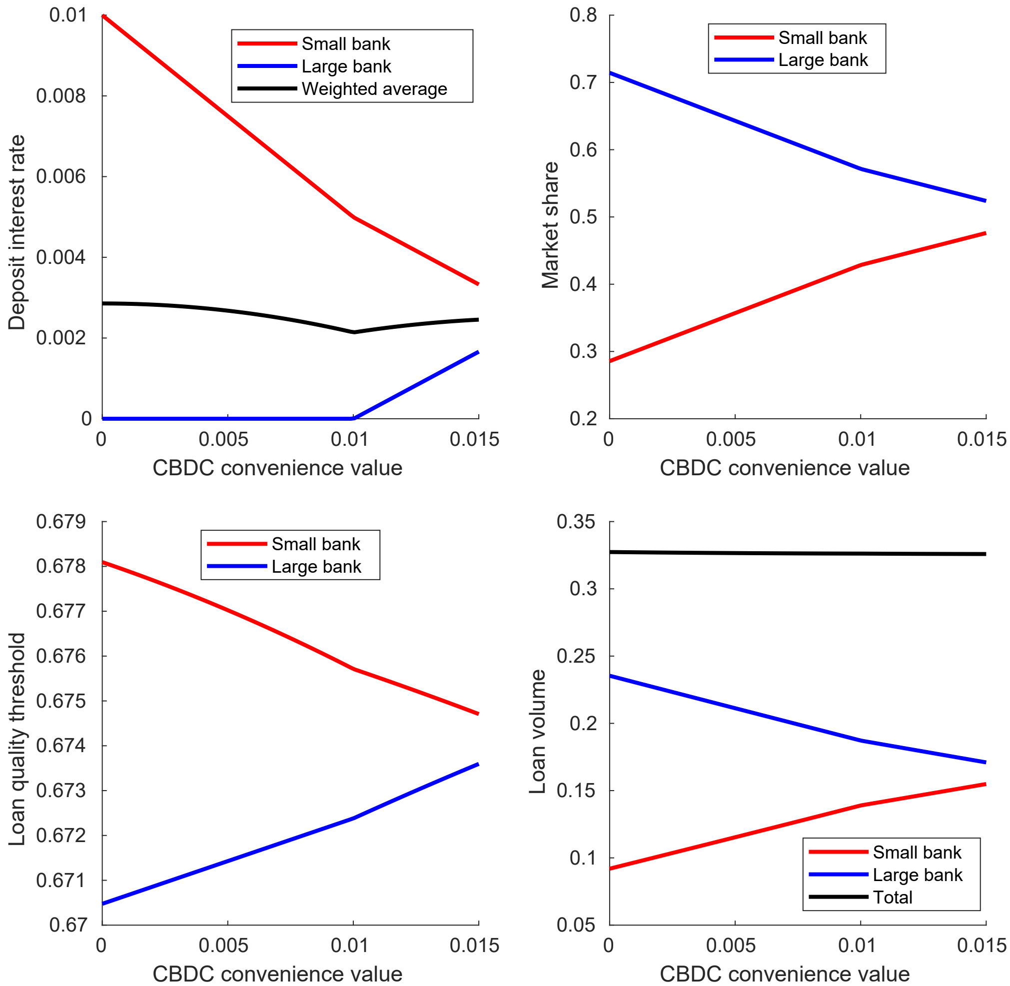
2.6 Figures

Figure 2-1: Impact of CBDC Interest Rate on Deposit and Lending Markets.



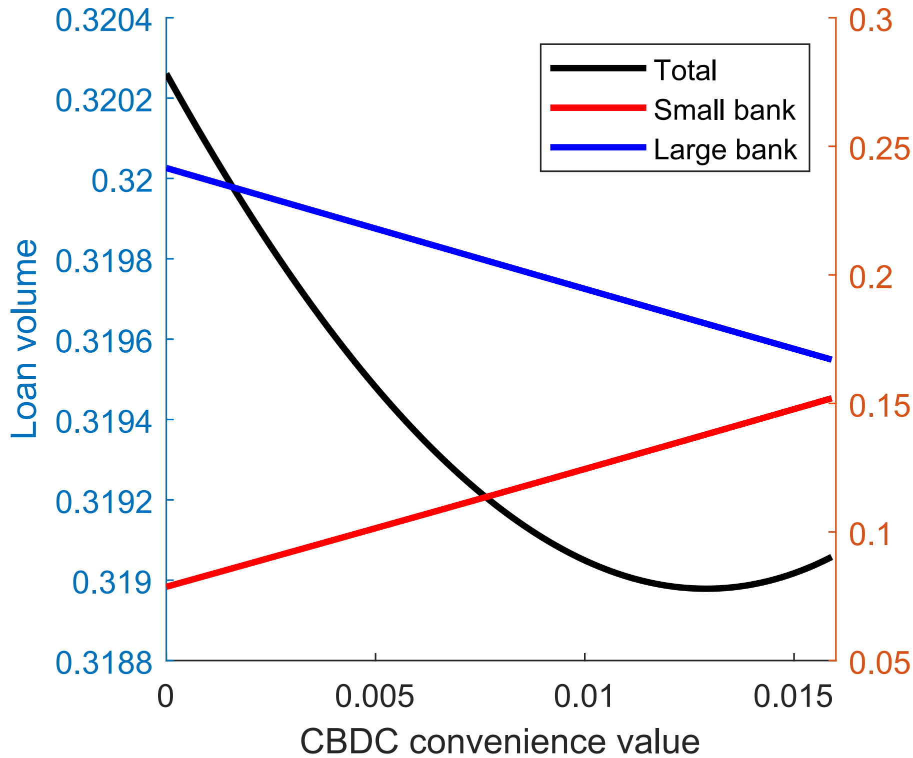
Numerical example: $G(\delta) = \delta/0.035, A = 1.5, X = 10, f = 0.02, v = 0.$

Figure 2-2: Impact of CBDC Convenience Value on Deposit and Lending Markets.



Numerical example: $G(\delta) = \delta/0.035, A = 1.5, X = 10, f = 0.02, s = 0$. The equilibrium transitions from constrained to unconstrained at $v = 0.01$.

Figure 2-3: CBDC Convenience Value and Loan Volume



Numerical example: $G(\delta) = \delta/0.035, A = 1.5, X = 10, f = 0.03, s = 0.0125$. All values of v in the depicted range correspond to a constrained equilibrium. Total loan volume is on the left axis. Loan volume of the two banks are on the right axis.

Chapter 3

Undercutting the Exchanges: Private Trading, Fee Competition, and Price Discovery at the Market Close

Over recent years, the trading volume of US equities has been shifting to the market close, partly driven by passive investment strategies benchmarked against indices, which seek to trade at the market close price to minimize tracking errors (Bogousslavsky and Muravyev, 2021). The market close price of a stock is determined in a special call auction – the close auction, held by its listing exchange. From 2012 to 2018, the total trading volume executed in the close auctions has increased by 120% and started to account for more than 8% of the total trading volume. Meanwhile, the lack of competition drove up the close auction fees – NYSE's base rate has gone up by 16% and Nasdaq's by 60%, adding to the cost of benchmarking strategies. [1]

"Guaranteed close", offered by investment banks like Goldman Sachs, executes clients' orders at the market close price set by the close auctions yet charges a lower transaction fee. Meanwhile, it allows the banks to trade and profit on the order flow

[1] See Tier F under Execution Fees for the Nasdaq Closing Cross in the Nasdaq fee schedule available at [link]; and Liquidity Indicator 7 in the NYSE fee schedule available at [link].

information – after pairing buyers with sellers, the banks hedge the imbalance and trade on their principal accounts. Trading volume through "guaranteed close" has reached almost 30% of that through close auctions in 2018. [2] Albeit this large trading volume, it is understudied how "guaranteed close" affects the formation of close prices, which are arguably the most important prices, widely used in net asset values, margin accounts, numerous financial contracts and risk metrics. Importantly, if "guaranteed close" makes close prices less informative, that is, farther from the fundamental values, it imposes higher costs for passive investment strategies that benchmark and trade at close prices, and undermines the role of stock market price discovery in allocating resources (Bakke and Whited, 2010; Bond, Edmans, and Goldstein, 2012). Hence, we undertake the task of formally studying the impact of "guaranteed close" on price discovery, both empirically and theoretically.

It is ex-ante unclear how "guaranteed close" might affect price discovery. In fact, close auctions are considered robust mechanisms in generating the close prices, and off-exchange venues like "guaranteed close" that siphon trading activities have raised public concerns. Executives at NYSE and Nasdaq suggested that "... if more trading moves to banks, it will make close prices less trustworthy...". Besides these incumbents, Credit Suisse Trading Strategy suggests in a market commentary that "...changes to end-of-day trading dynamics could indicate that there has been some impairment to the price discovery process". In comments submitted to SEC [3], 41 listing companies (85% of the respondents), including PayPal and FedEx, voiced against fragmentation and disruptions to the close auctions, suggesting that "... [fragmentation] will increase volatility and decrease precision in closing prices."

Our investigation is also well suited to speak to a more general question: how does trading on (retail) order flow affect price discovery? The theoretical literature on dual trading (Fishman and Longstaff, 1992; Röell, 1990; Sarkar, 1995) and back-

[2] Based on our calculation. Also documented by Credit Suisse Trading Strategy, WSJ (2018) and SEC report: Securities Exchange Act Release No. 34-80683 (May 6, 2017).

[3] To be precise, these are comments to the proposal of Bats Market Close, a closing match process that supplements NYSE and Nasdaq close auctions. The proposed process is similar to guaranteed close, charging lower fees and using prices set by close auctions. Notably, the public has general opposition to fragmentation and disruptions at market close.

running (Huddart, Hughes, and Levine, 2001; Yang and Zhu, 2020) shed lights on this question. It is, however, empirically challenging due to the lack of data on exogenous shocks to trading on order flow activities. This question becomes eminent as e-brokers like Robinhood flock to charge zero commission fees yet sell order flow data to sophisticated institutions for them to exploit and trade on. As will be discussed soon, our empirical setting provides a nice opportunity to study this question in the context of today's US brokerage industry.

We provide quasi-experimental evidence that "guaranteed close" improves price discovery. To start our analysis, we measure each stock's trading volume via "guaranteed close" by the volume of off-exchange trades between 4:00 p.m.–4:10 p.m. EST executed at the official close price using the NYSE Trade-and-Quote (TAQ) millisecond-level data.[4] We show that ETF and index fund ownership strongly correlate with the "guaranteed close" volume, indicating that "guaranteed close" is mainly used by passive investment strategies to save expenses on transaction costs.[5] We measure the informativeness of the close price by the "closeness" between the close price and the next day's open price. Specifically, we compute the mean of squared overnight return for individual stocks at a monthly frequency as a measure of close price's informativeness. This measure corresponds to the mean squared error measure of price informativeness used in the theoretical literature.

Next, we use the NYSE close auction fee cut in January 2018 as a quasi-experiment and use a standard difference-in-differences strategy to estimate the effects of banks' trading on order flow activities. In the fee cut, NYSE reduced the Tier 1 and Tier 2 rate charged to broker-dealers executing high trading volumes in the close auction significantly by around 40%, but kept rates for the remaining broker-dealers almost unchanged, with the non-tier rate reduced by only 9%. This fee cut is part of NYSE's effort to draw trading activities back to the close auctions, and this incurred differential impact on different stocks. We find that stocks ex-ante more heavily traded

[4] SEC DERA (2017) confirmed that these trades are almost surely from off-exchange market-on-close orders. "Guaranteed close" is the major venue that executes off-exchange market-on-close orders.

[5] This echoes anecdotal evidence that "guaranteed close" is used by index-fund managers including Vanguard Group and BlackRock Inc. (WSJ, 2018)

at "guaranteed close", and those with higher passive ownership are more exposed to the fee cut – the trading volume of these stocks at "guaranteed close" significantly dropped relative to the remaining stocks.

There are at least two reasons why the fee cut for broker-dealers passes through to individual stocks in such a differential manner. First, ETFs and index funds may actively choose the venue at which they trade and are sensitive to transaction fees. The fee cut induced them to actively choose to trade with broker-dealers that operate in the close auction. Their large trading volume easily helps broker-dealers reach the Tier 1 and Tier 2 thresholds, allowing them to take the lower transaction fees effectively. Second, institutional investors may delegate the venue choice entirely to their broker-dealers. They rely on broker-dealers to trade (Di Maggio, Egan, and Franzoni, 2021), build long-term relationships with their brokers, and concentrate their order flow with a relatively small set of broker-dealers (Goldstein, Irvine, Kandel, and Wiener, 2009). In that case, broker-dealers would have a different yet relatively stable client base. Some provide services mainly to passive investors like ETFs and index funds, while others provide services mainly to other types of investors. Broker-dealers' trading on behalf of passive investors is more exposed to the fee cut since they are more likely to reach the Tier 1 and Tier 2 thresholds. Once they go back to the close auctions, the stocks they usually facilitate to trade, that is, those with high passive ownership, will experience a larger drop in trading volume at "guaranteed close". These stocks are also ex-ante heavily traded at the "guaranteed close". Under both cases, and to the extent that banks trade on the order flow data they receive in the "guaranteed close", the fee cut is a plausibly exogenous shock to trading on order flow activities.

For each stock, we calculate the trading volume executed at "guaranteed close" as a fraction of the trading volume executed at the close price and term it as the "guaranteed close fraction". Trading volume executed at the close price includes those executed in the close auction and those in the "guaranteed close". We sort all NYSE stocks by their average "guaranteed close fraction" before the NYSE fee cut. The treated group consists of stocks that rank in the top 50%. The control group consists of the remaining stocks. Our difference-in-differences estimation finds that the NYSE fee cut significantly decreased the informativeness of the close prices for the treated stocks, relative to the control stocks, by 15.7% compared to the sample mean. Indeed, treated stocks may have vastly different characteristics than the control stocks – for example, they have larger market capitalization and higher passive ownership. To alleviate this concern, we control for a wide range of variables: stock fixed effects, time fixed effects, market cap, total trade volume, intra-day volatility, measures of total retail volume and institutional volume, after-market-close volume/total volume, close auction volume/total volume, and overnight betas of individual stocks. We also verified the parallel trend assumption: treated stocks are no different in price informativeness from control stocks before the NYSE fee cut. In addition, we show that intra-day volatility and quoted bid-ask spread during market hours did not respond to the NYSE fee cut shock. That means our result does not merely reflect structural changes in the general trading activities of treated stocks. Our result is also robust to different measures of price informativeness, such as the median of squared overnight returns and the mean absolute value of overnight returns.

We conduct additional exercises to further the robustness of our difference-in-differences estimation results. First, we adopt a different rule to designate the treatment group. Specifically, we rank stocks by their ETF and index fund ownership before the NYSE fee cut. The treated group consists of stocks that rank in the top 50%. The control group consists of the remaining stocks. A similar difference-in-differences estimation suggests that NYSE fee cut decreased the informativeness of treated stocks by 14.3%. Second, our price informativeness measure takes the next day's market open price as the fundamental value to which we compare the market close price with. Although a natural measure, the next day's market open price also incorporates overnight information, and it is possible that the strength of overnight information of treated and control stocks coincidentally changed after the NYSE fee cut. To alleviate this concern, we exclude data in the 3-day window around earnings announcement days and conduct the same difference-in-differences estimation, and arrive at similar results. Third, to further address the concern that treated and control stocks have different characteristics, we conduct a matched sample difference-in-differences estimation, using control group matched with treated group on pre-treatment values of market cap, trading volume, intra-day volatility, and overnight beta. The results are similar.

Our estimated effect of NYSE fee cut on mean absolute value of overnight returns shed light on how the fee cut affects the profit to investors of index funds and ETFs. The difference-in-differences estimated increase in mean absolute value of overnight returns is 5 bps. The relative fee cut between treated and control group is 2 bps. The impact on investor profits depends on the probability that a stock's overnight movement is against the index fund's trade, that is, the stock price goes up yet the index fund sold it at market close, and that the stock price goes down yet the index fund bought it at market close. For reasonable estimate of this probability, back-of-envelop calculation suggests that the benefit of the decline in fee is outweighed by the cost of increasing liquidity pressure in the close price for trades that are executed in the close auction. Each passive investor that chooses to go to the close auction for a lower fee ignores her impact on the price, and such an externality in the aggregate makes passive investors worse off.

How could the "guaranteed close" improve price discovery? Insights from existing studies that suggest dark pools[6] can improve price discovery (Zhu, 2014) cannot be directly applied to "guaranteed close". In Zhu (2014)'s model, since dark pools do not absorb excess order flows, informed traders, who are more likely to trade in the same direction as each other, face a higher execution risk in dark pools relative to uninformed traders. Hence dark pools would concentrate the informed traders on the exchange and improve price discovery. However, there is no execution risk in "guaranteed close". Also importantly, while dark pools typically passively match buyers and sellers and do not take their own positions, banks actively trade on the order flow information they received in "guaranteed close", which might also affect price discovery.

[6]Dark pools are equity trading systems that do not publicly display orders. They typically passively match buyers and sellers at exchange prices, such as the midpoint of the exchange bid and offer.

To explain our finding that "guaranteed close" improves price discovery, we build a model based on the single-period model in Kyle (1985). In our model, traders explicitly choose between two trading venues – that is, a bank and a market maker, based on transaction costs. The bank infers information from the net total orders it receives and trade. It trades in the same direction as the net total orders, when the proportion of informed orders in the orders sent to the bank is higher than those sent directly to the market maker. And it trades in the opposite direction otherwise. The bank's trading activity amplifies the informed orders eventually received by the market maker and hence improves price informativeness.

Uninformed traders and informed traders all have unit demand for a single asset with an uncertain liquidation value v. Each informed trader receives a signal about v and trades according to the signal. Each uninformed trader has to trade after receiving a liquidity shock. Both the market maker and the bank accept only market orders. Trading with the bank incurs a convenience cost, which is heterogeneous among traders. The market maker, being competitive, sets the price p. A trader can submit his order either to the market maker or the bank. While both venues execute his order at p, the transaction fees are different – the market maker charges φ_m per unit of asset traded and the bank charges φ_b. φ_m is exogenous and φ_b is chosen by the bank to maximize profit. While the bank executes orders for the traders, it also submits orders on its own account to the market maker for profit. The transaction fees and the distributions of traders' features are publicly known.

We solve the model in closed form, and find a linear Nash equilibrium where the net total orders of informed traders is linear in the asset's value v, the bank's trading strategy is linear in the net total orders it receives, and the market maker sets p as a linear function of the net total orders he receives. Under mild conditions, the bank participates with a fee $\varphi_b < \varphi_m$. The informativeness of p is higher than in an equilibrium without the bank. This is because the bank can trade profitably based on the net total orders received so long as the proportions of informed orders relative to uninformed orders are different between the two venues. The bank's trading activity amplifies the informed orders eventually received by the market maker, making the close price more informative. We decompose the effect of φ_m on the informativeness of p into two components – a volume effect and a ratio effect. The volume effect operates through the trading volume executed by the bank, and find that it is positive, meaning that higher trading volume at the bank leads to higher price informativeness. This is consistent with our empirical finding. The ratio effect instead operates through the ratio of informed relative to uninformed orders received by the bank. Finally, we find that dual trading of the bank unambiguously decreases the expected profit (before fees) of informed traders, and has a mixed effect on the expected profit (before fees) of liquidity traders.

Our empirical results and model are well-intended to study the effect of trading on order flow in today's US brokerage industry. Today's brokers offer assorted venue choices with different transaction fees. Some brokers like Robinhood charge zero transaction fee yet sell order flow data to sophisticated institutions for them to trade on. The combination of Robinhood and its partners who trade on order flow data is comparable to the "guaranteed close". Meanwhile, other brokers charge a higher transaction fee yet do not actively trade on order flow for profit, which is comparable to the "close auction". The brokerage firm Interactive Brokers even simultaneously offers two options: IBKR Lite and IBKR Pro. "If it is IBKR Lite with zero commissions ,..., we send them off to a market maker,..., and there is payment for order flow that comes back and you may not get as good of an execution ... If it is IBKR Pro, you will get better execution."(Steve Sanders, executive vice president of Interactive Brokers, link). In our empirical setting, designated market makers clear the market in a single auction and set the price. This fits nicely with the Kyle (1985) framework that has been successfully applied to understand the formation of prices generally. In our model, "guaranteed close" improves price discovery as long as the fraction of informed orders relative to uninformed order sent to the bank is different than that to the market maker. Without stipulating the reasons why the fractions of informed orders might be different, our model offers a powerful partial equilibrium result.

Let us finally caution that our model is stylized, with exogenous noise trading, and no entry of banks, among other assumptions. A dynamic model with multiple exchanges and multiple broker-dealers would be more realistic if extrapolating our result to other trading scenarios. In addition, our model is only meant to capture one aspect of "guaranteed close", namely its effect on price discovery. The other side of the coin, namely, the liquidity provision feature, especially whether it still provides liquidity when market conditions are volatile, is not studied. Future research may shed more light on these issues.

Literature Review

Our paper is related to three strands of literature. First, we add to the recent discussion on how the growth of passive investment strategies affects the market. Although market closures by themselves can generate endogenous time-variation in trading activity and price movements (Hong and Wang, 2000), Bogousslavsky and Muravyev (2021) find that the influx of ETF and index fund trades are key determinants of the volume at market close in recent years, and they adversely affect the informativeness of the close prices. Ben-David, Franzoni, and Moussawi (2018) find that ETF ownership is associated with higher volatility and more reversals for the index constituents. Baltussen, van Bekkum, and Da (2019) find that the growth of passive investment is associated with a decline in index return autocorrelation. Lines (2022) finds that when market volatility rises, portfolio tracking error also rises, which leads portfolio managers to rebalance their portfolios towards benchmark stocks, and this generates price effects. Baldauf, Frei, and Mollner (2021) study the manipulation of prices at market close. They build a model of financial contracting between a client, who wishes to trade a large position, and her dealer. Because of agency problem, market-on-close order is not the optimal contract for trading. Our paper focuses on the costs from transaction fees and price pressures, that are adversarial to passive investment strategies. We hope to inform policy attempts that design market structure to accommodate the growth of passive investment strategies.

Second, our paper contributes to the literature on dual trading, that is, broker-dealers strategically using customers' order flow information to trade on their own accounts. On the empirical side, Chakravarty and Li (2003) use proprietary audit trail transaction data to study dual trading in futures markets, and find that dual traders trade merely to supply liquidity and manage inventory, rather than trading against the customers. Barbon, Di Maggio, Franzoni, and Landier (2019) find that broker-dealers intermediating large stock portfolio liquidations spread order flow information to their clients. Using data provided by Robinhood, Kothari, Johnson, and So (2021) find that payment for order flow has saved retail investors unnecessary trading commissions, and improved the execution quality. We provide quasi-experimental evidence that trading on order flow can improve price discovery.

On the theory side, Röell (1990) builds a model where the broker observes only the trades of uninformed trader and trade on them. In Fishman and Longstaff (1992), the broker has private information about whether his customer is informed or not, and allow the customer and the dual-trading broker to trade at different prices. Sarkar (1995) finds that dual trading has no impact on discovery in a fully-revealing equilibrium, although it decreases net profits of informed traders and increases the utility of uninformed traders. Yang and Zhu (2020) and Huddart, Hughes, and Levine (2001) study the strategies of informed traders when there are back-runners, who partly infer informed traders' information from their order flow and exploit it in subsequent trading. The informed traders counteract back-runners by randomizing their orders (unless back-runners' signals are too imprecise), but back-runners unambiguously improve price discovery. These analyses are different from ours in terms of the economic questions and modeling approaches. In our model, the bank observes only the net orders and partially infer information from order flows. Also, we explicitly model the venue choices, and all the trading happens in one period and the price is set only once.

Finally, our paper is related to the literature on dark pools and alternative trading venues. Zhu (2014) shows that adding a dark pool alongside the exchange can improve price discovery due to the dark pool's execution risk. Buti, Rindi, and Werner (2017) find that dark pools may have adverse effects on market quality, since dark pools reduce the number of limit orders that provide liquidity on the exchange. Ernst, Sokobin, and Spatt (2021) find that market participants learn from the publication of off-exchange transactions, and the off-exchange orders are informationally-motivated and contribute to price discovery. Other models on trading venue choice include Hendershott and Mendelson (2000) and Ye (2010). Their models either do not model asymmetric information, or do not allow all the agents to freely select venues, and do not consider the commission fee difference. Chen and Duffie (2021) find that market fragmentation and exchange competition could lead to improvement in price discovery when all exchange prices are taken together. Brogaard and Pan (2021) provides evidence that that dark pool trading leads to greater information acquisition.

The remainder of this paper is organized as follows. In Section 3.1, we present the institutional background of close auctions on the primary exchanges and "guaranteed close" of the banks. Section 3.2 exhibits quasi-experimental evidence that "guaranteed close" improves price discovery. Section 3.3 introduces our model of dual trading and price discovery at market close. Section 3.4 concludes.

3.1 Institutional Background

3.1.1 Close auction

In this section, we introduce the mechanism of NYSE's close auction, whose characteristics are to a large extent shared by other exchanges like NYSE Arca and Nasdaq.[7]

Several types of orders can be used in NYSE's close auction, with the most common being market-on-close (MOC) and limit-on-close (LOC) orders. An MOC order is an unpriced order to buy or sell a security at the close price and is guaranteed to receive an execution. An LOC order sets the maximum price an investor is willing to pay, or the minimum price for which an investor is willing to sell. An LOC order priced better than the final close auction price is guaranteed to receive an execution. As shown in Table C.1 in the appendix, 65% of the orders in NYSE close auction are MOC orders and 14% are LOC orders.[8] The predominant use of MOC orders appears

[7]See Appendix A of Bogousslavsky and Muravyev (2021) for a detailed summary of Nasdaq's close auction mechanism, as well as NYSE's. Also, see "NYSE Open and Closing Auctions Fact Sheet", 2019 [link], "NYSE Arca Auctions Brochure", 2019 [link], and "Nasdaq Open Close Quick Guide", 2019 [link]

[8]Another 18% are Closing D Orders, a special order accessible only to NYSE floor brokers.

to be the consequence of benchmarking strategies conducting trades at market close price, regardless of what the price will be.

From 6:30 am in the morning, MOC and LOC orders can be entered. Existing MOC and LOC orders can be canceled until 3:50 pm. At 4:00 pm, the regular session trading ends and the close auction commences. The method for determining the close prices follows two principles: (1) maximize the number of shares that can be executed in the close auction; (2) minimize the difference between the close price and a reference price if multiple close prices satisfy principle (1). The auction effectively aggregates the supply and demand curve constituted by the MOC and LOC orders, and the transaction price and trade volume is determined by the intersection of the two curves.

The Designated Market Makers (DMMs) play an important role in the close auction. They set the closing price at a level that satisfies all interest that is willing to participate at a price better than the close auction price, and supply liquidity as needed to offset any remaining auction imbalances that exist at the closing bell. That means market-on-close orders are guaranteed to be executed.

3.1.2 "Guaranteed close" service

Investment banks such as Goldman Sachs, Morgan Stanley, Credit Suisse Group AG, and UBS Group AG, started a "guaranteed close" service around 2016.[9] Investors looking to buy or sell shares of a stock can get a guarantee from the bank to execute their orders at the close price set on the corresponding primary exchange, where the stock is listed. As an investor, using "guaranteed close" is equivalent to sending a MOC order to the close auction in execution outcomes, but paying a lower fee. People familiar with the matter told us that Goldman Sachs recently cut the fee charged to broker-dealers to zero (although buy-side clients still pay a fee).

At 4:00 p.m., the bank pairs the buyers with the sellers of the stock. For the unmatched orders, it can either send them to the exchange or take the other side

[9] This is basd on multiple sources. We do not, however, have a comprehensive list of the venues that conduct "guaranteed close". For the top ten alternative trading systems by total volume (during both regular hours and market close), see Table 4 in "Staff Report on Algorithmic Trading in U.S. Capital Markets", 2020, SEC.[link]

of the trade itself, storing the extra shares or short interest on its books overnight. The banks trade alongside the client orders for profit. This is one way to cover the bank's cost of providing liquidity, and is documented by the following excerpts from the documentation of "guaranteed close" Morgan Stanley and Goldman Sachs sent to their clients.

Morgan Stanley:[10]

> "When we accept an order for execution on a guaranteed benchmark basis (for example, a guaranteed opening, closing, volume weighted average price or other guaranteed transaction), we will typically attempt to offset the risk incurred as a result of such guarantee by transacting in the market on a principal basis, or accessing internal liquidity sources, in the benchmark security or a related instrument, although we may choose not to perfectly hedge our exposure."

Goldman Sachs:[11]

> "We offer client facilitation services, which are typically used by clients to obtain liquidity or a guaranteed execution price. When you use our client facilitation services, we may also effect transactions as agent, as principal (including trading as a market maker or liquidity provider to other clients and trading to manage risks resulting from client facilitation activities), or in a mixed capacity."

Also, Morgan Stanley claims to split the profit from dual trading with the client:

> "In accordance with market standards and best practices, we strive for allocations between client and principal orders that are fair and equitable.[...]Challenges presented by the current market structure and limitations of certain market centers and trading system may in some cases render a precisely even split impracticable."

[10]See "A Message to Morgan Stanley's U.S. Institutional Equity Division Sales & Trading Clients regarding U.S. Equity Order Handling Practices", 2019, [link]

[11]See "Cash Equities Order Handling Procedures of Goldman Sachs (Asia) L.L.C.", 2019,[link]

Who are the users of the "guaranteed close"? Anecdotal evidence suggests the "guaranteed close" is used by index-fund managers including Vanguard Group and BlackRock Inc., as well as some smaller broker-dealers (WSJ (2018)). In general, institutional investors and broker-dealers are primary users of alternative trading systems.[12] In the next section, we present evidence that stocks with higher ETF/index fund ownership are more heavily traded at the "guaranteed close".

"Guaranteed close" is different from dark pools in that it has no execution risk, while in the latter matching depends on the availability of counterparties and some orders from the "heavier" side of the market will fail to be executed (Zhu, 2014). Also, it is also in nature different from the recently approved Cboe Market Close program providing a lower-fee venue to trade at the market close price set by NYSE and Nasdaq. In Cboe Market Close, traders can enter, cancel or replace MOC orders only before 3:35pm. After that, the system would match for execution all buy and sell MOC orders entered with execution priority given based on time-received. But any remaining balance of unmatched shares would be cancelled and returned to the traders. That is, Cboe Market Close only pre-matches some non-price-forming orders.[13]

3.2 "Guaranteed Close" and Price Discovery: Evidence from NYSE Fee Cut

In this section, we empirically study the relationship between "guaranteed close" and the informativeness of market close price. We exploit the quasi-experimental setting of the NYSE close auction fee cut in January 2018 and use a difference-in-differences approach to establish causal evidence.

[12]See Section IV in "Staff Report on Algorithmic Trading in U.S. Capital Markets", 2020, SEC.[link]

[13]See WSJ article "SEC Decision on 4 p.m. Closing Trades Deals Blow to NYSE, Nasdaq " at [link], and SEC Release No. 34-88008 at [link].

3.2.1 NYSE close auction fee cut

In January 2018, the NYSE reduced the close auction fee for market-on-close (MOC) orders, which was seen as an attempt to keep clients from choosing the banks' low-cost "guaranteed close" service.[14] Indeed, the NYSE only reduced the fee for MOC orders but kept the fee for limit-on-close (LOC) orders unchanged. The NYSE's claimed intention is to encourage higher volumes of MOC orders at the close. This fee cut was given with short notice. On December 21, 2017, the NYSE announced the plan of fee changes intended to be effective January 2, 2018. On January 8, 2018, NYSE filed with the SEC about the change in close auction fees.[15]

In NYSE's close auctions, the amount of fee broker-dealers need to pay for the MOC orders depends on the trading volume – the per-share fee is lower for a higher trading volume. Specifically, there are three tiers. Tier 1 rates would be available for a broker-dealer who in the prior three billing months executed (1) an ADV (average daily volume) of MOC activity on the NYSE of at least 0.45% of NYSE CADV (consolidated average daily volume), (2) an ADV of total close activity (MOC/LOC and executions at the close) on the NYSE of at least 0.7% of NYSE CADV , and (3) whose MOC activity comprised at least 35% of the its total close activity. Tier 2 rates would require a lower ADV of MOC activity and ADV of total close activity. Those who don't meet the requirements for Tier 1 and Tier 2 rates are subject to the Non-Tier rate.

The fee cut reduced the Tier 1 rate and Tier 2 rate significantly, but left the Non-Tier rates almost untouched. Specifically, Tier 1 rate is reduced from $0.0007 to $0.0004 (a 42.9% drop). Tier 2 rate is reduced from $0.0008 to $0.0005 (a 37.5% drop). But Non-Tier rate is reduced from $0.0011 only to $0.0010 (a 9% drop).

Given the differential changes for different tiers, we expect the fee cut to have differential impact on individual stocks. There are at least two reasons why the fee cut for broker-dealers pass through to individual stocks in a differential manner. First, ETFs and index funds may actively choose the venue at which they trade and are

[14]Alternatively, the fee cut may be seen as a reaction to the threat of CBOE's entry into the close auction [link], which was awaiting SEC decision back in 2018.

[15]See SEC No. 34-82563 [link]. Also, see NYSE trader update [link].

sensitive to transaction fees. The fee cut induced them to actively choose to trade with broker-dealers that operate in the close auctions, this is because their large trading volume easily help broker-dealers reach the Tier 1 and Tier 2 thresholds, allowing them to effectively take the lower transaction fees. Second, institutional investors may delegate the venue choice entirely to their broker-dealers – this is a reasonable assumption as institutions rely on broker-dealers to trade (Di Maggio, Egan, and Franzoni, 2021), form long-term relationships with their brokers, and concentrate their order flow with a relatively small set of broker-dealers (Goldstein, Irvine, Kandel, and Wiener, 2009). In that case, broker-dealers would have different yet relatively stable client base. Some provide services mainly to passive investors like ETFs and index funds, while others provide service mainly to other types of investors. Broker-dealers trading on behalf of passive investors are more exposed to the fee cut since they more likely reach the Tier 1 and Tier 2 thresholds. Once they went back to the close auctions, the stocks they usually facilitate to trade, that is, those with high passive ownership, will experience a larger drop of trading volume at "guaranteed close". These stocks are also ex-ante heavily traded at the "guaranteed close". Under both cases, and to the extent that banks trade on the order flows data they receive in the "guaranteed close", the fee cut is a plausibly exogenous shock to trading on order flow activities. Stocks that are ex-ante heavily traded at the "guaranteed close", and those with higher passive ownership, are more exposed to the fee cut.

3.2.2 Data and descriptive findings

Data construction Our main data source is the NYSE millisecond-level trade and quote data (TAQ) spanning from 2012 to 2019. We also leverage the WRDS Intraday Indicator Dataset (WRDS IID) built from the TAQ data. Our sample contains common stocks listed on NYSE and Nasdaq, with a price greater than $5 and a market capitalization greater than $100 million at the end of a month. Our main results come from the difference-in-differences analysis, which uses the sample of 1,217 NYSE stocks, spanning from January 2015 to December 2019. Table 3.1 reports the summary statistics of all the variables we used in this sample.

The close price, open price, and close auction volume for each stock and each day are from WRDS IID. The close auction volume is measured by TAQ trades with the sale condition of 6 (Closing Print), that occurs on a stock's primary listing exchange (SEC DERA, 2017) – for example, an NYSE-listed stock's close volume executed in NYSE.[16] The close prices are the recorded transaction prices of these trades. The market open prices are measured by the recorded prices of the trades with the sale condition of O (Market Center Opening Trade). For each stock each day, we also calculate from TAQ the volume-weighted average price in the last 5 minutes, $p_t^{close5m}$, in the last 15 minutes, $p_t^{close15m}$, and in the first 5 minutes, p_t^{open5m}, of the regular session. For these calculations, we exclude invalid or erroneous trades that were later canceled or changed.

We proxy the trade volume via "guaranteed close" service by the trade volume using off-exchange market-on-close (MOC) orders, given that "guaranteed close" is anecdotally the major venue accepting off-exchange market-on-close orders. To be precise, we use the term "off-exchange MOC volume" in the formal analysis. To measure the off-exchange MOC volumes, we consider all the trades from TAQ that are not cancelled or corrected and occur between 4:00 p.m.–4:10 p.m. EST at the official market close price determined by the close auction. This off-exchange MOC volume has been validated against two regulatory datasets with more detailed information – the FINRA Trade Reporting Facility data, and the FINRA-provided Audit Trail data, to ensure that it is from MOC orders.[17] The two regulatory datasets identify off-exchange executions by venue and trace the executions back to the original orders. Our off-exchange MOC volume would be almost identical if measured by the regulatory datasets. As shown in Figure 3-1, the trading volume in "guaranteed close" has risen sharply from 2016, and reached almost 30% of the trading volume in close auctions.

ETF ownership data is obtained from ETF Global. Index fund ownership data is obtained from the CRSP Mutual Fund database. We closely follow Ben-David et al.

[16]NYSE stocks technically can also be traded in Nasdaq closing crosses, which is yet empirically rare.

[17]See SEC report SEC DERA (2017).

(2021)[18] and Dannhauser and Pontiff (2019) to identify passive index funds in the CRSP Mutual Fund database. Our sample contains 918 index funds in 2016.

For other variables, we obtain market capitalization and earnings announcement days from CRSP. Measures of volatility, liquidity, and retail and institutional order flows are from WRDS IID. They include the trade-based intraday volatility during market hours, total trade volume during market hours, total retail trade volume following Boehmer et al. (2021), total volume of trades $\geq$ 20K in value, and total volume of trades $\geq$ 50K in value. The last two variables are proxies for institutional trades. Lee and Radhakrishna (2000) show 53% of institutional trades are above $20,000 in value; Bhattacharya et al. (2007) and Shanthikumar (2003) use $50,000 dollar value-based cutoff. When trades with value exceeding these cut-offs are interpreted as institutional trades, not surprisingly, we find large institutional presence at banks' "guaranteed close", as large-value trades proliferate, although smaller-value trades also exist.[19] Overnight beta is individual stock's overnight return's loading on overnight market return, estimated quarterly using CAPM regression. Spread is time-weighted percent quoted spread during market hours, used as a placebo variable to validate the NYSE fee cut as an exogenous shock.

Finally, daily variables are all aggregated to monthly by taking averages. Given the presence of salient outliers in daily data, we winsorize the overnight returns at 5% tails before taking the monthly averages.[20] We further winsorize monthly observations of all the variables at 2% tails.

Price informativeness measure Price informativeness is usually measured by how well a price tracks the fundamental value of an asset. Close price is, however, unique. Being the last price of the day, it aggregates all the information of the day and is generated by auctions with substantial liquidity and turnovers. Prices right before the close are not good measures of the fundamental value since they

[18]We thank the authors for making their code publicly available as supplementary data to the Review of Financial Studies.

[19]Figure C-1 in the appendix plots the distribution of sizes of the off-exchange MOC trades for the stock AAPL. Both extremely large orders and smaller trades exist, and large orders are frequent.

[20]Our results are qualitatively similar if we winsorize at, for example, 2% tails, but noisier.

do not incorporate all the information of the day, and in some cases are prone to manipulation given the lower liquidity. Neither are prices generated by after hours trades since these trades are infrequent and slim. We recognize the next day's open price as the fundamental value of a stock, to which we compare the close price. In our model and most theoretical literature, fundamental value of the stock is realized in a future period at which investors can liquidate the stock. Aligning with this, for close price, a reasonable measure for the fundamental value it's reflecting would be the next day's open price.[21]

Therefore, our (inverse) measure of the informativeness of the close price is the mean squared error (MSE) of the close price relative to the next day's open price:

$$\text{MSE} = \mathbb{E}\left(\frac{p_{t+1}^{\text{open}} - p_t^{\text{close}}}{p_t^{\text{close}}}\right)^2 \tag{3.1}$$

where p_t^{close} is the close price of day t, p_{t+1}^{open} is the open price of day $t+1$. A lower MSE corresponds to better price informativeness.

Note that we scale the difference by the close price itself. This scaling takes into account that prices at higher levels mechanically vary more, and transforms the "closeness" to the fundamental value into return space. Our results are robust to the scaling factor, where we calculate the MSE's using volume-weighted price in the last 5 minutes (p_t^{last5m}), last 15 minutes ($p_t^{last15m}$) or first 5 minutes (p_{t+1}^{open5m}), as the scaling factor.

Descriptive findings Who trade at the banks' "guaranteed close"? In line with anecdotal evidence, we show that the off-exchange MOC volume of a stock is closely related to passive ETF/index fund ownership in the cross-section. Figure 3-2 shows the bin-scatter plots of off-exchange MOC volume as a fraction of the close price volume against ETF ownership, and against index fund ownership. The binscatter

[21] Indeed, next day's open price, beyond the existing information at the market close, also incorporates overnight information, hence is not a perfect measure for the fundamental value *at* the market close. Nevertheless, our results are robust to excluding data within 1 day from earnings announcements days, where overnight information is strong and potentially causes significant changes in the fundamental value.

plots are based on OLS regression, controlling for a battery of confounding factors including time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta. The results are robust to not including those controls, but noisier. The bin-scatter plots suggest that as ETF ownership or index fund ownership increases by 1 percent, off-exchange MOC volume as a fraction of the close price volume increases by 0.3 percent.

3.2.3 Difference-in-differences: The effects of NYSE fee cut

We use a standard difference-in-differences strategy to estimate the effect of the NYSE close auction fee cut in January 2018 on the MSE of close prices. As discussed in Section 3.2.1, stocks ex-ante more heavily traded at "guaranteed close", and those with higher ETF/index fund ownership are most exposed to the fee cut.

We first drop the NYSE stocks that participated as treated stocks in the 2016 Tick Size Pilot Program. The Pilot increased the tick size for select small stocks, which may force the close price to deviate from the fundamental price. The Pilot ends on September 28, 2018, when the tick size requirements are repealed. This may interfere with the NYSE fee cut quasi-experiment.[22] We rank the remaining stocks by the average fraction of close-price volume (i.e., volume traded at the close price) executed off exchange in 2017. The treated group consists of stocks that rank at top 50% – they are more exposed to the NYSE fee cut. The control group consists of stocks that rank at the bottom 50%. Treated-group stocks have on average a market cap of \$22 billion, and ETF/index fund ownership of 8.7% and 8.4%, respectively. Control-group stocks have on average a market cap of \$7.1 billion, and ETF/index fund ownership of 4.8% and 3.8%, respectively.

To validate that the fee cut induces large drop in off-exchange MOC volumes for treated stocks (i.e., the first stage), we run a regression of the following form and plot

[22]In fact, our results are robust to including those stocks.

the coefficients β_k's.

$$\frac{\text{off-ex MOC volume}_{i,t}}{\text{close price volume}_{i,t}} = \alpha_i + \lambda_t + \sum_k \beta_k \text{Treat}_i \cdot \mathbb{I}_{t=2017m11+k} + \Gamma X_{i,t} + \epsilon_{i,t}. \quad (3.2)$$

where Treat_i is a dummy variable that takes the value one if stock i is in the treated group, and takes the value zero otherwise. $\mathbb{I}_{t=2017m11+k}$ is a dummy variable that takes the value one if time t is k months after 2017m11, and takes the value zero otherwise. Stock fixed effects, time fixed effects, log(market cap), log(total trade volume), volatility, and other control variables, including log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta are controlled in this regression.

Indeed, it is salient in panel (a) of Figure 3-3, off-exchange MOC volume increased much more rapidly in the treated group prior to the NYSE close auction fee cut. But almost immediately after the fee cut, the off-exchange MOC volume of the treated group stopped growing and started declining, relative to the control group.

Now we study the effect on the informativness of the close price. The major difficulty in difference-in-differences analyses involves separating out pre-existing trends from the dynamic effects of a policy shock. To avoid confounding the two, we first test for pre-existing trends in the MSE measure of price informativeness. Specifically, we run a regression of the same form as Equation (3.2), but replacing the outcome variable with the MSE measure of price informativeness.

Panel (b) of Figure 3-3 plots the coefficients β_k and the corresponding confidence intervals. We have three findings. First, the MSEs of the treated and control group moved almost perfectly in tandem from 20 months before the NYSE fee cut, so the parallel trend assumption appears to hold well. Second, from September 2015 to March 2016, there appeared to be a drop in MSE for the treated stocks. This same time period was accompanied by the start of the ramp-up of off-exchange MOC volume of treated stocks relative to control stocks. This seems consistent with that off-exchange MOC activity improves close price informativeness, although a causal interpretation is unwarranted since low MSE stocks may select into the treated group. Third, the MSE of treated stocks increased substantially and remained differentially higher than the control stocks after the fee cut.

Finally, we run the following standard difference-in-difference regression.

$$\text{MSE}_{i,t} = \alpha_i + \lambda_t + \beta \text{Treat}_i \cdot \mathbb{I}_{t \geq 2018m1} + \Gamma X_{i,t} + \epsilon_{i,t} \tag{3.3}$$

where $\mathbb{I}_{t \geq 2018m1}$ is a dummy variable that takes the value one if time t is after January 2018, and all the other variables are as defined before in Equation (3.2).

The coefficient of interest is β, which measures the differential change in MSE for the treated stocks and control stocks, holding constant stock-level time-varying characteristics, as well as stock and time fixed effects. Besides MSE (close price as the scaling factor), we also used as a dependent variable the MSEs using volume-weighted price in the last 5 minutes (p_t^{last5m}), last 15 minutes ($p_t^{last15m}$) or first 5 minutes (p_{t+1}^{open5m}) as the scaling factor, as well as the median of $(\frac{p_{t+1}^{open} - p_t^{close}}{p_t^{close}})^2$ (median SE). We also adopt two placebo variables as dependent variables. They are intraday volatility during market hours, and quoted bid-ask spread during market hours. Both of them are important measures of market conditions for a stock, but given that they are measured during market hours, they should not be affected by the NYSE fee cut. To account for serial correlation and stock-specific random shocks, we cluster standard errors at the stock level in all specifications.

Table 3.2 shows the difference-in-difference regression results. We see that the fee cut increased significantly the MSE measures and the median SE of treated stocks, meaning the informativeness of treated group stocks relative to the control group have been worsened by the shock. Not surprisingly, the fee cut did not seem to affect the intraday volatility and quoted spread of the treated stocks.

3.2.4 Robustness checks

We take several additional steps to ensure the validity of our research design and the robustness of our estimates.

Alternative treatment designation One potential concern with the difference-in-difference results is about the designation of the treatment group. We argued that stocks more heavily traded ex-ante at the bank's venue would be more exposed to the fee cut, since they have higher ETF/index fund ownership, while ETF/index funds are more sensitive to fees and in the meantime received larger NYSE fee cuts. We now verify this idea by defining the treatment group with ownership by ETFs and index funds. Specifically, we classify stocks ranked in the top 50% in the sum of ETF and index-fund ownership into the treated group, and the remaining stocks into the control group. Table 3.3 gives the estimation results which are all consistent with the main results.[23]

Excluding earnings announcement days We take the next day's open price as the fundamental value to which we compare the close price. Next day's open price, beyond the existing information at the market close, also incorporates overnight information, hence is not the fundamental value *at* the market close. For most stocks, earnings announcements are the major overnight information that affects an individual stock's price. One potential concern is that treated stocks may experience stronger earnings news after 2018. To address this, we exclude the data within 1 day, that is, the [-1,0,1] days, from earnings announcements days, before conducting the difference-in-difference estimation. As shown in Table 3.4, our results are robust to excluding these dates.

Matching specification Another concern regarding our estimates is that stocks in the treated and control group might be very different in size and many other characteristics, although they are already all NYSE stocks. To alleviate this concern, we conduct a matched sample approach. We match the stocks based on pre-treatment values of log(market cap), log(trading volume), and intraday volatility. For each stock, the closest matching control stock is chosen (with replacement) according to the Mahalanobis distance of the three variables, to constitute the matched control

[23]Quoted spreads of treated stocks are estimated to increased by the fee cut, but not very statistically significant.

group. Table 3.5 shows that the matched sample yields quantitatively similar results compared to the nonmatched sample. Table 3.6 shows a set of balance test results for the nonmatched samples and the matched samples.

3.2.5 Discussion

How does the NYSE fee cut affect the welfare of index fund/ETF investors? We conduct the following back-of-envelop calculation to shed light on this question.

Suppose there is an index fund that invests in the universe of NYSE stocks using a value weighted portfolio. For simplicity, let us assume the index fund always submit all the orders to the NYSE close auction both before and after the NYSE fee cut. If we define $\Pr(reversal)$ as the probability that a stock's overnight movement is against the index fund's trade, that is, the stock price goes up yet the index fund sold it at market close, and that the stock price goes down yet the index fund bought it at market close. Let us denote the overnight return as r, then the overnight profit of the index fund is

$$-\mathbb{E}(|r||\text{reversal})\Pr(\text{reversal}) + \mathbb{E}(|r||\text{no reversal})\Pr(\text{no reversal}) \tag{3.4}$$

Empirically we define a stock has an overnight reversal for the index fund if the return from last 5min to close and the return from close to next day's open have opposite sign. The distribution of overnight return r conditional on reversal is empirically similar to the distribution conditional on no reversal. Hence we assume $\mathbb{E}(|r||\text{reversal}) = \mathbb{E}(|r||\text{no reversal})$, and the overnight profit is $[1 - 2\Pr(\text{reversal})]\mathbb{E}(|r|)$.

For each dollar of transaction, the effect of NYSE fee cut on the index fund's profit is

$$[1 - 2\Pr(\text{reversal})]\Delta\mathbb{E}(|r|) + \text{fee cut} \times \frac{\text{Total shares}}{\text{Total MV}} \tag{3.5}$$

where we estimate $\Delta\mathbb{E}(|r|)$ to be close to 5 bps, based on regression coefficient in Column 6 of Table 3.2. For the universe of NYSE stocks, our calculation suggests that $\frac{\text{Total shares}}{\text{Total MV}} = 0.18$. The relative fee cut between treated and control group is

(3bps - 1bps) = 2bps. As long as Pr($reversal$) > 53.6%, the welfare of the index fund declines. That is, the benefit of fee cut is outweighed by the cost of the increasing liquidity pressure in the close price.

In our data, Pr(reversal) is close to 60%. In fact, the deviation between the last bid-ask midpoint in regular session trading and the close price may better reflect the index fund's trading direction. Bogousslavsky and Muravyev (2021) find that the deviation between close price and last bid-ask midpoint almost always reverses overnight. In that case, Pr(reversal) = 1.

3.3 A Model of Dual Trading and Price Discovery at Market Close

In this section, we present a model to better understand the empirical findings. The model is closely tailored to our empirical setting, where traders explicitly choose whether to trade on an exchange, or with a bank, both offering guaranteed execution at the close price, which is set by the market maker on the exchange. We find that the informativeness of a stock's close price is improved when the bank trades based on total orders received, and is increasing in the total volume traded with the bank under mild conditions on parameters.

This result is driven by the bank's trading activity. As long as there is a difference between the two venues in the ratio of informed orders to uninformed orders, the bank can trade profitably solely by observing the net orders it received. In particular, it trades in the same(opposite) direction as the net orders received when the orders contain a higher(lower) proportion of informed orders than that of orders traded on the exchange. While informed traders are worse off due to the increase in price informativeness, the welfare changes of uninformed traders are more nuanced and discussed in details. In order to provide closed-form solutions and sharpen its predictions, our model adopts several simplifying assumptions. We end the section with a thorough discussion of these assumptions, and explain how our model can shed light on other scenarios such as the effect of Robinhood-like e-brokers on price discovery.

3.3.1 Model setup

There are three periods, denoted by $t = 0, 1, 2$. A single asset has an uncertain liquidation value $v \sim N(0, \sigma^2)$, which is realized and publicly revealed at $t = 2$.

There are m uninformed traders and n informed traders. Each trader buys or sells one unit of the asset when participating.[24] A trader can submit her order either to a market maker or a bank, both executing her order at the same close price p announced at $t = 2$ by the market maker. The market maker and the bank are considered as competitive sectors and break-even in expectation. The market maker determines the close price p. While executing ordering for traders, the bank also submits orders to the market maker itself. All market participants are risk neutral and use market orders only.

A trader is charged a transaction fee φ_m when trading with the market maker and φ_b when trading with the bank per unit of asset traded. φ_m is exogenous, representing the fee charged by the exchange that changes infrequently in practice. φ_b is determined by the bank's break-even condition.[25] Fees are announced at the beginning of $t = 0$. There are convenience costs for traders to trade with the bank, which represent unmodeled factors that deter investors from trading with the banks such as the cost of building connections and contracting with the bank.[26] Convenience costs are heterogeneous among traders, following cumulative distribution functions $G_u : [0, \infty) \to [0, 1]$ and $G_x : [0, \infty) \to [0, 1]$ for liquidity traders and informed

[24]The assumption of unit demand can reflect capital constraints that limit a trader's maximal trade size. The assumption of unit demand is frequently observed in models involving venue choice, for example, Zhu (2014); Hendershott and Mendelson (2000).

[25]These assumptions align with the institutional details in practice, where many banks are competing in the "guaranteed close" business while the primary exchanges are monopoly providers of the close auctions for stocks on each exchange. This does not conflict our assumption that the market maker break-even, since the fee is charged by the exchange instead of the market makers who compete to provide liquidity in the close auction.

[26]In practice, investors can use the "guaranteed close" service only on brokerage accounts at the banks that provide such service. Any incentives that motivate them to choose other brokers over these banks are considered as the convenience costs here. Brokers offer a variety of services to investors by providing efficient execution, market research, and order flow information (Di Maggio, Egan, and Franzoni, 2021). Services are heterogeneous among brokers and investors differ in their demand for these services, so they choose different brokers and are more or less willing to move their portfolios to an account served by these banks to get access to the "guaranteed close business".

traders respectively. Traders make their trading decisions after observing the fees.

The figure below shows the timeline of the model. At the beginning of $t = 0$, the bank announces fee φ_b. Uninformed traders receive liquidity shocks such that they have to buy or sell one unit of the asset.[27] Assume that each uninformed trader has equal probabilities $\frac{1}{2}$ to be a buyer or a seller and liquidity shocks are independent among traders. Then the net liquidity orders u approximately follows a normal distribution $N(0, \sigma_u^2)$ where $\sigma_u^2 = m$ when m is large enough. [28] Random variables v, u are independent and whether a liquidity trader receives a positive or negative shock is independent of her convenience cost for trading via the bank. Each informed trader receives a signal about the value of v that is uniformly distributed $s_i \sim U(v - \sigma_s, v + \sigma_s)$.

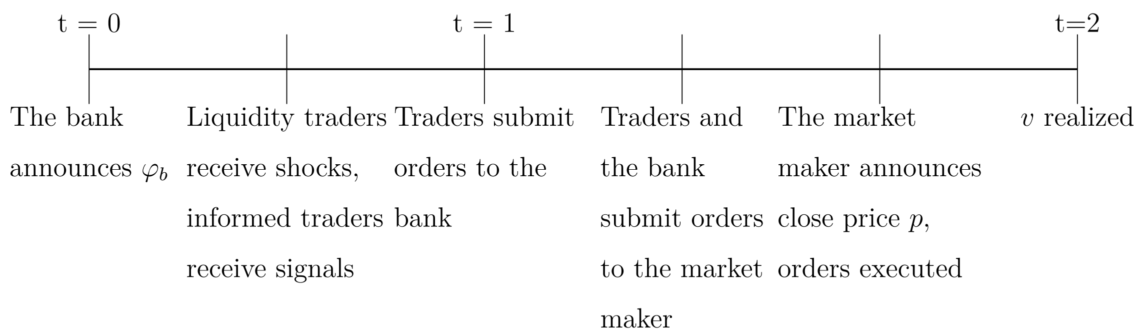


Figure: Model Time Line

At the beginning of $t = 1$, traders choose where to trade and submit their orders. After receiving orders from the traders, the bank can take its own position to trade along or against the net orders received. The bank cannot observe whether an order is informed or not, but knows the composition of its clients given the distributions of traders' features. The bank matches buy and sell orders, including its own position, and submits the remaining order imbalance to the market maker.

[27]We suppose it is very costly for an uninformed trader not to meet her liquidity need, so that she always trades when receiving a shock. The only decision left to be made is the venue choice.

[28]The random number of buyers can be written as $B(m, 1/2)$, where $B(,)$ denotes a binomial distribution. When m is large enough, it can be approximated by a normal distribution $N(m/2, m/4)$. Then the net total orders of these traders $B(m, 1/2) - (m - B(m, 1/2))$ can be approximated by $N(0, m)$. A common rule for this approximation to be appropriate is that everything within 3 standard deviations of its mean is within the range of possible values, which is satisfied in our case as long as $m > 9$.

At the end of $t = 1$, the close trading ends. The market maker does not observe whether an order is from an informed trader, a liquidity trader or the bank. After collecting all the orders from traders and the bank, the competitive market maker announces the close price p, which equals the expected asset value conditional on the net orders he receives. All orders are then executed at this close price p. Then at $t = 2$, the asset's value v is realized and all participants receive their payoffs.

3.3.2 Equilibrium

An equilibrium consists of the quoting strategy of the market maker, the fee charged by the bank, and the trading strategies of the traders and the bank. In equilibrium, the market maker breaks even in expectation, setting the price p that equals his expected asset value. The bank maximizes its trading profit and breaks even in total profit in expectation. Traders maximize their expected profits. We solve for the equilibrium backwards along the timeline, starting from trading strategies at $t = 1$ given the bank's fee φ_b announced at $t = 0$.

Traders' venue choices We first characterize the venue choices of the traders. A trader's expected payoff from trading $q_i = \pm 1$ unit of the asset with the market maker is

$$\mathbb{E}_i[v - p]q_i - \varphi_m \tag{3.6}$$

Likewise, the trader's expected payoff from trading via the bank is

$$\mathbb{E}_i[v - p]q_i - (\varphi_b + \gamma_i) \tag{3.7}$$

where $\mathbb{E}_i[\cdot]$ denotes the expectation of trader i, γ_i denotes the convenience cost for trader i to trade with the bank.

Traders choose the trading venue with higher expected payoff, or equivalently, lower total trading cost. Hence a trader trades via bank if

$$\gamma_i < \varphi_m - \varphi_b \tag{3.8}$$

and trades with the market maker otherwise. Let α be the equilibrium fraction of liquidity traders who choose to trade via the bank. The remaining fraction $1-\alpha$ of liquidity traders send their orders to the market maker. Similarly, let θ be the fraction of informed traders who trade via the bank, and the remaining fraction $1-\theta$ of informed traders send their orders to the market maker. Then we have $\alpha = G_u(\varphi_m - \varphi_b)$, and $\theta = G_x(\varphi_m - \varphi_b)$.

Let u_b, u_m be the net total orders submitted to the bank and the market maker by the liquidity traders. Since liquidity orders are independent among traders, u_b, u_m are independent and approximately normal, and the variances of the net liquidity orders are linear in the fractions of traders who trade via each venue. Then, since $u \sim N(0, \sigma_u^2)$, we have

$$\begin{aligned} u_b &\sim N(0, \alpha\sigma_u^2) \\ u_m &\sim N\left(0, (1-\alpha)\sigma_u^2\right) \end{aligned} \tag{3.9}$$

An informed trader buys if $\mathbb{E}[v-p|s_i] > \min(\varphi_m, \varphi_b+\gamma_i)$ and sells if $-\mathbb{E}[v-p|s_i] > \min(\varphi_m, \varphi_b+\gamma_i)$. When σ_s is large enough, the total informed orders is approximately $x = \eta v$, where $\eta = \frac{m}{\sigma_s}$.[29] Then given the fractions of informed traders trading via each venue shown above, the net total informed orders submitted to the bank (x_b) and the market maker(x_m) are

$$\begin{aligned} x_b &= \theta\eta v \\ x_m &= (1-\theta)\eta v \end{aligned} \tag{3.10}$$

[29]Let $\pi(s_i) = \mathbb{E}[|v-p||s_i]$ and guess it is strictly increasing in $|s_i|$, which can be verified in the following equilibrium. Then $\pi(s_i)$ is reversible for $s_i \geq 0$. Given σ_s, the total informed orders is $x = \frac{m}{\sigma_s}v$ for $v \in (-\sigma_s + \pi^{-1}(\mu_m), \sigma_s - \pi^{-1}(\mu_m))$. Since v is normal, the tail probabilities converges to zero when σ_s goes to infinity so the approximation is valid.

Bank's trading strategy Given the liquidity and informed orders submitted to each trading venue, we now characterize the bank's trading strategy. The net total orders received by the bank is $y = u_b + \theta\eta v$, with variance $\sigma_y^2 = \alpha\sigma_u^2 + \theta^2\eta^2\sigma^2$. When $\sigma_y^2 > 0$, that is, the bank receives a positive measure of orders, it takes its own position $d(y)$ based on its expectations given the net orders received and the distributions of traders' features. $d(y)$ can be either along or against the direction of the net orders the bank receives. The bank's expected trading profit is

$$\mathbb{E}[v - p|y]d(y) \tag{3.11}$$

Here we assume that the bank's position $d(y)$ does not incur changes in transaction costs to preserve tractability of the model. [30]

We restrict our attention to rational expectations equilibria in which the price is linear in the net orders received by the market maker. Guess that the market maker applies a linear price setting rule $p = \lambda z$, where $z = x+u+d$ is the net orders received by the market maker. Then we can write the bank's trading problem as

$$\max_d E[v - p|y]d = \left(\left[(1-\lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2} - \lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right]y - \lambda d\right)d \tag{3.12}$$

By solving the bank's optimal trading strategy and combining it with the market maker's price setting rule, $p = E[v|z]$, where $z = u + x + d$ is the net orders received by the market maker, we get a linear equilibrium as in the following.

Proposition 3.1 *If $\sigma_y^2 > 0$, a linear Nash equilibrium of the model described above is given by*

$$p(z) = \lambda z \tag{3.13}$$

$$d(y) = Ky \tag{3.14}$$

where $\lambda = \frac{(1+K\theta)\eta\sigma^2}{((1+K)^2\alpha+(1-\alpha))\sigma_u^2+(1+K\theta)^2\eta^2\sigma^2}$, $K = \frac{-B+\sqrt{B^2-4AC}}{2A}$, *with* $A = \sigma_y^2, B = \theta\eta^2\sigma^2 +$

[30] The incentive to save transaction fees paid to the exchange may make the bank more willing to provide liquidity with its own holdings and trade against the net orders received.

$(2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$ *when* $\theta > 0$*, and* $K = -\frac{1}{2}$ *when* $\theta = 0, \alpha > 0$.[31]

Proofs are in the Appendix.

In this equilibrium, the trading strategy of the bank and the quoting strategy of the market maker are both linear in the net total orders they receive, and determined only by the exogenous parameters and the bank's fee announced at $t = 0$. Some interesting results about the bank's trading strategy are worthwhile to note here. First, K has the same sign as $\theta - \alpha$. When $\theta > \alpha$, that is, the fraction of informed traders trading via the bank exceeds that of the liquidity traders, then $K > 0$ — the bank takes its own position in the same direction as the net orders it receives. When $\theta < \alpha$, that is, the fraction of liquidity traders trading via the bank exceeds that of the informed traders, then $K > 0$ — the bank takes its position in the opposite direction to the net orders it receives. This is because if $\theta > \alpha$, the bank perceives the orders received to be more informative than orders on the whole market, so following them is profitable. If $\theta < \alpha$, the bank perceives the orders received to be more noisy than orders on the whole market, so trading against them is profitable due to the price impact of liquidity traders. Second, when $\theta = 0$, the bank's trading strategy is $K = -\frac{1}{2}$. In this case, the bank only receives liquidity orders, and makes a profit by providing liquidity to them. Third, when $\alpha = 0$, $d = \frac{1}{2}\left(\sqrt{\eta^2 + 4\frac{\sigma_u^2}{\sigma^2}} - \eta\right)v$, the bank's trading position does not depend on θ as long as it is positive. This is because when the bank only receives informed orders, it can perfectly infer the value of the asset from just a small mass of informed orders. Then it behaves in the same way as a large informed trader who perfectly observes the asset value.

Fee setting of the bank Since the bank is competitive, the fee φ_b announced at $t = 0$ is determined by the bank's break-even condition, given the equilibrium trading

[31]When $\alpha = \theta = 0$, we get the trivial equilibrium with no bank where $\lambda = \frac{a\sigma^2}{\sigma_u^2 + a^2\sigma^2}$ and K is not defined.

strategies of the market participants at $t = 1$.

$$0 = \mathbb{E}\left[\mathbb{E}[v-p|y]d(y)|\varphi_b\right] + \mathbb{E}\left[|u_b^+| + |u_b^-| + |\theta\eta v||\varphi_b\right]\varphi_b - \mathbb{E}\left[|u_b + \theta\eta v||\varphi_b\right]\varphi_m \tag{3.15}$$

where u_b^+, u_b^- are buying and selling orders from uninformed traders received by the bank. It is hard to solve for φ_b in closed form generally. However, in order to get the equilibrium results at $t = 1$ solved above, it is enough to show that the bank does participate with $\varphi_b < \varphi_c$ in equilibrium, which is shown below.

Proposition 3.2 *If $G_u(0) > 0$, there exists $\varphi_b < \varphi_m$ such that the above break-even condition is satisfied.* [32]

Proofs are in the Appendix.

3.3.3 Price informativeness

The MSE measure of price informativeness Now we measure the informativeness of the close price p, in terms of how well it reveals the fundamental asset value v. We define that the close price is more informative when it is "closer" to v, and we use the scaled mean square error to measure this closeness.

We calculate the MSE of p and scale it by the variance of the asset value σ^2 to get the scaled MSE

$$\frac{MSE}{\sigma^2} = \frac{\mathbb{E}\left[\mathbb{E}[(v-p)^2|v]\right]}{\sigma^2} = \frac{1}{\xi + 1} \tag{3.16}$$

where $\xi = \frac{(1+K\theta)^2\eta^2\sigma^2}{[(1+K)^2\alpha+1-\alpha]\sigma_u^2}$.

Improvement in price informativeness In order to analyze the effect of the bank's guaranteed close service on close price informativeness, we compare the scaled MSE in the equilibrium where bank conducts the guaranteed close service, with the

[32]The condition $G_u(0) > 0$ is sufficient and means that a positive proportion of uninformed traders have zero convenience cost. They can represent existing clients of the bank. We are assuming that m is large enough such that the normal approximation applies at $G_u(0)$.

scaled MSE in the equilibrium without the bank. Observe that the equilibrium without the bank coincides the equilibrium in which the fractions of traders trading via the bank is 0, i.e., $\alpha = \theta = 0$. So the scaled MSE in the "no bank" equilibrium is

$$\frac{MSE_{\text{no bank}}}{\sigma^2} = \frac{1}{\xi_{\text{no bank}} + 1} \tag{3.17}$$

where $\xi_{\text{no bank}} = \frac{\eta^2\sigma^2}{\sigma_u^2}$. Comparing this to the above equilibrium result with the bank, we get the following result.

Proposition 3.3 *The variance of informed orders increases more than the variance of liquidity orders due to dual trading of the bank, that is,*

$$(1 + K\theta)^2 \geq (1 + K)^2\alpha + 1 - \alpha \tag{3.18}$$

Having the bank conducting guaranteed close service improves the informativeness of p.

Proofs are in the appendix.

The above inequality is strict as long as $\alpha \neq \theta$ and $\alpha > 0$, so that the bank does trade for profit in equilibrium. Such trading activity improves the informativeness of the close price because it increases the ratio of informed orders relative to uninformed orders, measured by their variances, received by the market maker.

Comparative statics Having shown that having the bank improves the informativeness of the close price, we next discuss how the bank's share of orders received from traders affects the price informativeness. We do this by comparative statics on how the above measure of price informativeness changes with the parameters and get the following result.

Proposition 3.4 *Let $\delta = \frac{\theta}{\alpha}$ be the ratio of the proportion of informed traders trading with the bank relative to the proportion of uninformed traders trading with the bank, $\zeta = \frac{\alpha\sigma_u^2 + \theta\eta^2\sigma^2}{\sigma_u^2 + \eta^2\sigma^2}$ be a proxy of the proportion of traders' trading volume executed by the* *bank, then we can rewrite $\xi(\alpha(\varphi_m), \theta(\varphi_m))$ as $\xi(\delta(\varphi_m), \zeta(\varphi_m))$, and*

$$\frac{\partial \xi}{\partial \varphi_m} = \frac{\partial \xi}{\partial \delta}\frac{\partial \delta}{\partial \varphi_m} + \frac{\partial \xi}{\partial \zeta}\frac{\partial \zeta}{\partial \varphi_m} \tag{3.19}$$

where $\frac{\partial \xi}{\partial \zeta} \geq 0$ and the inequality is strict when $\delta \neq 1$. The effect of a fee change by the exchange on the price informativeness can be decomposed into a ratio effect and a volume effect, where the volume effect, i.e., the effect of the bank's volume share on price informativeness is positive, and the ratio effect is mixed.

Proofs are in the appendix.

The result here exactly aligns with our empirical findings, which is not surprising. Since uninformed traders are independent with each other, the bank infers more information about the overall orders when it receives more orders and trades more. Since the bank's trading is beneficial for the informativeness of the close price, as we have analyzed above, the increase in such trading improves the price informativeness. For example, when $\theta = 0$ and $\alpha > 0$, the bank provides liquidity to the liquidity orders it receives by trading against half of the trading demand ($K = -\frac{1}{2}$). In this case, the bank provides more liquidity when it receives more orders, making the price more informative. When there is a fee cut on the exchange, the bank's volume share decreases and the volume effect reduces the price informativeness. To see that the ratio effect is mixed, consider when δ is slight above or below 1. Since the informativeness with $\delta = 1$ is equivalent to the no bank result that is lower than informativeness with $\delta \neq 1$, an increase in δ makes it further away from 1 and improves the price informativeness when δ is slightly above 1, while making it closer to 1 and reduces the price informativeness when δ is slightly below 1. Our empirical finding that the fee cut reduces price informativeness can be interpreted as the two effects have the same sign or the volume effect is dominating.

3.3.4 Changes in the profits of the traders

Now we discuss the effect of dual trading by the bank on the profits of the traders, that is, the expected gains of the informed traders and the expected losses of the liquidity traders.

Proposition 3.5 *The bank's business decreases informed traders' expected gain before fees. It improves market depth and decreases the expected loss before fees of liquidity traders who still directly trade with the exchange if and only if* $\theta > \alpha$ *and* $(\theta - 3\alpha)\sigma_u^2 < 2\theta\eta^2\sigma^2$. *It decreases the expected loss before fees of liquidity traders who trade with the bank if and only if* $\theta < \alpha$ *and* $(1-\theta)(1+K\theta)\eta^2\sigma^2 + [1+\theta-2\alpha+(\theta-\alpha)K]\sigma_u^2 > 0$.

Proofs are in the appendix.

It is not surprising that the expected gain of the informed traders decrease given our previous result that price informativeness improved. If the bank receives a larger fraction of informed orders than uninformed orders, it chooses $K > 0$ and competes with the informed traders, reducing their profits. If the bank receives a smaller fraction of informed orders than uninformed orders, it chooses $K < 0$ and reduces market depth, still hurting the profits of the informed even if they do not trade with the bank.

The welfare effects on the liquidity traders are mixed. For a liquidity trader, the welfare effect largely depends on whether she trades with the bank and whether the bank receives a larger fraction of liquidity orders than informed orders. A liquidity trader benefit from the bank's business if she trades with the bank and it provides liquidity to trades with $K < 0$, or if she does not trade with the bank and it amplifies orders received with $K > 0$. Notice that such difference does not affect the venue choices of the liquidity traders since they are small and price taking. The results are calculated for all liquidity traders who trade with the bank and all who do not, instead of each individual trader, and they are not able to coordinate.

3.3.5 Discussion

Lastly, we discuss the robustness of our results to assumption changes and the generality of our model implications on scenarios beyond trading at the market close.

The venue choices are modeled in a largely exogenous way in our model, driven by a heterogeneous convenience cost drawn from arbitrary cumulative distribution functions G_u and G_x. That allows us to generally analyze any arbitrary sorting of traders between the two venues with $(\alpha, \theta) \in [0,1] \times [0,1]$. Therefore, our results apply to any specific setting that models the venue choices endogenously, which can be treated as a special case of our model. The improvement of price informativeness due to the introduction of a bank is strict if and only if $\frac{\theta}{\alpha} \neq 1$, that is, the fraction of informed traders trading with the bank is different from that of uninformed traders. Such condition can be satisfied as long as the incentives driving traders' venue choices are not identical between informed and uninformed traders. This is usually true in models with endogenous venue choices due to price or execution related incentives. For example, if we apply our model to a regular darkpool like in Zhu (2014), execution risks make uninformed traders more likely to trade in the darkpool compared to informed traders.

Our model is built on a stylized setting following the one period model in Kyle (1985) and can be applied generally to trading scenarios with multiple trading venues where the brokers are able to trade, which is quite common in most markets. In order to focus on the role of the bank in our specific empirical setting, we assume in our model that only the bank conducts dual trading and traders can directly trade with the market maker. Such assumption does not drive our results on price informativeness, which still apply if everyone trades via dual trading brokers, as long as the brokers receive different fractions of informed orders relative to uninformed order. In practice, the variety of service offered by brokers can make some of them more attractive to informed traders while others more attractive to uninformed traders.

3.4 Conclusion

In this paper, we formally study the impact of "guaranteed close" on the informativeness of the close price, both empirically and theoretically. Using the NYSE close auction fee cut in January 2018 as a policy shock, we provide quasi-experimental evidence that "guaranteed close" improves price discovery. Our unique empirical setting is ideal for testing the effect of dual trading: banks trade on their own accounts after viewing order flows of customers, and a quasi-experimental shock reduces the order flow to the banks; designated market makers clear the market in a single auction and set the price – these features correspond to the framework that the dual-trading literature usually builds on. Our empirical finding cannot be explained by the predictions of the dual-trading literature.

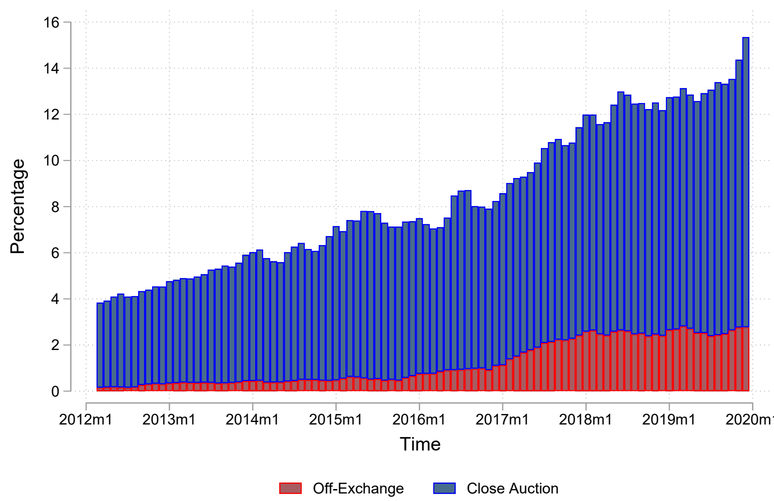
We build a model and provide a novel mechanism through which dual-trading improves price discovery. In our model, traders explicitly choose between two trading venues (i.e., the bank and the close auction) based on transaction costs. As long as the proportions of informed orders relative to uninformed orders are different between two venues, the bank can infer information from the net total orders and trade profitably. Such trading activity amplifies the proportion of informed orders received by the market maker and improves price informativeness.

The US brokerage industry has been undergoing rapid changes in its landscapes nowadays. While some brokers still charge commission fee and offer explicit execution price, e-brokers like Robinhood flock to charge zero commission fee and sell order flow data to sophisticated investment firms. The former brokers would be comparable to the close auction in our paper, and the latter the bank's "guaranteed close". Our model and empirical evidence shed light on the effect of these changes on price discovery. But of course, the jury is still out when it comes to the impact of these changes on liquidity, fairness, and aggregate investor welfare.

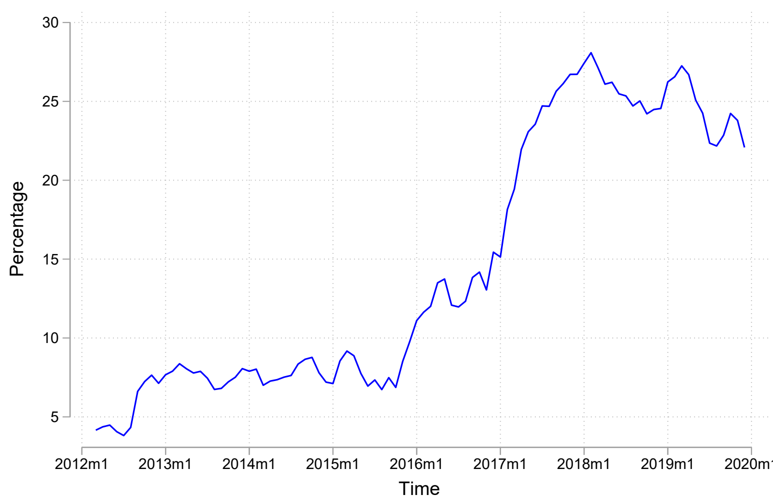
3.5 Figures and Tables

Figure 3-1: Trade Volume of S&P 500 Stocks in "Guaranteed Close" and Close Auctions

(a) Volume in "guaranteed close" and close auctions



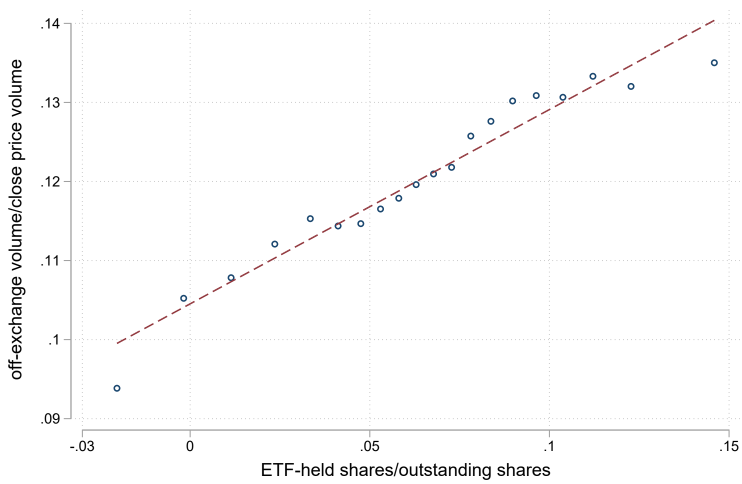
(b) Volume in "guaranteed close"/ volume in close auctions



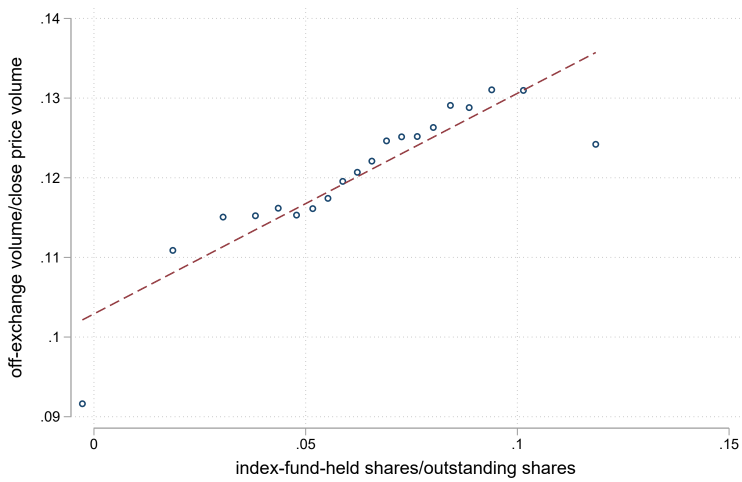
Notes. Panel (a) shows the trade volume in "guaranteed close" and trade volume in close auctions as a percentage of total trade volume for S&P 500 stocks from 2012m1 to 2019m12. Panel (b) compares trade volume in "guaranteed close" with trade volume in close auction. In both panels, we first aggregate the volumes of each stock to monthly observations and calculate the percentages, then smooth the time series by taking three-month moving average.

Figure 3-2: Cross-sectional Evidence: Relationship between ETF/Index-fund Ownership and Off-exchange MOC Volume

(a) ETF ownership vs. Off-exchange MOC volume/close price volume



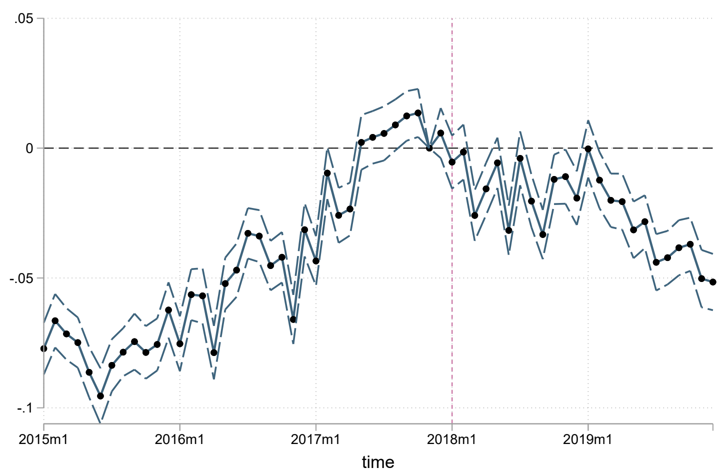
(b) Index fund ownership vs. Off-exchange MOC volume/close price volume



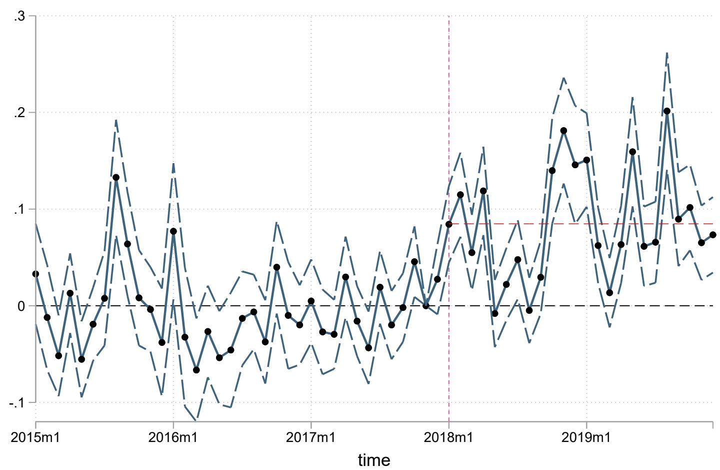
Notes. Panel (a) shows the bin-scatter plot between off-exchange MOC volume as a fraction of total volume traded at close price (the "off-exchange fraction"), and ETF ownership of stocks. Panel (b) shows the bin-scatter plot between the "off-exchange fraction" , and index fund ownership of stocks. The sample consists of 4,091 stocks (common shares), from 2015m1 to 2019m12. For each stock, we aggregate the trade volumes from daily data to monthly observations, and then calculate the fractions. ETF ownership and index fund ownership are the average ownerships in 2016. In both panels, we controlled for time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta in the regressions.

Figure 3-3: Trends and Dynamic Responses of Off-exchange MOC Volume and Close Price Informativeness

(a) First stage: off-exchange MOC volume/close price volume



(b) MSE of close price



Notes. We rank NYSE stocks by the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017. The treated group consists of stocks that rank at top 50% – they are more exposed to the NYSE fee cut. The control group consists of the remaining stocks. Panel (a) shows the treatment effect of the NYSE fee cut on the fraction of close price volume executed off exchange. Panel (b) plots treatment effect of the NYSE fee cut on price informativeness, that is the DID coefficients β_k and 95% confidence intervals estimated from the model: $\text{MSE}_{i,t} = \alpha_i + \lambda_t + \sum_k \beta_k \text{Treat}_i \cdot \mathbb{I}_{t=2017m11+k} + \Gamma X_{i,t} + \epsilon_{i,t}$. We controlled for stock fixed effects, time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta. Standard errors are clustered at the stock level. The horizontal red dashed line shows the average post-treatment effect, that is, the average of the β_k's.

Table 3.1: Summary Statistics: NYSE Sample

	N	Mean	SD	10th	50th	90th
MSE ($\times 10^4$)	69,273	0.56	0.54	0.09	0.39	1.30
MSE1 ($\times 10^4$)	69,186	0.76	0.73	0.12	0.52	1.74
MSE2 ($\times 10^4$)	69,100	0.54	0.50	0.09	0.38	1.22
MSE3 ($\times 10^4$)	69,159	0.54	0.51	0.09	0.38	1.24
Median SE ($\times 10^4$)	69,273	0.21	0.28	0.03	0.11	0.51
MAD (percent)	69,273	0.51	0.26	0.23	0.44	0.87
Index fund ownership	69,294	0.07	0.04	0.00	0.08	0.11
ETF ownership	69,294	0.06	0.05	0.00	0.07	0.13
Off-ex MOC volume/close price volume	69,291	0.11	0.09	0.00	0.09	0.25
Volatility ($\times 10^6$)	69,288	0.87	2.34	0.03	0.15	1.72
Market cap ($ bil.)	69,290	13.68	27.13	0.32	3.79	35.27
Total volume (in $1,000$ shares)	69,293	1606.41	2447.13	56.05	690.57	4185.92
Total retail volume (in $1,000$ shares)	69,258	92.52	170.06	4.78	29.89	238.30
Total volume of trades $\geq$ \$20K (in $1,000$ shares)	69,027	415.31	773.41	11.31	123.90	1069.77
Total volume of trades $\geq$ \$50K (in $1,000$ shares)	68,457	257.97	485.67	9.28	76.26	661.82
After-close volume/total volume	67,652	0.04	0.04	0.01	0.03	0.09
Close auction volume/total volume	69,291	0.08	0.05	0.01	0.08	0.15
Overnight beta	68,769	0.91	0.50	0.24	0.91	1.56
Spread (percent)	69,292	15.56	15.75	3.50	10.16	33.20

Notes. This table reports the summary statistics of all the variables used in this paper. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. Variables with daily observations are aggregated at a monthly frequency by calculating averages. MSE, MSE1, MSE2, MSE3 are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. Median SE is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$. MAD is the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$. Index fund ownership and ETF ownership are the fraction of outstanding shares held on average in 2016 by index funds and ETFs, respectively. Close price volume is volume traded at close price (i.e., close auction volume + Off-ex MOC volume). Volatility is trade-based intrayday volatility during market hours. Total volume is total trade volume during market hours. Total retail volume is total volume of retail trades during market hours. After-close volume is trade volume after market close and before next day's market open. Overnight beta is individual stock's overnight return's loading on overnight market return, estimated quarterly using CAPM regression. Spread is time-weighted percent quoted spread during market hours.

Table 3.2: DID Estimated Effects of NYSE Fee Cut on Price Informativeness: Main Results

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.088***	0.122***	0.084***	0.085***	0.044***	0.052***	0.017	-0.306
	(10.738)	(10.880)	(10.990)	(10.976)	(9.648)	(13.271)	(0.413)	(-1.954)
Volatility	0.051***	0.116***	0.047***	0.047***	0.018***	0.024***		3.812***
	(16.642)	(24.107)	(16.595)	(16.438)	(11.788)	(18.988)		(26.790)
log(Total Volume)	0.144***	0.215***	0.129***	0.132***	0.056***	0.060***	-1.659***	-3.823***
	(11.016)	(11.774)	(10.644)	(10.698)	(7.836)	(10.525)	(-12.615)	(-12.707)
N	66580	66579	66580	66580	66580	66580	66580	66580
Adj. R^2	0.710	0.700	0.723	0.719	0.649	0.782	0.767	0.918
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the difference-in-differences regressions estimating the effect of NYSE fee cut on close price informativeness. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017 . Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Table 3.3: DID Estimated Effects of NYSE Fee Cut on Price Informativeness: Designating Treatment Group by Passive Ownership

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.080***	0.107***	0.076***	0.077***	0.036***	0.045***	-0.047	0.288
	(9.868)	(9.643)	(10.061)	(10.048)	(8.056)	(11.590)	(-1.189)	(1.889)
Volatility	0.051***	0.117***	0.047***	0.048***	0.018***	0.024***		3.813***
	(16.564)	(24.014)	(16.512)	(16.355)	(11.780)	(18.752)		(26.806)
log(Total Volume)	0.143***	0.213***	0.128***	0.131***	0.055***	0.059***	-1.670***	-3.718***
	(10.870)	(11.594)	(10.484)	(10.540)	(7.625)	(10.172)	(-12.595)	(-12.328)
N	66569	66568	66569	66569	66569	66569	66569	66569
Adj. R^2	0.710	0.700	0.722	0.719	0.649	0.782	0.767	0.918
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the difference-in-differences regressions estimating the effect of NYSE fee cut on close price informativeness. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of shares held by ETFs and index funds in 2017. Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Table 3.4: DID Estimated Effects of NYSE Fee Cut on Price Informativeness: Excluding Earnings Announcement Days

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.068***	0.100***	0.066***	0.067***	0.042***	0.047***	0.019	-0.313*
	(11.076)	(11.544)	(11.114)	(11.158)	(9.347)	(13.125)	(0.476)	(-1.994)
Volatility	0.035***	0.085***	0.034***	0.034***	0.017***	0.020***		3.870***
	(16.953)	(23.477)	(16.705)	(16.618)	(11.791)	(18.654)		(26.970)
log(Total Volume)	0.085***	0.146***	0.081***	0.081***	0.054***	0.044***	-1.602***	-3.753***
	(9.246)	(10.967)	(8.981)	(8.999)	(7.581)	(8.943)	(-12.631)	(-12.601)
N	66530	66529	66530	66530	66530	66530	66530	66530
Adj. R^2	0.759	0.733	0.759	0.759	0.647	0.794	0.764	0.917
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the robustness of the main difference-in-differences regressions estimates to excluding earnings announcement days from the sample. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data, excluding data within 1 day, that is, the [-1,0,1] days, from earnings announcement days. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017 . Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Table 3.5: DID Estimated Effects of NYSE Fee Cut on Price Informativeness: Matching Specification

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.068***	0.102***	0.065***	0.066***	0.032***	0.037***	0.028	0.163
	(7.555)	(8.693)	(7.767)	(7.737)	(6.194)	(8.940)	(1.811)	(1.512)
Volatility	0.078***	0.157***	0.071***	0.073***	0.023***	0.035***		4.972***
	(6.942)	(7.738)	(6.935)	(6.937)	(5.304)	(6.933)		(11.969)
log(Total Volume)	0.197***	0.291***	0.181***	0.185***	0.082***	0.089***	-0.505***	-3.625***
	(13.106)	(14.026)	(12.944)	(12.988)	(9.686)	(12.938)	(-7.367)	(-14.842)
N	68186	68186	68186	68186	68186	68186	68186	68186
Adj. R^2	0.734	0.719	0.746	0.743	0.671	0.799	0.620	0.846
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the robustness of the main difference-in-differences estimates to using matched treated and control stocks. We perform the matching on market cap, total trade volume, volatility, and overnight beta. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data, excluding data within 1 day, that is, the [-1,0,1] days, from earnings announcement days. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017. Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Table 3.6: Distribution of Variables and Balance of Matching

	Treated			Control			Comparison	
	A. Pre-Matching							
	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
log(Market Cap)	2.327	1.480	0.225	0.428	2.574	0.674	1.334	0.575
log(Total Volume)	13.933	1.250	-0.051	12.576	2.294	0.104	1.019	0.545
Volatility	0.154	0.181	13.679	1.094	5.894	4.005	-0.539	0.031
Overnight Beta	1.015	0.262	0.280	0.882	0.376	0.254	0.235	0.697
	B. Post-Matching (Mahalanobis distance with replacement)							
	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
log(Market Cap)	2.327	1.480	0.225	2.222	1.360	0.049	0.089	1.089
log(Total Volume)	13.933	1.250	-0.051	13.872	1.076	-0.030	0.056	1.162
Volatility	0.154	0.181	13.679	0.150	0.195	13.831	0.009	0.929
Overnight Beta	1.015	0.262	0.280	1.042	0.243	0.240	-0.055	1.078

Notes. This table reports distributional test statistics of four variables (pre-treatment values) we use in matching treated stocks to control stocks: log(market cap), log(total volume), volatility during market hours, and overnight beta. The table also assesses balance between treated and control group in the means (using the standardized difference) and in the variances (using the variance ratio).

Appendix A

Appendix for Chapter 1

A.1 Appendix: Data

Accounts Receivable Based Credit Lines

I start the data collection process by retrieving all SEC filings filed from 2001 to 2020 that contain the words "eligible accounts receivable" or "eligible receivable". To do this, I utilize the SEC Edgar Full Text Search, a new search tool that allows users to search keywords through full text of electronic filings since 2001. Filings retrieved are from 8-K, 10-K, 10-Q, S-1 and other SEC forms.

I use the help of regular expressions to collect loan terms in the first pass. For example, I use a regular expression algorithm to locate the text that surrounds the keywords "eligible accounts receivable" and "eligible receivable" in each document. Then, advance rate can be located by applying identifiers "(% | percent) of eligible accounts receivable" or "advance up to %.*" to the text. The other loan terms, namely interest rate, credit limit, origination/amendment date and lender name, can be accurately identified only through manual reading of the documents. Manual reading also makes sure we collect data from the correct loan contracts.

We also double check the documents filed by firms that have accounts receivable credit lines covered by DealScan. Eventually, we read through 17,951 documents filed by 2,663 Compustat non-financial non-utility firms. These documents include both the SEC filings' main text which contains short description of the loan contracts, and their exhibits which are the original documents of loan contracts.

Many documents repetitively report the same A/R based credit line a firm entered into a few years ago. In that case, we remove the duplicates. If a document mentions any changes in credit limit, advance rate, or interest rate, we pay special attention to record them and the dates of changes. Some documents (exhibits) are amendment to an existing contract. We only record the changes in terms that are explicitly stated, and assume the remaining terms keep unchanged.

Sometimes a borrower's interest rate depends on its covenant status, for example, its leverage ratio, debt/ebitda ratio. The contract specifies the interest rate contingent on each covenant status. Sometimes only the range of interest rates is known. In such cases, we compute and record the average interest rate. When there are multiple lenders, we record the administrative agent or the primary lender.

Here is an example of the description of an A/R credit line in the 8-K main text:[1]

> On November 21, 2013, the Company entered into a Loan and Security Agreement with Silicon Valley Bank (the "Credit Agreement"). The Credit Agreement provides a revolving credit facility of $12 million, including a $2 million sub-limit for letters of credit. Borrowings under the credit facility are limited to eighty percent (80%) of eligible accounts receivable. The Credit Agreement has interest rates ranging from LIBOR + 2.25% to Prime + 0.75%, expires on November 21, 2015, and is secured by substantially all of the assets of the Company. The Credit Agreement requires that the Company maintain a minimum tangible net worth and provides for an annual limit on capital expenditures. The foregoing description of the Credit Agreement does not purport to be complete and is qualified in its entirety by reference to the full text of the Credit Agreement (and the exhibits thereto), which is attached hereto as Exhibit 10.1, and is incorporated herein by reference.

For an example of the original loan contract in SEC filing's exhibit, see Exhibit 10.6 of the 10-K report of Radyne Comstream Inc. filed on April 1, 2002: https://www.sec.gov/Archives/edgar/data/0000718573/000095015302000629/p66289ex10-6.txt. And see Exhibit 10.11 of the 10-K report of Avistar Communications filed on 2003-03-25:

[1] See 8-K report of Planar Systems, Inc. filed on November 21, 2013 at https://www.sec.gov/Archives/edgar/data/0000722392/000119312513457047/d636079d8k.htm

https://www.sec.gov/Archives/edgar/data/1111632/000089161803001414/f88236exv10w11.txt.

Factoring Agreements

I utilize the SEC Edgar Full Text Search to retrieve all SEC filings filed from 2001 to 2020 that contain the words "factoring agreement", "factoring agreements", "factoring arrangement", "factoring facility", "factoring program", "receivable purchase agreement" or "receivables financing agreement". I restrict the search to be within 8-K, 10-K, 10-Q, S-1 forms. The search results in 13,701 documents. To further economize the manual effort, I restrict documents to those filed by Compustat firms. Then we manually read through the documents and collect the terms of the factoring agreements. Most of the documents are actually formal syndicated loan agreements, which restrict the borrower from entering into subsequent factoring agreements to dilute the lender, hence mentions "factoring agreement" in its covenant section. We exclude these syndicated loans manually.

The remaining documents contain information on factoring agreements. We collect advance rate, interest rate, credit limit, origination/amendment date, lender name, and recourse status. Usually the borrowing firm discloses whether a factoring agreement has recourse or non-recourse, and we record as is. There are cases where a factor offers both recourse and non-recourse factoring, at the borrower's choice. Borrowers often disclose the fraction of the loan they borrow that has recourse. We record a factoring agreement as having recourse if more than 50% of the loan has recourse. In such cases, we also compute and record the averages of the advance rate and the interest rate. According to our observations, firms tend to use recourse factoring more often than non-recourse factoring, because the former is cheaper.

Below is an example of non-recourse factoring, described in a 10-Q main text: [2]

> GAAV is a party to a factoring agreement, dated as of May 22, 2007 (the "Factoring Agreement") with FCC LLC, d/b/a First Capital Western Region, LLC (the "Factor"). The Factoring Agreement, which provides for an initial term of two years and a one-year automatic extension unless GAAV provides

[2] See Great American Group, Inc. 10-Q (Quarterly report), dated 2010/5/17: https://www.sec.gov/Archives/edgar/data/1464790/000119312510122301/d10q.htm

> written notice of termination to the Factor, will expire on May 22, 2010. The Factor, at its discretion, purchases on a nonrecourse basis, all of the GAAV's customer receivables. The Factor is responsible for servicing the receivables. The Factor pays 90% of the net receivable invoice amount upon request by GAAV and retains the remaining 10% in a reserve. The Factor, at its discretion, may offset the reserve for amounts not collected or outstanding at the end of the term of the Factoring Agreement. GAAV may request releases from the reserve for any excess over a minimum balance set by the Factor. The Factor charges a factoring commission equal to 0.25% of the gross invoice amount of each account purchased, or five dollars per invoice, whichever is greater, with a minimum commission of $24 per year, prorated for the first year. The Factor also charges interest at prime plus 1% with a floor of 8% on the net uncollected outstanding balance of the receivables purchased. Effective December 1, 2009, the interest charge by the Factor was reduced to London Interbank Offered Rate ("LIBOR") plus 4.5% on the net uncollected outstanding balance of the receivables purchased. One of the members of the GAAV personally guarantees up to a maximum of $500 plus interest and certain fees for accounts receivables sold pursuant to the Factoring Agreement.

For an example of the original factoring agreement contract, see Exhibit 10.1 of the 8-K report of Viscount Systems, Inc. filed on March 30, 2015 at: https://www.sec.gov/Archives/edgar/data/1158387/000106299315001622/exhibit10-1.htm.

Accounts Receivable from Individual Customers

Firms disclose accounts receivable from individual customers to abide by FASB No.105. I collect these accounts receivable data from 10-K filings. This task is made possible by feeding the names of a firm's major customers available in Compustat historical customer segment file into a textual scraping algorithm. Specifically, I look for any of the following word patterns:

"major customer" or "concentration of risk"; customer names + "$" + any numbers; any numbers + "receivable";"customer" + "receivable";"amounts due from customer"; "amounts due from" + customer names + any numbers; "receivable"+ customer names; customer names + "receivable".

There are 7,420 10-K filings that contain some of these word patterns. Then we read through the text surrounding these word patterns to collect the accounts receivable data. Sometimes a firm does not report the name of its customer the accounts receivable is due from, but identifies them as "the largest customer", or "the second largest customer". In such cases, we again rely on the Compustat historical customer segment file to infer names of these customers based on the amount of sales made to them.

Below are some examples of the data source.

Example 1

> Receivables from Ingram and Tech Data accounted for 32% and 18%, respectively, of our total receivables at December 1, 2000. As of December 3, 1999, receivables from Tech Data accounted for 11% of our total receivables, and in fiscal 1998 receivables from Ingram at year-end accounted for 14% of our total receivables.
>
> – ADOBE System Inc. Form 10-K, dated February 26, 2001.
>
> https://www.sec.gov/Archives/edgar/data/796343/000091205701006700/a2039309z10-k.htm

Example 2

The Company performs ongoing credit evaluations of its customers and generally requires no collateral. Customers who accounted for 10% or more of net accounts receivable are as follows:

	December 31,					
	2008		2007		2006	
Excelpoint Systems Pte Ltd (1)	36	%	31	%	*	%
Answer Technology, Inc.	18	%	14	%	*	%
Avnet, Inc.	10	%	13	%	13	%
Metatech (1)	-		-		31	%

– PLX Technology, Inc. Form 10-K, dated March 6, 2009. https://www.sec.gov/Archives/edgar/data/850579/000085057909000035/plx_body10k.htm

Payment Terms and Supply Agreements

The payment terms I follow Costello (2013, 2019) to collect supply agreements from SEC filings. The sample includes all long-term supply contracts entered into between 2000 and 2020. I search SEC filings (10-K, 10-Q, 8-K and S-1) for exhibits with "supply" or "procurement" in the title and "buyer" and "supplier" or "seller" in the first few paragraphs, and I search for trade credit details in the body of the contract. When the supply agreement adopts milestone payment schedules, for example, down payment, mid-stage payment, and final payment, we only collect the payment terms for the final payment.

Below is an example of the payment terms, described in a supply agreement between Hyaluron Inc. and Hemispherx Biopharma, Inc.:

> All invoices are due and payable upon receipt and past due after thirty (30) days from the date of invoice. All amounts past due shall incur interest at the rate of 1.5% per month or the highest rate permitted by law (whichever is less). All payments shall be made to Hyaluron Inc. at the address specified on the front of the invoice.
>
> – Hemispherx Biopharma, Inc. Form 10-K, Exhibit 10.47, dated April 3,2006.
>
> https://www.sec.gov/Archives/edgar/data/946644/000114420406013365/v039320_ex10-47.txt

My data are consistent with the payment terms data used in existing studies in terms of the low presence of early discounts. For US firms, Ng, Smith, and Smith (1999) find that 25.5% of the firms in a sample drawn from Compustat mainly offer two-part contracts, whereas Giannetti, Burkart, and Ellingsen (2011), using the 1998 National Survey of Small Business Finances, document that firms on average are offered early payments discounts from 21.3% of their suppliers. Moreover, Giannetti, Burkart, and Ellingsen (2011) find that only 7.8% of firms operate in industries in which discounts are common. Using more recent data from 2005, Klapper, Laeven, and Rajan (2011) (using PrimeRevenue data) find that 13% of the contracts extended to large US and European buyers included early payment discounts.

Do buyers return bad products to sellers after delivery? Yes. Most supply agreements in my sample specify a period of testing time after product delivery. Take the same supply agreement we have seen between Hyaluron Inc. and Hemispherx Biopharma, Inc as an example:

> Hemispherx [buyer] will conduct release testing on quality control samples obtained from each Batch of Hemispherx Product shipped by Hyaluron [seller] hereunder to confirm that such quality control samples conform to the Manufacturing Standards. [When the quality control samples are available and the

Batch Record is done], Hemispherx shall request that the Hemispherx Product be shipped immediately to Hemispherx ... After 45 days, Hemispherx will be deemed to have accepted the Batch, unless Hemispherx, by written notice ("Notice of Rejection/Nonconformance") to Hyaluron within the 45-day period initiates an investigation into the reasons for the failure to allegedly conform to the Manufacturing Standards by returning allegedly non-conforming Product to Hyaluron within 14 days after giving notice of such non-conformance. Once Hemispherx has been deemed to accept the Product, Hyaluron's responsibilities and liabilities for the Product will be null and void.

– Hemispherx Biopharma, Inc. Form 10-K, Exhibit 10.47, dated April 3, 2006.

https://www.sec.gov/Archives/edgar/data/946644/000114420406013365/v039320_ex10-47.txt

Can buyers terminate supply agreements? Yes. In the majority of supply agreements I collected, buyers have the right to terminate supply agreements before the contracted duration. Buyers are required to file an termination notice in advance of the intended termination date.

The original term of this Agreement shall be for a period of three (3) years, commencing on the first day of November, 2004, and terminating on the 31st day of October, 2007, unless extended or sooner terminated, as hereinafter provided. If the Buyer wishes to terminate this Agreement sooner then three years, Buyer may terminate this Agreement at any time during the term of the Agreement by delivering Seller a written termination notice six (6) months in advance of the termination date.

– Exhibit 10.1 of Form 8-K, dated March 18, 2005 of Unifi, Inc.

https://www.sec.gov/Archives/edgar/data/100726/000095014405003042/g94086exv10w1.htm

Due to this reason, in my model, the duration of supply agreement is not a choice variable for buyers. Instead of abiding by the contracted supply agreement's duration, buyers can terminate the supply agreement when they find the sellers are bad.

Does a supply agreement contract on delivery quantity? The supply agreements I collected usually negotiate on the quantity a seller needs to produce and deliver in a sophisticated manner, to prevent over-production or under-production.

When a seller needs to invest significantly to build the facility for the production, buyers use multiple approaches to negotiate on the production quantities. Buyers often provide demand forecast to guide the seller's production and to some extent commit to purchase an amount close to the forecast. Buyers also often choose to set up a joint committee with the seller to communicate the quantity of goods that needs to be produced. See the following excerpts from two supply agreements as examples.

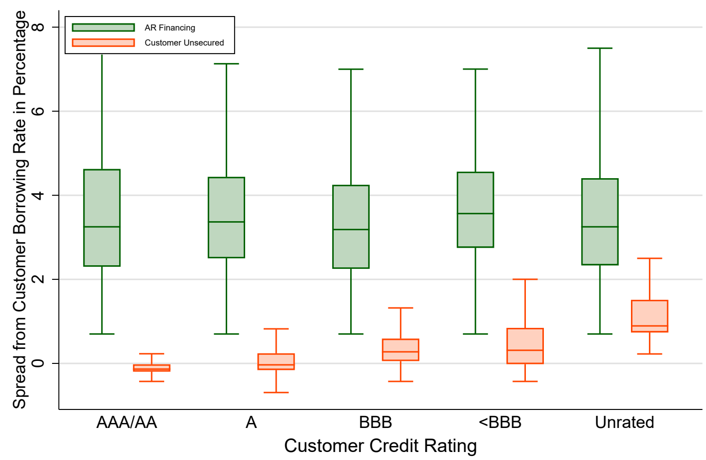
> Millennium will order, pursuant to the forecasts agreed upon by the JPC, Product from the Third Party Product Contractor(s).
>
> – Millennium Pharmaceutical, Inc. Exhibit 10.3, Form 10-Q, dated September 30, 2005.
>
> https://www.sec.gov/Archives/edgar/data/1002637/000100263705000038/fm10q09302005exhibit10_3.txt

> During the term of this Agreement and any non-exclusive period hereunder, CryoCath will purchase the Systems from ENDOCARE at the prices set forth in Schedule 2 and at minimum volumes as set forth in Schedule 3. CryoCath will supply on a monthly basis a rolling 12-month demand forecast, the first two months of each rolling 12-month forecast is binding and constitutes a firm purchase commitment. ENDOCARE agrees that it will use reasonable efforts to supply products in accordance with the forecast.
>
> – Endocare, Inc. Exhibit 10.45, Form 10-Q, dated September 30, 2001.
>
> https://www.sec.gov/Archives/edgar/data/1003464/000093639201500219/a76830ex10-45.txt

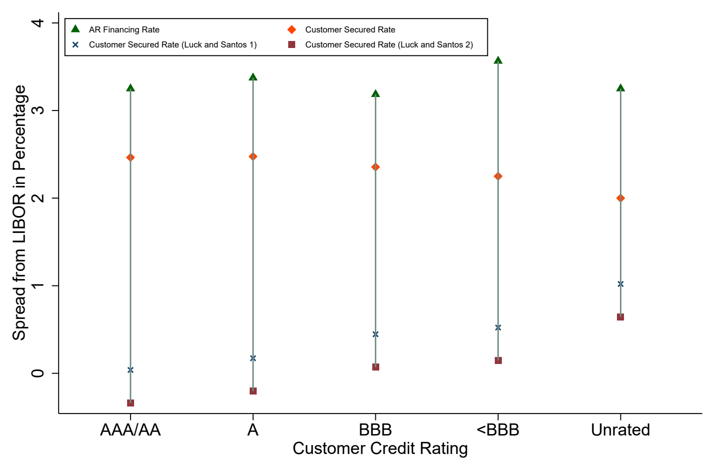
A.2 Appendix: Figures and Tables

Figure A-1: Interest Rates of A/R Financing and Customer Firms' Borrowing Rates: Alternative Measures

(a) Customer Unsecured Borrowing Rates Weighted by Imputed A/R Shares



(b) Comparison with Customer Secured Borrowing Rates



Notes: Panel (a) compares the interest rates a seller pays when using A/R financing versus the interest rates a seller's customers on average pay in unsecured borrowings. The difference from Figure 1-4 is that customers' borrowing rates are weighted by seller's imputed accounts receivable due to individual customers. Panel (b) compares the A/R financing interest rates versus the interest rates seller's customers on average pay in secured borrowings. Medians of the spreads of interest rates over LIBOR are plotted for each group. In the plot, Customer Secured Rate means the interest rates on secured term loans and credit lines in DealScan, that happen within 3 years of the unsecured borrowings (to reduce selection bias). Customer Secured Rate (Luck and Santos 1) and Customer Secured Rate (Luck and Santos 2) equal customer's average unsecured rate plus 14.5 bps and minus 23.1 bps, respectively. These numbers are the upper and lower bound of the spread between a firm's secured borrowings and unsecured borrowings estimated by Luck and Santos (2022).

Table A.1: Impute Accounts Receivable from Individual Customers

Dep Var: **Accounts receivable from individual customers (% of total)**				
	(1)	(2)	(3)	(4)
Sales to individual customers (% of total)	0.676***	0.693***	0.683***	0.697***
	(48.748)	(48.373)	(47.620)	(48.168)
log(Total Asset)	-0.012***	-0.011***	-0.012***	-0.011***
	(-7.791)	(-7.516)	(-7.713)	(-7.325)
log(Customer Total Asset)	-0.009**	-0.003	-0.012**	0.004
	(-2.041)	(-0.859)	(-2.545)	(0.978)
log(Accounts Payable)	0.012***	0.002	0.015***	-0.004
	(2.771)	(0.857)	(3.311)	(-0.903)
Constant	0.150***	0.149***	0.156***	0.128***
	(6.730)	(7.773)	(6.674)	(6.021)
N	3470	3437	3429	3433
Adj. R^2	0.499	0.467	0.486	0.474
Supplier Industry FE	Yes			
Customer Industry FE	Yes		Yes	
Year FE	Yes			
Supplier Industry × Year FE		Yes	Yes	Yes
Customer Industry × Year FE				Yes

Notes: In this table, I regress a seller's accounts receivable due from individual customers on observables. Regressions are at the seller-buyer-year level. Fitted values from Column (4)'s regression are used to impute the accounts receivable due from individual customers. Standard errors are clustered at the firm level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Comparing the A/R Financing Data with Supervisory Fed Y14 data

Firm Size	Committed Credit (in $m)							Total Committed Credit (in $b)	
(Assets in $m)	1st Percentile	10th Percentile	Mean	Median	90th Percentile	99th Percentile	Firms in Category	My data	Y14 data
0-50	0.25	1	9.9	4	15	200	116	1.0	2.5
50-250	0.25	4	28.0	20	55	175	119	3.2	8.0
250-1000	0.25	4.5	97.8	50	250	500	109	9.7	37.2
1000-5000	0.25	10	303.3	177.5	750	1100	115	32.8	149.8
5000-	0.25	0.25	694.5	282	2100	3700	53	34.0	106.7

Notes: This table reports the distribution and the sum of committed credit as of 2019Q4 in my A/R financing data, and the sum of committed credit as of 2019Q4 in Y14 data, both for Compustat firms. Firm size means firm's average total asset during 2000 to 2020. My data is based on disclosure. We track firms with disclosed loan originations or loan amendments during 2015–2019, and further verify the eventual year firms stop using A/R financing by reading through all of their 10-K filings until 2019. To get the aggregate outstanding committed credit in A/R financing from the Y14 data, I combine the committed credit in credit lines to Compustat firms in Table 1 and the breakdown of collateral types in Table A.26 of Chodorow-Reich et al. (2022) downloaded from https://scholar.harvard.edu/files/chodorow-reich/files/crdlp__bank__liquidity__provision.pdf. A few reasons explain why total committed credit in my data is lower. First, the Y14 data include utility and financial firms, while I exclude them. Second, my data is based on firms' disclosures, so it covers small firms' credit lines better than large firms.

Table A.3: Summary Statistics of Payment Terms Collected from Supply Agreements

Industry group	Net terms (% adoption)	Payment days (mean)	Late interest (mean,%)	No. of Obs.
Mining	net 30 (35%)	25.9	7.6	83
Manufacturing –				
Food and tabacco products	net 30 (40%), net 15 (17%)	28.4	13.1	30
Fabrics, carpet and apparel	net 30 (56%), net 45 (22%)	31.9	14.3	9
Wood and paper products	net 30 (55%)	32.0	12.2	22
Chemicals and allied products	net 30 (54%), net 45 (10%)	30.9	12.5	841
Petroleum, rubber and plastics	net 90 (28%), net 30 (21%)	52.8	7.7	68
Metals	net 30 (67%), net 45 (14%)	32.2	15.0	57
Machinery and equipment	net 30 (45%), net 45 (17%)	42.5	16.7	47
Electrical equipment	net 30 (61%), net 45 (15%)	37.7	15.1	151
Transportation equipment	net 30 (40%), net 60 (27%)	38.0	18.0	15
Instruments	net 30 (57%), net 60 (19%)	38.3	12.4	201
Other manufacturing	net 30 (50%), net 45 (13%)	34.3	9.1	8
Transportation and public utilities	net 10 (36%), net 30 (30%)	19.7	8.8	250
Wholesale trade	net 30 (42%), net 90 (22%)	45.3	12.3	69
Other	net 30 (49%), net 45 (10%)	33.9	11.2	286

Notes: This table tabulates the payment terms from 2,137 bilateral supply agreements. In Column 2, I report the payment terms most commonly used in each industry group and their adoption rates. Only 4.3% of the agreements offer two-part terms, i.e., early discount. 48% of the agreements charge interest on late payments, and the mean late payment interest rates are separately reported in Column 4. The most popular late payment interest rate terms with the adoption rates in parentheses are: 2% + prime (27%) , 1.5% per month (21%), 1% per month (17%). Industry groups are defined as in Ng, Smith, and Smith (1999) based on SIC code. No. of Obs. in the last column refers to the number of supply agreements observed.

Table A.4: Propensity of A/R Financing Borrowing

	(1)	(2)	(3)	(4)
	Logit		Linear Probability	
Net Trade Credit (months)	-0.039***	-0.039***	-0.003***	-0.005***
	(-5.11)	(-5.08)	(-6.31)	(-9.49)
log(Total Asset)	-0.104***	-0.105***	-0.014***	-0.028***
	(-6.42)	(-6.43)	(-6.81)	(-17.67)
Cash/Total Asset	-1.264***	-1.256***	-0.147***	-0.225***
	(-8.04)	(-7.98)	(-9.60)	(-15.27)
EBITDA/Total Asset	0.268***	0.269***	0.023***	0.042***
	(3.22)	(3.20)	(4.29)	(7.50)
Q	-0.043***	-0.045***	-0.007***	-0.003**
	(-3.54)	(-3.67)	(-4.84)	(-2.02)
Leverage	0.041	0.045	0.009	-0.004
	(0.49)	(0.53)	(1.00)	(-0.48)
Debt/EBITDA	0.011***	0.011***	0.002***	0.002***
	(4.83)	(4.92)	(5.21)	(5.24)
N	80,117	79,686	80,117	80,117
Industry FE	Yes		Yes	
Year FE	Yes		Yes	
Industry × Year FE		Yes		Yes

Notes. The table reports logit regression and linear probability regression that study the determinants of A/R financing borrowing. The dependent variable is a dummy that indicates whether a firm has outstanding A/R financing loans. I regard a firm as having outstanding A/R financing loans in a year if it enters into A/R financing contracts (originations and amendments) within 2 years. Standard errors are clustered at the firm level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Pricing and Renegotiations of A/R Financing Terms: Credit Limit, Advance Rate and Interest Rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Credit Limit/Sales		Advance Rate		Interest Rate	
EBITDA/Sales	-0.054***	-0.027**	0.300***	0.023	-0.286***	-0.056
	(-7.84)	(-2.12)	(2.86)	(0.14)	(-3.79)	(-0.27)
Receivable/Sales	0.338***	0.336***				
	(3.06)	(3.14)				
Net Trade Credit (years)		0.112***	0.404	0.200	1.147***	1.257**
		(2.77)	(0.74)	(0.43)	(4.83)	(2.50)
Uncollectible AR/total AR			-1.994	0.727	2.596***	0.317
			(-1.21)	(0.36)	(5.95)	(0.51)
N	3429	2838	2517	1804	3001	1870
adj. R^2	0.512	0.526	0.009	0.654	0.090	0.430
Firm FE	Yes	Yes		Yes		Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The table reports regressions of A/R financing loan terms on observables. Advance rate and interest rate (annualized) are in percent. Other Controls include size, leverage, cash balance (quick ratio), and market to book ratio (Tobin's Q). Standard errors are clustered at the firm level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Examine the First Stage of Instruments

	(1)	(2)	(3)	(4)	(5)
Customer Refinancing Needs	13.53***	4.224*	19.66***		
	(5.61)	(1.92)	(7.81)		
Lagged Industry Receivable Days				0.463***	0.395***
				(27.28)	(22.91)
log(Total Asset)			9.112***		17.92***
			(33.78)		(98.61)
log(Sales)			-10.66***		-17.90***
			(-40.36)		(-115.56)
Cash/Total Asset			-12.43***		-28.71***
			(-8.46)		(-36.94)
EBITDA/Total Asset			3.943***		1.750***
			(6.04)		(9.12)
Q			-0.160		0.0304
			(-1.47)		(0.68)
F-stat	31.45	3.70	281.24	744.17	2642.44
N	39014	37740	32686	190572	163121
Adj. R^2	0.011	0.530	0.059	0.506	0.563
Firm FE		Yes		Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes. The table reports first-stage regressions of receivable days on two instrument variables. The dependent variable is a firm's receivable days. Customer Refinancing Needs is the a firm's customers' average (weighted by customer size) fraction of debt that needs to be refinanced in the subsequent year (Compustat item dd1/dlc+dltt). Standard errors are clustered at the firm level. t statistics are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Appendix: Proofs

A.3.1 Proof of Lemma 1.1

After simplifying Equation (1.10), bank's profit equals

$$\begin{aligned}\Pi =&\alpha e^{rD} M + \Pr(bad)\phi(b)\Big\{(1-\delta)\min[1-\nu D/M - \alpha e^{rD}, 0] \\ &+ \delta \min[1 + C/M - \nu D/M - \alpha e^{rD}, 0]\Big\} M - \alpha e^{gD} M \end{aligned} \tag{A.1}$$

We consider two cases depending on the values of advance rate and interest rate.

Case 1: $1 - \alpha e^{rD} \geq \nu D/M$. In this case, let $\pi_1 \triangleq \Pi/M = \alpha e^{rD} - \alpha e^{gD}$.

Case 2: $1 - \alpha e^{rD} \in [\nu D/M - C/M, \nu D/M)$. In this case, let $\pi_2 \triangleq \Pi/M = \alpha e^{rD} + \Pr(bad)\phi(b)(1-\delta)(1-\nu D/M - \alpha e^{rD}) - \alpha e^{gD}$

When $1 - \alpha e^{rD} = \nu D/M$, we have $\pi_1 = \pi_2$. And π_2 is increasing in the adjusted haircut $1 - \alpha e^{rD}$. So $\pi_2 > \pi_1$ and π_2 achieves maximum when $1 - \alpha e^{rD} = \nu D/M - C/M$. Intuitively, this is because setting adjusted haircut lower than $\nu D/M - C/M$ does not fully exploit the value of recourse against seller. So we have the optimality condition for loan terms: $1 - \alpha e^{rD} = \nu D/M - C/M$.

The maximum of π_2 is $1 - \nu D/M + C/M - \Pr(bad)\phi(b)(1-\delta)C/M - \alpha e^{gD}$. And the break-even advance rate that sets $\pi_2 = 0$ is given by

$$\alpha = \Big[1 - \nu D/M + C/M - \Pr(bad)\phi(b)(1-\delta)C/M\Big] e^{-gD} \tag{A.2}$$

Plugging advance rate into the optimality condition, the break-even interest rate is given by:

$$e^{rD} = \frac{1 - \nu D/M + C/M}{1 - \nu D/M + C/M - \Pr(bad)\phi(b)(1-\delta)C/M} e^{gD} \tag{A.3}$$

A.3.2 Proof of Lemma 1.2

Plugging the parametric assumptions in Assumption 1.2 into the expression of loan terms, i.e., Equation (1.11) and (1.12), we have

$$1 - \alpha e^{rD} = vD - c \tag{A.4}$$

$$e^{rD} = \frac{1 - vD + c}{1 - vD + c - \Pr(bad)\phi(b)(1-\delta)c} e^{gD} \tag{A.5}$$

When bad seller's misbehavior is at level b, the buyer's profit is

$$\max_{Q,D} \ \Gamma = \Pr(good)(Q^{\rho}A - pQ) + \Pr(bad)\Big[(Q^{\rho}A - pQ)[1-\phi(b)] + (-pQ + \nu D)\phi(b)\Big], \tag{A.6}$$

$$\text{subject to } \ \phi'(b)[\delta\nu D + (1-\delta)(1-\alpha e^{rD})pQ] = \psi'(b). \tag{A.7}$$

Simplifying, plugging in the parametric assumptions in Assumption 1.2, and plugging the loan terms expression, the buyer's profit is

$$\max_{Q,D} \Gamma = Q^{\rho}A\Big[\Pr(good) + \Pr(bad)(1-\phi(b))\Big] - pQ + vD\Pr(bad)\phi(b)pQ, \tag{A.8}$$

$$\text{subject to } \ \phi'(b)[vD - (1-\delta)c]pQ = \psi'(b), \tag{A.9}$$

where, given by Equation (1.7), under the parametric assumptions, input price p is given by

$$pQ = \frac{(\eta + Q)e^{rD}}{f(\alpha)(1 - \bar{i}e^{rD})} \tag{A.10}$$

And the incentive compatible constraint, given the parametric assumptions that

$\phi(b) = b^2, \psi(b) = zbpQ$, implies

$$b = \frac{1}{2}\frac{z}{vD - (1-\delta)c} \tag{A.11}$$

As we can see, misbehavior level b is related to only one choice variable of the buyer, that is the payment delay D. Let us denote the probability of bad product as $m(D) = \Pr(bad)\phi(b) = \Pr(bad)(\frac{z}{2[vD-(1-\delta)c]})^2$. Buyer's profit in Equation (A.8) is further simplified to $\Gamma = Q^\rho A[1 - m(D)] - [1 - vDm(D)]pQ$. Taking derivatives of Γ with respect to Q, we have

$$\frac{\partial \Gamma}{\partial Q} = \rho Q^{\rho-1} A\Big[1 - m(D)\Big] - \frac{e^{rD}[1 - vDm(D)]}{f(\alpha)(1 - \bar{i}e^{rD})} = 0 \tag{A.12}$$

Taking derivatives of Γ with respect to D, we have

$$\frac{\partial \Gamma}{\partial D} = Q^\rho A \frac{d(1 - m(D))}{dD} - [1 - vDm(D)]\frac{dpQ}{dD} + vpQ\frac{dDm(D)}{dD} = 0. \tag{A.13}$$

A.3.3 Proof of Proposition 1.1

By Bayes rule, the conditional probability of seller's type given historical signals is $h_t = \Pr(good|\theta^t = 1, \gamma^t = 1) = \frac{h}{h+(1-h)(\sigma_b\sigma_c)^t}$. In every period, $f(\alpha)$ is the probability of continuation for the good seller. Let $\hat{f} = f(\alpha)[1 - \phi(b)(1-\delta)]$ be the probability of continuation of the bad seller. The bad seller can default either due to not being able to meet the investment need, or due to the exercise of recourse by A/R lenders when it does not have enough additional cash flow. The expected relationship length is

$$
\begin{aligned}
&\mathbb{E}(T|\theta^t = 1, \gamma^t = 1) \\
&= h_t \sum_{\tau=1}^{\infty} \tau f(\alpha)^\tau + (1-h_t) \sum_{\tau=1}^{\infty} \tau \hat{f}^\tau \Pr(\theta_{t+\tau} = 0 \text{ or } \lambda_{t+\tau} = 0, \theta^{t+\tau-1} = \lambda^{t+\tau-1} = 1|bad) \\
&= h_t \sum_{\tau=1}^{\infty} \tau f(\alpha)^\tau + (1-h_t) \sum_{\tau=1}^{\infty} \tau \hat{f}^\tau (\sigma_b\sigma_c)^{\tau-1}(1-\sigma_b\sigma_c) \\
&= h_t \frac{f(\alpha)}{[1-f(\alpha)]^2} + (1-h_t)(1-\sigma_b\sigma_c)\frac{\hat{f}}{[1-\hat{f}\sigma_b\sigma_c]^2}. \qquad \text{(A.14)}
\end{aligned}
$$

The first term is the probability of the seller being good times the expected relationship length conditional on the seller is good, and the second term is the probability of the seller being bad times the expected relationship length conditional on the seller is bad.

Appendix B

Appendix for Chapter 2

B.1 Appendix: Proofs

B.1.1 Proof of Proposition 2.1

Taking the difference of the two FOCs, we have $(r_S - r_L)G'(r_S - r_L + v) = 1 - 2G(r_S - r_L + v)$. If $r_S \leq r_L$, then the left-hand side is non-positive but the right-hand side is $1 - 2G(r_S - r_L + v) \geq 1 - 2G(v) > 0$, a contradiction. So $r_S > r_L$. This implies that $G(r_S - r_L + v) < 0.5$ in equilibrium, i.e., $\alpha_L > \alpha_S$.

Let

$$B \equiv \frac{\frac{1}{A}\alpha_L \alpha_S (f - r_L)(f - r_S)G'(r_S - r_L + v)}{X + \alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)} > 0. \tag{B.1}$$

The two FOCs are separately written as

$$(f - r_L)G'(r_S - r_L + v) = \alpha_L + B, \tag{B.2}$$

$$(f - r_S)G'(r_S - r_L + v) = \alpha_S + B. \tag{B.3}$$

So both r_L and r_S are below f. Take the ratio:

$$\frac{f - r_L}{f - r_S} = \frac{\alpha_L + B}{\alpha_S + B} > 1 > \frac{\alpha_S}{\alpha_L}. \tag{B.4}$$

Hence, $(f - r_L)\alpha_L > (f - r_S)\alpha_S$, and $q_L^* < q_S^*$.

B.1.2 Proof of Proposition 2.2

By the assumption that the function Π_S is well behaved to admit a unique global maximum, the derivative $d\Pi_S/dr_S$ should be strictly decreasing in r_S. To show that $r_S > s$, it is sufficient that the right-hand side of (2.20) is positive at $r_S = s$, i.e.,

$$[X+\alpha_L(1-q_L^*)+\alpha_S(1-q_S^*)][(f-s)G'(v)-G(v)]-\frac{1}{A}G(v)(1-G(v))(f-s)^2G'(v) > 0. \tag{B.5}$$

Clearly, the above equation holds at $v = 0$. By continuity, it also holds if v is below a cutoff, say $\bar{v}$. If $v \in [0, \bar{v})$, we have $r_S > s = r_L$.

Let $B \equiv \frac{\frac{1}{A}\alpha_L\alpha_S(f-s)(f-r_S)G'(r_S-s+v)}{X+\alpha_L(1-q_L^*)+\alpha_S(1-q_S^*)} > 0$. The two FOCs are separately written as

$$(f-s)G'(r_S-s+v) < \alpha_L + B, \tag{B.6}$$

$$(f-r_S)G'(r_S-s+v) = \alpha_S + B. \tag{B.7}$$

Since $B > 0$, $G'(r_S - s + v) > 0$, we know $r_S < f$. Take the difference, we have $0 < (r_S - s)G'(r_S - s + v) < \alpha_L - \alpha_S$. That is, $\alpha_L > \alpha_S$. It follows that $(f-s)\alpha_L > (f-r_S)\alpha_S$ and $q_L^* < q_S^*$.

B.1.3 Proof of Proposition 2.3

Since the large bank is constrained by the lower bound, its deposit rate rises in step with the CBDC interest rate s. Meanwhile, the small bank adjusts its equilibrium deposit rate at a slower pace, continuing to balance its ability to maintain depositors while its profit margin shrinks. To see how r_S is affected by s, let $\Gamma_S = d\Pi_S/dr_S$, $\Gamma_L = d\Pi_L/dr_L$, and start with the expression

$$0 = \frac{\partial \Gamma_S}{\partial s} + \frac{\partial \Gamma_S}{\partial r_S}\frac{dr_S}{ds}. \tag{B.8}$$

Because $\partial\Gamma_S/\partial r_S < 0$, a sufficient condition for $dr_S/ds > 0$ is $\partial\Gamma_S/\partial s > 0$. Writing total loan volume as $V = \alpha_L(1-q_L^*) + \alpha_S(1-q_S^*)$, we have

$$\begin{aligned}\frac{\partial\Gamma_S}{\partial s} = &\left(-\alpha_L\frac{\alpha_L}{A}\right)[(f-r_S)G'(r_S-s+v) - G(r_S-s+v)]\\ &+ (X+V)[-(f-r_S)G''(r_S-s+v) + G'(r_S-s+v)]\\ &+ \frac{1}{A}(f-r_S)\frac{\partial}{\partial s}[\alpha_L\alpha_S(f-s)G'(r_S-s+v)].\end{aligned} \tag{B.9}$$

On any closed region of f and s, the first and third term are bounded, by G being twice-differentiable. So if X is sufficiently large, the second term dominates. Under the assumption that $G''(\delta) < G'(\delta)/f$ for $\delta \in [0, f-s+v]$, we have $-(f-r_S)G''(r_S-s+v) + G'(r_S-s+v) > 0$, so a sufficiently large X would imply that $\partial\Gamma_S/\partial s > 0$, and so is dr_S/ds.

Next, we show that $r_S - s$ decreases in s. We have

$$\begin{aligned}\frac{\partial\Gamma_S}{\partial r_S} = &X\frac{\partial}{\partial r_S}[(f-r_S)G'(r_S-s+v) - G(r_S-s+v)]\\ &+ \frac{\partial}{\partial r_S}\left\{[\alpha_L(1-q_L^*) + \alpha_S(1-q_S^*)]\cdot[(f-r_S)G'(r_S-s+v) - G(r_S-s+v)]\right\}\\ &- \frac{\partial}{\partial r_S}\left[\frac{1}{A}\alpha_L\alpha_S(f-s)(f-r_S)G'(r_S-s+v)\right].\end{aligned} \tag{B.10}$$

The second and the third term are bounded on any closed region of r_S. The first term equals $X[(f-r_S)G''(r_S-s+v) - 2G(r_S-s+v)]$. Hence,

$$\frac{dr_S}{ds} = -\frac{\partial\Gamma_S/\partial s}{\partial\Gamma_S/\partial r_S} \to \frac{(f-r_S)G''(r_S-s+v) - G'(r_S-s+v)}{(f-r_S)G''(r_S-s+v) - 2G'(r_S-s+v)}, \tag{B.11}$$

as X becomes sufficiently large. We also have $\frac{d(r_S-s)}{ds} = \frac{G'(r_S-s+v)}{(f-r_S)G''(r_S-s+v)-2G'(r_S-s+v)}$, whose denomintor is negative under the condition that $G''(\delta) < G'(\delta)/f$ Hence, $d(r_S - s)/ds < 0$. This implies that $\alpha_S = G(r_S-s+v)$ is decreasing in s, and α_L is increasing in s.

The weighted average interest rate is $\alpha_S r_S + \alpha_L s$. Its derivative with respect to s

is

$$\frac{d(\alpha_S r_S + \alpha_L s)}{ds} = \frac{d\alpha_S}{ds} r_S + \alpha_S \frac{dr_S}{ds} + \frac{d\alpha_L}{ds} s + \alpha_L = [(r_S - s)G'(r_S - s + v) + \alpha_S]\left(\frac{dr_S}{ds} - 1\right) + 1. \tag{B.12}$$

By the calculation earlier, as X becomes large, $\frac{dr_S}{ds} - 1 \to \frac{G'(r_S - s)}{(f - r_S)G''(r_S - s + v) - 2G'(r_S - s + v)} > -\frac{f}{f + r_S}$, where the inequality follows from $G''(\delta) < G'(\delta)/f$ for any $\delta \in [0, f - s + v]$. So, as X becomes large,

$$\frac{d(\alpha_S r_S + \alpha_L s)}{ds} > 1 - \frac{f}{f + r_S}[(r_S - s)G'(r_S - s + v) + \alpha_S] \geq 1 - \frac{f}{f + r_S} > 0, \tag{B.13}$$

where the second last inequality follows from the large bank's FOC that, $\lim_{X \to \infty}(f - s)G'(r_S - s + v) + G(r_S - s + v) \leq 1$.

Now we turn to loan market outcomes. Since α_S decreases in s and r_S increases in s, $\alpha_S(f - r_S)$ is decreasing in s and q_S^* is increasing in s. The small bank's loan volume, $\alpha_S(1 - q_S^*)$, is then decreasing in s.

For the large bank's loan quality q_L^*, we have

$$\frac{dq_L^*}{ds} = -\frac{1}{A}\Big[G'(r_S - s + v)(1 - \frac{dr_S}{ds})(f - s) - 1 + G(r_S - s + v)\Big]. \tag{B.14}$$

For the first term in the brackets, we know that $G'(r_S - s + v)(1 - \frac{dr_S}{ds})(f - s) < (f - s)G'(r_S - s + v)$, since $dr_S/ds > 0$. Also, from the large bank's optimality condition, as X is sufficiently large, we know that $(f - s)G'(r_S - s + v) - 1 + G(r_S - s + v) \leq 0$. That means $dq_L^*/ds > 0$ and q_L^* is increasing in s. However, the impact of s on the large bank's loan volume $\alpha_L(1 - q_L^*)$ is ambiguous.

The total loan volume is $\alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*)$. Its derivative with respect to s is

$$\frac{1}{A}[2\alpha_S(f - r_S) - 2\alpha_L(f - s)]G'(r_S - s + v)\left(\frac{dr_S}{ds} - 1\right) - \frac{1}{A}\alpha_L^2 - \frac{1}{A}\alpha_S^2\frac{dr_S}{ds}. \tag{B.15}$$

While the first term is positive, the last two terms are negative. It is, however, possible to show that this derivative is negative if $G''(\delta) \leq 0$ and X is sufficiently large. As X becomes large, the two first-order conditions imply that

$$\lim_{X\to\infty}(f-s)G'(r_S-s+v)-\underbrace{(1-G(r_S-s+v))}_{\alpha_L}\le 0,$$
$$\lim_{X\to\infty}(f-r_S)G'(r_S-s+v)-\underbrace{G(r_S-s+v)}_{\alpha_S}=0. \tag{B.16}$$

Multiplying α_L to the first equation and α_S to the second equation, we have

$$\lim_{X\to\infty}\alpha_L(f-s)G'(r_S-s+v)-\alpha_L^2\le 0,$$
$$\lim_{X\to\infty}\alpha_S(f-r_S)G'(r_S-s+v)-\alpha_S^2=0. \tag{B.17}$$

Plugging these in Equation (B.15), we have, as X becomes large,

$$\lim_{X\to\infty}\frac{d}{ds}(\alpha_L(1-q_L^*)+\alpha_S(1-q_S^*))\le\frac{1}{A}(2\alpha_L^2-2\alpha_S^2)\left(1-\frac{dr_S}{ds}\right)-\frac{1}{A}\alpha_L^2-\frac{1}{A}\alpha_S^2\frac{dr_S}{ds}$$
$$=\frac{1}{A}\left[\left(1-2\frac{dr_S}{ds}\right)\alpha_L^2+\left(\frac{dr_S}{ds}-2\right)\alpha_S^2\right], \tag{B.18}$$

Because $\frac{dr_S}{ds}<1$, $(\frac{dr_S}{ds}-2)\alpha_S^2<0$. If $G''(\delta)\le 0$ and X is sufficiently large, we know from the expression of $\frac{dr_S}{ds}$ above that $\frac{dr_S}{ds}\ge\frac{1}{2}$. That means $(1-2\frac{dr_S}{ds})\alpha_L^2\le 0$ as well. So the total loan is decreasing in s in the limit. Because the limit is strictly negative, it is also negative for finite but large enough X.

B.1.4 Proof of Proposition 2.4

First we consider the unconstrained equilibrium and then the constrained one.

The unconstrained equilibrium We know that $r_L<r_S<f$, and $\alpha_S<\frac{1}{2}<\alpha_L$. Let $\Gamma_S=d\Pi_S/dr_S$, $\Gamma_L=d\Pi_L/dr_L$. To calculate how r_L and r_S are affected by v, we take derivative of Γ_L and Γ_S at the equilibrium values and obtain

$$0=\frac{\partial\Gamma_L}{\partial v}+\frac{\partial\Gamma_L}{\partial r_L}\frac{dr_L}{dv}+\frac{\partial\Gamma_L}{\partial r_S}\frac{dr_S}{dv}, \tag{B.19}$$
$$0=\frac{\partial\Gamma_S}{\partial v}+\frac{\partial\Gamma_S}{\partial r_L}\frac{dr_L}{dv}+\frac{\partial\Gamma_S}{\partial r_S}\frac{dr_S}{dv}. \tag{B.20}$$

We solve for $\frac{dr_L}{dv}$ and $\frac{dr_S}{dv}$ from above equations. Denote $A_v = \frac{\partial \Gamma_L}{\partial v}$, $A_L = \frac{\partial \Gamma_L}{\partial r_L}$, $A_S = \frac{\partial \Gamma_L}{\partial r_S}$, $B_v = \frac{\partial \Gamma_S}{\partial v}$, $B_L = \frac{\partial \Gamma_S}{\partial r_L}$, and $B_S = \frac{\partial \Gamma_S}{\partial r_S}$. Then we have,

$$\frac{dr_L}{dv} = \frac{A_S B_v - B_S A_v}{A_L B_S - B_L A_S} \tag{B.21}$$

$$\frac{dr_S}{dv} = \frac{B_L A_v - A_L B_v}{A_L B_S - B_L A_S} \tag{B.22}$$

When X is sufficiently large, A_v is dominated by $X[(f - r_L)G''(r_S - r_L + v) + G'(r_S - r_L + v)]$, so $A_v \approx X[(f - r_L)G''(r_S - r_L + v) + G'(r_S - r_L + v)]$. Similarly, $B_v \approx X[(f - r_S)G''(r_S - r_L + v) + G'(r_S - r_L + v)]$, $A_L \approx X[-(f - r_L)G''(r_S - r_L + v) - 2G'(r_S - r_L + v)]$, $A_S \approx X[(f - r_L)G''(r_S - r_L + v) + G'(r_S - r_L + v)]$, $B_L \approx X[-(f - r_S)G''(r_S - r_L + v) + G'(r_S - r_L + v)]$, and $B_S \approx X[(f - r_S)G''(r_S - r_L + v) - 2G'(r_S - r_L + v)]$. Hence, $\frac{dr_L}{dv} \to \frac{(f - r_L)G''(r_S - r_L + v) + G'(r_S - r_L + v)}{(r_S - r_L)G''(r_S - r_L + v) + 3G'(r_S - r_L + v)}$, and $\frac{dr_S}{dv} \to \frac{(f - r_S)G''(r_S - r_L + v) - G'(r_S - r_L + v)}{(r_S - r_L)G''(r_S - r_L + v) + 3G'(r_S - r_L + v)}$. Since $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$, $(r_S - r_L)G''(r_S - r_L + v) + 3G'(r_S - r_L + v)$ is positive, and $(f - r_L)G''(r_S - r_L + v) + G'(r_S - r_L + v)$ is positive, so $\frac{dr_L}{dv} > 0$. Also, $(f - r_S)G''(r_S - r_L + v) - G'(r_S - r_L + v)$ is negative, so $\frac{dr_S}{dv} < 0$. So r_L is increasing and r_S is decreasing in v.

For deposit market share $\alpha_S = G(r_S - r_L + v)$, we take the difference of the two FOCs, and have

$$(r_S - r_L)G'(r_S - r_L + v) + 2G(r_S - r_L + v) = 1. \tag{B.23}$$

Write $y = r_S - r_L + v$, and take derivative of the above equation with respect to v, then we have

$$[3G'(y) + (r_S - r_L)G''(y)]\frac{dy}{dv} - G'(y) = 0 \tag{B.24}$$

Since $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$, we know that $3G'(y) + (r_S - r_L)G''(y) > 0$, hence $\frac{dy}{dv} > 0$. So α_S is increasing in v, and α_L is decreasing in v.

The weighted average deposit rate is $\alpha_S r_S + \alpha_L r_L = \alpha_S(r_S - r_L) + r_L$. Its derivative with respect to v is

$$\begin{aligned}\frac{d(\alpha_S r_S + \alpha_L r_L)}{dv} &= \frac{d\alpha_S}{dv}(r_S - r_L) + \alpha_S \frac{d(r_S - r_L)}{dv} + \frac{dr_L}{dv} \\ &> \frac{d\alpha_S}{dv}(r_S - r_L) + \frac{1}{2}\frac{d(r_S - r_L)}{dv} + \frac{dr_L}{dv} = \underbrace{\frac{d\alpha_S}{dv}}_{>0}(r_S - r_L) + \frac{1}{2}\frac{d(r_L + r_S)}{dv},\end{aligned} \tag{B.25}$$

where the inequality follows from $\alpha_S < \frac{1}{2}$ and $r_S - r_L$ decreasing in v. As X becomes large,

$$\frac{d(r_S + r_L)}{dv} \to \frac{(2f - r_L - r_S)G''(r_S - r_L + v)}{(r_S - r_L)G''(r_S - r_L + v) + 3G'(r_S - r_L + v)} \tag{B.26}$$

The denominator is positive as $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$. As $r_L + r_S < 2f$, when $G''(\delta) \geq 0$, we have $\frac{d(r_L + r_S)}{dv} \geq 0$, and hence $\alpha_S r_S + \alpha_L r_L$ increases in v.

For loan quality thresholds, since α_L is decreasing in v and r_L is increasing in v, q_L^* is increasing in v. Since α_S is increasing in v and r_S is decreasing in v, q_S^* decreasing in v.

For loan volumes, $\alpha_L(1 - q_L^*)$ is decreasing in v, since α_L is decreasing and q_L^* is increasing. Similarly, $\alpha_S(1 - q_S^*)$ is increasing in v.

Total loan volume equals $\alpha_L(1 - q_L^*) + \alpha_S(1 - q_S^*) = 1 - \frac{1+f}{A} + \frac{\alpha_L^2(f - r_L) + \alpha_S^2(f - r_S)}{A}$. Its derivative with respect to v is

$$\frac{1}{A}\left\{[2\alpha_S(f - r_S) - 2\alpha_L(f - r_L)]\frac{d\alpha_S}{dv} - \alpha_L^2\frac{dr_L}{dv} - \alpha_S^2\frac{dr_S}{dv}\right\}. \tag{B.27}$$

where $2\alpha_S(f - r_S) - 2\alpha_L(f - r_L) < 0$, $\frac{d\alpha_S}{dv} > 0$, $\frac{dr_L}{dv} > 0$, and $\frac{dr_S}{dv} < 0$. We know $-\alpha_L^2\frac{dr_L}{dv} - \alpha_S^2\frac{dr_S}{dv} < -\alpha_S^2\frac{d(r_L + r_S)}{dv}$. If $G''(\delta) \geq 0$, we know from above that $\frac{d(r_L + r_S)}{dv} \geq 0$, so $-\alpha_L^2\frac{dr_L}{dv} - \alpha_S^2\frac{dr_S}{dv} \leq 0$, and so Equation (B.27) is negative. If $G''(\delta) < 0$, however, the sign of the equation is ambiguous.

The constrained equilibrium To calculate how r_S is affected by v, we take derivative of Γ_S at the equilibrium values and obtain

$$0 = \frac{\partial \Gamma_S}{\partial v} + \frac{\partial \Gamma_S}{\partial r_S}\frac{dr_S}{dv}. \tag{B.28}$$

When X is sufficiently large, the term $X[(f-s)G''(r_S-s+v) - G'(r_S-s+v)]$ dominates $\frac{\partial \Gamma_S}{\partial v}$. Since $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$, we know that $\frac{\partial \Gamma_S}{\partial v} < 0$. The second-order condition implies that $\partial\Gamma_S/\partial r_S < 0$. Hence $\frac{dr_S}{dv} < 0$, i.e., r_S is decreasing in v.

For deposit market share $\alpha_S = G(r_S - s + v)$, when X becomes sufficiently large, we have

$$\frac{dr_S}{dv} = -\frac{\partial \Gamma_S}{\partial v}\Big/\frac{\partial \Gamma_S}{\partial r_S} \to -\frac{(f-r_S)G''(r_S-s+v) - G'(r_S-s+v)}{(f-r_S)G''(r_S-s+v) - 2G'(r_S-s+v)}. \tag{B.29}$$

Hence,

$$\frac{d(r_S-s+v)}{dv} = \frac{dr_S}{dv} + 1 = \frac{-G'(r_S-s+v)}{(f-r_S)G''(r_S-s+v) - 2G'(r_s-s+v)}, \tag{B.30}$$

where the numerator and the denominator are both negative. So $\frac{d(r_S-s+v)}{dv} > 0$, i.e., α_S is increasing in v and α_L is decreasing in v.

For weighted average deposit rate, take derivative with respect to v:

$$\begin{aligned}\frac{d}{dv}(\alpha_L s + \alpha_S r_S) &= \frac{d\alpha_S}{dv}(r_S - s) + \alpha_S\frac{dr_S}{dv} \\ &= \frac{(f-s)dr_S/dv + r_S - s}{f - r_S}\alpha_S\end{aligned} \tag{B.31}$$

where the second equality uses $d\alpha_S/dv \to \frac{\alpha_S}{(f-r_S)}(dr_S/dv+1)$, implied by the small bank's FOC when X is sufficiently large. The derivative is negative if and only if $(f-s)dr_S/dv + r_S - s < 0$. Write $y = r_S - r_L + v$. Plugging in $\frac{dr_S}{dv} = -\frac{(f-r_S)G''(y)-G'(y)}{(f-r_S)G''(y)-2G'(y)}$, the derivative is negative if and only if $G''(r_S-s+v) \leq \frac{f+s-2r_S}{(f-r_S)^2}G'(r_S-s+v)$. We now show that $f + s - 2r_S \geq 0$. We know that r_S is decreasing in v. So we only need to show, given s, $f+s-2r_S \geq 0$ when $v = 0$. Let $l(x) = (f-x)G'(x-s) - G(x-s)$, then $l(r_S)=0$, and $\frac{dl(x)}{dx}=(f-x)G''(x-s)-2G'(x-s)<0$ under the condition that $-G'(\delta)/f<G''(\delta)<G'(\delta)/f$ for any $\delta\in[0,f-s+v]$. To show that $f+s-2r_S\geq 0$, we only need $l(\frac{1}{2}(f+s))\leq 0$. That is, $\frac{f-s}{2}G'(\frac{f-s}{2})-G(\frac{f-s}{2})\leq 0$. This is true because if we let $m(x)=xG'(x)-G(x)$, then $\frac{dm(x)}{dx}=xG''(x)\leq 0$. And since $m(0)=0$, we have $m(\frac{f-s}{2})\leq 0$.

For loan quality thresholds and individual banks' loan volumes, the same proofs for the unconstrained equilibrium apply and are omitted.

Total loan volume equals $\alpha_L(1-q_L^*)+\alpha_S(1-q_S^*)=1-\dfrac{1+f}{A}+\dfrac{\alpha_L^2(f-s)+\alpha_S^2(f-r_S)}{A}$. Its derivative with respect to v is

$$\frac{1}{A}\left\{[2\alpha_S(f-r_S)-2\alpha_L(f-s)]\frac{d\alpha_S}{dv}-\alpha_S^2\frac{dr_S}{dv}\right\}. \tag{B.32}$$

Its sign is ambiguous because while the first term in the brackets is negative, the second term is positive.

B.1.5 Proof of Proposition 2.5

In the unconstrained equilibrium, the difference of the two banks' FOCs leads to Equation (2.16) . So r_S-r_L does not vary with f.

Further, let $B\equiv\frac{\frac{1}{A}\alpha_L\alpha_S(f-r_L)(f-r_S)G'(r_S-r_L+v)}{X+\alpha_L(1-q_L^*)+\alpha_S(1-q_S^*)}$, then the two FOCs are separately written as

$$(f-r_L)G'(r_S-r_L+v)=1-G(r_S-r_L+v)+B \tag{B.33}$$

$$(f-r_S)G'(r_S-r_L+v)=G(r_S-r_L+v)+B \tag{B.34}$$

When X is sufficiently large, $B\to 0$. Hence as f changes, r_L and r_S move one-for-one with f.

In the constrained equilibrium, let $\Gamma_S=d\Pi_S/dr_S$. To study the sensitivity of r_S to f, we take derivative of Γ_S at the equilibrium values and obtain

$$0=\frac{\partial\Gamma_S}{\partial f}+\frac{\partial\Gamma_S}{\partial r_S}\frac{dr_S}{df} \tag{B.35}$$

When X becomes sufficiently large, we have

$$\frac{dr_S}{df} = -\frac{\partial \Gamma_S}{\partial f} / \frac{\partial \Gamma_S}{\partial r_S} \rightarrow \frac{1}{-(f - r_S)\frac{G''(r_S - s + v)}{G'(r_S - s + v)} + 2} \tag{B.36}$$

We know that $(f - r_S)$ is decreasing in s and increasing in v, and that $r_S - s + v$ is decreasing in s and increasing in v. If G satisfies that $\frac{G''(\delta)}{G'(\delta)}$ is increasing in δ for any $\delta \in [0, f - s + v]$, then the denominator is increasing in s and decreasing in v . Also, since $-G'(\delta)/f < G''(\delta) < G'(\delta)/f$ for any $\delta \in [0, f - s + v]$, the denominator is positive. So $\frac{dr_S}{df}$ is decreasing in s and increasing in v.

Given v and s, we solve for the cut-off value f^* that separates the constrained equilibrium and the unconstrained equilibrium. Let r_S^* be the equilibrium value of r_S when $f = f^*$. When X is sufficiently large, the small bank's FOC and the large bank's FOC are

$$(f^* - r_S^*)G'(r_S^* - s + v) - G(r_S^* - s + v) = 0 \tag{B.37}$$

$$(f^* - s)G'(r_S^* - s + v) - 1 + G(r_S^* - s + v) = 0 \tag{B.38}$$

Taking the difference of the two equations, we have

$$\Delta = (r_S^* - s)G'(r_S^* - s + v) - 1 + 2G(r_S^* - s + v) = 0 \tag{B.39}$$

So $r_S^* - s = m(v)$ is a function of v. By the small bank's FOC, we have

$$f^* = s + \frac{G(m(v) + v)}{G'(m(v) + v)} + m(v) \tag{B.40}$$

Hence f^* increases one-for-one with s. To show that f^* is decreasing in v, we take derivative of Δ to solve for $\frac{dm(v)}{dv}$

$$0 = \frac{\partial \Delta}{\partial v} + \frac{\partial \Delta}{\partial m(v)} \frac{dm(v)}{dv} \tag{B.41}$$

Hence, $\frac{dm(v)}{dv} = -\frac{\partial \Delta}{\partial v} / \frac{\partial \Delta}{\partial m(v)} = -\frac{m(v)G''(m(v)+v)+2G'(m(v)+v)}{m(v)G''(m(v)+v)+3G'(m(v)+v)}$. Denoting $y = m(v) + v =$

$r_S^* - s + v$, we have

$$\begin{aligned}\frac{df^*}{dv} &= \frac{d}{dv}\left\{\frac{G(y)}{G'(y)}\right\} + \frac{dm(v)}{dv} \\ &= \frac{G'(y)^2[-m(v)G''(y) - G'(y)] - G''(y)G'(y)G(y)}{G'(y)^2[m(v)G''(y) + 3G'(y)]}\end{aligned} \tag{B.42}$$

where $r_S^* - s = m(v) > 0$. For G that satisfies $0 \leq G''(\delta) < G'(\delta)/f$ for any $\delta \in [0, f - s + v]$, we know the denominator is positive, and the numerator is negative, so $\frac{df^*}{dv} < 0$.

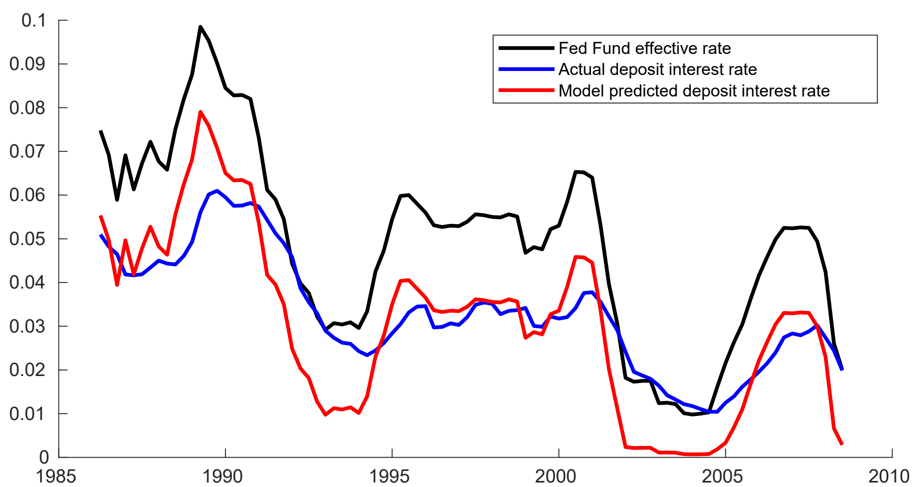
B.2 Appendix: Fitting U.S. Deposit Rates Data

In this appendix, we illustrate the predictive performance of our model using U.S. data on deposit rates from 1986 to 2021. While our model is highly stylized and features only two banks, it captures qualitative features of deposit rates: they are lower than and somewhat non-responsive to the interest rate on reserves.

We use the model to predict deposit rates. The opportunity cost of funds for banks is determined in part by either the IOR rate or the federal funds rate, whichever is larger. In the period before the 2008-09 crisis the relevant rate was the federal funds rate. In the period after the crisis the relevant rate was (generally) the IOR rate. We use the higher of the two rates in each period as f, and apply the model under a specific parameterization. Specifically, we assume the convenience value for deposit at the large bank, δ, is distributed uniformly from 0 to 3.5%. The resulting predicted data series seems to fit the actual data reasonably well. Figure B-1 shows the actual and predicted U.S. deposit rates from 1986Q1 to 2008Q2 relative to the federal fund rate. Figure B-2 shows the actual and predicted U.S. deposit rates from May 2009 to February 2021 relative to the IOR rate.

In Figure B-1, predicted deposit rates match the levels of actual deposit rates quite well and move almost one-for-one with the federal funds rate. The main deviation from the actual data is that predicted rates are too sensitive to changes in the federal funds rate. This is not surprising, as in our model, deposit rates of both the large and small bank move one-for-one with the federal funds rate in the unconstrained equilibrium. A noticeable deviation occurs during the pre-crisis period from 2001Q3 to 2004Q3. This is when the large bank's predicted deposit rate is constrained at the zero lower bound. The transition to low deposit rates associated with the constrained equilibrium is immediate in our model, but not in the data. Our model does not build in any "stickiness" into the deposit rate that would be necessary to match the data more closely.

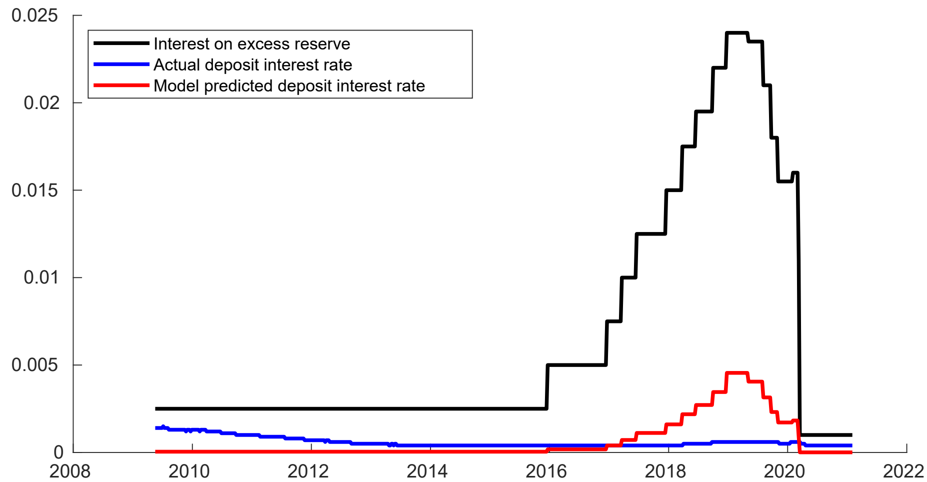
Figure B-1: Actual and Predicted U.S. Deposit Rates from 1986Q1 to 2008Q2.



Domestic deposit rates are quarterly, calculated from call reports, as total interest expense on domestic deposits divided by total domestic deposits, multiplied by 4. The model-implied interest rate is the weighted average of the large bank's and the small bank's deposit rates, weighted by their market shares. Model parameters: $G(\delta) = \delta/0.035, A = 1.5, X = 10, s = 0$.

The time period in Figure B-2 is characterized by a long stretch of near zero rates in the Federal Funds market and an IOR rate of 25 basis points. As the IOR rate was typically higher than the federal funds rate, the IOR rate is the relevant variable for predicting deposit rates. Deposit rates fell slowly during this period toward zero until the Fed began to raise the IOR rate in December 2015. The Fed raised the IOR rate multiple time reaching a peak of 2.40% from December 2018 to April 2019, but deposit rates reacted very slowly. Our model's predicted deposit rate captures this non-responsiveness. It is still too sensitive to changes in IOR compared to the data,

Figure B-2: Actual and Predicted U.S. Deposit Rates from May 18, 2009 to February 1, 2021



Weekly deposit rates for amounts less than $100,000 are obtained from the FDIC through FRED. The model-implied interest rate is the weighted average of the large bank's and the small bank's deposit rates, weighted by their market shares. Model parameters: $G(\delta) = \delta/0.035, A = 1.5, X = 10, s = 0$.

but the deviation is not large. The low deposit rates that are predicted by our model occur because at the low IOR rates that existed during most of this period the zero lower bound is binding in our model and hence average market rates are determined largely by large bank deposit rates which are constrained at zero.

Appendix C

Appendix for Chapter 3

C.1 Appendix: MOC Orders and Off-Exchange Trades

C.1.1 Predominant use of MOC orders in close auctions

Table C.1: Order Type Usage in NYSE Exchanges as of March 2020: Percentage of Matched Volumes

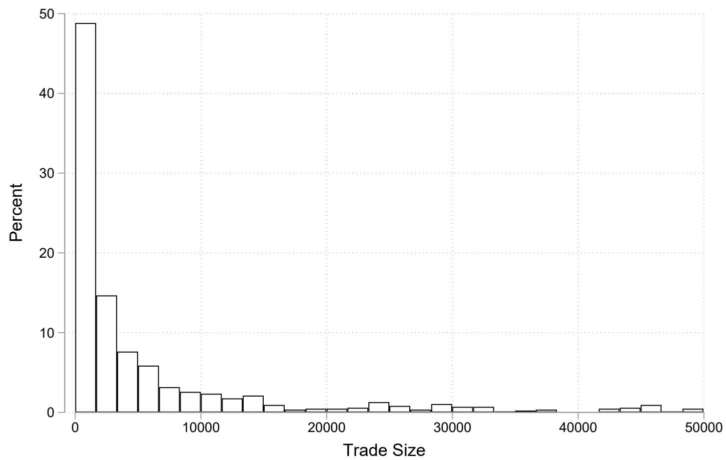
	NYSE	NYSE Arca	NYSE American
Auction	**28.21%**	**5.95%**	**8.07%**
Market-on-Close	16.37%	3.72%	4.61%
Limit-on-Close	3.63%	1.18%	1.59%
Market-on-Open	1.88%	0.55%	0.99%
Limit-on-Open	1.22%	0.49%	0.88%
Closing D-Orders	5.07%	0.00%	0.00%
Closing Offset	0.04%	0.00%	0.00%

Notes: The table reports the percentage of matched total daily volumes constituted by different order types. The table should be interpreted as: in NYSE in Mar 2020, 28.21% of total daily volume happens in auctions, and 16.37% of total daily volume are triggered by market-on-close orders. Source: https://www.nyse.com/publicdocs/nyse/NYSE_Group_Executed_Order_Type_Usage.xlsx

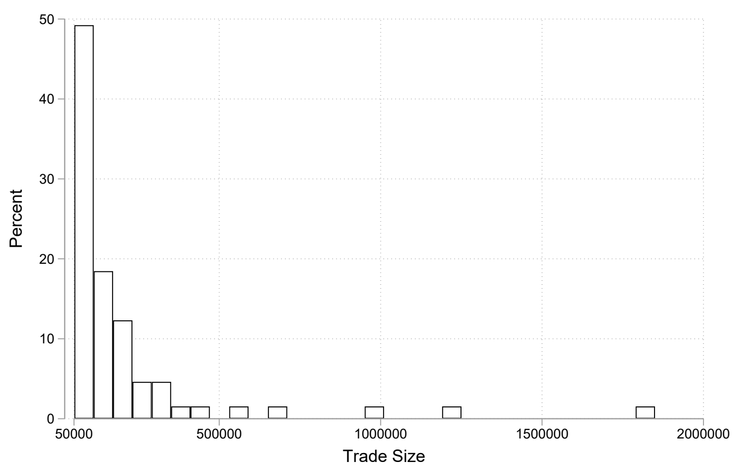
C.1.2 Distribution of off-exchange trade size

Figure C-1: Histogram of the Sizes of Off-exchange MOC Trade for Stock AAPL in Jan 2018

(a) Histogram of off-exchange MOC trade size (AAPL Jan 2018) ($\leq$ 50,000 Shares)



(b) Histogram of off-exchange MOC trade size (AAPL Jan 2018) ($>$ 50,000 Shares)



Notes: Off-exchange MOC trades are all the trades in the TAQ data that are not canceled or corrected and occur between 4:00 p.m. - 4:10 p.m. EST at the official market close price determined by the close auction. The figures plot the distribution of the off-exchange MOC trades of the common stock of Apple, Inc in January 2018.

C.2 Appendix: Proofs

C.2.1 Proof of Proposition 3.1

The bank chooses $d(y)$ that maximizes its expected trading profits, given the net orders received $y = u_b + \theta\eta v$. The bank solves the following optimization problem

$$\max_d E[v - p|y]d(y) \tag{C.1}$$

when taking the fee φ_b as given. Here α, θ are the fractions of uninformed and informed traders that trade through the bank. Guess that the market maker applies a linear price setting rule $p = \lambda z = \lambda\left[u + \eta v + d(y)\right]$. For simplicity, we write $d(y)$ as d in the following proof.

Knowing that $v \sim N(0, \sigma^2), u \sim N(0, \sigma_u^2)$, the bank's expectation of the asset value is

$$\begin{aligned} E[v|y] &= E[v] + \frac{Cov(v,y)}{Var(y)}(y - E[y]) \\ &= \frac{\theta\eta\sigma^2}{\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2} y \end{aligned} \tag{C.2}$$

And similarly the bank's expectation of net liquidity orders is

$$E[u|y] = \frac{\alpha\sigma_u^2}{\theta^2\eta^2\sigma^2 + \alpha^2\sigma_u^2} y \tag{C.3}$$

Then we can solve for the bank's problem as

$$\begin{aligned} \max_d E[v - p|y]d &= E[v - \lambda\left[u + \eta v + d\right]|y]d \\ &= \left((1 - \lambda\eta)E[v|y] - \lambda E[u|y] - \lambda d\right) d \\ &= \left(\left[(1 - \lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2} - \lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right] y - \lambda d\right) d \end{aligned} \tag{C.4}$$

where $\sigma_y^2 = \theta^2\eta^2\sigma^2 + \alpha\sigma_u^2$. Taking first order conditions, we get

$$d = \frac{1}{2\lambda}\left[(1-\lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2} - \lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right]y$$
$$= \left[\frac{1}{2\lambda}\frac{\theta\eta\sigma^2}{\sigma_y^2} - \frac{1}{2}\frac{\theta\eta^2\sigma^2 + \alpha\sigma_u^2}{\sigma_y^2}\right]y \tag{C.5}$$

Hence

$$d = K(\lambda)y \tag{C.6}$$

where

$$K(\lambda) = \frac{1}{2\lambda}\frac{\theta\eta\sigma^2}{\sigma_y^2} - \frac{1}{2}\frac{\theta\eta^2\sigma^2 + \alpha\sigma_u^2}{\sigma_y^2} \tag{C.7}$$

Next we derive the price setting rule of the market maker and show that it takes the linear form as we guessed. The market maker receives net order

$$z = u + \eta v + d(y) = u_m + (1 + K(\lambda))u_b + (1 + K(\lambda)\theta)\eta v \tag{C.8}$$

and knows that $v \sim N(0, \sigma^2), u \sim N(0, \sigma_u^2)$. Since the market maker is competitive, the close price equals the market maker's expected value of the asset. Then the close price is

$$p = E[v|z] = \frac{(1 + K(\lambda)\theta)\eta\sigma^2}{\left((1 + K(\lambda))^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1 + K(\lambda)\theta)^2\eta^2\sigma^2}z \tag{C.9}$$

which can be written as

$$p = \lambda z \tag{C.10}$$

where

$$\lambda = \frac{(1 + K(\lambda)\theta)\eta\sigma^2}{\left((1 + K(\lambda))^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1 + K(\lambda)\theta)^2\eta^2\sigma^2} \tag{C.11}$$

Therefore, the equilibrium is characterised by the bank's optimal strategy (C.7) and the market maker's price setting rule (C.11). Substituting λ as a function of K, i.e., Equation (C.11) into Equation (C.7), we can solve for K as a result of a quadratic equation:

$$0 = \sigma_y^2 K^2 + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right) K + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2 \tag{C.12}$$

where $\sigma_y^2 = \theta^2\eta^2\sigma^2 + \alpha\sigma_u^2$. Then

$$K = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{C.13}$$

where $A = \sigma_y^2, B = \theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$.[1] And

$$\lambda = \frac{(1+K\theta)\eta\sigma^2}{\left((1+K)^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.14}$$

Both K, λ are constants that only depend on parameters. The proposition is hence proved.

C.2.2 Proof of Proposition 3.2

By substituting previous results into the first term, the expected trading profit can be written as $\lambda(\varphi_b)K^2(\varphi_b)\sigma_y(\varphi_b)^2 \geq 0$, which is non-negative.

When the above sufficient conditions are satisfied, the bank makes a positive profit by setting a fee slightly lower than φ_m and match some liquidity orders. Note that the bank's total profit is negative when φ_b goes to $-\infty$. Since the bank's total profit is a continuous function of φ_b, there exists $\varphi_b < \varphi_m$ such that the break-even conditions holds.

C.2.3 Proof of Proposition 3.3

We first show that

$$\frac{(1+K\theta)^2}{(1+K)^2\alpha + 1 - \alpha} \geq 1 \tag{C.15}$$

[1]The other solution of the quadratic equation $\frac{-B-\sqrt{B^2-4AC}}{2A}$ does not optimize the bank's profit here.

where

$$K = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{C.16}$$

where $A = \sigma_y^2, B = \theta a^2 \sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$.

If $\theta = \alpha$, then $K = 0$ and the LHS equals 1.

If $\theta > \alpha$, then $K > 0$, the problem is equivalent to

$$\theta^2 K + 2\theta > \alpha K + 2\alpha \tag{C.17}$$

The inequality immediately holds when $\theta^2 \geq \alpha$. When $\theta^2 < \alpha$, the inequality can be written as

$$K < \frac{2(\theta - \alpha)}{\alpha - \theta^2} \tag{C.18}$$

By substituting K as a function of the parameters and rearranging, we can write the inequality as

$$0 < D_1 a^2 \sigma^2 + D_2 \sigma_u^2 \tag{C.19}$$

where

$$D_1 = 2\frac{(\theta - \alpha)\theta}{(\alpha - \theta^2)^2}(\theta^2 - 2\alpha\theta + \alpha) \tag{C.20}$$

$$D_2 = 4\frac{(\theta - \alpha)^2}{(\alpha - \theta^2)^2}\alpha + \frac{1}{\alpha - \theta^2}\left(3\alpha - 2\alpha\theta + \theta^2\right)\left(1 - \frac{\alpha}{\theta}\right) \tag{C.21}$$

Since $D_1 > 0$ and $D_2 > 0$, the inequality holds.

If $\theta < \alpha$, then $K < 0$, the problem can be written as

$$K > \frac{2(\theta - \alpha)}{\alpha - \theta^2} \tag{C.22}$$

By substituting K as a function of the parameters and rearranging, we can write the inequality as

$$0 > D_1 \eta^2 \sigma^2 + D_2 \sigma_u^2 \tag{C.23}$$

where

$$D_1 = 2\frac{(\theta-\alpha)\theta}{(\alpha-\theta^2)^2}(\theta^2 - 2\alpha\theta + \alpha) \tag{C.24}$$

$$D_2 = \frac{\theta-\alpha}{(\alpha-\theta^2)^2}\left[3\frac{\alpha^2}{\theta} - 6\alpha^2 + 2\alpha\theta + 2\alpha\theta^2 - \theta^3\right] \tag{C.25}$$

Since $D_1 < 0$ and $D_2 < 0$, the inequality holds.

Then

$$\xi = \frac{(1+K\theta)^2\eta^2\sigma^2}{[(1+K)^2\alpha + 1 - \alpha]\,\sigma_u^2} \geq \frac{\eta^2\sigma^2}{\sigma_u^2} = \xi_{\text{no bank}} \tag{C.26}$$

Hence $MSE_{\text{with bank}} \leq MSE_{\text{no bank}}$. Having the bank conducting the service improves price informativeness.

C.2.4 Proof of Proposition 3.4

Since the MSE measure is strictly decreasing in ξ, it is sufficient to show that $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} \geq 0$, where $\delta = \frac{\theta}{\alpha}$ and $\zeta = \frac{\alpha\sigma_u^2+\theta\eta^2\sigma^2}{\sigma_u^2+\eta^2\sigma^2}$.

First, notice that $\zeta = \alpha\frac{\sigma_u^2+\delta\eta^2\sigma^2}{\sigma_u^2+\eta^2\sigma^2}$, so if we rewrite $\xi(\delta,\zeta)$ as $\xi(\delta,\alpha)$, $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} = \frac{\sigma_u^2+\delta\eta^2\sigma^2}{\sigma_u^2+\eta^2\sigma^2}\frac{\partial\xi(\delta,\alpha)}{\partial\alpha}$. Therefore, showing $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} \geq 0$ is equivalent to showing that $\frac{\partial\xi(\delta,\alpha)}{\partial\alpha} \geq 0$. We write $K(\alpha,\delta)$ as K for simplicity.

$$\frac{\partial\xi(\alpha,\delta)}{\partial\alpha} = \frac{\eta^2\sigma^2}{\sigma_u^2}\frac{\partial}{\partial\alpha}\frac{[1+\alpha\delta K(\alpha,\delta)]^2}{[1+K(\alpha,\delta)]^2\alpha + 1 - \alpha} \tag{C.27}$$

$$= \frac{\eta^2\sigma^2}{\sigma_u^2}\frac{1}{[(1+K)^2\alpha+1-\alpha]^2}\Bigg\{2(1+\alpha\delta K)\left(\delta K + \alpha\delta\frac{\partial K}{\partial\alpha}\right)\left[(1+K)^2\alpha + 1 - \alpha\right]$$
$$- (1+\alpha\delta K)^2\left[2K + K^2 + 2\alpha(1+K)\frac{\partial K}{\partial\alpha}\right]\Bigg\} \tag{C.28}$$

Since $\frac{\eta^2\sigma^2}{\sigma_u^2}\frac{(1+\theta K)}{[(1+K)^2\alpha+1-\alpha]^2} > 0$, $\frac{\partial\xi(\alpha,\delta)}{\partial\alpha}$ takes the same sign as

$$\tilde{\xi} = 2\left(\frac{\theta}{\alpha}K + \theta\frac{\partial K}{\partial\alpha}\right)\left[(1+K)^2\alpha + 1 - \alpha\right] - (1+\theta K)\left[2K + K^2 + 2\alpha(1+K)\frac{\partial K}{\partial\alpha}\right] \tag{C.29}$$

$$= \theta K^3 + (2\theta - 1)K^2 + 2\left(\frac{\theta}{\alpha} - 1\right)K + 2\left[(\theta-1)\alpha K + \theta - \alpha\right]\frac{\partial K}{\partial\alpha} \tag{C.30}$$

Rewrite (C.13) as a function of α and δ and take its partial derivative to α, we get

$$\frac{\partial K}{\partial \alpha} = -K\frac{(2\alpha\delta^2\eta^2\sigma^2+\sigma_u^2)K+\delta\eta^2\sigma^2-\sigma_u^2}{\sqrt{\left[\alpha\delta\eta^2\sigma^2+(\frac{2}{\delta}-\alpha)\sigma_u^2\right]^2-4(\alpha^2\delta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{1}{\delta}-1)\sigma_u^2}} \tag{C.31}$$

$$= -K\frac{(2\frac{\theta^2}{\alpha}\eta^2\sigma^2+\sigma_u^2)K+\frac{\theta}{\alpha}\eta^2\sigma^2-\sigma_u^2}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}} \tag{C.32}$$

$$= -\frac{K}{\alpha}+\sigma_u^2K\frac{K+\frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}} \tag{C.33}$$

Then (C.30) can be written as

$$\tilde{\xi} = \theta K^3+(2\theta-1)K^2+2\left(\frac{\theta}{\alpha}-1\right)K-2\left[(\theta-1)\alpha K+\theta-\alpha\right]\frac{K}{\alpha}$$
$$+2\left[(\theta-1)\alpha K+\theta-\alpha\right]\sigma_u^2K\frac{K+\frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}}$$
$$= (1+\theta K)K^2 \tag{C.34}$$
$$+2\left[(\theta-1)\alpha K+\theta-\alpha\right]\sigma_u^2K\frac{K+\frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}} \tag{C.35}$$

Since $K \geq -\frac{1}{2}$ and $\theta \in [0,1]$, the first term $(1+\theta K)K^2$ is non-negative and is positive when $K \neq 0$. The second term has the same sign as $\left[(\theta-1)\alpha K+\theta-\alpha\right]K$. Next we show that it is positive when $K \neq 0$.

When $\theta = 1$, it becomes $(\theta-\alpha)K > 0$. When $\theta < 1$, let $\tilde{K} = \frac{\theta-\alpha}{(1-\theta)\alpha}$. Since K and $\tilde{K}$ have the same sign as $\theta-\alpha$, $\left[(\theta-1)\alpha K+\theta-\alpha\right]K$ is positive as long as $|\tilde{K}| > |K|$. Since K is the larger root of (C.12), $\sigma_y^2 > 0$, $\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2 > 0$ and $\tilde{K}$ has the same sign as K, $|\tilde{K}| > |K|$ as long as

$$0 < K\left[\sigma_y^2\tilde{K}^2 + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right)\tilde{K} + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2\right] \tag{C.36}$$

$$= K\left[\frac{\left(\frac{\theta}{\alpha} - 1\right)^2}{(1-\theta)^2}(\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2) + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right)\frac{\frac{\theta}{\alpha} - 1}{1-\theta} + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2\right] \tag{C.37}$$

$$= \frac{K(\theta-\alpha)}{(1-\theta)^2}\left[\left(\frac{\theta^2}{\alpha} - 2\theta + 1\right)\frac{\theta}{\alpha}\eta^2\sigma^2 + \left(\frac{\theta}{\alpha} - 2 + \frac{1}{\theta}\right)\sigma_u^2\right] \tag{C.38}$$

The inequality holds since $K(\theta-\alpha) > 0$, $\left(\frac{\theta^2}{\alpha} - 2\theta + 1\right) > 0$, and $\left(\frac{\theta}{\alpha} - 2 + \frac{1}{\theta}\right) > 0$ when $\alpha, \theta \in (0,1)$. Then (C.33) is non-negative and is positive when $K \neq 0$. Therefore, $\frac{\partial\xi(\alpha,\delta)}{\partial\alpha} \geq 0$ and the inequality is strict when $\theta \neq \alpha$.

C.2.5 Proof of Proposition 3.5

The expected trading profit before fees of informed traders is

$$\Pi_x = \mathbb{E}[\mathbb{E}[(v-p)\eta v|v]] \tag{C.39}$$

$$= \eta\sigma^2\frac{\left((1+K)^2\alpha + (1-\alpha)\right)\sigma_u^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.40}$$

The expected trading profit before fees of liquidity traders who trade directly with the exchange is

$$\Pi_{um} = \mathbb{E}[\mathbb{E}[(v-p)u_m|u_m]] \tag{C.41}$$

$$= -(1-\alpha)\sigma_u^2\frac{(1+K\theta)\eta\sigma^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.42}$$

The expected trading profit before fees of liquidity traders who trade with the bank is

$$\Pi_{ub} = \mathbb{E}[\mathbb{E}[(v-p)u_b|u_b]] \tag{C.43}$$

$$= -\alpha\sigma_u^2\frac{(1+K)(1+K\theta)\eta\sigma^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.44}$$

Note again that the equilibrium results with no bank coincides with the results when we impose $K = 0$, that is, the bank do not trade. Then we compare the profits of the traders with and without the bank and get the following results.

Informed traders' expected trading profit before fees in the "no bank" equilibrium coincides with the profit when $K = 0$, that is

$$\Pi_{x,\text{no bank}} = \eta\sigma^2 \frac{\sigma_u^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.45}$$

Following Proposition 3.3, we have $\Pi_x < \Pi_{x,\text{no bank}}$.

Similarly, liquidity traders' expected trading profit before fees in the "no bank" equilibrium is

$$\Pi_{u,\text{no bank}} = -\sigma_u^2 \frac{\eta\sigma^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.46}$$

For traders who still trade with the exchange in the equilibrium with the bank, they are better of if and only if

$$\Pi_{um} > (1-\alpha)\Pi_{u,\text{no bank}} \tag{C.47}$$

That is equivalent to

$$\lambda < \lambda_{\text{no bank}} = \frac{\eta\sigma^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.48}$$

When $\theta < \alpha$ and $K < 0$, $1 > 1 + K\theta > (1 + K\theta)^2 > (1 + K)^2\alpha + (1 - \alpha)$, then $\Pi_{um} < (1 - \alpha)\Pi_{u,\text{no bank}}$.

When $\theta > \alpha$ and $K > 0$, substitute K from Proposition 3.1 into the inequality (C.47) and reorganize, we get

$$(\theta - 3\alpha)\sigma_u^2 < 2\theta\eta^2\sigma^2 \tag{C.49}$$

For traders who trade with the bank in the equilibrium with the bank, they are better off if and only if

$$\Pi_{ub} > \alpha \Pi_{u,\text{no bank}} \tag{C.50}$$

When $\theta > \alpha$ and $K > 0$, $(1+K)(1+K\theta) > (1+K\theta)^2 > (1+K)^2\alpha + (1-\alpha)$, then $\Pi_{ub} < \alpha \Pi_{u,\text{no bank}}$.

When $\theta < \alpha$ and $K < 0$, substitute K from Proposition 3.1 into the inequality (C.50) and reorganize, we get

$$(1-\theta)(1+K\theta)\eta^2\sigma^2 + [1+\theta-2\alpha+(\theta-\alpha)K]\sigma_u^2 > 0 \tag{C.51}$$

Bibliography

Agur, I., A. Ari, and G. Dell'Ariccia. 2022. Designing central bank digital currencies. *Journal of Monetary Economics* 125:62–79.

Alfaro, L., P. Antràs, D. Chor, and P. Conconi. 2019. Internalizing global value chains: A firm-level analysis. *Journal of Political Economy* 127:508–59. doi:10.1086/700935.

Alfaro, L., M. García-Santana, and E. Moral-Benito. 2021. On the direct and indirect real effects of credit supply shocks. *Journal of Financial Economics* 139:895–921. ISSN 0304-405X. doi:https://doi.org/10.1016/j.jfineco.2020.09.004.

Almeida, H., M. Campello, B. Laranjeira, and S. Weisbenner. 2011. Corporate debt maturity and the real effects of the 2007 credit crisis. *Critical Finance Review* 1:3–58.

Amberg, N., T. Jacobson, and E. von Schedvin. 2021. Trade credit and product pricing: The role of implicit interest rates. *Journal of the European Economic Association* 19:709–40.

Andolfatto, D. 2021. Assessing the impact of central bank digital currency on private banks. *The Economic Journal* 131:525–40.

Bakke, T.-E., and T. M. Whited. 2010. Which firms follow the market? an analysis of corporate investment decisions. *The Review of Financial Studies* 23:1941–80.

Bakker, M.-R., G. F. Udell, and L. Klapper. 2004. *Financing small and medium-size enterprises with factoring: Global growth and its potential in Eastern Europe*, vol. 3342. World Bank Publications.

Baldauf, M., C. Frei, and J. Mollner. 2021. Principal trading arrangements: When are common contracts optimal? *Management Science* .

Baltussen, G., S. van Bekkum, and Z. Da. 2019. Indexing and stock market serial dependence around the world. *Journal of Financial Economics* 132:26–48.

Bank of England. 2020. Central bank digital currency: Opportunities, challenges and design.

Bank of England Quarterly Bulletin. 2014. Money creation in the modern economy.

Barbon, A., M. Di Maggio, F. Franzoni, and A. Landier. 2019. Brokers and order flow leakage: Evidence from fire sales. *The Journal of Finance* 74:2707–49.

Barrot, J.-N. 2016. Trade credit and industry dynamics: Evidence from trucking firms. *The Journal of Finance* 71:1975–2016.

Barrot, J.-N., and R. Nanda. 2020. The employment effects of faster payment: Evidence from the Federal Quickpay reform. *The Journal of Finance* 75:3139–73.

Beaumont, P., and C. Lenoir. 2019. Building a customer base under liquidity constraints. *Available at SSRN 3232638* .

Ben-David, I., F. Franzoni, and R. Moussawi. 2018. Do ETFs increase volatility? *The Journal of Finance* 73:2471–535.

Ben-David, I., J. Li, A. Rossi, and Y. Song. 2021. What do mutual fund investors really care about? *The Review of Financial Studies* forthcoming. ISSN 0893-9454. doi:10.1093/rfs/hhab081.

Benmelech, E., and N. K. Bergman. 2009. Collateral pricing. *Journal of Financial Economics* 91:339–60.

Benmelech, E., M. J. Garmaise, and T. J. Moskowitz. 2005. Do liquidation values affect financial contracts? Evidence from commercial loan contracts and zoning regulation. *The Quarterly Journal of Economics* 120:1121–54.

Bhattacharya, N., E. L. Black, T. E. Christensen, and R. D. Mergenthaler. 2007. Who trades on pro forma earnings fnformation? *The Accounting Review* 82:581–619. ISSN 0001-4826. doi:10.2308/accr.2007.82.3.581.

Biais, B., and C. Gollier. 1997. Trade credit and credit rationing. *The Review of Financial Studies* 10:903–37.

Bigio, S., and J. La'O. 2020. Distortions in Production Networks. *The Quarterly Journal of Economics* 135:2187–253. ISSN 0033-5533. doi:10.1093/qje/qjaa018.

Billett, M. T., K. Freeman, and J. Gao. 2021. Access to debt and the provision of trade credit. *Available at SSRN 3966713* .

BIS. 2020. Central bank digital currencies: Foundational principles and core features.

———. 2021. Gaining momentum – results of the 2021 BIS survey on central bank digital currencies.

Boehmer, E., C. Jones, X. Zhang, and X. Zhang. 2021. Tracking retail investor activity. *The Journal of Finance* 76:2249–305. doi:https://doi.org/10.1111/jofi.13033.

Bogousslavsky, V., and D. Muravyev. 2021. Who trades at the close? implications for price discovery and liquidity. *Available at SSRN* .

Bond, P., A. Edmans, and I. Goldstein. 2012. The real effects of financial markets. *Annu. Rev. Financ. Econ.* 4:339–60.

Breza, E., and A. Liberman. 2017. Financial contracting and organizational form: Evidence from the regulation of trade credit. *The Journal of Finance* 72:291–324.

Brick, I. E., and W. K. Fung. 1984. Taxes and the theory of trade debt. *The Journal of Finance* 39:1169–76.

Broda, C., and D. E. Weinstein. 2006. Globalization and the gains from variety. *The Quarterly Journal of Economics* 121:541–85.

Brogaard, J., and J. Pan. 2021. Dark pool trading and information acquisition. *The Review of Financial Studies, Forthcoming* .

Brunnermeier, M. K., and D. Niepelt. 2019. On the equivalence of private and public money. *Journal of Monetary Economics* 106:27–41.

Burkart, M., and T. Ellingsen. 2004. In-kind finance: A theory of trade credit. *American Economic Review* 94:569–90.

Buti, S., B. Rindi, and I. M. Werner. 2017. Dark pool trading strategies, market quality and welfare. *Journal of Financial Economics* 124:244–65.

Capponi, A., W.-S. A. Cheng, S. Giglio, and R. Haynes. 2022. The collateral rule: Evidence from the credit default swap market. *Journal of Monetary Economics* 126:58–86.

Cen, L., E. L. Maydew, L. Zhang, and L. Zuo. 2017. Customer–supplier relationships and corporate tax avoidance. *Journal of Financial Economics* 123:377–94.

Chakravarty, S., and K. Li. 2003. An examination of own account trading by dual traders in futures markets. *Journal of Financial Economics* 69:375–97. ISSN 0304-405X. doi:https://doi.org/10.1016/S0304-405X(03)00117-X.

Chava, S., and M. R. Roberts. 2008. How does financing impact investment? The role of debt covenants. *The Journal of Finance* 63:2085–121.

Chen, D., and D. Duffie. 2021. Market fragmentation. *American Economic Review* 111:2247–74. doi:10.1257/aer.20200829.

Chernenko, S., I. Erel, and R. Prilmeier. 2022. Why do firms borrow directly from nonbanks? *The Review of Financial Studies* ISSN 0893-9454. doi:10.1093/rfs/hhac016. Hhac016.

Chiu, J., M. Davoodalhosseini, J. H. Jiang, and Y. Zhu. 2019. Bank market power and central bank digital currency: Theory and quantitative assessment. Working Paper, Bank of Canada.

Chodorow-Reich, G., O. Darmouni, S. Luck, and M. Plosser. 2022. Bank liquidity provision across the firm size distribution. *Journal of Financial Economics* 144:908–32.

Cohen, G. J., J. Dice, M. Friedrichs, K. Gupta, W. Hayes, I. Kitschelt, S. J. Lee, W. B. Marsh, N. Mislang, and M. Shaton. 2021. The US syndicated loan market: Matching data. *Journal of Financial Research* 44:695–723.

Cohen, L., and A. Frazzini. 2008. Economic links and predictable returns. *The Journal of Finance* 63:1977–2011.

Costello, A. M. 2013. Mitigating incentive conflicts in inter-firm relationships: Evidence from long-term supply contracts. *Journal of Accounting and Economics* 56:19–39.

———. 2019. The value of collateral in trade finance. *Journal of Financial Economics* 134:70–90.

———. 2020. Credit market disruptions and liquidity spillover effects in the supply chain. *Journal of Political Economy* 128:3434–68.

Cunat, V. 2007. Trade credit: suppliers as debt collectors and insurance providers. *The Review of Financial Studies* 20:491–527.

Dannhauser, C. D., and J. Pontiff. 2019. Flow. *Working Paper, Boston College* .

Desai, M. A., C. F. Foley, and J. R. Hines Jr. 2016. Trade credit and taxes. *Review of Economics and Statistics* 98:132–9.

Di Maggio, M., M. Egan, and F. Franzoni. 2021. The value of intermediation in the stock market. *Journal of Financial Economics* .

Diamond, D. W. 1984. Financial intermediation and delegated monitoring. *The Review of Economic Studies* 51:393–414.

Diamond, D. W., and P. H. Dybvig. 1983. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy* 91:401–19.

Drechsler, I., A. Savov, and P. Schnabl. 2017. The deposits channel of monetary policy. *The Quarterly Journal of Economics* 132:1819–76.

Duffie, D., and A. Krishnamurthy. 2016. Passthrough efficiency in the fed's new monetary policy setting. Jackson Hole Symposium, Federal Reserve Bank of Kansas.

Ernst, T., J. Sokobin, and C. Spatt. 2021. The value of off-exchange data. *Working Paper* .

Ersahin, N., M. Giannetti, and R. Huang. 2022. Trade credit and the stability of supply chains. CEPR Discussion Paper No. DP17282.

Fabbri, D., and L. F. Klapper. 2008. Market power and the matching of trade credit terms. *World Bank Policy Research Working Paper* .

Feenstra, R. C. 1994. New product varieties and the measurement of international prices. *The American Economic Review* 157–77.

Fernández-Villaverde, J., D. Sanches, L. Schilling, and H. Uhlig. 2021. Central bank digital currency: Central banking for all? *Review of Economic Dynamics* 41:225–42.

Fishman, M. J., and F. A. Longstaff. 1992. Dual trading in futures markets. *The Journal of Finance* 47:643–71.

Frank, M., and V. Maksimovic. 1998. Trade credit, collateral, and adverse selection. *Unpublished manuscript, University of Maryland* .

Freeman, K. 2020. The economics of trade credit: Risk and power. *Available at SSRN 3235838* .

Frost, J., L. Logan, A. Martin, P. E. McCabe, F. M. Natalucci, and J. Remache. 2015. Overnight RRP operations as a monetary policy tool: Some design considerations. Federal Reserve Bank of New York Staff Report, No. 712.

Garleanu, N., and J. Zwiebel. 2009. Design and renegotiation of debt covenants. *The Review of Financial Studies* 22:749–81.

Garratt, R., and M. J. Lee. 2021. Monetizing privacy. Federal Reserve Bank of New York Staff Report, No. 958.

Giannetti, M., M. Burkart, and T. Ellingsen. 2011. What you sell is what you lend? Explaining trade credit contracts. *The Review of Financial Studies* 24:1261–98.

Giannetti, M., and F. Saidi. 2019. Shock propagation and banking structure. *The Review of Financial Studies* 32:2499–540.

Giannetti, M., N. Serrano-Velarde, and E. Tarantino. 2021. Cheap trade credit and competition in downstream markets. *Journal of Political Economy* 129:1744–96. doi:10.1086/713731.

Glode, V., and C. C. Opp. 2021. Private renegotiations and government interventions in debt chains. *Available at SSRN 3667071* .

Gofman, M., and Y. Wu. 2022. Trade credit and profitability in production networks. *Journal of Financial Economics* 143:593–618.

Goldstein, M. A., P. Irvine, E. Kandel, and Z. Wiener. 2009. Brokerage commissions and institutional trading patterns. *The Review of Financial Studies* 22:5175–212.

Greenwald, D. L., J. Krainer, and P. Paul. 2021. The credit line channel. Working Paper.

Gustafson, M. T., I. T. Ivanov, and R. R. Meisenzahl. 2021. Bank monitoring: Evidence from syndicated loans. *Journal of Financial Economics* 139:452–77. ISSN 0304-405X. doi:https://doi.org/10.1016/j.jfineco.2020.08.017.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 153–61.

Hendershott, T., and H. Mendelson. 2000. Crossing networks and dealer markets: Competition and performance. *The Journal of Finance* 55:2071–115.

Holmstrom, B., and J. Tirole. 1997. Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics* 112:663–91.

Holmström, B., and J. Tirole. 1998. Private and public supply of liquidity. *Journal of Political Economy* 106:1–40.

———. 2000. Liquidity and risk management. *Journal of Money, Credit and Banking* 295–319.

Hong, H., and J. Wang. 2000. Trading and returns under periodic market closures. *The Journal of Finance* 55:297–354.

Honka, E., A. Hortaçsu, and M. A. Vitorino. 2017. Advertising, consumer awareness, and choice: Evidence from the U.S. banking industry. *The RAND Journal of Economics* 48:611–46.

Huddart, S., J. S. Hughes, and C. B. Levine. 2001. Public disclosure and dissimulation of insider trades. *Econometrica* 69:665–81.

Keister, T., and D. R. Sanches. 2021. Should central banks issue digital currency? Working Paper, Federal Reserve Bank of Philadelphia.

Kim, S.-J., and H. S. Shin. 2012. Sustaining production chains through financial linkages. *American Economic Review* 102:402–06.

Kiyotaki, N., and J. Moore. 1997. Credit chains. ESE Discussion Paper no.118, Edinburgh School of Econ., Univ. Edinburgh.

Klapper, L. 2006. The role of factoring for financing small and medium enterprises. *Journal of Banking & Finance* 30:3111–30. ISSN 0378-4266.

Klapper, L., L. Laeven, and R. Rajan. 2011. Trade Credit Contracts. *The Review of Financial Studies* 25:838–67. ISSN 0893-9454. doi:10.1093/rfs/hhr122.

Kothari, S., T. L. Johnson, and E. C. So. 2021. Commission savings and execution quality for retail trades. *Available at SSRN* .

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society* 1315–35.

Lagarde, C. 2018. Winds of change: The case for new digital currency. Speech at Singapore Fintech Festival.

Lee, C. M., and B. Radhakrishna. 2000. Inferring investor behavior: Evidence from torq data. *Journal of Financial Markets* 3:83–111. ISSN 1386-4181. doi:https://doi.org/10.1016/S1386-4181(00)00002-1.

Lee, Y. W., and J. D. Stowe. 1993. Product risk, asymmetric information, and trade credit. *Journal of Financial and Quantitative Analysis* 28:285–300.

Lemmon, M., L. X. Liu, M. Q. Mao, and G. Nini. 2014. Securitization and capital structure in nonfinancial firms: An empirical investigation. *The Journal of Finance* 69:1787–825.

Li, Y., and Y. Li. 2021. Payment risk and bank lending. Fisher College of Business Working Paper No. 2021-03-017.

Lines, A. 2022. Do institutional incentives distort asset prices? *Available at SSRN 2873739* .

Liu, E. 2019. Industrial policies in production networks. *The Quarterly Journal of Economics* 134:1883–948.

Long, M. S., I. B. Malitz, and S. A. Ravid. 1993. Trade credit, quality guarantees, and product marketability. *Financial Management* 117–27.

Luck, S., and J. A. Santos. 2022. The valuation of collateral in bank lending. *Available at SSRN 3467316* .

Martin, A., J. McAndrews, and D. Skeie. 2016. Bank lending in times of large bank reserves. *International Journal of Central Banking* 12:193–222.

Mian, S. L., and C. W. J. Smith. 1992. Accounts receivable management policy: theory and evidence. *The Journal of Finance* 47:169–200.

Murfin, J., and K. Njoroge. 2013. The implicit costs of trade credit borrowing by large firms. *Working Paper, Available at SSRN: https://ssrn.com/abstract=2023409* .

———. 2015. The implicit costs of trade credit borrowing by large firms. *The Review of Financial Studies* 28:112–45.

Ng, C. K., J. K. Smith, and R. L. Smith. 1999. Evidence on the determinants of credit terms used in interfirm trade. *The Journal of Finance* 54:1109–29.

Office of the Comptroller of the Currency. 2000. Accounts receivable and inventory financing. Handbook, March https://www.occ.gov/publications-and-resources/publications/comptrollers-handbook/files/accts-rec-inventoryfinancing/pub-ch-accts-rec-inventory-financing.pdf.

Parlour, C. A., U. Rajan, and J. Walden. 2022. Payment system externalities. *Journal of Finance* 77:1019–53.

Petersen, M. A., and R. G. Rajan. 1997. Trade credit: theories and evidence. *The Review of Financial Studies* 10:661–91.

Piazzesi, M., and M. Schneider. 2020. Credit lines, bank deposits or CBDC? Competition & efficiency in modern payment systems. Working Paper, Stanford University.

Röell, A. 1990. Dual-capacity trading and the quality of the market. *Journal of Financial Intermediation* 1:105–24.

Sarkar, A. 1995. Dual trading: Winners, losers, and market impact. *Journal of Financial Intermediation* 4:77–93. ISSN 1042-9573. doi:https://doi.org/10.1006/jfin.1995.1004.

Schott, P. K. 2008. The relative sophistication of Chinese exports. *Economic Policy* 23:6–49.

SEC DERA. 2017. Bats market close: Off-exchange closing volume and price discovery. *SEC memorandum* .

Shanthikumar, D. 2003. Small and large trades around earnings announcements: Does trading behavior explain post-earnings-announcement drift? *Working Paper* .

Smith, J. K. 1987. Trade credit and informational asymmetry. *The Journal of Finance* 42:863–72.

Soderbery, A. 2015. Estimating import supply and demand elasticities: Analysis and implications. *Journal of International Economics* 96:1–17.

Sufi, A. 2009. Bank lines of credit in corporate finance: An empirical analysis. *The Review of Financial Studies* 22:1057–88.

Tunca, T. I., and W. Zhu. 2018. Buyer intermediation in supplier finance. *Management Science* 64:5631–50.

Weintraub, B., and A. N. Resnick. 1982. From the bankruptcy courts: The use of cash collateral in reorganization cases. *Uniform Commercial Code Law Journal* 15:168–.

Wilner, B. S. 2000. The exploitation of relationships in financial distress: The case of trade credit. *The Journal of Finance* 55:153–78.

WSJ. 2018. Goldman cashes in on passive-investing boom with big 4 p.m. trade. *The Wall Street Journal (August 26, 2018)* .

Yang, L., and H. Zhu. 2020. Back-running: Seeking and hiding fundamental information in order flows. *The Review of Financial Studies* 33:1484–533.

Ye, M. 2010. Non-execution and market share of crossing networks. *Available at SSRN 1719016* .

Zhu, H. 2014. Do dark pools harm price discovery? *The Review of Financial Studies* 27:747–89.