From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2021

Essays on Liquidity, Information, and Macroecono

Abadi, Joseph Allen

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp011g05ff68c

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON LIQUIDITY, INFORMATION, AND MACROECONOMICS

JOSEPH ABADI

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF ECONOMICS

ADVISOR: MARKUS BRUNNERMEIER

MAY 2021

© Copyright by Joseph Abadi, 2021.

All rights reserved.

Abstract:

Chapter 1 proposes a theory of credit cycles driven by the private production of opaque, liquid assets (e.g., ABS or CLOs). Opacity enhances assets' liquidity, permitting greater issuance volumes, but prevents investors from determining whether the underlying projects are of low quality. Strong macroeconomic fundamentals give rise to credit booms characterized by opaque asset origination and pervasive credit misallocation. As bad projects build up in the economy, investors begin to question the value of opaque assets and eventually refuse to finance them altogether, precipitating a collapse in investment. The bust has a cleansing effect: opaque origination is abandoned, and investors no longer finance projects whose quality they cannot evaluate. I show that a policymaker would limit opaque intermediation during booms in order to moderate the subsequent bust.

Chapter 2 presents a model in which agents with heterogeneous beliefs borrow by using a physical asset and the liabilities of other agents as collateral. In equilibrium, a chain of lending emerges: each agent lends to the next-most optimistic agent and borrows from the next-most pessimistic agent. Intermediation allows optimists to lever up while pessimists invest in safe assets. In extensions of the benchmark model, I examine the implications of this arrangement for financial stability and relate the model's predictions to stylized facts.

Chapter 3, which was co-authored with Markus Brunnermeier, develops a model of digital record-keeping. Traditional centralized record-keeping systems establish a consensus based on trust in the record-keeper. Trust arises from the ability to incentivize honest reporting. Rents extracted by the record-keeper create an internal source of trust, allowing the system to be self-sufficient. Blockchains decentralize record-keeping, dispensing with the need for trust in a single entity. Some build a consensus based on externally verifiable resource costs (proof-of-work), whereas others do not (proof-of-stake). We prove a Blockchain Trilemma: it is impossible for any

digital record-keeping system to simultaneously be (i) self-sufficient, (ii) rent-free, and (iii) resource-efficient. Record-keeping systems without rents or resource costs must ultimately rely on some external source of trust.

Acknowledgments:

I am deeply indebted to my advisor, Markus Brunnermeier, for his constant support and encouragement throughout my time at Princeton. My collaboration and conversations about research with Markus were the most formative part of my graduate school experience and have been invaluable in my development as an economist. I would also like to thank Stephen Morris and Mark Aguiar, the other members of my dissertation committee, for many helpful conversations over the past years as well as for their advice and backing while I was on the job market. Special thanks go to Jonathan Payne and Moritz Lenel, too. They spent an enormous amount of time talking to me about my research, preparing me for the job market, and guiding me through the process with the wisdom they picked up during their time as graduate students.

I am also grateful for several people in my personal life. My family was there for me through all the highs and lows of graduate school. My parents, Barbara and Carlos, my brother Elliott, and Marina made sure to be there for me even when I was away during the most difficult periods of the job market. I could not have completed this dissertation without my partner Emily, either. She was with me my entire last year of graduate school and always knew how to support me (and surprise me), even during that unusual time.

Contents

Abstract . . . iii
Acknowledgments . . . v

1 Opaque Intermediation and Credit Cycles **1**
1.1 Introduction . . . 1
1.2 Static Model: Environment . . . 6
1.2.1 Discussion of setting and assumptions . . . 11
1.3 Static Model: Decision problems and equilibrium . . . 12
1.3.1 Afternoon: Asset market . . . 12
1.3.2 Morning: Intermediary disclosure policy . . . 16
1.3.3 Equilibrium . . . 17
1.4 Dynamics . . . 24
1.4.1 Model setup . . . 24
1.4.2 Dynamic equilibrium . . . 26
1.4.3 Steady states and endogenous cycles . . . 29
1.5 Optimal Opacity . . . 34
1.5.1 The planner's problem . . . 35
1.5.2 The externality and optimal policy . . . 36
1.5.3 Implementation . . . 38
1.6 Conclusion . . . 39
References . . . 40

2 A Model of Collateralized Lending Chains **55**
2.1 Introduction . . . 55
2.1.1 Literature Review . . . 58
2.2 Model . . . 60
2.2.1 Environment . . . 60
2.2.2 Optimization Problem . . . 62
2.2.3 Equilibrium . . . 63

2.3 Properties of Equilibrium . . . 63
2.3.1 Social Planner's Problem . . . 64
2.3.2 Tranched Payoffs in Equilibrium . . . 66
2.3.3 Equilibrium Prices, Interest Rates, and Leverage . . . 69
2.3.4 Equilibrium Characterization . . . 71
2.4 What Does Intermediation Do? . . . 72
2.5 Applications of the model . . . 75
2.5.1 Contagion in Margins . . . 75
2.5.2 Leverage and Safe Asset Production Cycles . . . 78
2.5.3 The Impact of Financial Innovation . . . 82
2.6 Discussion . . . 85
2.6.1 Alternative Asset Markets . . . 85
2.6.2 Beliefs as a Motive for Trade . . . 86
2.7 Conclusion . . . 87
References . . . 88

3 Blockchain Economics 102
3.1 Introduction . . . 102
3.2 Digital Ledger Technology . . . 108
3.2.1 The digital record-keeping problem . . . 108
3.2.2 What is a blockchain? . . . 110
3.2.3 Consensus algorithms . . . 111
3.2.4 Blockchain security . . . 112
3.3 Examples of consensus algorithms . . . 113
3.3.1 Private blockchain . . . 114
3.3.2 Proof-of-Stake blockchain . . . 115
3.3.3 Proof-of-Work blockchain . . . 117
3.4 Model . . . 118
3.4.1 Communication and the consensus algorithm . . . 120
3.4.2 Payoff structure and token-based mechanisms . . . 122
3.4.3 Social interactions . . . 125
3.4.4 Equilibrium concept . . . 125
3.5 The Blockchain Trilemma . . . 127
3.5.1 Double-spending . . . 127
3.5.2 Proof of the Trilemma . . . 130
3.6 Discussion . . . 133

3.6.1 Extension to arbitrary set of agents 133
3.6.2 Extension to an infinite-horizon environment 134
3.6.3 Enforcement 134
3.7 Conclusion 136
References 137

Chapter 1

Opaque Intermediation and Credit Cycles

1.1 Introduction

Credit booms are often fueled by the creation of liquid securities backed by assets that investors know little about. However, these booms sometimes run out of steam and end in busts: investors begin to question the quality of the assets underlying the new securities, liquidity dries up, and investment collapses.[1] The run-up to the Great Recession provides a striking example: intermediaries financed a surge in mortgage lending by issuing an array of complex securities, such as ABS and CDOs. As perceptions of the underlying mortgages' quality soured, markets for those securities froze, causing a crash in new lending and a recession. Similarly, recent years have witnessed the reemergence of opaque assets, such as CLOs, as well as a loosening of lending standards, but the onset of the Covid-19 crisis disrupted the functioning of markets for even the highest-rated securities.[2] While the literature has stressed the role of opacity in enhancing assets' liquidity,[3] less attention has been directed towards the macroeconomic side effects of the private production of opaque, liquid

[1]Beyond the examples of the Great Recession and the Covid-19 crisis, such episodes go back to at least the boom in farm-adjacent mortgages of the 1850s in the United States (Riddiough and Thompson, 2012). Other examples include the rise of commercial real estate loan securitzation in the period preceding the Great Depression (White, 2009) and the boom-bust cycles of loan syndication in emerging markets in the 1980s and 1990s (Kaminsky, 2008).

[2]Foley-Fisher, Gorton, and Verani (2020) argue that safe CLO tranches command a liquidity premium due to their opacity and document the increase in spreads even for AAA tranches during the Covid crisis.

[3]Dang, Gorton, and Hölmstrom (2015), among others, highlight this role of "symmetric ignorance."

assets. Under what circumstances will opaque asset production emerge, and how does it shape the dynamics of credit and financial market liquidity? Should policy seek to increase the transparency of asset origination and curb private liquidity provision, or should a public provider of liquidity instead aim to supplement it?

To address these questions, I develop a macroeconomic model of credit booms and busts driven by the private production of opaque assets. I begin in a static setting to highlight a novel tradeoff between *asset liquidity* and *credit misallocation* governing the transparency of claims produced by the financial system and demonstrate that strong macroeconomic fundamentals incentivize the production of opaque assets. I then embed the key mechanism of the static setting into a dynamic model in order to understand its positive implications for credit cycles as well as to derive normative prescriptions for private and public liquidity provision. I show that the misallocation caused by opaque asset production during transitory credit booms leads to slumps featuring depressed credit and persistent illiquidity, and I characterize conditions under which such cycles are fully endogenous (in the sense that they arise even in the absence of exogenous shocks). I further prove that the production of opaque, liquid assets is always excessive in equilibrium and outline realistic policy tools that can be used to implement the social optimum.

In the static model, intermediaries lend to firms of heterogeneous quality (good or bad) and sell assets backed by firms' projects to investors with varying ability to evaluate their quality (skilled or unskilled). To capture the idea that asset originators differ in the quantity of information they disclose about the underlying projects, I assume there are two types of intermediaries that may enter to channel funds from investors to firms. *Transparent* intermediaries (e.g., IPO underwriters) disclose key details about firms' projects to investors, whereas *opaque* intermediaries (e.g., shadow banks) keep those details secret.[4] Expertise is scarce in this economy– only skilled investors are sophisticated enough to interpret any information revealed by transparent intermediaries, but they sometimes lack the funds to finance good firms' projects.

The central mechanism that gives rise to a role for opaque intermediation is that information permits two opposing types of selection. On the one hand, information allows for *virtuous selection*: a skilled investor who knows more about a firm's quality is able to direct investment towards good firms' projects and avoid misallocation towards bad ones. On the other hand, information creates *adverse selection*: to the

[4]Arora et al. (2009) argue that, in fact, asset-backed securities are typically so complex that it is computationally infeasible to compute their fundamental value.

extent that investors are not equally able to interpret signals of an asset's quality, information will put unskilled investors at a disadvantage when making investments.

The tension between virtuous and adverse selection yields a tradeoff between financial assets' liquidity and the efficient allocation of credit. Transparent projects are illiquid because unskilled investors will compete with skilled investors over claims on good projects, but they will be alone in demanding claims on bad ones. Claims against a transparent project must then be sold at a discount to attract investment from unskilled investors. If the illiquidity discount is large enough, it may prevent the intermediary from being able to profitably finance the project altogether. The cost of adverse selection is then *under*-investment in good projects. Opaque intermediation renders investors' information symmetric, which resolves the adverse selection problem and creates liquid assets that can be freely issued to any investor. Opacity comes at a cost, however: it deprives skilled investors of the benefits of virtuous selection. Thus, opaque projects are financed regardless of their quality, causing misallocation, i.e., *over*-investment in bad projects. The key result of the static model is therefore that opaque liquidity creation is attractive when investment in firms' projects is profitable on average, whereas asset origination tends to be transparent when the expected profitability of investment is low.

After establishing this result, I proceed to the dynamic model. The key state variables that govern the model's dynamics are the exogenous productivity of firms' projects and the average quality of the pool of firms in the economy, which evolves endogenously. In particular, I assume that credit is essential to firms' survival: the average quality of the pool deteriorates when bad firms are financed and improves when they are discovered and fail to raise additional funds.

I show that the economy can exhibit both amplification of fundamental shocks and fully endogenous credit cycles. The economy transits through two regimes in the dynamic equilibrium: an "opaque boom" regime and a "transparent bust" regime. Opaque booms occur when macroeconomic fundamentals are strong; that is, when either productivity or the average quality of firms in the economy is high. It is costly to miss out on good investment opportunities, so opaque intermediaries enter to issue liquid assets against firms' projects. Financial markets are highly liquid in the boom regime. Good firms' projects are always financed, and output expands. Misallocation is widespread, though, and bad firms are able to attract financing as well. This misallocation causes a gradual build-up of bad firms in the economy. Nevertheless, as long as it remains profitable for unskilled investors to finance opaque projects, the boom continues.

The economy enters the transparent bust regime and experiences a sharp reversal when the fundamental profitability of firms' projects falls below a critical threshold, which can be triggered by an exogenous reversion in productivity or the endogenous deterioration of firms' quality during the boom. In particular, this occurs when the benefits of virtuous selection exceed the costs of adverse selection. Investors become unwilling to finance new investment in opaque projects, and private liquidity creation is abruptly abandoned. Financial markets become fragmented, and credit contracts: assets are originated by transparent intermediaries, which are able to finance projects only when skilled capital is available. The disruption in financial markets pushes the economy into a recession, which has a cleansing effect due to the greater prevalence of transparency. Bad firms are identified and fail to raise funds, forcing them to exit. As misallocation is undone, the economy can eventually re-enter the opaque boom regime, allowing the credit cycle to restart. Thus, financial cycles in this economy feature recurring episodes of high liquidity, loose lending standards, and opaque liquidity creation followed by periods of low liquidity, tight lending standards, and a return to more traditional asset origination.

The intrinsic link between information, liquidity, and misallocation in the model has distinctive policy implications. I study the problem of a social planner who possesses the same information as unskilled investors. The environment features a *dynamic* information externality: investors do not internalize that by purchasing opaque assets, they continue bad firms' projects and allow them to remain in the pool to borrow from others in the future. By contrast, the benefits of liquidity creation are fully internalized by intermediaries. The constrained optimal degree of liquidity provision is therefore lower than what competitive markets deliver. I show that private liquidity provision can be reined in using three equivalent tools: transparency regulation, which alters the information structure available to investors by restricting the quantity of opaque projects in the economy, macroprudential policy (taxes on opaque asset origination), which reduce the incentives to produce opaque assets, or monetary policy, which increases the rate of return investors can earn on safe assets and reduces demand for risky ones. Nevertheless, it is not always optimal to fully eliminate opacity: despite the fact that opacity allows bad projects to be financed, it can be socially beneficial because it also permits greater investment in good projects.

Related literature. The central mechanism in my paper highlights that intermediaries may create opacity to mitigate adverse selection and enhance assets' liquidity. Similarly to my paper, Dang, Gorton, Hölmstrom, and Ordoñez (2017) argue that

opacity is an essential function of banks. In their model, though, the key tradeoff is between a lower quantity of stable liquidity (provided by banks) and a greater quantity of risky liquidity (provided by capital markets). Hence, in that model, banking eliminates risk rather than adverse selection. Closer to my tradeoff, Pagano and Volpin (2012) show that opaque securitization can alleviate adverse selection in the primary market for debt securities, but they argue that the other side of the tradeoff is decreased liquidity in the secondary market if experts choose to acquire information on their own later on. My paper is the first to incorporate a role for opaque liquidity creation in a macroeconomic model. I obtain a tradeoff that is new to the literature: when liquidity creation requires opacity, it necessarily goes hand-in-hand with the misallocation of credit.

My paper also relates to others that model credit booms and busts as changes in the prevailing informational regime in the economy. Gorton and Ordoñez (2014) show that small fundamental shocks can cause abrupt shifts to a regime in which lenders inspect collateral, triggering adverse selection between borrowers and lenders and leading to a financial crisis. In a related model, Gorton and Ordoñez (2020) demonstrate that their mechanism can generate fully endogenous credit cycles. In those models, however, information has no social value– all investment projects have positive present value, while information concerns the quality of an exogenous stock of collateral that serves only to facilitate borrowing. As a result, there is a sense in which it would be optimal in those models to ban information acquisition (i.e., impose total opacity). By contrast, in my model, information concerns real investment opportunities rather than an exogenous supply of collateral. Thus, opacity generates misallocation, which is a force absent from theirs. This distinction has welfare consequences: transparency has a socially beneficial cleansing effect in my model, thereby allowing me to analyze the tradeoff entailed by private liquidity creation. Farboodi and Kondor (2020) study a model of endogenous credit cycles in which investors choose whether to be bold or cautious when evaluating entrepreneurs (rather than information being concealed by intermediaries for liquidity creation motives). While the externality in their model is similar to mine in the sense that investors collect too little information about borrowers, their environment is quite different: their focus is investors' information acquisition, whereas mine is the production of liquid assets as a driver of cycles. There are other theories of information acquisition over the credit cycle that are relevant to my work. Dell'Ariccia and Marquez (2009) provide an early model in which lax screening is optimal during economic booms but sub-optimal in busts. Asriyan, Laeven, and Valencia (2019) build a model of the

amplification of busts in which information deteriorates during booms as financial collateral substitutes for costly screening.

There is a literature on securitization that also relates to my paper. Vanasco (2017) builds a model in which screening by an intermediary reduces the liquidity of claims it sells in secondary markets, providing an alternative channel through which information acquisition and more efficient credit allocation affect an asset's liquidity. Chemla and Hennessy (2014) provide a role for opacity in the creation of securities when investors are risk-averse: opacity prevents uninformed investors from facing adverse selection in asset markets, allowing them to more efficiently insure against shocks.

Finally, in macroeconomics, my paper builds off of a literature on information asymmetries and the cyclicality of liquidity. Eisfeldt (2004) shows how liquidity can respond procyclically to productivity shocks and Kurlat (2013) examines the transmission of exogenous shocks to macroeconomic aggregates in a model with asymmetric information, showing that information frictions tend to amplify shocks to the economy. Bigio (2015) calibrates a related model and shows that asymmetric information can provide a reasonable quantitative explanation of the Great Recession. While my model predicts that liquidity will be procyclical, as in these papers, the mechanism is different. In this literature, the information structure is exogenous and information does not play an allocative role. In my paper, the key driver of procyclical liquidity is the endogenously opaque nature of assets created by the financial sector during booms. In a sense, then, my model endogenizes the emergence of information asymmetries. Caramp (2017) also models an economy in which the demand for liquidity during booms leads to misallocation, but his focus is on moral hazard in securitization rather than the intrinsic tradeoff between virtuous and adverse selection that features in my model.

Outline: The paper is structured as follows. Section 1.2 presents the static model. Section 1.3 characterizes the static equilibrium and outlines the conditions under which opaque intermediation emerges. Section 1.4 introduces dynamics. Section 1.5 studies optimal policies in the dynamic model. Section 1.6 concludes. All proofs can be found in the Appendix.

1.2 Static Model: Environment

I begin with a static model to highlight the role of opacity and understand the circumstances under which intermediaries finance investment through the issuance

of opaque, liquid claims. I will derive implications for liquidity and misallocation and precisely characterize the tradeoff between virtuous selection and adverse selection faced by investors when choosing to finance transparent or opaque projects. Then, I will show that opacity will tend to dominate transparency when the average firm's project is profitable, which is the key result that will drive the dynamic model.

There are three periods, $\tau = 0, 1, 2$ (morning, afternoon, and evening). There is a single good. The economy consists of a continuum of "islands" $n \in [0, 1]$, as in Lucas (1973). On each island, there is a continuum of investors and firms. Investors are indexed by i, and firms are indexed by j. The islands inhabited by an investor or firm are denoted by $n(i)$ and $n(j)$, respectively. An island should be thought of, in this model, as comprising a collection of investors that have the expertise to evaluate a collection of firms. All agents are risk-neutral and do not discount. There is also a unit mass of intermediaries on each island that match randomly with firms. Intermediaries on an island are owned by investors on the same island and therefore maximize their profits with respect to the discount factor used by those investors.

Firms: Each firm runs a *project* in which it can invest x units of goods in the afternoon in order to produce $\varphi(x) = \min\{x, 1\}z$ goods if the project succeeds.[5] I will refer to parameter z as the productivity of firms' projects. In the dynamic model, I will allow z to vary exogenously over time.

Each firm may be good ($\Theta_j = G$) or bad ($\Theta_j = B$). On each island, a fraction $1-\delta$ of firms are good, and the remaining fraction δ are bad. The probability θ_j that a firm's project succeeds is type-dependent:

$$\theta_j = \begin{cases} \theta_G & \Theta_j = G \\ \theta_B & \Theta_j = B \end{cases} \tag{1.1}$$

I assume that bad projects have negative net present value, whereas good projects have positive present value: $\theta_B z < 1 < \theta_G z$. I will refer to θ_j as the firm's quality or the quality of its project interchangeably. The ex ante probability that a project succeeds will be denoted

$$\overline{\theta} = (1 - \delta)\theta_G + \delta\theta_B. \tag{1.2}$$

Intermediaries: The role of intermediaries will be to issue securities to investors backed by firms' projects and to mediate the information about the underlying projects that is available to investors. On each island n, intermediaries are matched with a single firm. Intermediaries have access to a *disclosure technology* they can

[5] That is, the maximum size of a project is one unit of investment.

use to design the information they will reveal about firms to investors. I describe this technology in greater detail below. Firms are incapable of otherwise credibly communicating their types to other agents in the economy.

After matching with a firm, intermediaries make take-it-or-leave-it offers to firms specifying the amount they are to repay in the evening (as a function of their output).[6] Firms then pledge their entire output to the intermediary, so I will sometimes refer to the firm's project as the intermediary's project. In the afternoon, intermediaries will have to raise additional funds to finance interim investment in firms' projects. They do so by selling claims on projects in an asset market described in Section 1.3.1.

Investors: Investors receive an endowment of goods e in the morning and lend it to intermediaries on the same island. In the afternoon, investors will face preference shocks that are heterogeneous across islands. Their preferences are

$$U_n = \mathbb{E}[\tilde{\lambda}_n c_1 + c_2],$$

where $\tilde{\lambda}_n \in \{1, \lambda\}$ (with $\lambda > 1$) reflects an *impatience shock* that is iid across islands, with $\Pr(\tilde{\lambda} = \lambda) = \alpha$. That is, on a fraction α of islands, investors are impatient and discount final consumption at a high rate $\frac{1}{\lambda}$, whereas on the remaining islands, investors are patient and do not discount final consumption. Islands that receive the impatience shock will be called *impatient islands*, and those that do not will be *patient islands*. After receiving these shocks, investors will purchase additional claims on firms' projects from intermediaries in the asset market.

Information: The information available to investors about a firm j in the afternoon, when it needs to raise funds for investment, will consist of a market price and a signal $s_j \in [0, 1]$ revealed by the intermediary matched with the firm.[7] The structure of the signal will be determined by the *disclosure policy* of the intermediary that lends to firm j. In the morning, intermediaries do not know the type of the firms to which they are matched, but in the afternoon, they receive information about the firm in a *file* and learn its type.

Intermediaries are free to choose a disclosure policy in the morning, which consists of a CDF $F(\cdot|\Theta)$ for $\Theta \in \{G, B\}$ (defined over the unit interval $[0, 1]$). The signal s disclosed by the intermediary will be distributed according to F (conditional on

[6] In principle, there are other ways to specify the split of surplus between firms and intermediaries. In any case, firms and intermediaries would act to maximize their joint surplus, however, so the surplus split will not affect aggregate outcomes.

[7] That is, investors do not see the island $n(j)$ on which an asset was originated. This assumption allows me to embed an adverse selection problem in a model with Walrasian markets rather than a more complicated setting with asymmetric information, as in Kurlat (2016).

the firm's true type). For simplicity, I assume that the CDF also must have finite support. Otherwise, I allow for fully flexible information design– the intermediary may perfectly reveal the firm's type (transparency), reveal nothing about the firm (opacity), or reveal a noisy signal. I denote the class of feasible CDFs by $\mathcal{F}$.

I make the key assumption that only a subset of investors have the expertise to infer a firm j's quality from its file: when (possibly noisy) information in a firm's file is disclosed, it is *asymmetrically revealed* across islands. In particular, only investors on the same island, $n(j)$, have the expertise to interpret the file's contents. Investors on other islands, by contrast, will gain no information by observing information in the file. Hence, investors on each island n are *skilled* at evaluating firms on the same island, whereas others are *unskilled*.

Formally, I assume that investor i receives a signal s_{ij} of firm j's quality of the form

$$s_{ij} = \begin{cases} s_j & n(i) = n(j) \\ N & n(i) \neq n(j) \end{cases} \tag{1.3}$$

where s_j is the signal disclosed under the intermediary's policy and N denotes an uninformative signal.

In addition to potentially having private information about the quality of firms on the same island, all agents on an island n are also privately informed of their impatience shocks. Other investors on islands $n' \neq n$ will therefore be unable to infer whether a firm on n lacks access to capital from its own island (skilled capital) because investors on n are impatient or because they have identified the firm as bad. This inference problem will give rise to adverse selection in equilibrium.

Everyone can observe the disclosure policy used by an intermediary. That is, even though some agents may not be able to interpret the information disclosed by an intermediary, all agents can observe whether it did in fact disclose information. Opacity (full or partial) will then serve as a public signal that no investor knows the quality of a firm's project.

Frictions: Agents face limited commitment: all financial contracts across islands must be collateralized by claims on firms' projects. In turn, this assumption implies that unsecured borrowing across islands will be infeasible. Heterogeneity in islands' preference shocks will therefore generate a need to trade claims on projects across islands: goods will need to flow from patient islands to impatient ones. Information asymmetries across islands will have the potential to hinder trade, however.

Further, I assume that only a fraction $\xi \in (0,1)$ of a project's output can be pledged across islands.[8] Under this assumption, illiquidity will be costly: if claims on a project sell at a low enough price, investment in that project may not proceed even when it has positive present value in expectation. Nevertheless, I assume the average project is of high enough quality that

$$\xi \cdot \overline{\theta} z \geq 1 \tag{1.4}$$

That is, the pledgeable output of the average project is high enough to sustain the maximum scale of investment.

Financial assets: There is only one type of financial asset in the economy: claims on intermediaries that pay z if their projects succeed, which trade in the afternoon and may be sold by intermediaries across islands. These assets are backed by intermediaries' claims on firms, which have the same structure. Given that projects pay either z or 0, the assumption that assets take this form is without loss of generality. In what follows, I will refer to an asset issued by an intermediary that uses disclosure policy F as an asset of type F.

Timing: In the morning, intermediaries choose their disclosure policies. In the afternoon, intermediaries match with firms, disclose information and attempt to raise additional funds by selling claims to investors under (potentially) asymmetric information. Investment in projects takes place. In the evening, output is realized and investors are repaid. The timeline is illustrated in Figure 1.1.

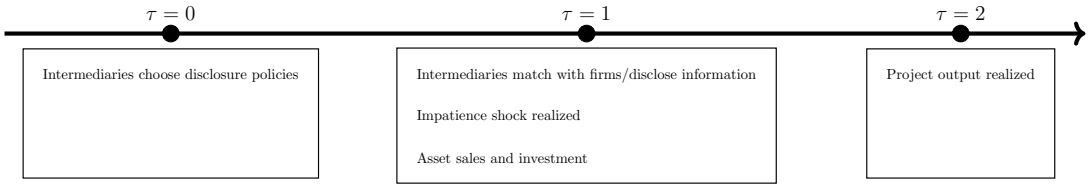


Figure 1.1: Timeline of the static model.

I solve the static model by backwards induction. I begin with agents' problems in the afternoon and work back to the morning.

[8] This assumption can be micro-founded, for example, by assuming that agents on an island possess specific, inalienable human capital required to extract a fraction $1 - \xi$ of a project's output.

1.2.1 Discussion of setting and assumptions

I introduce a model with intermediaries that issue securities to investors and differ in the information they disclose about the physical projects backing those securities. Furthermore, for each project, there are some investors who have the expertise to evaluate its quality (those on the same "island") and others who do not. I interpret intermediaries as entities that issue different types of securities in the primary market. For instance, the issuance of a "transparent" asset could correspond to an IPO issuance, in which key details of the issuer's balance sheet are publicly disclosed. In this case, the intermediary would be the IPO underwriter. I interpret "opaque" assets as encompassing complex securitized products such as an ABS or CLO, for which less information is available to investors at the time of issuance.[9] The corresponding opaque intermediary would be the securitizer issuing the security. Investors with the expertise to evaluate the underlying project map to highly sophisticated financial institutions with limited ability to quickly raise funds to finance investment opportunities, such as hedge funds (specializing in stocks or securitized products, respectively).

The key feature of this environment is that islands are heterogeneous in both their preference shocks and in their information. This combination of assumptions captures that while there are investors who have the specific expertise required to evaluate certain assets, they may sometimes have more profitable investment opportunities elsewhere. In the context of securitized products, this could, for example, reflect that hedge funds specialized in evaluating the prospects of real estate management companies in a particular region may not be able to raise sufficient capital to fully meet mortgage demand in that region. Hence, it may be necessary for less expert investors to meet that demand. Importantly, the assumption that preference shocks are private information of investors on an island implies that less expert investors do not know which assets are passed up by skilled investors because they are bad and which are passed up because skilled investors did not have sufficient funds to buy them. This inference problem will be at the heart of the model's main mechanism, since it causes unskilled investors to face adverse selection. Even in markets with highly efficient price discovery, this type of adverse selection can reduce the price at which a security can be issued.[10]

[9] See Arora et al. (2009), who argue that valuing such securities is effectively computationally infeasible.

[10] For example, see the seminal work of Rock (1986), which explains the anomalous underpricing of IPOs as a symptom of adverse selection.

There are other assumptions in the model that are convenient for the exposition but inessential. The most salient of these is that investors cannot trade the initial claims they hold on intermediaries– if these claims could be traded across islands, it would be possible for skilled investors to purchase some assets even when their endowments are zero. However, the main mechanism would still operate under these circumstances provided that the these claims did not provide skilled investors with sufficient liquidity to finance all good projects on an island.

1.3 Static Model: Decision problems and equilibrium

1.3.1 Afternoon: Asset market

At $\tau = 1$, firms' types are realized and intermediaries use their disclosure technologies. Intermediaries attempt to issue claims to investors in order to finance investment. Intermediaries will repay investors in the evening only if their projects succeed. Investors do not initially know the probability θ_j that an intermediary's project will succeed, but they draw inferences about this probability from the information disclosed by the intermediary about its project (if any) as well as the prices at which assets trade. They purchase claims issued by intermediaries with their endowments and profits π_n rebated to them by intermediaries (which depend on the island n on which they reside).

There are individual markets for claims on each firm j's project with prices p_j. Intermediaries sell claims in these markets, and investors buy. In each market, an investor i sees the firm's index j, the price p_j, the disclosure policy F_j used by the associated intermediary, and her signal s_{ij}, but she does not see the island $n(j)$ on which the firm resides. This assumption captures the idea that less expert investors cannot tell whether demand for an asset is low because the asset is of low quality or because skilled investors lack the funds to purchase it.[11] Investors are price-takers, but as in Gârleanu, Panageas, and Yu (2019), intermediaries are monopolists in the market for claims on their projects, so they take into account the impact of their asset sales on the price at which they sell.

[11]Importantly, it does not matter that investors *on the same island* cannot identify the island on which the firm resides– they know their own liquidity shock, so they can disentangle why a particular transparent asset they have the skill to evaluate is traded at a low price.

Investor's problem

Investors solve a portfolio choice problem. They form expectations $\mathbb{E}[\theta_j|s_{ij}, F_j, p_j]$ of the probability that claims on firms' projects pay off z in the evening based on their signals and the market price. Their signals are informative for projects that they have the expertise to evaluate (those on their own island), but uninformative for all other assets. They choose how much to consume, c, the quantity of funds to store until $\tau = 2$, b, and the quantity of claims to purchase on each project j, a_{Bj}, subject to a budget constraint. Consistent with the assumption of limited commitment, they may not short assets. The problem of an investor on island n is

$$\max_{c,b,a_{Bj}} \tilde{\lambda}_n c + b + \int_j \mathbb{E}[\theta_j z|s_{ij}, F_j, p_j] a_{Bj} dj \tag{1.5}$$
$$\text{s.t. } c + b + \int_j p_j a_{Bj} dj \leq e + \pi_n,\ b \geq 0,\ a_{Bj} \geq 0\ \forall\ j.$$

As usual, investors purchase the assets that deliver the highest returns in expectation, given their signals and publicly available information. The solution to investors' problem will determine the demand schedule faced by intermediaries.

Lemma 1. *Investors are indifferent between all assets j that maximize their returns on wealth,*

$$R_i^* = \max_j \frac{\mathbb{E}[\theta_j z|s_{ij}, F_j, p_j]}{p_j}.$$

If $R_i^* > \tilde{\lambda}$, they invest their entire endowment among such assets. If $R_i^* = 1$, they are indifferent between consuming and investing in those assets. Otherwise, they consume their endowment.

I make the following assumption about investors' impatience, which ensures they will consume all of their funds when they are impatient.

Assumption 1. *The marginal value of consumption for impatient investors satisfies $\lambda > \theta_G z$.*

Under this assumption, the marginal returns to immediate consumption, λ, exceed the maximum return on investment in any project, $\theta_G z$.

Intermediary's problem

Intermediaries choose asset sales a_S and investment x. They take into account the impact of their asset sales on the price at which they sell, which is summarized by

a price schedule $p_j(a_S)$ that the intermediary takes as given. In general, demand for an intermediary's assets will depend on the information it discloses (determined by its disclosure policy F_j), the type of its project θ_j, and the impatience shock faced by skilled investors on its island $\tilde{\lambda}_j$.[12] I will show that in equilibrium, the form of this price schedule will be unimportant, but the fact that intermediaries internalize their price impact will give rise to pooling in equilibrium. In particular, opaque assets will always sell at a single price regardless of their quality, as will transparent assets for which skilled capital is unavailable (either because the investors on the corresponding island are impatient or because they have identified the underlying project as bad). This will be the key feature of the model that causes investors who do not have the expertise to evaluate transparent assets' quality to face adverse selection.

Intermediaries cannot invest more funds than they raise through asset sales, so intermediary j's investment satisfies $x \leq \min\{1, p_j(a_S)a_S\}$ (since the project's investment capacity is equal to one unit of goods per unit of capital). Furthermore, each intermediary faces limited commitment, and may sell no more than a fraction ξ of the firm's output. Hence, intermediaries effectively face a collateral constraint, and attempted asset sales per unit of capital must satisfy $a_S \leq \xi x$.

The problem of an intermediary j is then

$$v_j = \max_{x, a_S} \ \tilde{\lambda}(p_j(a_S)a_S - x) + \theta_j z(x - a_S), \text{ s.t. } 0 \leq a_S \leq \xi x \qquad (1.6)$$
$$x \leq \min\{1, p_j(a_S)a_S\}.$$

Here v_j denotes the value achieved by intermediary j. It consists of an intermediary's expected profits from investment. This object will be important in transparent or opaque intermediaries' entry decisions, since only those intermediaries that can attain the highest value (in expectation) will be able to enter and successfully raise funds from investors in the morning.

The solution to the intermediary's problem is simple.

Lemma 2. *Intermediaries sell the quantity a_S of assets that maximizes $(p_j(a_S)-1)a_S$, independently of the type of the underlying project. They invest whenever there exists $a_S > 0$ such that $p_j(a_S) \geq \frac{1}{\xi}$.*

The intermediary is always willing to sell assets if it is able to finance its investment by doing so. In equilibrium, demand for an intermediary's assets will be elastic at

[12]Note that this implies the intermediary has all the information required to back out the price schedule.

a given price p_j. The collateral constraint faced by the intermediary implies that it can raise ξp_j per unit invested, so it will be possible for the intermediary to invest whenever $\xi p_j \geq 1$ (the cost of investment).

The reason that the intermediary's decision depends only on the market price of its assets (rather than the type of the underlying project) is that the intermediary does not have funds of its own in the afternoon– the entire endowment of goods is owned by investors. Intermediaries' role is therefore mechanical. Essentially, investment decisions are made by the marginal investor who prices the asset sold by an intermediary. Note that the collateral constraint implies that in some cases, an investment may not go forward even when investors think it is socially efficient in expectation: the marginal investor believes the project is profitable if $p_j \geq 1$, but the intermediary can invest only if $p_j \geq \frac{1}{\xi}$. As in Hölmstrom and Tirole (1998), this implies that there is a role for *liquidity creation*. Opacity will allow intermediaries to resolve their liquidity needs by mitigating the adverse selection problem that arises in equilibrium when information is disclosed, raising the price at which their liabilities sell and allowing investment to proceed more often.

Market clearing and asset market equilibrium

I now characterize the equilibrium in the asset market, which is effectively a subgame of the full model. An equilibrium in asset markets is standard. It consists of solutions to investors' and intermediaries' problems (Problems 1.5 and 1.7), price schedules for each asset $p_j(a_S)$, and expectations of projects' quality for both investors and intermediaries such that agents optimize, markets clear, and expectations are consistent with Bayes' rule whenever possible.

I look for symmetric asset market equilibria in which asset prices depend only on the underlying project's type θ_j, the intermediary's disclosure policy F_j, and the impatience shock $\tilde{\lambda}_{n(j)}$ of investors on the corresponding island. The following proposition characterizes the unique equilibrium satisfying these properties.

Proposition 1. *Let $\hat{\theta}_j$ denote skilled investors' posterior belief of a firm j's quality, and let $\hat{F}_j$ denote the prior distribution of $\hat{\theta}_j$ given the corresponding intermediary's disclosure policy. Then there exists a symmetric equilibrium such that for any $\hat{F}$, there is $\hat{\theta}^*(F)$ such that*

- *When the investors with the skill to evaluate a firm are patient, they are willing to buy claims on its project whenever* $\hat{\theta} \geq \hat{\theta}^*(F)$ *at price*

$$p^S(\hat{\theta}, F) = \hat{\theta} z$$

. The project is financed if $\xi \hat{\theta} z \geq 1$.

- *When skilled investors are impatient or* $\hat{\theta} < \hat{\theta}^*(F)$*, skilled investors do not buy claims on the project. Unskilled investors price it at the margin, and claims on the project trade at price*

$$p^U(F) = \frac{\alpha \overline{\theta} + (1-\alpha)\mathbb{E}[\hat{\theta}|\hat{\theta} < \hat{\theta}^*(F)]}{\alpha + (1-\alpha)F(\hat{\theta}^*(F))} z$$

. The project is financed if $\xi p^U(F) \geq 1$.

The threshold satisfies $\hat{\theta}^*(F) z = p^U(F)$.

This proposition shows how adverse selection arises in markets for claims on firms' projects due to information disclosure. Skilled investors will finance a project only when they are patient and they perceive its quality to be sufficiently high. Then, when unskilled investors see that skilled investors did not finance a project, they will rationally draw a negative inference: they know skilled investors may have not invested in the project because it was bad rather than because they were simply impatient. Thus, information disclosure allows skilled investors to more efficiently allocate capital towards good projects. However, information disclosure also forces intermediaries to issue claims to unskilled investors at a lower price, making it more difficult to finance the project when skilled capital is unavailable. The consequence of illiquidity can therefore be *under*-investment in good projects.

On the other hand, opacity (concealing information) prevents skilled investors from allocating capital efficiently, but it also mitigates the adverse selection problem faced by unskilled investors. This is why the issuance of opaque assets will generate booms featuring misallocation of credit towards bad projects in this model.

1.3.2 Morning: Intermediary disclosure policy

In the morning, before matching with firms, intermediaries choose their disclosure policies. They do so in order to maximize their expected profits with respect to the discount factor of investors on the same island. The main tradeoff in disclosing

information is that by doing so, an intermediary causes unskilled investors to face adverse selection, lowering the price at which it can sell claims to them and perhaps preventing investment from taking place, whereas transparency increases the price at which it can sell claims to skilled investors.

Formally, the intermediary's information design problem is

$$\max_{F \in \mathcal{F}} \ \mathbb{E}[v_j | F_j]. \tag{1.7}$$

In principle, an intermediary can choose any disclosure policy. While this problem is thus quite abstract, later I show that in equilibrium, intermediaries will make use of only two simple disclosure policies: full transparency and full opacity.

1.3.3 Equilibrium

I now define and characterize a symmetric equilibrium of the static model. In a symmetric equilibrium, asset prices depend only on a project's quality, the disclosure policy of the associated intermediary, and the local impatience shock. Agents on all islands that receive the same impatience shock behave in the same way. Their decisions in the afternoon depend on the local shock as well as their information.

Definition 1. *A **symmetric static equilibrium** consists of a disclosure policy F, intermediary returns $v(s, F, \tilde{\lambda})$ investors' decisions $\{c(\tilde{\lambda}), b(\tilde{\lambda}), a_B(\tilde{\lambda}, s, F, p)\}$, intermediaries' decisions $\{x(\theta, F, p), a_S(\theta, F, p)\}$, a price schedule $p(a_S|\theta, F, \tilde{\lambda})$, and expectations $\mathbb{E}[\cdot|s, F, p]$ for investors such that*

1. *Investors' choices c, b, a_B solve Problem 1.5 taking prices as given;*
2. *Intermediaries' decisions are optimal given the price schedule p they face at $\tau = 1$, and their returns are defined by Equation 1.7;*
3. *Markets clear at $\tau = 1$, and the price schedules faced by intermediaries' are consistent with investors' demand;*
4. *Investors' expectations are consistent with Bayes' Rule whenever possible.*

I focus on symmetric equilibria in which asset prices are as in Proposition 1.

The choice of information structure

The key determinant of equilibrium outcomes will be intermediaries' choice of information disclosure. This decision, in turn, will determine the information structure in

the economy when it comes time to invest goods in projects in the afternoon. I will show that there are two possible outcomes: an *opaque regime* in which intermediaries disclose no information, and a *transparent regime* in which intermediaries fully disclose the quality of their projects.

The central feature of this economy that gives rise to a role for opacity is its information structure. Within an island, investors always know the quality of all transparent projects. Across islands, however, investors cannot tell whether an asset sells at a low price because (1) those with the expertise to evaluate the underlying project's quality are impatient, or (2) those with the expertise to evaluate it know it is of bad quality. Assets will sometimes sell at a low price even when the underlying project is good, which can cause intermediaries to forgo ex-post efficient investments.

I now describe the intermediary's information design problem. The intermediary's disclosure policy induces a distribution F over skilled investors' posterior belief $\hat{\theta}$ of the firm's quality. When skilled investors are not impatient (with probability $1-\alpha$), they will finance the project if and only if $\hat{\theta} \geq \theta^* = \frac{1}{z}$ (that is, if they perceive the project to be positive-NPV). If skilled investors are impatient (with probability α), they never finance the project. Then, when skilled investors do not finance the project, unskilled investors think its expected quality is

$$\theta^U(F) = \frac{\alpha\overline{\theta} + (1-\alpha)\mathbb{E}[\theta|\hat{\theta} < \theta^*]}{\alpha + (1-\alpha)F(\theta^*)}$$

Intermediaries can sell claims on the project at price $p(\hat{\theta}) = \max\{\hat{\theta}, \theta^U(F)\}z$. Equivalently, the project is financed at an interest rate $R(\hat{\theta}) = \frac{z}{p(\hat{\theta})} = \frac{1}{\max\{\hat{\theta}, \theta^U(F)\}}$. The project is financed as long as $R(\hat{\theta}) \leq \xi z$ (that is, as long as the interest repayment R is less than the project's collateral, ξz). The intermediary's profits are $p(\hat{\theta}) - 1 = \frac{z}{R(\hat{\theta})} - 1$ if the project is financed, so its problem is then

$$\begin{aligned} \max_{F \in \mathcal{F}} \ & \alpha\lambda\left(\frac{z}{R(\theta^U(F))} - 1\right) \cdot \mathbf{1}\{R(\theta^U(F)) \leq \xi z\} \\ & + (1-\alpha)\int_0^1 \left(\frac{z}{R(\hat{\theta})} - 1\right) \cdot \mathbf{1}\{R(\hat{\theta}) \leq \xi z\} dF(\hat{\theta}). \end{aligned} \tag{1.8}$$

The first result that I prove is that in equilibrium, intermediaries choose either a policy of full transparency or full opacity.

Proposition 2. *In equilibrium, the solution to intermediaries' problem can be implemented by choosing one of two disclosure policies: transparency (in which case skilled*

investors learn the project's quality in the afternoon) or opacity (in which case they learn nothing).

I provide a proof sketch, with a more complete proof in the Appendix. There are two possibilities for the intermediary. It may wish to finance the project by issuing claims to unskilled investors, or it may not. If it issues claims to unskilled investors, then the project is always financed (regardless of its quality). Investors break even, so the average price at which the intermediary sells claims cannot exceed the average project cashflow, $\overline{\theta}z$. Hence, the intermediary cannot do better than a policy of full opacity, which maximizes the price unskilled investors are willing to pay (and thus prevents a situation in which the price those investors pay is insufficient to cover the intermediary's cost of investment).

On the other hand, the intermediary may design an information disclosure policy such that only skilled investors finance the project, forgoing investment when unskilled capital is needed. This occurs when the adverse selection faced by unskilled investors is so severe that they are not willing to finance the project. The price paid by skilled investors is $\hat{\theta}z$, where $\hat{\theta}$ denotes their expectation of the project's quality. The project is financed if $\xi \cdot \hat{\theta}z \geq 1$, the cost of investment. Let F denote the distribution of $\hat{\theta}$, investors' posterior expectation of the project's quality. The intermediary's problem in this case is

$$v = \max_{F} (1-\alpha) \int_{\frac{1}{\xi z}}^{1} (\hat{\theta}z - 1) dF(\hat{\theta}) \text{ s.t. } \int_{0}^{1} \hat{\theta} dF(\hat{\theta}) = \overline{\theta}.$$

Note that the right-hand side is a convex function of $\hat{\theta}$, since it is linear for $\hat{\theta} \geq \frac{1}{\xi z}$ and identically zero otherwise. Thus, the right-hand side is increasing under a mean-preserving spread of F. The intermediary can therefore do no better than providing all information to skilled investors (i.e., a policy of full transparency). Intuitively, when the intermediary wants to sell claims to unskilled investors, it maximizes the price at which it does so through a policy of full opacity, which mitigates adverse selection concerns. On the other hand, if an intermediary issues to skilled investors only, it can do no better than a policy of full transparency, which maximizes the average price at which it sells claims to them. This differs from traditional solutions to information design problems (e.g., Kamenica and Gentzkow, 2011) because the intermediary's payoffs do not depend only on whether the project is financed, which is a binary action. Rather, payoffs depend on the *price* paid by investors, which is a continuous variable.

If an intermediary uses a policy of full transparency, it can sell claims on the project to skilled investors when they are patient *and* the project is good (at price $\theta_G z$). On the other hand, if skilled investors are impatient *or* the project is bad, it will have to attempt to sell to unskilled investors. Those investors will value the project at

$$\theta_U z \equiv \frac{\alpha(1-\delta)\theta_G + \delta\theta_B}{\alpha(1-\delta)+\delta} z. \tag{1.9}$$

Parameter θ_U represents the project's expected quality to unskilled investors when skilled investors do not finance the project under transparency. The (expected) value of an intermediary that uses a policy of full transparency is then

$$V^T = \big(\alpha\lambda + (1-\alpha)\delta\big)(\theta_U z - 1)\mathbf{1}\{\theta_U z \geq \xi^{-1}\} + (1-\alpha)(1-\delta)(\theta_G z - 1). \tag{1.10}$$

If the price $\theta_U z$ at which a transparent intermediary can sell assets is too low when skilled capital is unavailable, it may be prevented from investing entirely. That is, illiquidity can hamper investment even when it is efficient. Nevertheless, illiquidity also prevents inefficient investments from proceeding: if a transparent intermediary is able to sell claims and invest only when skilled capital is available, it never finances a bad project.

An intermediary that uses a policy of full opacity, on the other hand, always sells claims at the same price $p^O = \overline{\theta} z$. Its value is

$$V^O = (1 + \alpha(\lambda - 1))(\overline{\theta} z - 1)\mathbf{1}\big\{\overline{\theta} z \geq \xi^{-1}\big\}. \tag{1.11}$$

From this equation, it is clear how opacity differs from transparency. Opacity raises the price at which intermediaries can sell assets when skilled capital is unavailable. Hence, opacity may allow the intermediary to invest when it would be unable to do so under transparency. Opacity enhances the liquidity of an intermediary's liabilities, allowing it to finance its project by attracting unskilled capital. On the other hand, it also causes greater misallocation towards bad projects– no agent is able to determine the project's quality, so the intermediary faces the same financing terms when the project is good and when it is bad. Opaque assets are, in a sense, the mirror image of transparent ones. They are more liquid and therefore easier to trade across islands, but they deprive skilled investors of the information required to ensure they do not finance bad projects.

Adverse selection vs. virtuous selection

Having characterized the tradeoff between transparency and opacity, I now describe the equilibrium outcome in the economy.

Proposition 3. *The unique symmetric static equilibrium consists of two regimes.*

1. ***Opaque regime**: When*

$$(1+\alpha(\lambda-1))(\overline{\theta}z-1) \geq (1-\alpha)(1-\delta)(\theta_G z-1),$$

all intermediaries choose a policy of opacity. All projects, both good and bad, are financed in the afternoon, independently of the preference shock on each island. There is trade in financial assets across islands.

2. ***Transparent regime**: When*

$$(1+\alpha(\lambda-1))(\overline{\theta}z-1) < (1-\alpha)(1-\delta)(\theta_G z-1),$$

all intermediaries choose a policy of transparency. Good projects are financed only when skilled capital is available (i.e., investors on the same island are patient). Bad projects are never financed. Investors purchase only assets whose quality they can evaluate.

This proposition draws a sharp distinction between two types of regimes that may arise in equilibrium: an opaque regime and a transparent regime. In the opaque regime, assets are highly liquid and can be sold across islands to investors who lack the skills to evaluate the quality of the underlying projects. Unskilled investment, however, goes hand-in-hand with misallocation, and bad projects are always financed. That is, the opaque regime essentially corresponds to a situation in which lending standards are lax.

In the transparent regime, investors are reluctant to finance projects through intermediaries that will conceal information. Rather, each investor finances only projects whose quality they can observe. This improves allocative efficiency– bad projects are never financed in the transparent regime. The resulting information asymmetries hamper liquidity, though. In fact, financial markets are fully illiquid in the sense that assets are never traded across islands. Therefore, many good projects go unfunded (in particular, those for which skilled capital is unavailable).

In the opaque regime, there is *over*-investment relative to the first-best benchmark. In the transparent regime, by contrast, there is *under*-investment. What determines

which regime will arise? There are two forces that play a role in determining the equilibrium outcome: the benefit of *virtuous selection* (VS), which I define as the value of forgoing negative-present value investments, and the cost of *adverse selection* (AS), which I define as the cost of the inability to finance a good project when skilled capital is unavailable.

Transparent intermediation grants skilled investors the benefits of virtuous selection but comes at the cost of adverse selection: in the transparent regime, projects are financed only when they are good and skilled investors are patient. By allowing skilled investors to see a signal of the project's quality, the intermediary lowers its cost of funds when those investors finance the project, but by giving information to skilled investors, it increases its cost of raising funds from unskilled investors to the point that doing so is unprofitable. Opaque intermediation, by contrast, gives up the benefits of virtuous selection in order to mitigate adverse selection costs. When the underlying project is good, skilled investors are no longer willing to pay such a high price for assets issued by the intermediary, since they cannot see the project's quality, but unskilled investors no longer require an illiquidity discount in order to finance the intermediary's project.

The cost of adverse selection is forgone investment when a project is good but skilled investors lack funds: with probability $\alpha(1-\delta)$, an investment opportunity worth $\theta_G Z - 1$ is passed up. The benefit of virtuous selection (to investors) is avoiding investment in bad projects and not incurring a loss of $1 - \theta_B Z$ when the project is bad (with probability δ). Investors factor this cost into the price they are willing to pay for claims on an intermediary's project, so it is passed on to intermediaries through an elevated cost of funds. Thus, the cost of adverse selection is greater than the benefit of virtuous selection whenever

$$\underbrace{\alpha\lambda(1-\delta)(\theta_G Z - 1)}_{\text{AS cost}} \geq \underbrace{(1+\alpha(\lambda-1))\delta(1-\theta_B Z)}_{\text{VS benefit}}. \tag{1.12}$$

From Proposition 3, this is precisely the condition required for the economy to be in the opaque regime.

Before moving to the dynamic model, it will be useful to derive comparative statics results.

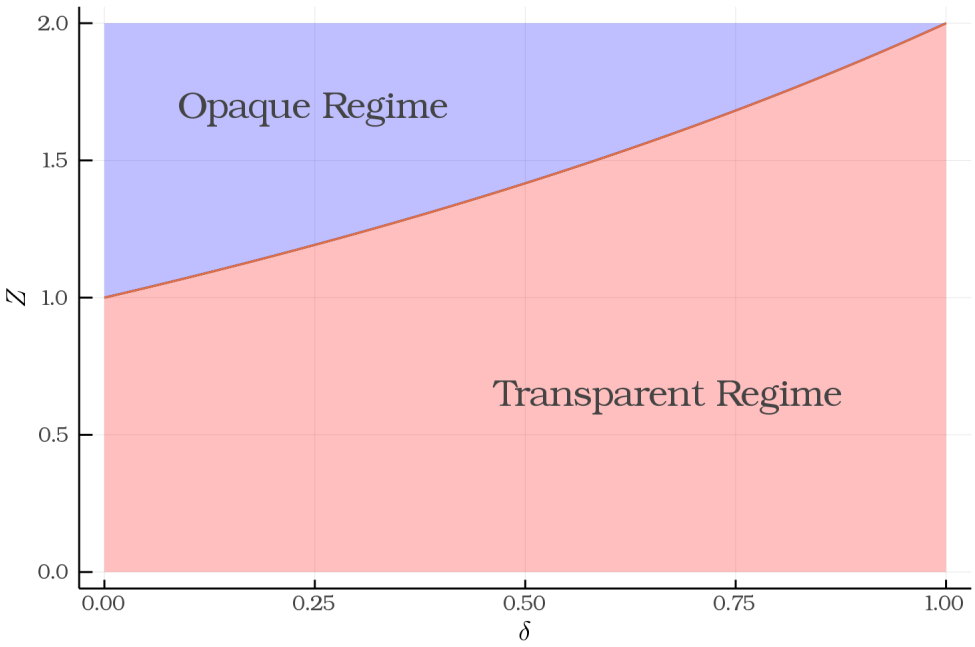


Figure 1.2: Opaque and transparent regimes described in Proposition 3. This example uses parameters $\alpha = 0.3$, $\theta_G = 1.0$, $\theta_B = 0.5$.

Proposition 4. *There exists δ^* such that the economy is in the opaque regime whenever $\delta \leq \delta^*$ and in the transparent regime otherwise, where*

$$\delta^* = \frac{\alpha\lambda(\theta_G z - 1)}{\alpha\lambda(\theta_G z - 1) + (1 + \alpha(\lambda - 1))(1 - \theta_B z)}. \tag{1.13}$$

Likewise, holding other parameters fixed, there exists z^* such that the economy is in the opaque regime for $z \geq z^*$ and in the transparent regime otherwise.

The logic underlying this result is that the cost of adverse selection is high, and the benefit of virtuous selection is low, when the average project in the economy is profitable. This is because the cost of adverse selection relates to passing up good investment opportunities, whereas the benefit of virtuous selection relates to avoiding up bad ones. The critical threshold δ^* that demarcates the boundary between the opaque and transparent regimes will be crucial in governing endogenous cycles in the dynamic model. Figure 1.2 depicts the opaque and transparent regimes in (z, δ) space. In the dynamic model, exogenous productivity z and the endogenous fraction of bad firms δ will be the key state variables.

1.4 Dynamics

In this section, I extend the static model to a dynamic setting to study the macroeconomic conditions under which opacity can arise as well as its role in generating and amplifying credit cycles. I embed opaque intermediaries in a dynamic production economy and introduce persistence in the quality of projects. The dynamic model, which features long-lived investors, will effectively reduce to a repeated version of the static one, with periods linked only by the evolution of productivity and the fraction of bad projects, as well as the accumulation of physical capital. Opacity will keep inferior projects in the pool, whereas transparency weeds out bad projects. Under these assumptions, the economy will feature both amplification of busts following long booms and fully endogenous credit cycles.

1.4.1 Model setup

Time is discrete and infinite, $t = 0, 1, 2, \ldots$, and each period consists of two subperiods, $\tau = 0, 1$ (morning and afternoon). As in the static model, there is a single good that can be consumed or invested as capital, and the economy consists of islands $n \in [0, 1]$. The economy is populated by investors, firms, and intermediaries that match randomly with firms in each period. Agents are randomly sent to new islands at the beginning of each period. The consumption good can be stored across periods (from $\tau = 1$ of t to $\tau = 0$ of $t + 1$) to earn returns $1 + r$. This implies an aggregate resource constraint in each period,

$$y_t = c_t + \overline{k}_t, \tag{1.14}$$

where $\overline{k_t}$ denotes the aggregate capital stock.

Investors: The economy is populated by investors who face impatience shocks. They have preferences that are an infinite-horizon analogue of those in the static model,

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \tilde{\lambda}_{nt} c_t\right], \tag{1.15}$$

where $\beta \in (0, 1)$ denotes the subjective discount factor used by investors. Impatience shocks $\tilde{\lambda}_{nt} \in \{1, \lambda\}$ are I.I.D. across islands and time but perfectly correlated within an island, with $\Pr(\tilde{\lambda} = \lambda) = \alpha$. Investors are endowed with the economy's entire capital stock at $t = 0$. In the afternoon of each period, investors finance projects, consume, and choose how many goods to store until the next period.

Firms and intermediaries: Each firm has a type Θ_{jt} that may evolve over time according to a process specified later. Firms are sent to random islands at the beginning of the morning of each period. Then, intermediaries enter and match anonymously with firms, as in the static model. The production technology available to firms is as in the static model, with one minor modification: the investment capacity of a project is $\omega\overline{k}_t$ rather than one unit, where ω is a constant.[13] I allow the productivity z_t of projects to vary exogenously over time.

Intermediaries choose a disclosure policy $F \in \mathcal{F}$ in the morning and anonymously meet with a firm in the afternoon of each period. In the afternoon, they sell claims on the firm's project to investors in order to finance it. Firms and intermediaries rebate all profits to investors on the same island in proportion to their wealth.[14]

The only differences from the static model are that (1) firms' types Θ_{jt} may change over time, (2) firms randomly match with an intermediary in each period, and (3) the productivity of firms' technology, z_t, follows a stochastic process.

Shocks and state variables: There are three aggregate state variables: the productivity of firms' technology z_t, the fraction of bad firms at t, which I denote δ_t, and the aggregate capital stock $\overline{k}_t$. Shocks to z_t are the only source of aggregate uncertainty in this economy. For now, I simply assume that z_t follows an arbitrary stochastic process.

I now specify the process followed by firms' types Θ_{jt} (which gives rise to the process followed by δ_t). Aggregate firm quality evolves through exit and entry. When a firm exits (or "dies"), it is replaced by a newborn firm that is good with exogenous probability $1-\overline{\delta}$. The key assumption is that firms exit when either (1) their projects fail, or (2) they fail to raise additional funds at $\tau = 1$. An exiting firm is replaced by a newborn good firm.[15] In order to survive into the next period, a firm must raise funds and successfully complete its investment project. These assumptions correspond to a setting in which a firm's failure to raise funds is, to a certain extent, observable. In addition, in every period, a firm may die with exogenous probability $\kappa \in (0,1)$. This specification captures mean reversion in firm quality and ensures that the fraction of bad projects (when only good firms are financed) has a nondegenerate stationary distribution.

[13] I assume $\omega\theta_G z \leq 1$, so that investors financing projects always break even.

[14] That is, the fraction of profits received by an investor on an island is proportional to the wealth of that investor.

[15] It is not important that newborn firms are good rather than drawn from some distribution containing both good and bad firms. I make this assumption only for simplicity of exposition.

Timing: At the beginning of the morning, agents are sent off to random islands. Then, intermediaries publicly commit to their disclosure policies. The afternoon is the exact analogue of $\tau = 1$ in the static model. Intermediaries first match randomly and anonymously with firms.[16] Firms then get investment opportunities, and intermediaries attempt to raise funds to finance investment in their firms' projects. They follow their disclosure policies and trade with investors in the same markets as in the static model. Investors consume and make their investment decisions. All output that is not consumed or invested in firms' projects is stored until the next period. Finally, at the beginning of the morning in the next period, firms return output to intermediaries, who in turn pay back investors.

1.4.2 Dynamic equilibrium

The dynamic model is essentially identical to a repeated version of the static one. Just as in the static model, intermediaries choose their disclosure policies and then attempt to raise additional funds in financial markets subject to asymmetric information. The output of a good project that receives investment is z_t at the beginning of the next period (per unit invested). The information structure within a period is also unchanged: intermediaries and investors enter a period with a common prior δ_t that any given firm is bad, and further information is revealed in the afternoon according to intermediaries' disclosure technologies.

The only difference between the static and dynamic models occurs when investors make their consumption-savings decisions, since they now solve a dynamic optimization problem rather than a static one. I outline that decision here.

Investors' optimization problem: Investors on each island n enter a period with their net worth w_t and receive an impatience shock $\tilde{\lambda}_{nt}$ in the afternoon. Then, they choose their consumption $c_t \geq 0$, how many claims $a_{Bj,t}$ to purchase on each project j, and the amount b_t they want to store until the next period, subject to the budget constraint

$$c_t + b_t + \int_j p_{jt} a_{Bj,t} \leq (1 + \pi_{nt}) w_t, \tag{1.16}$$

[16] Intermediaries therefore have no knowledge of the firm's type within a period. This assumption circumvents a situation in which investors and intermediaries need to keep track of a distribution of beliefs about firm quality.

where $\pi_{n,t}$ denotes intermediary profits rebated to investors on island n per unit of net worth. Investors accumulate net worth according to

$$w_{t+1} = (1+r)b_t + \int_j \tilde{z}_{jt} a_{Bj,t} dj, \tag{1.17}$$

where $\tilde{z}_{j,t+1} = z_t$ if firm j's project succeeds and 0 if it fails. An investor's problem can then be written as

$$\max_{c_t, b_t, a_{Bj,t}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \tilde{\lambda}_{nt} c_t\right] \text{ s.t. (1.16), (1.17), } c_t \geq 0, \ w_t \geq 0. \tag{1.18}$$

I make the following assumption to guarantee that investors consume if and only if they are impatient, thereby simplifying the characterization of their optimization problem greatly.

Assumption 2. *The marginal utility of consumption λ conditional on an impatience shock satisfies*

$$\lambda > (1 + \alpha(\lambda - 1)) \max\{\theta_G \overline{z}, 1 + r\}$$

where $\overline{z}$ denotes the supremum of the support of z_t. Furthermore, $\beta(1-\alpha)\max\{\theta_G\overline{z}, 1+r\} < 1 < \beta(1+r)$.

With risk-neutral preferences, Problem 1.18 is particularly simple to solve.

Proposition 5. *Investor i consumes if and only if she is impatient. When she does so, she consumes all of her net worth. Otherwise, she invests. She is indifferent between all assets that maximize her expected returns on investment,*

$$\max\left\{1 + r, \ \max_j \frac{\mathbb{E}_{it}[\theta_{jt} z_t]}{p_{jt}}\right\}.$$

Investors choose how to allocate their funds between consumption, storage, and investment in projects. They consume only when they are hit by impatience shocks. Otherwise, they invest in projects and storage, which are perfect substitutes.

Equilibrium concept: I now define a dynamic equilibrium of this economy. I will show that in the dynamic equilibrium, the only link between periods will be the evolution of the state variables $(z_t, \delta_t, \overline{k}_t)$.

Definition 2. *A **symmetric dynamic equilibrium** consists of a disclosure policy F_t, intermediary returns $v_t(s, F, \tilde{\lambda})$ investors' decisions $\{c_t(w, \tilde{\lambda}), b_t(w, \tilde{\lambda}), a_{Bt}(w, \tilde{\lambda}, s, F, p)\}$,*

intermediaries' decisions $\{x_t(\theta, F, p), a_{St}(\theta, F, p)\}$*, a price schedule* $p_t(a_S|\theta, F, \tilde{\lambda})$*, expectations* $\mathbb{E}_t[\cdot|s, F, p]$ *for investors, and an aggregate state process* $\{z_t, \overline{k}_t, \delta_t\}$ *such that*

1. *Investors' individual decisions are optimal, taking prices, returns earned by intermediaries, and the law of motion of the aggregate state as given;*
2. *Intermediaries' decisions maximize their single-period returns, taking price schedules* p_t *as given, and their returns are given by the analogue of Equation 1.6;*
3. *Investors' and intermediaries' expectations are consistent with Bayes' Rule whenever possible;*
4. *The sequence* δ_t *is consistent with investment decisions at* $\tau = 1$*, and the sequence* $\overline{k}_t$ *is consistent with the family's consumption-savings decision;*
5. *All markets clear, and the price schedules faced by intermediaries are consistent with investors' demand.*

The following proposition allows me to treat the dynamic equilibrium as a repeated sequence of static equilibria. The morning and the afternoon of each period will be exactly as in the static model, but the state variables $(z_t, \delta_t, \overline{k}_t)$ that determine outcomes in that model will vary over time.

Proposition 6. *In the symmetric dynamic equilibrium, within a period, the equilibrium features two regimes, given the state* $(z_t, \delta_t, \overline{k}_t)$*. Let*

$$\delta_t^* = \frac{\alpha\lambda(\theta_G z_t - (1+r))}{\alpha\lambda(\theta_G z_t - (1+r)) + (1+\alpha(\lambda-1))(1+r-\theta_B z_t)} \tag{1.19}$$

The economy is in an opaque regime (in which all intermediaries choose opacity) if and only if $\alpha\lambda(1-\delta_t)(\theta_G z_t - (1+r)) \geq (1+\alpha(\lambda-1))\delta_t(1+r-\theta_B z_t)$*; otherwise, the economy is in the transparent regime. Investors consume if and only if they are hit by an impatience shock. The state* δ_t *follows the process*

$$\delta_{t+1} = \begin{cases} \kappa\overline{\delta} + (1-\kappa)\overline{\delta}((1-\delta_t)(1-\theta_G) + \delta_t(1-\theta_B)) + (1-\kappa)\theta_B\delta_t & \textit{Opaque regime} \\ \kappa\overline{\delta} + (1-\kappa)\overline{\delta}\big((\alpha + (1-\alpha)(1-\theta_G))(1-\delta_t) + \delta_t\big) & \textit{Transparent regime} \end{cases} \tag{1.20}$$

Crucially, this proposition implies that the opaque and transparent regimes correspond to those in the static model: whenever (z_t, δ_t) are such that

$$\alpha\lambda(1-\delta_t)(\theta_G z_t - (1+r)) \geq (1+\alpha(\lambda-1))\delta_t(1+r-\theta_B z_t),$$

the dynamic equilibrium will feature an opaque regime. The only difference from the static model is that now the exogenous return on storage $1+r$, which can be thought of as an interest rate, enters into the expression determining whether the economy is in the opaque or transparent regime. Intuitively, when the interest rate is higher, the substitute safe asset is more attractive, reducing the incentives for intermediaries to produce opaque assets in order to invest in large volumes.

1.4.3 Steady states and endogenous cycles

In this section, I study the behavior of the economy in the absence of shocks to productivity. I prove that the economy may either converge to a steady state or experience endogenous cycles of transparency and opacity generated by the dynamics of firm quality. I characterize the conditions under which steady states or cycles emerge and analyze their implications for macroeconomic outcomes. For the remainder of this section, I assume that productivity is fixed at some constant level $z_t = z$.

Recall that firms exit when either their projects fail or they fail to raise funds at $\tau = 1$. Denote the probability that a firm of type $\Theta \in \{G, B\}$ dies at time t by

$$\mu_{\Theta t} \equiv \Pr_t(j \text{ dies} \mid \Theta_{jt} = \Theta).$$

Under these assumptions, the fraction of bad firms in the economy, δ_t, follows the process

$$\delta_{t+1} = \big(\mu_{Gt}(1-\delta_t) + \mu_{Bt}\delta_t\big)\overline{\delta} + (1-\mu_{Bt})\delta_t. \tag{1.21}$$

The second term corresponds to the fraction of bad firms that survive after getting financed. This endogenous survival mechanism is the key channel that will generate cycles. Note that this channel operates only when some bad firms are financed, which, in turn, is possible only in the opaque regime. The first term simply represents firms that die and are reborn as bad firms with probability $\overline{\delta}$. This equation highlights that the survival of bad firms endogenously worsens the overall pool.

The economy is in the opaque regime whenever

$$\delta_t \leq \delta^* = \frac{\alpha\lambda(\theta_G z - (1+r))}{\alpha\lambda(\theta_G z - (1+r)) + (1+\alpha(\lambda-1))(1+r-\theta_B z)}.$$

As in the static model, when average firm quality is sufficiently high, intermediaries finance investment by originating opaque assets. In particular, this occurs whenever the fraction of bad projects δ_t is below its critical value δ^*. When this is the case, all

bad projects receive additional financing in the afternoon. On the other hand, when $\delta_t > \delta^*$, investors no longer want to finance projects they know nothing about, and asset origination becomes transparent. Due to the adverse selection problem, investors require a large premium in order to purchase new claims on transparent projects whose quality they do not have the skills to evaluate (i.e., those originated on other islands). This premium is so large that intermediaries cannot raise sufficient funds to complete their investment by selling claims to unskilled investors. Only skilled investors participate in markets for transparent assets, meaning investors purchase only assets originated on their own islands. There is no trade across islands, so good projects go unfunded when those with the skill to evaluate them are impatient. Investors never misallocate financing towards bad projects in the transparent regime.

The probability that a bad firm dies in either regime is

$$\mu_{Bt} = \begin{cases} \kappa + (1-\kappa)(1-\theta_B) & \delta_t \leq \delta^* \\ 1 & \delta_t > \delta^* \end{cases} \tag{1.22}$$

Similarly, the death rate of good firms is discontinuous at the threshold δ^*. The discontinuity in the survival rate of firms can generate cycles. The long-run dynamics depend on the basins of attraction of the law of motion, which differ depending on whether δ_t is greater or less than δ^*. The basins of attraction are δ_h for $\delta_t \leq \delta^*$ and δ_l for $\delta_t > \delta^*$, which are defined as

$$\delta_h = \frac{\mu_G^O \overline{\delta}}{\mu_G^O \overline{\delta} + \mu_B^O (1-\overline{\delta})}, \ \delta_l = \frac{\mu_G^T \overline{\delta}}{\mu_G^T \overline{\delta} + \mu_B^T (1-\overline{\delta})}. \tag{1.23}$$

where $\mu_G^O = \kappa + (1-\kappa)(1-\theta_G)$ and $\mu_G^T = \kappa + (1-\kappa)(\alpha + (1-\alpha)(1-\theta_G))$ are the death rates of good firms in the opaque and transparent regimes, respectively, and μ_B^O and μ_B^T are the death rates of bad firms given in Equation 1.22.

Cycles will arise when $\delta_l < \delta^* < \delta_h$. Even when the fraction of bad firms δ_t is slightly less than the critical value δ^* that triggers a switch to the transparent regime, the fraction of bad firms can continue to climb towards δ_h. The intuition is that whenever the fraction of bad firms is not critically high, a project's value is maximized if it is financed by an opaque intermediary, since claims on the project will be liquid and will always sell at a high price. Even as the critical threshold δ^* is approached, then, intermediation remains opaque. The economy suddenly exits the opaque regime when there are many bad firms and the benefit of avoiding misallocation towards bad projects (virtuous selection) begins to exceed the cost of passing up good ones

(adverse selection). After the critical threshold δ^* is crossed, a project's value is instead maximized by transparency, which, despite precluding financing when skilled capital is unavailable, allows the intermediary to signal the project's quality to skilled investors and sell claims at a higher price when they can identify the project as good. In the transparent regime, only skilled investors participate in financial markets at $\tau = 1$ when projects require investment, and they identify all bad projects. This forces a large quantity of bad firms to exit, thereby cleansing the pool and causing the fraction of bad projects to converge down towards δ_l. Subsequently, the economy can re-enter the opaque regime.

Let $\delta^+(\cdot)$ represent the transition law in Equation 1.21, so that $\delta^+(\delta_t) = \delta_{t+1}$. I now formally define the notions of a steady state and an equilibrium cycle.

Definition 3. *A **steady state** is a value of δ such that $\delta^+(\delta) = \delta$ in the recursive dynamic equilibrium. An **equilibrium cycle** consists of a sequence $\{\delta_n\}_{n=1}^m$ (for some $m \geq 2$) such that in the dynamic equilibrium, $\delta^+(\delta_n) = \delta_{n+1}$ for $n \leq m-1$ and $\delta^+(\delta_m) = \delta_1$.*

The following result characterizes the conditions determining whether the economy converges to a steady state or experiences recurring cycles of transparency and (at least partial) opacity.

Proposition 7. *The recursive dynamic equilibrium features a steady state in the transparent regime at δ_l if $\delta_l > \delta^*$. There is a steady state in the opaque regime at δ_h if $\delta_h < \delta^*$. If $\delta_l < \delta^* < \delta_h$, the equilibrium features a cycle.*

When there is a steady state, the model's dynamics are standard. For instance, consider the case in which there is a steady state in the opaque regime, $\delta_h < \delta^*$. When $\delta_t < \delta_h$, the economy is in an opaque regime (since $\delta_t < \delta^*$ as well) and the fraction of bad firms steadily increases. However, for $\delta_t > \delta_h$, the rate at which bad firms exit (due to failure of their projects) is high enough to offset the entry of new bad firms. Intuitively, such a steady state can arise when the success rate of bad firms, θ_B, is low, because then bad firms will tend to exit even when they obtain financing. In this case, the fraction of bad firms never grows so large that the economy enters the transparent regime, and there is a steady state in the opaque or mixed regime. On the other hand, a steady state in the transparent regime can arise when the rate κ at which good firms go bad is large relative to the critical threshold δ^*. In this case, the steady-state fraction of bad firms is high even when no bad firm is financed. Hence, the economy remains in the transparent regime: the fraction of bad firms

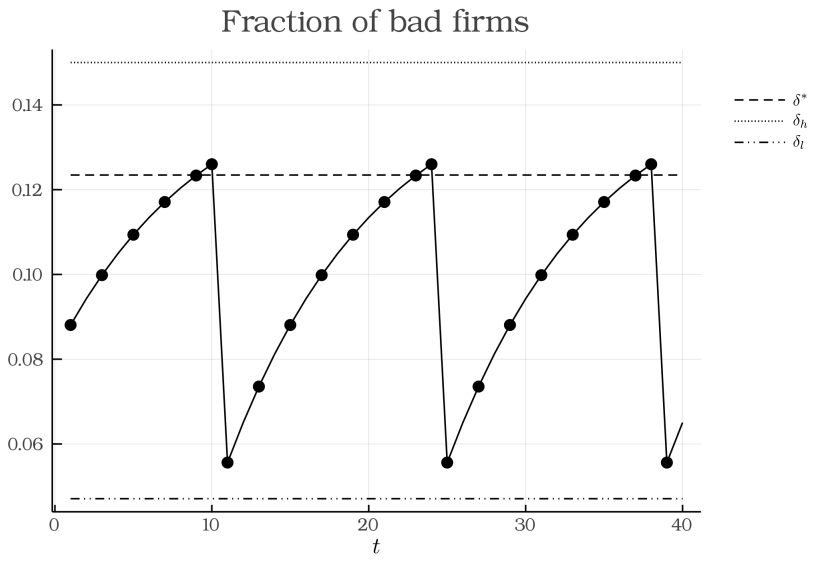


Figure 1.3: This figure illustrates a typical path of the fraction of bad firms δ_t. In this case, there is an eight-period cycle: seven boom periods in which investors are willing to finance opaque intermediaries (while $\delta_t \leq \delta^*$) and one period in which the fraction of bad firms is high enough that opaque intermediation is abandoned.

tends to remain high enough that investors never want to finance projects through opaque intermediaries (because the value of virtuous selection is high at the steady state relative to the cost of adverse selection).

Throughout the rest of this section, I will focus on the case in which cycles emerge. I interpret an equilibrium cycle as a medium-run phenomenon capturing recurrent shifts in the structure of the financial system and the types of claims it produces. Booms will correspond to times in which large volumes of credit are intermediated through entities that issue opaque financial claims, and busts will be times in which those opaque claims have fallen out of favor among investors, leading to lower credit volumes and financial market liquidity.

Under the conditions outlined in the proposition, the fraction of bad firms δ_h to which the economy converges in the opaque regime is greater than the critical threshold δ^* that triggers the transition from the opaque to the transparent regime. Once in the transparent regime, the point δ_l towards which the fraction of bad firms starts to converge is less than δ^*. Figure 1.3 depicts the dynamics in this situation. The economy stays in the opaque regime for seven periods, during which bad firms build up, and then there is a shift to the transparent regime in which unskilled investors are no longer wiling to participate in financial markets.

Economic outcomes differ markedly across the two regimes. The opaque regime features much greater investment in firms' risky projects due to participation by unskilled investors. In turn, this generates higher output $y_t^R \overline{k}_t$ produced by risky projects, where

$$y_t^R = \begin{cases} \omega \overline{\theta}_t z & \text{Opaque regime} \\ \omega(1-\alpha)(1-\delta_t)\theta_G z & \text{Transparent regime} \end{cases} \tag{1.24}$$

Higher output increases the growth rate of the economy, since investors save some output for the future. Therefore, the opaque regime corresponds to a credit boom and, more broadly, an economic boom. In the transparent regime, these patterns reverse. Importantly, the transition from the opaque to the transparent regime causes a crash, which manifests as a sharp dip in output and investment.

Financial markets also exhibit starkly different behavior in booms and busts. In the opaque boom phase, liquidity is high, and the returns achievable for both skilled and unskilled investors are low. However, misallocation towards bad projects progressively worsens over the course of the boom, and the default premium on assets sold by opaque intermediaries (which is just $\frac{\theta_G}{\overline{\theta}_t} - 1$) increases. The transition to the transparent regime triggers a reversal in credit markets. The collapse in opaque intermediation creates an informational environment in which adverse selection is pervasive, and investors are no longer willing to finance projects they cannot evaluate. The adverse selection problem manifests as an *illiquidity* premium for claims on good transparent projects. The wedge between the value of a good transparent project and the price at which claims on such projects can be sold to unskilled investors is

$$\frac{\delta_t(\theta_G - \theta_B)}{(1-\alpha)(1-\delta_t)\theta_G + \delta_t \theta_B}.$$

The transparent regime has a cleansing effect: after opaque intermediation is abandoned, skilled investors finance only those projects they recognize as good, forcing bad firms out of the pool. Although the quantity of credit provided to firms shrinks in the transparent regime, the *quality* of credit is higher: all firms that obtain credit end up repaying investors. These dynamics are consistent with a common view of financial cycles: booms are periods of ignorance featuring rampant misallocation towards inefficient projects, which ultimately causes them to end in busts during which investors seem to exercise an overabundance of caution. The key channel that generates these dynamics in this model is the endogenous structure of financial claims issued over the cycle: claims are optimally more opaque during booms in

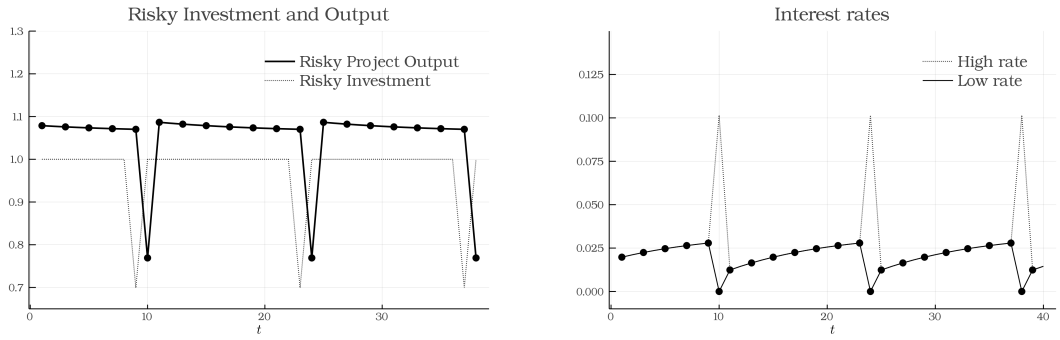


Figure 1.4: The left panel depicts investment into and output from risky projects (per unit of capital) over the course of a cycle with seven opaque boom periods and a transparent bust period. The right panel shows the low interest rate at which projects are financed in equilibrium (solid line) and the high interest rate faced by some firms in the transparent regime when there is adverse selection (dashed line).

order to overcome adverse selection problems, consistent with the boom-bust nature of securitization and the production of other types of liquid liabilities backed by risky assets of unknown quality. The dynamics of real and financial outcomes are illustrated in Figure 1.4.

1.5 Optimal Opacity

The possibility of endogenous cycles generated by opaque intermediation in this model raises the question of whether these cycles can be efficient. Would a policymaker want to increase the degree of transparency in the economy or would it be better to hide more information and increase financial market liquidity? What tools should be used to implement the optimal quantity of opaque projects? Would a policymaker ever allow for cycles?

In order to answer these questions, I begin by considering an abstract problem faced by a social planner who faces the same inference problem as the model's agents. There are two differences between the planner and private agents: first, the planner internalizes the effects of keeping bad firms in the pool, and second, the planner can circumvent the collateral constraints facing intermediaries.[17] I will analyze the properties of the planner's problem in the absence of shocks and then relate the planner's solution to realistic policies such as transparency regulations, public liquidity

[17]Implicit in this assumption is that the planner has a better ability to commit than private agents do. Any sort of tax or subsidy scheme requires some ability on behalf of the planner to promise repayments to agents.

provision, and macroprudential policy. Analyzing the economy without shocks will permit me to analytically characterize the externality in this model and the policy that corrects it.

1.5.1 The planner's problem

The planner in this economy plays an *information design* role: he chooses how much information intermediaries will disclose about their projects. That is, for each intermediary j, the planner chooses a disclosure policy F_j (which in principle allows the planner to mix between information structures rather than choosing just one). The information available to the planner is the same that would be available to an unskilled investor in the transparent regime. The planner can distinguish two types of projects: those for which skilled capital is available and those for which it is not. In turn, skilled capital is available to a project when skilled investors believe its net present value is positive, $\hat{\theta}z \geq 1 + r$, and those investors are patient. Otherwise, the planner does not observe the project's type. Hence, the planner chooses

1. A finite subset $\mathcal{F}_t^* \subset \mathcal{F}$ of disclosure policies, and a measure m_{Ft} of intermediaries using disclosure policy F for each $F \in \mathcal{F}_t^*$;

2. An investment policy $x_t(\tilde{\theta}|F) \in [0, 1]$, where

$$\tilde{\theta} = \begin{cases} \hat{\theta} & \text{Skilled investors are patient and } \hat{\theta}z \geq 1 + r \\ \alpha\overline{\theta} + (1-\alpha)\mathbb{E}\big[\theta\big|\hat{\theta}z \leq 1 + r\big] & \text{Otherwise} \end{cases}$$

Claims on projects sell at price $\tilde{\theta}z$. Hence, consumption on islands hit by the impatience shock is

$$c_{nt} = \begin{cases} \overline{k}_t(1 + \omega\mathbb{E}[x(\tilde{\theta}|F)(\tilde{\theta}z - 1)]) & \tilde{\lambda}_{nt} = \lambda \\ 0 & \tilde{\lambda}_{nt} = 1 \end{cases} \tag{1.25}$$

The capital stock grows according to the investment policy chosen by the planner. Thus

$$\overline{k}_{t+1} = (1 + r)(\overline{k}_t - c_t) + \underbrace{\mathbb{E}[\omega(x_t(\tilde{\theta}|F)\tilde{\theta}z - (1 + r))]\overline{k}_t}_{\text{Risky investment}}. \tag{1.26}$$

The economy grows due to investment in the risk-free technology and risky investment, which is determined by the planner's optimal policy.

The planner uses a utilitarian social welfare function to rank outcomes. Formally, the problem faced by the planner is

$$W = \max_{\mathcal{F}_t^*, m_{Ft}, x_t(\tilde{\theta}|F)} \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t \int_0^1 \tilde{\lambda}_{nt} c_{nt} dn\Big] \text{ s.t.(1.25), (1.26).} \tag{1.27}$$

That is, W is just the weighted sum of the utilities achieved by all agents. I first show that at an optimum, the planner chooses a disclosure policy either transparency or opacity for each project.

Proposition 8. *At an optimum, the only disclosure policies that may be chosen for any project are a policy of full transparency or one of full opacity. Hence, the planner's problem reduces to choosing a fraction of opaque projects $m_t^O \in [0,1]$.*

1.5.2 The externality and optimal policy

I write the planner's problem in recursive form in order to characterize the externality in this environment. The main state variables in this problem are the aggregate capital stock $\overline{k}$ and the fraction of bad firms δ. Due to the linearity of the technology, the planner's value function can be written as $V(\overline{k}, \delta) = v(\delta)\overline{k}$. The value function v satisfies

$$\begin{aligned} V(\overline{k}, \delta) = \max_{\mu^O} \; & \lambda\big(1 + \omega m^O(\overline{\theta} z - 1)\big)\overline{k} + \beta \mathbb{E}[V(k', \delta')] \\ \text{s.t. } & k' = (1+r)(1 - \alpha(1 + \omega m^O(\overline{\theta} z - 1)))k + \omega\big(m^O \overline{\theta} z + (1 - m^O)(1-\alpha)\theta_G z\big)k \\ & \delta' = \kappa + (1-\kappa)\big(m^O(1-\theta_B) + (1-m^O)(\alpha + (1-\alpha)(1-\theta_G))\big)\overline{\theta} + (1-\kappa) m^O \theta_B \delta. \end{aligned}$$

This problem captures two tradeoffs. First, the planner understands that by choosing a greater fraction of opaque projects, investment is increased but at the cost of greater misallocation. This is essentially the same tradeoff faced by intermediaries: opacity increases their investment volume but prevents them from selling to skilled investors at a high price. This tradeoff is captured in the growth of the capital stock and in investors' consumption. However, the planner also understands that greater opacity has an effect on the fraction of bad firms in the economy. In particular, it allows more bad firms to survive, which can increase the fraction of bad firms going forward. There will be a role for policy because this tradeoff is not internalized by agents in the economy, who simply take the fraction of bad firms as given.

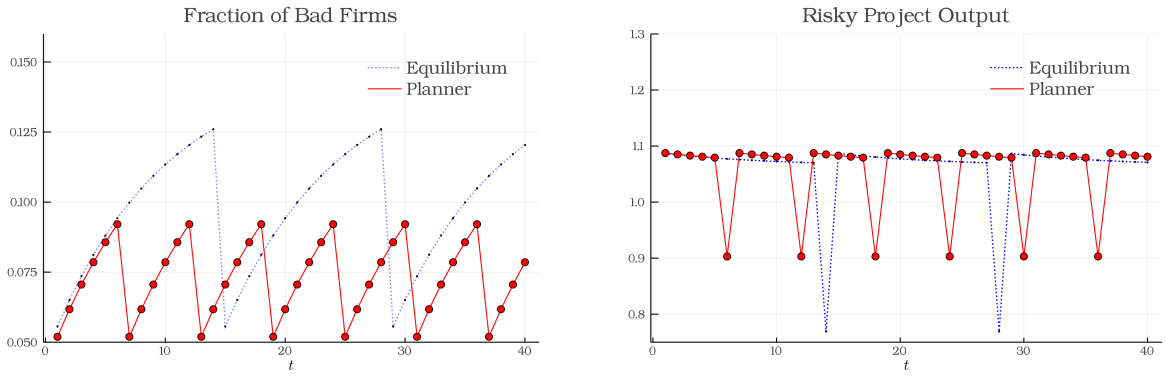


Figure 1.5: Fraction of bad firms (left panel) and output produced by risky projects per unit of capital in the economy (right panel). The equilibrium outcome is shown in blue, whereas the planner's constrained optimum is in red.

Intuitively, investors do not internalize that when they finance an opaque project whose quality they cannot evaluate, they may be keeping a bad firm alive to be financed by other investors later on, thereby lowering those investors' return on capital. This implies that investors are willing to pay a price for opaque assets that is too high, giving intermediaries too great of an incentive to produce such assets. Therefore, the equilibrium will feature too much creation of opaque, liquid assets, as formalized by the following proposition.

Proposition 9. *The optimum may feature a cycle with opaque and transparent regimes. When both the constrained optimum and the equilibrium feature a cycle, the cutoff $\delta^{*,sp}$ at which the planner stops the production of opaque assets is less than the equilibrium cutoff $\delta^{*,eqm}$.*

I provide a numerical example of the optimal policy compared to the equilibrium outcome in Figure 1.5. Just as in the equilibrium, the economy undergoes boom-bust cycles driven by private liquidity creation in the planner's social optimum. The reason such a "bang-bang" policy may be optimal is that a sharp bust can greatly cleanse the economy of bad firms for the following boom, allowing the economy to recover more. However, the optimal length of the boom is shorter than in equilibrium. This prevents crashes from being overly severe, since fewer bad firms accumulate in the economy.

The result that the constrained optimum may feature a cycle depends on agents' risk-neutrality, but the prescription that private liquidity creation should be reduced rather than increased is robust. This result can be interpreted as saying that the planner should lean against booms driven by private liquidity creation rather than

financing firms that cannot borrow in busts, which resembles policies such as quantitative easing, when a central bank purchases risky assets on the open market to provide *public* liquidity to firms or financial intermediaries. Such a policy could keep bad firms alive for *longer* than they would survive in equilibrium, which would actually exacerbate the externality rather than counteracting it.

The externality in this model comes from short-term contracting: investors finance one of a firm's projects, but if the firm survives, it will raise funds in the future from other investors. Thus, this is an information externality– one intermediary's decision to reveal information about a particular firm impacts the investment opportunities of investors who will finance the firm in the future under a different contract.

1.5.3 Implementation

I now discuss how the constrained optimum in the planner's optimization problem can be implemented using common tools. In fact, I show that there are several equivalent tools that the planner could use in order to achieve it. These tools are

1. *Transparency regulation*: A restriction $\overline{m}^O(\delta)$ on the fraction of projects that are financed through the issuance of opaque assets. That is, a fraction $1-\overline{m}^O(\delta)$ of intermediaries are required to disclose information in aggregate state δ.

2. *Macroprudential policy*: A Pigouvian tax $\tau(\delta)$ on the origination of opaque assets (whose revenue is rebated back to investors in proportion to wealth).

3. *Monetary policy*: A subsidy to the interest rate $1+r$ such that the return on storage faced by agents is $1+\tilde{r}(\delta) \geq 1+r$ (financed by a proportional wealth tax on agents).

A transparency regulation in the model is similar to what is seen in reality: the planner simply forces intermediaries to disclose information to investors about the projects they are financing. This generates adverse selection and reduces issuance volumes, but it prevents investors from financing projects whose quality they cannot identify and thus counteracts the externality in this environment.

I show that all of these tools are equivalent in terms of their ability to achieve the constrained optimum.

Proposition 10. *The constrained optimum can be achieved through a transparency regulation, macroprudential policy, or monetary policy. Transparency regulation may need to fluctuate over time, but it is always sufficient to use a macroprudential policy or monetary policy that is constant over time.*

Hence, each tool can be used to achieve the planner's policy objective. The common thread among all three tools is that they should be used to discourage the production of opaque, liquid assets. The one difference between the tools is that while a constant macroprudential policy or monetary policy suffices to moderate booms driven by private liquidity creation, transparency regulation must vary over time. In particular, the optimal transparency regulation may need to allow a boom to develop for some time before requiring information disclosure about projects. That is, opacity has social value that the planner would like to exploit. One interpretation of such a cyclical transparency policy is that an asset class must be allowed to develop for some time before bad projects build up to the point that transparency is required. After the implementation of transparency, issuance volumes and liquidity will decrease, but the boom is cut off before it could cause an even larger bust. The economy then proceeds towards a new boom, which can be interpreted as being driven by a different asset class (set of firms) that will later require regulation as well.

1.6 Conclusion

In this paper, I study the macroeconomic effects of the private production of opaque, liquid assets. I characterize the tradeoff governing the conditions under which this type of opaque intermediation can arise: on the one hand, opacity mitigates asymmetric information among investors, allowing unskilled investors to finance projects without fear of facing adverse selection. On the other, opacity prevents investors from uncovering bad projects, causing them to misallocate credit towards those projects. During good times, when firms' projects tend to be profitable, opacity maximizes the expected value of a project because it ensures the project will always be financed when it is good. During bad times, by contrast, project value is maximized by transparency because it ensures the project will never be financed when it is bad.

I embed this mechanism in a dynamic macroeconomic model in order to study its implications for credit booms and busts. When economic fundamentals are strong, opaque intermediation is prevalent, permitting an expansion in the supply of credit and an increase in financial market liquidity but also leading to a deterioration of the quality of firms in the economy. As such, booms featuring an expansion in the supply of liquid assets, high credit volumes, and lax lending standards are followed by persistent slumps with fragmented financial markets, depressed credit, and tight lending standards. These dynamics amplify busts following transitory booms and can even produce fully endogenous cycles.

The optimal policy in the face of opaque intermediation involves restricting the quantity of opaque, private liquidity creation. This can involve transparency regulations, macroprudential tools, or even monetary policy. Increased transparency and taxes on the issuance of opaque assets lean against unskilled investment during booms, combating the build-up of bad projects brought about by opacity. The benefits of private liquidity creation are fully internalized by intermediaries in this model, so there is no particular reason for a policymaker to provide public liquidity. I interpret this result as a statement about the medium run, which is the frequency of financial cycle fluctuations that the model intends to capture.

A fruitful avenue to extend this paper's results would be a quantitative examination of the main mechanisms. In particular, the model implies that the issuance of opaque securities (i.e., the rise of certain types of intermediation) should predict subsequent credit busts during which issuance of those securities collapses. This prediction could confronted with data on privately produced liquid assets and the default rates of the underlying projects. Furthermore, the model implies a strong comovement between liquidity premia and the excess returns to skilled capital that could be tested by measuring the correlation between common measures of the liquidity premium and the return on sophisticated financial institutions' equity.

References

Arora, Sanjeev, Boaz Barak, Markus Brunnermeier, and Rong Ge, "Computational Complexity and Information Asymmetry in Financial Products," 2009. Working Paper.

Asriyan, Vladimir, Luc Laeven, and Alberto Martín, "Collateral Booms and Information Depletion," 2019. NBER Working Paper.

Bigio, Saki, "Endogenous Liquidity and the Business Cycle," *American Economic Review*, 2015, *106* (6), 1883–1927.

Caramp, Nicolas, "Sowing the Seeds of Financial Crisis: Endogenous Asset Creation and Adverse Selection," 2017. Working Paper.

Chemla, Gilles and Christopher Hennessy, "Skin in the Game and Moral Hazard," *Journal of Finance*, August 2014, *69* (4), 1597–1641.

Chernenko, Sergey, Samuel Hanson, and Adi Sunderam, "Who Neglects Risk? Investor Experience and the Credit Boom," *Journal of Financial Economics*, 2016, *122* (2), 248–269.

Dang, Tri Vi, Gary Gorton, and Bengt Hölmstrom, "Ignorance, Debt and Financial Crises," 2015. Working Paper.

_ , _ , _ , and Guillermo Ordoñez, "Banks as Secret Keepers," *American Economic Review*, 2017, *107* (4), 1005–1029.

Diamond, William, "Safety Transformation and the Structure of the Financial System," 2020. Working Paper.

Eisfeldt, Andrea, "Endogenous Liquidity in Asset Markets," *Journal of Finance*, 2004, *59* (1), 1–30.

Farboodi, Maryam and Peter Kondor, "Rational Sentiments and Economic Cycles," 2020. Working Paper.

Foley-Fisher, Nathan, Gary Gorton, and Stéphane Verani, "Adverse Selection Dynamics in Privately-Produced Safe Debt Markets," 2020. NBER Working Paper.

Gârleanu, Nicolae, Stavros Panageas, and Jianfeng Yu, "Impediments to Financial Trade: Theory and Applications," 2019. Working Paper.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "A Model of Shadow Banking," *Journal of Finance*, 2013, *58* (4), 1331–1363.

Gorton, Gary and Guillermo Ordoñez, "Collateral Crises," *American Economic Review*, 2014, *104* (2), 343–378.

_ and _ , "Good Booms, Bad Booms," *Journal of the European Economic Association*, 2018, *18* (2), 618–665.

Hölmstrom, Bengt and Jean Tirole, "Private and Public Supply of Liquidity," *Journal of Political Economy*, 1998, *106* (1), 1–40.

Kaminsky, Graciela, "Crises and Sudden Stops: Evidence from International Bond and Syndicated-Loan Markets," 2008. NBER Working Paper.

Kiyotaki, Nobuhiro and John Moore, "Credit Cycles," *Journal of Political Economy*, 1997, *105* (2), 211–248.

_ **and** _ , "Inside Money and Liquidity," 2018. Working Paper.

_ **and** _ , "Liquidity, Business Cycles, and Monetary Policy," *Journal of Political Economy*, 2019, *127* (6), 2926–2966.

_ **and Mark Gertler**, "Financial Intermediation and Credit Policy in Business Cycle Analysis," *Handbook of Monetary Economics*, 2010, *3*, 547–599.

Kopytov, Alexandr, "Booms, Busts, and Common Risk Exposures," 2020. Working Paper.

Kurlat, Pablo, "Lemons Markets and the Transmission of Aggregate Shocks," *American Economic Review*, 2013, *103* (4), 1463–1489.

_ , "Asset Markets with Heterogeneous Information," *Econometrica*, 2016, *84* (1), 33–85.

Lee, Michael Junho, "Uncertain Booms and Fragility," 2018. Working Paper.

Li, Ye, "The Procyclicality of Intermediated Credit," 2020. Working Paper.

Lucas, Robert, "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 1973, *63* (3), 326–334.

Moreira, Alan and Alexi Savov, "The Macroeconomics of Shadow Banking," *Journal of Finance*, 2017, *72* (6), 2381–2432.

Neuhann, Daniel, "Inefficient Asset Price Booms," 2019. Working Paper.

Pagano, Marco and Paolo Volpin, "Securitization, Transparency, and Liquidity," *Review of Financial Studies*, 2012, *25* (8), 2418–2453.

Riddiough, Timothy and Howard Thompson, "Deja vu All Over Again: Agency, Uncertainty, Leverage, and the Panic of 1857," 2012. Working Paper.

Rock, Kevin, "Why New Issues are Underpriced," *Journal of Financial Economics*, 1986, *15* (1-2), 187–212.

Vanasco, Victoria, "The Downside of Screening for Asset Market Liquidity," *Journal of Finance*, October 2017, *72* (5), 1937–1982.

White, Eugene, "Lessons from the Great American Real Estate Boom of the 1920s," 2009. NBER Working Paper.

Appendix

Proofs for Section 1.3

Proof of Lemma 1:

Proof. Let λ be the multiplier on investors' budget constraint, and let μ_j be the multiplier on the short-sale constraint for asset j (with μ_0 being the multiplier on the no-borrowing constraint $b \geq 0$). The first-order condition for investors' demand for asset j is

$$\mathbb{E}[\theta_j Z | s_{ij}, \sigma_j, p_j] + \mu_j = \lambda p_j,$$

and the first-order condition on storage b is

$$1 + \mu_0 = \lambda.$$

Each multiplier μ_j is either positive or zero (where positivity indicates that the corresponding constraint binds, $a_{Bj} = 0$), so the investor holds only assets such that

$$\frac{\mathbb{E}[\theta_j Z | s_{ij}, \sigma_j p_j]}{p_j} = \lambda,$$

where $\lambda \geq 1$ by the first-order condition for storage. If $R_i^* < 1$, then, the investor is never willing to purchase a positive quantity of any asset. □

Proof of Lemma 2

Proof. Note that the intermediary will always choose $x = \min\{1, p_j(a_S)a_S\}$ whenever it is profitable to invest. If the intermediary sells assets a_S at price $p_j(a_S)$, its profits are

$$\theta_j(p_j(a_S) - 1)a_S.$$

Hence, the intermediary's problem reduces to

$$\max_{a_S} (p_j(a_S) - 1)a_S,$$

which does not depend on the intermediary's type. The intermediary must raise at least 1 per unit of assets that it sells, but it cannot pledge more than a fraction ξ of each unit of assets created. Thus, it can invest only if there exists a_S such that $\xi p_j(a_S) \geq 1$. Note that it is always optimal for the intermediary to invest a positive amount in this case, since the above simplification of the intermediary's problem shows that it is profitable for the intermediary to invest whenever it can sell at $p_j \geq 1$. $\square$

Proof of Proposition 1:

Proof. In this section, I construct the model's equilibrium. I assume a symmetric equilibrium such that investors' demand schedules depend only on their signals s, their liquidity shocks $\tilde{\lambda}$, and the intermediary's disclosure policy F. Observe that under these conditions, aggregate demand for any particular asset will depend only on the signals of investors who are not hit by a liquidity shock. Therefore, the demand for an asset j will depend on the signal $\hat{\theta}_j$ disclosed to skilled investors, the intermediary's disclosure policy F_j, and the liquidity shock $\tilde{\lambda}_{n(j)}$ of investors on its island.

Observe that skilled investors will purchase an asset j only if (1) they are not hit by the liquidity shock, $\tilde{\lambda}_{n(j)} = 1$, and (2) their posterior of its quality is sufficiently high, $\hat{\theta}_j \geq \hat{\theta}^*(F_j)$ for some cutoff value $\hat{\theta}^*(F_j)$ that depends on the intermediary's disclosure policy F_j. In this case, these investors are willing to buy the entire stock of the asset as long as $p_j \leq \hat{\theta} z$ (their expected value of the asset). Thus, their demand curve is given by

$$a_B^S(p|\hat{\theta}, F, \tilde{\lambda} = 1) = \begin{cases} 0 & p > \hat{\theta} z \\ a_B^S(F) & p = \overline{\theta} z \\ \frac{e}{p} & p < \overline{\theta} z \end{cases}$$

for some constant $a_B^S(F)$ (which will be set such that asset markets clear).

Skilled investors do not finance the project when either their posterior of the project's quality is low enough, $\hat{\theta}_j < \hat{\theta}^*(F_j)$, or when they are impatient, $\tilde{\lambda}_{n(j)} = \lambda$. Then unskilled investors' expected value of the project's quality is

$$\hat{\theta}^U(F_j) = \frac{\alpha\overline{\theta} + (1-\alpha)\mathbb{E}[\hat{\theta}|\hat{\theta} < \theta^*(F_j)]}{\alpha + (1-\alpha)F_j(\hat{\theta}^*(F_j))}.$$

Unskilled investors' demand curve is given by

$$a_B^U(p|F) = \begin{cases} 0 & p > \hat{\theta}^U(F) \\ a_B^U(F) & p = \hat{\theta}^U(F) \\ 0 & p > \hat{\theta}^U(F) \end{cases}$$

for a constant $a_B^U(F)$. There is an indeterminacy in who actually purchases these assets, since everyone will observe the price in equilibrium and observe that the asset is good. However, this indeterminacy is irrelevant in determining aggregate outcomes.

Now I show that there is pooling in equilibrium: assets such that $\hat{\theta}_j < \theta^U(F_j)$ are pooled with all assets such that the corresponding skilled investors are impatient, $\tilde{\lambda}_{n(j)} = \lambda$). Investors hit by a liquidity shock must submit a demand of zero. Furthermore, skilled investors must submit a demand of zero at any price $p \geq \hat{\theta} z$. Hence, the demand for these assets must come from unskilled investors on surplus islands. This immediately implies that the demand schedules faced by intermediaries selling such assets must be identical. By Lemma 2, all these intermediaries must then supply the same quantity of assets. The entire supply is absorbed by investors on patient islands. Out of the assets available, a fraction $\frac{(1-\alpha)(1-\delta)}{(1-\alpha)(1-\delta)+\delta}$ are actually good. Therefore, these investors will break even if

$$p^U(F) = \frac{\alpha\overline{\theta} + (1-\alpha)\mathbb{E}[\hat{\theta}|\hat{\theta} < \hat{\theta}^*(F)]}{\alpha + (1-\alpha)F(\hat{\theta}^*(F))}.$$

It is simple to check that these demand schedules are optimal for investors. Further, they imply that each intermediary effectively faces a perfectly elastic demand schedule, making their problems relatively simple. When skilled investors are willing to finance the project, intermediaries can sell claims at price $p = \hat{\theta} z$, so the project is financed if $\xi\hat{\theta} z \geq 1$, and intermediaries sell $a_S(\hat{\theta}, F, \tilde{\lambda} = 1) = \frac{1}{\hat{\theta} z}$ assets and invest $x = 1$. When skilled investors are not willing to finance a project, intermediaries are able to sell claims to unskilled investors at price $p^U(F)$, so the project is financed whenever $\xi p^U(F) \geq 1$. Then intermediaries invest $x = 1$ and supply a quantity $a_S(\hat{\theta}, F, \tilde{\lambda}) = \frac{1}{p^U(F)}$ of assets.

Finally, I derive the quantities of assets demanded at each market-clearing price. First, I assume that agents on patient islands absorb the entire supply of assets from their islands, so if $\omega(F)$ is the fraction of projects with disclosure policy F, they must purchase a quantity $a_B^S(F) = \omega(F) \cdot F(\hat{\theta}^*(F))$ of assets. They also purchase all available assets from other islands, so $a_B^U(F) = \omega(F) \cdot \frac{\alpha + (1-\alpha)F(\hat{\theta}(F))}{1-\alpha}$.

□

Proof of Proposition 2

Proof. Note that there are two possibilities in equilibrium. Either the information structure chosen by the intermediary is such that the project can be financed by unskilled investors (so the intermediary can sell to those investors at price $p^U(F) \geq \xi^{-1}$), or it is such that the project is not financed by unskilled investors (so $p^U(F) < \xi^{-1}$). In the former case, the project is always financed. The intermediary sells claims on the project at price $p^U(F)$ when skilled investors are impatient and at price $\max\{p^U(F), p^S(\hat{\theta}, F)\}$. The value function of an intermediary can be written as

$$\begin{aligned} v(F) &= \alpha\lambda\mathbb{E}[(p^U(F) - 1) \mid \tilde{\lambda} = \lambda] + (1-\alpha)\mathbb{E}[(\max\{p^U(F), p^S(\hat{\theta}, F)\} - 1) \mid \tilde{\lambda} = 1] \\ &\leq (1 + \alpha(\lambda - 1))\mathbb{E}[\max\{p^U(F), p^S(\hat{\theta}, F)\} - 1] \\ &= (1 + \alpha(\lambda - 1))(\overline{\theta}z - 1). \end{aligned}$$

The last line uses the fact that $\mathbb{E}[\max\{p^U(F), p^S(\hat{\theta}, F)\}] = \overline{\theta}z$, since investors break even and the project is financed in every state (regardless of its quality). Hence, if the intermediary wishes to issue claims to unskilled investors, it can do no better than a policy of full opacity.

On the other hand, if the intermediary wishes to issue claims only to skilled investors, the project will be financed only when those investors expect it to have positive net present value and are patient. Then the intermediary always sells claims at $p^S(\hat{\theta}, F) = \hat{\theta}z$, and its value in this case is

$$v(F) = (1-\alpha)\mathbb{E}[(\hat{\theta}z - 1)^+].$$

The function inside the expectation on the right-hand side is a convex function of $\hat{\theta}z$. Thus, a distribution F that provides a mean-preserving spread of $\hat{\theta}$ increases the intermediary's value. Therefore, in this case, a policy of full transparency is optimal.

□

Proof of Proposition 3:

Proof. First, observe that any symmetric static equilibrium must define a symmetric equilibrium of the asset market in the afternoon. This equilibrium is characterized in Proposition 1, and the relevant objects in the asset market equilibrium ($b(\tilde{\lambda})$, $a_B(p|\hat{\theta}, F, \tilde{\lambda})$, $x(p|\hat{\theta}, F, \tilde{\lambda})$, $a_S(p|\hat{\theta}, F, \tilde{\lambda})$, $p(a_S|\hat{\theta}, F, \tilde{\lambda})$, $\mathbb{E}[\cdot|\hat{\theta}, F, p]$) are derived in the proof of that proposition.

Let O denote a disclosure policy of full opacity ($\hat{\theta} = \overline{\theta}$ with probability 1), and let T denote one of full transparency ($\hat{\theta} = \theta$). In this equilibrium, the value of an opaque intermediary, $v_j = v(\theta_j, F_j = O, \tilde{\lambda}_{n(j)})$ is

$$v(\theta_j, F_j = O, \tilde{\lambda}_{n(j)}) = \tilde{\lambda}(p^O \xi - 1) + \theta_j z(1 - \xi),$$

where $p^O = \overline{\theta}$. This implies that

$$V^O = \mathbb{E}[v_j | F_j = O] = \big(1 + \xi\alpha(\lambda - 1))\big)(\overline{\theta} z - 1).$$

Let

$$\overline{p}^T = \theta_G z, \ \underline{p}^T = \frac{\alpha\overline{\theta} + (1 - \alpha)\delta\theta_B}{\alpha + (1 - \alpha)\delta}.$$

The value of a transparent intermediary is

$$v(\theta_j, F_j = T, \tilde{\lambda}_{n(j)}) = \begin{cases} \theta_G z - 1 & \theta_j = \theta_G \text{ and } \tilde{\lambda}_{n(j)} = 1 \\ \big(\lambda(\underline{p}^T \xi - 1) + \theta_G z(1 - \xi)\mathbf{1}\{\underline{p}^T \geq \xi^{-1}\}\big) & \theta_j = \theta_G \text{ and } \tilde{\lambda}_{n(j)} = \lambda \\ \big(\tilde{\lambda}(\underline{p}^T \xi - 1) + \theta_B z(1 - \xi)\big)\mathbf{1}\{\underline{p}^T \geq \xi^{-1}\} & \theta_j = \theta_B \end{cases}$$

After some algebra, this implies

$$V^T = \mathbb{E}[v_j | F_j = T] = (1-\alpha)(1-\delta)(\theta_G z - 1) + \big(\alpha + (1-\alpha)\delta\big)\big(1 + \xi\alpha(\lambda - 1)\big)(\theta_U z - 1)\mathbf{1}\{\theta_U z \geq \xi^{-1}\}.$$

Equations 1.10 and 1.11 imply that, whenever

$$\big(1 + \xi\alpha(\lambda - 1)\big)(\overline{\theta} - 1) \geq (1 - \alpha)(1 - \delta)(\theta_G z - 1),$$

opacity dominates transparency: as shown in Proposition 2, it is never optimal to choose transparency if the intermediary raises funds from unskilled investors.

In the opaque regime, there is trade in financial assets across islands, since opaque projects are always financed even when investors on the same island have no funds. Projects are financed regardless of quality because no investor can tell if opaque projects are good or bad.

If

$$\big(1 + \xi\alpha(\lambda - 1)\big)(\overline{\theta} - 1) < (1 - \alpha)(1 - \delta)(\theta_G z - 1),$$

on the other hand, the value of a transparent intermediary is strictly higher than that of an opaque intermediary, so in that region, investors will lend capital only to transparent intermediaries. Furthermore, since $\theta_U z < 1 < \xi^{-1}$, transparent interme-

diaries will never be able to profitably finance their investments by selling assets to unskilled investors, since that would require instead that $\theta_U z \geq \xi^{-1}$ (Proposition 1). In the transparent regime, then, there is no trade in financial assets across islands. All intermediaries are transparent, so they are able to finance investment by selling assets to investors on their own islands only when their projects are good.

□

Proof of Proposition 4:

Proof. The opaque regime consists precisely of the pairs (Z, δ) such that

$$(1 + \xi\alpha(\lambda - 1))((1 - \delta)\theta_G z + \delta\theta_B z - 1) \geq (1 - \alpha)(1 - \delta)(\theta_G z - 1).$$

Letting $\hat{\lambda} = 1 + \xi(\lambda - 1)$ and rearranging, we obtain

$$\alpha\hat{\lambda}(1-\delta)(\theta_G z-1) \geq (1-\alpha)\delta(1-\theta_B z) \Rightarrow \delta \leq \frac{\alpha\hat{\lambda}(\theta_G z - 1)}{\alpha\hat{\lambda}(\theta_G z - 1) + (1 + \alpha(\hat{\lambda} - 1))(1 - \theta_B z)},$$

as claimed. Note that, equivalently, the transparent regime obtains whenever

$$z \leq z^* = \frac{\alpha\hat{\lambda} + (1 - \alpha)\delta}{\alpha\hat{\lambda}(1 - \delta)\theta_G + (1 + \alpha(\hat{\lambda} - 1))\delta\theta_B}.$$

□

Proofs for Section 1.4

Proof of Proposition 5:

Proof. We first outline properties that the solution to the investor's problem must satisfy if it exists. In fact, we will show that the solution to the investor's problem is essentially unique. When an investor is patient, she will never consume because the rate of return on the risk-free technology satisfies $\beta(1 + r) > 1$. On the other hand, when the investor is impatient at t, in order to justify postponing consumption until a later date $\tau > t$ it must be that

$$\max_{\tau} \mathbb{E}_t[\beta^{\tau - t} w_\tau] \geq \lambda w_t.$$

Note first that $\tilde{\lambda}_\tau$ must be equal to λ, since the investor cannot consume at any time such that $\lambda = 1$. Next, define $\zeta = \max\{\theta_G \overline{z}, 1 + r\}$. An investor's expected wealth at

time τ must satisfy $\mathbb{E}_t[w_\tau] \leq \zeta^{\tau-t}$ (since the price of a claim on a project cannot be less than one if an intermediary is able to finance its investment, and the maximum possible payoff of such a claim is $\theta_G \overline{z}$). It is enough to check that the investor is not better off simply by waiting until the next time a liquidity shock arrives. We have

$$\begin{aligned}
\mathbb{E}_t[\beta^{\tau-t} w_\tau] &= \sum_{\tau=t+1}^{\infty} (1-\alpha)^{\tau-t-1} \beta^{\tau-t} \alpha \mathbb{E}_t[w_\tau] \\
&\leq \sum_{\tau=t+1}^{\infty} (1-\alpha)^{\tau-t-1} \beta^{\tau-t} \zeta^{\tau-t} \alpha w_t \\
&= \frac{\alpha}{1-(1-\alpha)\beta\zeta} w_t
\end{aligned}$$

Then, if

$$\frac{\alpha}{1-(1-\alpha)\beta\zeta} < \lambda,$$

it is optimal for the investor to consume immediately.

Next, we show that the investor's problem is well defined. Our previous result showed that the investor consumes at the first time t such that $\tilde{\lambda}_{nt} = \lambda$. Denote this time by $\tau \geq 0$. Then, the value achieved by an investor with initial wealth w_0 is

$$\lambda \mathbb{E}[\beta^\tau w_\tau] = \lambda \sum_{t=0}^{\infty} (1-\alpha)^t \beta^t \alpha \mathbb{E}[w_t | \tau = t].$$

Next, observe that no matter what sequence of states arises, the investor's return on wealth within a period cannot exceed $\zeta \equiv \max\{\theta_G \overline{z}, 1+r\}$ (since the price of a claim on a project cannot be less than one if an intermediary is able to finance its investment, and the maximum possible payoff of such a claim is $\theta_G \overline{z}$). Hence, $\mathbb{E}[w_t | \tau = t] \leq \zeta^t w_0$. We then have

$$\lambda \mathbb{E}[\beta^\tau w_\tau] \leq \left(\sum_{t=0}^{\infty} (1-\alpha)^t \beta^t \zeta^t \right) \alpha \lambda w_0 = \frac{1}{1-(1-\alpha)\beta\zeta} \alpha \lambda w_0,$$

since we have assumed $(1-\alpha)\beta\zeta < 1$ in the statement of the proposition. Thus, the investor's value is finite and her problem is well-defined.

□

Proof of Proposition 6:

Proof. As shown in the proof of the previous proposition, within a period, investors simply try to maximize their returns on wealth. Investors' problems in a period are

therefore identical to those in the static model, with the exception of the fact that the returns on storage are equal to $1+r$ rather than 1. The solutions to their decision problems in the dynamic model are then identical to the solutions to those problems in the static model. Likewise, intermediaries make their decisions so as to maximize returns in the same way as in the static model.

Thus, all decision-makers act in the same way within a period. Plugging in the decisions, prices, and expectations from the static equilibrium, then, we again obtain optimization and market clearing in the dynamic model, so the equilibrium within a period coincides with the symmetric static equilibrium.

□

Proof of Proposition 7:

Proof. Observe that if $\delta_h < \delta^*$, then when $\delta_t = \delta_h$, the economy is in the opaque regime, and the fraction of bad firms in the next period satisfies

$$\begin{aligned}\delta_{t+1} &= (\mu_G(1-\delta_h) + \mu_B^O \delta_h)\overline{\delta} + (1-\mu_B^O)\delta_h \\ &= \delta_h - (\mu_B^O(1-\overline{\delta}) + \mu_G\overline{\delta})\delta_h + \mu_G\overline{\delta} \\ &= \delta_h - \mu_G\overline{\delta} + \mu_G\overline{\delta} = \delta_h\end{aligned}$$

so δ_h is a steady state.

Similarly, if $\delta_l > \delta^*$, whenever $\delta_t = \delta_l$, the economy is in the transparent regime, and we have

$$\begin{aligned}\delta_{t+1} &= (\mu_G(1-\delta_l) + \mu_B^T \delta_l)\overline{\delta} + (1-\mu_B^T)\delta_l \\ &= \delta_l - (\mu_B^T(1-\overline{\delta}) + \mu_G\overline{\delta})\delta_l + \mu_G\overline{\delta} \\ &= \delta_l - \mu_G\overline{\delta} + \mu_G\overline{\delta} = \delta_l\end{aligned}$$

so δ_l is a steady state.

Finally, if $\delta_l < \delta^* < \delta_h$, neither δ_l nor δ_h is a steady state. For $\delta_t \in [\delta_l, \delta^*)$, the law of motion is

$$\delta_{t+1} = \left(1 - (\mu_G(1-\overline{\delta}) + \mu_B^O\overline{\delta})\right)\delta_t + \mu_G\overline{\delta} \equiv \zeta^O\delta_t + \Delta$$

whereas for $\delta_t \in (\delta^*, \overline{\delta}]$, the law of motion is

$$\delta_{t+1} = \left(1 - (\mu_G(1-\overline{\delta}) + \mu_B^T\overline{\delta})\right)\delta_t + \mu_G\overline{\delta} \equiv \zeta^T\delta_t + \Delta.$$

Hence, we get a linear law of motion for δ with $0 < \zeta^T < \zeta^O < 1$. Whenever δ_t is in the opaque regime, $\delta_{t+1} > \delta_t$, and whenever it is in the transparent regime, $\delta_{t+1} < \delta_t$.

In order to proceed, it will be useful to define the following:

$$\hat{\delta}_{T,1} = \frac{\delta^* - \Delta}{\zeta^T},\ \hat{\delta}_{T,n+1} = \frac{\hat{\delta}_{T,n} - \Delta}{\zeta^T},$$

$$\hat{\delta}_{O,1} = \frac{\delta^* - \Delta}{\zeta^O}, \hat{\delta}_{O,n+1} = \frac{\hat{\delta}_{O,n} - \Delta}{\zeta^O}.$$

Note that since $\zeta^T\delta^* + \Delta < \delta^* < \zeta^O\delta^* + \Delta$ by assumption, it must be the case that $\hat{\delta}_{O,1} < \delta^* < \hat{\delta}_{T,1}$, and by the continuity of the law of motion on either side of δ^*, $\hat{\delta}_{O,n}$ is decreasing in n and $\hat{\delta}_{T,n}$ is increasing in n. We will define the intervals $I_{O,n} = [\hat{\delta}_{O,n}, \hat{\delta}_{O,n-1}]$ and $I_{T,n} = [\hat{\delta}_{T,n-1}, \hat{\delta}_{T,n}]$ for $n \geq 1$, where it is understood that $\hat{\delta}_{O,0} = \hat{\delta}_{T,0} = \delta^*$.

By definition, whenever $\delta_t \in I_{O,n}$ for $n > 1$, then $\delta_{t+1} \in I_{O,n-1}$. Furthermore, $\delta_{t+1} > \delta^*$ when $\delta_t \in I_{O,1}$. Similarly, whenever $\delta_t \in I_{T,n}$ for $n > 1$, then $\delta_{t+1} \in I_{T,n-1}$, and when $\delta_t \in I_{T,1}$, we have $\delta_{t+1} < \delta^*$.

We first look for a cycle that transits the opaque regime for only one period. In order for this to occur, it must be that for all $\delta > \delta^*$, $\zeta^T\delta + \Delta \in I_{O,1}$. We will show that there exists a cycle with an initial point $\delta_0 \in I_{O,1}$. As noted previously, $\delta^+(\delta_0) > \delta^*$. There must be some finite number $n(\delta_0) + 1$ such that $\delta^{+(n(\delta_0)+1)}(\delta_0) \in I_{O,1}$, then, since if $\delta^+(\delta_0) \in I_{T,n}$ for some n, then, $\delta^{+k}(\delta_0) \in I_{T,n(\delta_0)-k+1}$ for $k \leq n(\delta_0)+1$. Denote $G(\delta) = \delta^{+(n(\delta)+1)}$ for $\delta \in I_{O,1}$. We have that $G(\delta)$ is continuous in δ because $\delta^+(\delta_0)$ is continuous and the law of motion $\delta^+(\cdot)$ is continuous in the region $\delta > \delta^*$. Then observe that $G(\delta^*)$ must be less than δ^*, since it lies in the interval $I_{O,1}$, and by the same token $G(\hat{\delta}_{O,1}) > \hat{\delta}_{O,1}$. There must then exist some δ_0^{**} such that $G(\delta_0^{**}) = \delta_0^{**}$. The sequence $\{\delta_0^{**}, \delta^+(\delta_0^{**}), \ldots, \delta^{+k}(\delta_0^{**}), \ldots, \delta^{+n(\delta_0)}(\delta_0^{**}), \delta_0^{**}\}$ is then an equilibrium cycle of length $n(\delta_0^{**})$.

I omit the proof, but by exactly analogous reasoning it follows that whenever $\delta^+(\delta^*) \in I_{T,1}$, then there is a cycle that transits the transparent regime for one period and remains in the opaque regime in all other periods.

Observe that as z increases, δ^* increases as well, but δ_h and δ_l remain fixed. In particular, δ^* is an increasing function of z, so there is $\underline{z}$ such that $\delta_l = \delta^*$ and $\overline{z}$ such that $\delta_h = \delta^*$. These threshold values of z satisfy the conditions required by the proposition.

□

Proofs for Section 1.5

Proof of Proposition 8

Proof. When the planner chooses an investment policy $x(\tilde{\theta}|F)$, the relevant tradeoff is investment in the project versus investment in the risk-free technology. This is because impatient investors consume all of their funds, and patient investors must save all of their funds, so there is no tradeoff between investment in a particular project and consumption. Then it is optimal to finance a project only if its expected return (from the planner's perspective) is at least $1 + r$.

The expected returns on a project given the planner's information $\tilde{\theta}$ are just $\tilde{\theta}z$. It is therefore optimal to fully finance a project whenever $\tilde{\theta}z \geq 1 + r$, so

$$x(\tilde{\theta}|F) = \begin{cases} 1 & \tilde{\theta}z > 1 + r \\ \in [0, 1] & \tilde{\theta}z = 1 + r \\ 0 & \tilde{\theta}z < 1 + r \end{cases}$$

Furthermore, observe that the lowest possible value that $\tilde{\theta}$ can take is $\tilde{\theta}_{\min} \equiv \alpha\overline{\theta} + (1 - \alpha)\mathbb{E}_F[\theta|\hat{\theta}z < 1 + r]$.

There are two possibilities. First, assume that F is such that $\tilde{\theta}_{\min}z < 1 + r$. Then the project is not financed whenever $\hat{\theta}z < 1 + r$ or whenever skilled investors are impatient. The expected returns on the project are

$$\mathbb{E}[\theta z \cdot x(\tilde{\theta}|F)] = (1 - \alpha) \int_{\frac{1+r}{z}}^{\theta_G} \hat{\theta}z dF(\hat{\theta}) \leq (1 - \alpha)(1 - \delta)\theta_G z.$$

The probability that a bad firm survives is

$$\mathbb{E}[\Pr(\theta = \theta_B)\mathbf{1}\{\hat{\theta} \geq \frac{1 + r}{z})\}] = \int_{\frac{1+r}{z}}^{\theta_G} \frac{\theta_G - \hat{\theta}}{\theta_G - \theta_B} dF(\hat{\theta}) \geq 0.$$

Notice that it is possible to earn returns $(1 - \alpha)(1 - \delta)\theta_G z$ under full transparency while ensuring that no bad firm survives. Therefore, an information structure of full transparency dominates any information structure F such that it is optimal to finance the project when skilled investors do not choose to do so.

Second, it may be that F is such that $\tilde{\theta}_{\min}z \geq 1 + r$. In this case, the project is financed regardless of its quality. Thus, a policy of opacity must at least weakly

dominate such an information structure F. Along with the previous result, this observation implies that we may restrict attention to policies in which the planner chooses only fully opaque or fully transparent projects.

□

Proof of Proposition 9:

Proof. Figure 1.5 shows by example that cycles can emerge as an outcome under the planner's solution. If the constrained optimum features a cycle, there exists $\delta^{*,\text{sp}}$ such that the planner chooses opaque projects only for $\delta < \delta^{*,\text{sp}}$ and transparent projects only for $\delta \geq \delta^{*,\text{sp}}$.

First, we define a fictitious planner's problem in which the planner does not take into account the evolution of δ, and we prove that the outcome under the solution to this problem coincides with the equilibrium outcome. Then, we show that since the planner takes into account the evolution of δ, it must be optimal to switch to the transparent regime for a lower value of δ^* than under the fictitious planner's problem.

In recursive form, the fictitious planner's problem for an exogenous threshold δ^* is

$$\begin{aligned} V(k,\delta) = \max_{m^O} \ & \lambda\alpha k(1+\omega m^O(\overline{\theta}z-(1+r))) + \beta V(k',\delta') \text{ s.t. } \delta' = \zeta(\delta|\delta^*)\delta + \Delta, \\ & k' = (1+r)(1-\alpha(1+\omega m^O(\overline{\theta}z-(1+r))))k \\ & + \omega(m^O(\overline{\theta}z-(1+r)) + (1-m^O)(1-\alpha)(\theta_G z-(1+r)))k \end{aligned}$$

where $\zeta(\delta|\delta^*) = \zeta_T$ if $\delta \geq \delta^*$ and $\zeta(\delta|\delta^*) = \zeta_O$ otherwise.

□

Proof of Proposition 10:

Proof. I outline how each of the three types of policy tools is formally represented in the model and then prove that each can achieve the constrained optimum when the planner's optimum and the equilibrium both feature a cycle.

Transparency regulation: Under a transparency regulation, the planner can directly ban the creation of opaque, liquid assets. That is, the planner can specify regions of the state space $\delta \in [0,1]$ such that creating opaque assets is not allowed. If the planner bans the production of opaque assets for all $\delta \geq \delta^{*,\text{sp}}$, then the constrained optimum is implemented. The optimal policy must fluctuate over time, since there are states in which no transparency regulation is needed and others in which it is.

Macroprudential policy: Macroprudential policy corresponds to the imposition of a tax $\tau(\delta)$ on the origination of opaque assets (that is, a tax on the profits of intermediaries that produce such assets). I will show that a flat tax τ is sufficient to implement the constrained optimum.

With a tax τ, an intermediary's value under opacity is

$$v^O(\delta, \tau) = (1 - \tau)(1 + \alpha(\lambda - 1))(\overline{\theta}z - (1 + r)),$$

whereas its value under transparency is

$$v^T(\delta) = (1 - \alpha)(\theta_G z - (1 + r)).$$

The threshold $\delta^*(\tau)$ at which the economy transitions from the opaque regime to the transparent one satisfies $v^O(\delta, \tau) = v^T(\delta)$. For $\tau = 1$, an opaque intermediary's profits are equal to zero, so $\delta^*(1) = 0$. For $\tau = 0$, the equilibrium outcome arises. Since $\delta^*(\tau)$ is continuous, there must exists some $\tau^* \in (0, 1)$ such that $\delta^*(\tau^*) = \delta^{*,\text{sp}} < \delta^{*,\text{eqm}}$.

Monetary policy: Monetary policy is a subsidy to investment in the risk-free technology, so that investors perceive a return $\tilde{r}(\delta) > r$. Again, I show that it is possible to implement the constrained optimum under a fixed interest rate $\tilde{r}(\delta) = \tilde{r}$. In this case, the intermediary's value under opacity is

$$v^O(\delta, \tilde{r}) = (1 + \alpha(\lambda - 1))(\overline{\theta}z - (1 + \tilde{r}))$$

and its value under transparency is

$$v^T(\delta, \tilde{r}) = (1 - \alpha)(\theta_G z - (1 + \tilde{r})).$$

As shown previously, the threshold value $\delta^*(\tilde{r})$ at which the economy transitions from the opaque regime to the transparent regime satisfies

$$\delta^*(\tilde{r}) = \frac{\alpha\lambda(\theta_G z - (1 + \tilde{r}))}{\alpha\lambda(\theta_G z - (1 + \tilde{r})) + (1 + \alpha(\lambda - 1))(1 + \tilde{r} - \theta_B z)}.$$

Observe that the right-hand side is continuous, decreasing in $\tilde{r}$, and can become arbitrarily negative. Thus, there must exist some $\tilde{r}^*$ such that $\delta^*(\tilde{r}) = \delta^{*,\text{sp}} < \delta^{*,\text{eqm}}$.

□

Chapter 2

A Model of Collateralized Lending Chains

2.1 Introduction

Recent innovations have led to the use of collateral to secure promises among multiple parties in complex financial arrangements. For example, a collateralized mortgage obligation (CMO) is a debt contract backed by mortgages, which in turn are debt contracts backed by houses. This type of arrangement is termed "pyramiding" by Geanakoplos (1997). In repo markets, lenders can finance themselves by rehypothecating the borrowed security, thereby spreading the collateral further. Finally, tranched securities allow for the collateral to be split between multiple parties at once: the owners of a given tranche are entitled to the residual cash flows from the collateral whenever the tranche begins to take losses. During the financial crisis of 2007-2009, margins on collateralized lending rose dramatically and markets involving these types of complex arrangements collapsed. In order to address the emergence of these arrangements and the the consequences for financial stability, this paper studies the implications of collateralized lending among multiple parties.

I present a model of lending chains in which an asset is sold to N types of agents who are permitted to use the asset or debt contracts as collateral. Gains from trade stem from differences in beliefs, as in Geanakoplos (2010) or Simsek (2013). In this model, the most optimistic agents always purchase the asset, and other agents engage in a chain of borrowing and lending: each type lends to the next most optimistic type and borrows from the next most pessimistic type, using their long position in debt contracts as collateral. This financing structure leads to an endogenous tranching

of the collateral. Agents walk away from their obligations whenever the value of the assets used to collateralize a loan falls below the repayment due on that loan. When an agent's payoffs are wiped out, the asset is passed on to the next agent in the lending chain. In fact, in this model lending chains lead to the exact same outcome as direct tranching of the asset, which suggests that heterogeneity may lead to tranched payoffs independently of the precise institutional details of financial markets.

The first contribution of this work is to outline how belief heterogeneity can lead to endogenous intermediation. In the model, optimists receive payoffs only in good states because they are relatively more optimistic about those states. Optimists therefore take risky loans from other agents. Intermediate types are relatively more optimistic about intermediate states, so to receive payoffs in these states, they must default only in bad states. To accomplish this, these types act as intermediaries: they lend to optimistic types to secure safe payoffs in good states and fund themselves by borrowing from pessimistic agents who demand relatively safe assets. As discussed in the literature review (Section 1.1), belief heterogeneity is a novel explanation of intermediation and is to my knowledge the only rationalization of *endogenous* chains of lending in the literature.

In order to understand the role of intermediation in determining margins and asset prices simultaneously, it is critical to have a model in which intermediation is endogenous. I study these questions by characterizing the dependence of leverage, safe asset issuance, and asset prices on intermediaries' net worth and the length of lending chains. Long intermediation chains result in the alignment of agents' portfolios with their risk preferences. When an intermediary's capitalization increases, all downstream borrowers are able to lever up and take riskier loans at lower interest rates. By contrast, all upstream lenders benefit from the additional cash buffer provided by the intermediary and make safer loans. The role of intermediation is thus to simultaneously increase leverage and create safe assets. This model illustrates how the rise of leverage and the supply of safe assets and the subsequent fall in these quantities after the most recent financial crisis can be seen as a reduction in the financial sector's intermediation capacity.

Agents' ability to reuse debt contracts as collateral and create arbitrarily long lending chains in the model yields additional theoretical insights. I show that the reusability of collateral reduces the problem of pricing the underlying asset to pricing state-specific contracts that pay off 1 if the asset's payoff (the realized state of nature s) is above a certain cutoff $\hat{s}$ and zero otherwise. Agents can produce these securities by lending against a certain promised repayment $\hat{s}$ and using that contract

as collateral to promise a slightly smaller repayment $\hat{s} - \epsilon$ to a lender. The fact that the asset's payoff can be deconstructed in this way makes my setting analogous to an Arrow-Debreu economy in which the asset's price is just a sum of the prices of assets that pay off 1 in each possible future state. In my model, the analogue of the Arrow-Debreu security that pays 1 in state s is the portfolio that pays 1 in all states greater than s. The link between re-pledgeability of collateral and the decomposition of the underlying asset's price into the prices of state-specific securities is also new to the literature.

I apply and extend the model to study several features of markets that feature chains of collateralized lending. First, I ask whether the model can produce contagion in which margins spike in several markets simultaneously. Geanakoplos (2010) faults a "double leverage cycle", in which lending conditions in housing markets and repo markets for mortgage-backed securities fed back on each other, for much of the turmoil in financial markets during the crisis of 2007-2009. Gorton and Metrick (2012) show that margins in the bilateral repo market spiked in essentially all risky asset classes. The model of collateralized lending chains can generate both contagion *along* the chain of lending (i.e., double leverage cycles) and *across* asset classes, even when the payoffs of those assets are uncorrelated. The simple intuition behind my results is that when bad news scares some agents and tightens lending conditions in one market, intermediaries become constrained in their ability to borrow and therefore lend less against any security. This explanation of contagion in margins as a pure wealth effect is more parsimonious than others offered in the literature, as discussed in the literature review.

Second, I show that an extension of the model with an investment technology can produce cycles in which leverage, safe asset production, asset prices, and investment all comove positively with intermediary net worth. Empirically, Fostel and Geanakoplos (2016) document that leverage and investment in housing markets tracked real estate prices during the period 2000-2009, and Lenel (2017) shows that production of private-label long-maturity safe assets reached record levels before the financial crisis and then collapsed after 2008. My model rationalizes these trends as an increase and subsequent decrease in the intermediary (or shadow banking) sector's ability to act as a cash buffer and protect risk-averse investors from crashes in mortgage-backed security prices while providing end-borrowers with low-margin loans. I present a three-period version of the model in which the intermediary's net worth fluctuates over time and show using numerical examples that even when intermediaries see the potential of a great buying opportunity after a crash, they still expose themselves to

aggregate risk, so the empirical patterns listed above are replicated in the dynamic version of the model.

Third, I present an alternative model in which debt contracts cannot be used as collateral in order to understand how the transition to a world in which long chains of lending become possible affects risk-taking. I show that this specific type of financial innovation either increases the risk taken by agents who become intermediary lenders or increases the risk taken by end-borrowers who benefit from the additional leverage provided by levered intermediaries. Volatility in the economy with debt contracts as collateral is much higher because lenders can lever up and go bankrupt precisely in the states in which intermediation is needed most.

The paper is organized as follows. The rest of this section is dedicated to a review of the relevant literature. Section 2 presents the model and defines an equilibrium. In Section 3, I characterize and describe the properties of equilibria (including the endogenous tranched payoff structure). Section 4 studies the consequences of intermediation in detail. Section 5 discusses several applications of the model. I discuss the model's assumptions in Section 6. Section 7 concludes.

2.1.1 Literature Review

This paper is related to the literature on collateralized lending with heterogeneous beliefs, as in Geanakoplos (2003, 2009, 2010) and Simsek (2013). In particular, the assumptions on beliefs and the structure of the economy are very similar to those in Simsek (2013). This model departs from the models in those works in that it allows for the use of debt contracts as collateral. This assumption allows agents to act as intermediaries and gives rise to the endogenous multi-level tranching of the asset. Without allowing for the reuse of collateral, it would be impossible to analyze how intermediaries' beliefs and wealth affect asset prices and leverage. My results regarding how beliefs affect the equilibrium price are similar to those in Geanakoplos (2003, 2010) and Simsek (2013) in that I find increased disagreement can increase margins, thereby lowering the asset price. In this model, however, there is more than one margin because there is a chain of lending rather than a single debt contract, so my result has additional qualifications. Geerolf (2017) also shows that chains of lending can emerge in equilibrium, but that paper relies on dogmatic beliefs under which agents are completely certain that a given asset payoff will be realized. My paper is the first to illuminate the nature of endogenous chains of lending and show

that the allocation with reusable collateral coincides with that in a dual Arrow-Debreu economy.

This paper is also a part of the theoretical literature on collateralized lending, securitization, and lending chains. Bottazzi, Luque, and Páscoa (2012) show the existence of equilibrium in an economy with rehypothecation of collateral and the liquidity premium associated with pledgeable assets. Muley (2016) studies the optimality of rehypothecation and securitization in an environment with limited commitment. Dang, Gorton, and Hölmstrom (2013) demonstrate that haircuts arise in order to solve an adverse selection problem faced by lenders and study how haircuts affect credit in lending chains. Other recent papers, such as Di Maggio and Tahbaz-Salehi (2015), Infante (2015), and Kahn and Park (2015) attempt to understand the implications of collateralized lending in a setting where credit relationships are exogenously given. In my model, by contrast, lending chains form endogenously as agents attempt to align their portfolios with their appetites for risk.

This literature on asset pricing with margin constraints is relevant to my paper as well. For example, Garleanu and Pedersen (2011) and Rytchkov (2014) model economies with credit constraints that affect asset prices. Unlike those papers, in my model credit constraints are endogenous and come from beliefs rather than being given exogenously or resulting from limited commitment. Brunnermeier and Pedersen (2009) also examine asset markets with endogenous margins and concludes, as I do, that spikes in margins can be contagious across asset classes, but their explanation of how margins are set in equilibrium is more involved and contagion does not reduce to a simple wealth effect, as in my model.

There is a large body of empirical work related to the recent financial crisis and the collapse of repo (i.e., collateralized lending) markets. Gorton and Metrick (2012) provides a timeline of the financial crisis and examines the increase in margins on asset-backed securities during that period. Krishnamurthy, Nagel, and Orlov (2014) argue that the run on aggregate repo borrowing was small but concentrated on a handful of systematically important borrowers. Copeland, Martin, and Walker (2014) provide some evidence from the triparty repo market to support this hypothesis. Shin (2009) studies the effects of deleveraging in a complex, interconnected financial system. The consequences of a reduction in "collateral velocity" are examined in Singh (2011). Reinhart and Rogoff (2008) emphasize the role of mistaken beliefs in serious financial crises.

2.2 Model

2.2.1 Environment

The economy exists for two periods, $t = 0$ and $t = 1$. There is a single consumption good, which is referred to as a dollar. There is a continuum of each of N types of agents who invest their endowments at $t = 0$ in order to consume at $t = 1$ and a continuum of agents who sell an asset a in order to consume at $t = 0$. The only source of uncertainty is the terminal payoff of the asset: there is a continuum of states $s \in [0, \overline{s}] = S$ that index the payoff of the asset in terms of the consumption good at $t = 1$.

Agents differ in their degree of optimism about the asset's payoff, meaning each type n has a different subjective probability measure on $[0, \overline{s}]$ defined by a density function $f_n(s)$. The following assumption formalizes the notion of optimism:

Assumption 1. *The types of agents $n \in \{1, 2, \ldots, N\}$ are ordered by their optimism, with type $n = 1$ being the most optimistic and $n = N$ being the most pessimistic, in the sense that the hazard rate inequality*

$$\frac{f_n(s)}{1 - F_n(s)} < \frac{f_{n+1}(s)}{1 - F_{n+1}(s)}$$

is satisfied for $n = 1, \ldots, N - 1$ (where F_n is the CDF corresponding to f_n).

This property implies that F_n first-order stochastically dominates F_{n+1}, but it is weaker than the monotone likelihood ratio property.

All agents invest in order to maximize expected consumption $E_n[c]$ at $t = 1$. Agents of type n can invest their endowments w_n at $t = 0$ in one of three ways: they may purchase the asset, lend to other agents, or hold cash, which yields a safe return equal to one dollar at $t = 1$. In this economy agents are permitted to default, so borrowers must use collateral in order to provide agents an incentive to lend. That is, in order to agree to pay a lender θ units of the consumption good at $t = 1$, a borrower must set aside one unit of some asset (which could be the asset a or a debt contract) as collateral in case of default. In principle, the amount of money raised by the borrower could depend on the collateral.

Let

$$A : \{\phi(s) : \phi : S \to \mathbb{R}_+ \textit{ measurable and bounded}\}$$

be the set of assets. Formally, a debt contract is defined as a pair (θ, ϕ), where θ is the amount the borrower agrees to repay at $t = 1$ and $\phi \in A$ is the asset used

as collateral. The debt contract is itself an asset with payoff $\min\{\theta, \phi(s)\}$, since the borrower chooses to default whenever the collateral's payoff is less than the amount owed. In this setting, the space of assets is restricted to $\tilde{A} = \bigcup_k A^k$, where the sets A^k are defined inductively as

$$A^0 = a, \;\; A^k = \{\min\{\theta, \phi(s)\} : \theta \in \mathbb{R}_+, \phi \in A^{k-1}\}$$

The set A^1 consists of debt contracts that use a unit of the asset a as collateral. The set A^2 consists of debt contracts that use debt contracts in A^1 as collateral, and so forth. The set $\tilde{A}$ is therefore the set of all assets that can be constructed using only a and collateralized debt contracts.

Proposition 2.1. *The set of assets is* $\tilde{A} = A^0 \cup A^1$.

All proofs are relegated to the Appendix. This proposition implies that no matter what collateral is used, the payoff of a debt contract in which the borrower agrees to pay θ at $t = 1$ is $\min\{\theta, s\}$ (since the lender would never accept collateral that pays less than θ in every state). The space of assets therefore consists of functions of the type $\min\{\theta, s\}$ with $\theta \in S$. In equilibrium, then, the amount of money a borrower can raise by promising θ at $t = 1$ will be independent of the asset used as collateral so long as that asset does not deliver less than θ in every state. Such a contract will henceforth be referred to as a loan of riskiness θ.

Note that whenever a borrower defaults on a loan of riskiness θ in a state $s < \theta$, the lender's payoff at $t = 1$ is s. It is as if lenders receive the asset itself in the event of default even if the collateral was a loan of riskiness $\theta' > \theta$. Thus there is a sense in which using a debt contract as collateral is equivalent to rehypothecating the asset, so in this model lending chains resemble an arrangement in which lenders take possession of collateral and are permitted to reuse it, as in a repo market.

2.2.2 Optimization Problem

Let p be the price of the asset and $q(\theta)$ be the amount borrowed at $t = 0$ when taking a loan of riskiness θ. The problem of an agent of type n is

$$\max_{a,\mu_+,\mu_-,c} aE_n[s] + c + \int E_n[\min\{\theta, s\}]d\mu_+(\theta) - \int E_n[\min\{\theta, s\}]d\mu_-(\theta)$$
$$s.t.\ w_n = pa + c + \int q(\theta)d\mu_+(\theta) - \int q(\theta)d\mu_-(\theta)$$
$$a + \int_{\theta\geq\hat{\theta}} d\mu_+(\theta) \geq \int_{\theta\geq\hat{\theta}} d\mu_-(\theta)\ \forall\hat{\theta} \in S,\ \mu_+(\theta) \geq 0,\ \mu_-(\theta) \geq 0$$

That is, agents choose the quantity a of the asset to purchase, measures $\mu_+(\theta)$ and $\mu_-(\theta)$ corresponding to the amount of lending and borrowing they choose for each $\theta \in S$, respectively, and an amount c of cash to hold. Their choices are subject to their budget constraints (the second line) and collateral constraints (the third line). Every contract in which the agent agrees to pay θ at $t = 1$ must be backed by one unit of the asset or a lending contract in which the agent receives $\theta' \geq \theta$. Absence of arbitrage will imply $q(\overline{s}) \leq p$.

Consider the alternative formulation

$$\max_{\mu,c} \int_{\theta'\leq\theta} E_n[\min\{\theta, s\} - \min\{\theta', s\}]d\mu(\theta, \theta') + c$$
$$s.t.\ w_n = \int_{\theta'\leq\theta} (q(\theta) - q(\theta'))d\mu(\theta, \theta') + c,\ \ \mu(\theta, \theta') \geq 0$$

It is possible to show the constraint set in this problem is the same as that in the original problem (after a change of variables). The two problems are thus equivalent. Henceforth this formulation will be used for analytical tractability. Note that in this problem, agents choose a measure μ defined on the space

$$\hat{A} = \{\phi_1(s) - \phi_2(s) : \phi_1(s), \phi_2(s) \in \tilde{A}, \phi_2 \leq \phi_1\} = \{\min\{\theta, s\} - \min\{\theta', s\} : \theta' \leq \theta\}$$

In this problem, agents choose $\mu(\theta, \theta')$, which represents lending funded by borrowing in which the agent receives θ and pays θ' at $t = 1$ or asset purchases funded by borrowing when $\theta = \overline{s}$. This problem is isomorphic to a portfolio choice problem in which agents purchase assets (θ, θ') with expected payoffs $\pi(\theta, \theta') = E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ and prices $\tilde{q}(\theta, \theta') = q(\theta) - q(\theta')$. Clearly, due to the linearity of the

objective function, collateralizing a loan of riskiness θ' with a loan of riskiness θ is optimal for type n agents only if

$$\frac{E_n[\min\{\theta, s\} - \min\{\theta', s\}]}{q(\theta) - q(\theta')} \geq \max\left\{1, \max_{(\tilde{\theta},\tilde{\theta}')} \frac{E_n[\min\{\tilde{\theta}, s\} - \min\{\tilde{\theta}', s\}]}{q(\tilde{\theta}) - q(\tilde{\theta}')}\right\} \equiv r_n$$

Let $\Theta^* = \{(\theta, \theta') : \frac{E_n[\min\{\theta,s\}-\min\{\theta',s\}]}{q(\theta)-q(\theta')} = r_n\}$. A choice of measure μ is optimal if and only if $\mu(\hat{A}\backslash\Theta^*) = 0$.

2.2.3 Equilibrium

In equilibrium, agents will maximize their objective functions and asset markets will clear. The following definition formalizes the notion of general equilibrium in this economy:

Definition 2.2. *A general equilibrium of this economy consists of measures* $(\mu_1, \mu_2, \ldots, \mu_N)$*, cash holdings* $(c_1, \ldots, c_N)$ *and prices* $q(\theta) : S \to \mathbb{R}_+$ *such that*

- *Taking* $q(\theta)$ *as given,* (μ_n, c_n) *solves agent* n*'s optimization problem for each* $n \in \{1, 2, \ldots, N\}$*.*
- *Debt markets and the market for the asset a clear, meaning*

$$\sum_{i=1}^{N} \int_{\theta' \leq \theta} (\min\{\theta, s\} - \min\{\theta', s\}) d\mu_n(\theta, \theta') = s \; \forall s \in S$$

The market clearing condition merits further explanation. The condition says that state by state, the sum of agents' payoffs resulting from borrowing, lending, and direct purchases of the asset is equal to the asset's payoff. This embeds the assumption that agents who consume at $t = 0$ supply one unit of the asset inelastically and the observation that in equilibrium, if debt markets clear, the amount of gross debt outstanding has no effect on the sum of agents' payoffs. One agent's payoff from lending at $t = 1$ is another agent's repayment.

2.3 Properties of Equilibrium

Given the need to compute the price function $q(\theta)$ for each $\theta \in S$, it may seem a daunting task to find the set of equilibria. In this section it will be shown that equilibrium allocations correspond to constrained social optima. The solution to the

social planner's problem will shed light on the form of the price function $q(\theta)$ and the properties of equilibrium allocations.

2.3.1 Social Planner's Problem

Consider the problem of a benevolent social planner who partitions asset a's payoff among N agents using assets with payoffs of the form $E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ and places weight λ_n on the subjective *ex-ante* utility of type n agents. The planner's problem is

$$\max_{\mu_1,\ldots,\mu_N} \sum_{n=1}^{N} \lambda_n \int_{\theta' \leq \theta} E_n[\min\{\theta, s\} - \min\{\theta', s\}] d\mu_n(\theta, \theta')$$

$$s.t. \sum_{n=1}^{N} \int_{\theta' \leq \theta} (\min\{\theta, s\} - \min\{\theta', s\}) d\mu_n(\theta, \theta') = s \; \forall s \in S, \; \mu_n \geq 0$$

Note that the constraints on the measures μ_n are exactly the general equilibrium market clearing conditions.

The social planner's problem is thus to optimally assign assets to agents in such a way that the asset market clearing conditions are replicated. However, in the social planner's problem the constraints should not be interpreted as having anything to do with market clearing, as there is no sense in which agents borrow from or lend to each other. Rather, these conditions guarantee that the solution represents some partition of the payoff of one unit of the asset a into payoffs of the form $\min\{\theta, s\} - \min\{\theta', s\}$. These payoffs already resemble those of tranched securities in that each such asset's payoff is constant over $s \geq \theta$, declines over $\theta > s \geq \theta'$, and is zero below θ'. It will soon be shown that each agent's payoff will also take this form at an optimum.

It is optimal for the social planner to assign an asset (θ, θ') to agent n only if

$$\lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] = \max_m \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$$

In order to determine when this condition holds, it will be useful to determine for which pairs (θ, θ') the inequality $\lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] > \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$ holds (for fixed $n < m$). The following lemma definitively answers this question.

Lemma 3.1. *Fix $n < m$. The following three statements hold:*

1. *There exists a weakly decreasing function* $\overline{g}_{nm}(\theta) : S \to S$ *such that* $\lambda_n E_n[\min\{\theta', s\} - \min\{\theta, s\}] > \lambda_m E_m[\min\{\theta', s\} - \min\{\theta, s\}]$ *iff* $\theta' > \overline{g}_{nm}(\theta)$.

2. *There exists a weakly decreasing function* $\underline{g}_{nm}(\theta)$ *such that* $\lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] > \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$ *iff* $\theta' > \underline{g}_{nm}(\theta)$.

3. *There exists* $\theta^*_{nm} \in [0, \overline{s}]$ *such that* $\overline{g}_{nm}(\theta) = \theta$ *if* $\theta > \theta^*_{nm}$ *and* $\underline{g}_{nm}(\theta) = \theta$ *if* $\theta < \theta^*_{nm}$.

Let $\theta^*_N = 0$, $\theta^*_1 = \max_{m>1} \theta^*_{1m}$, and inductively define $\theta^*_n = \min\{\max_{m>n} \theta^*_{nm}, \theta^*_{n-1}\}$ for $2 \leq n \leq N-1$. Then $\theta^*_N \leq \theta^*_{N-1} \leq \cdots \leq \theta^*_1$. Lemma 3.1 implies that

$$\theta^*_n \leq \theta' \leq \theta \leq \theta^*_{n-1} \Rightarrow \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] = \max_m \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$$

Define $\Theta^*_n = \{(\theta, \theta') : [\theta', \theta] \subset [\theta^*_n, \theta^*_{n-1}]\}$. It is always suboptimal for the social planner to choose a μ_n such that $\mu_n(\hat{A} \backslash \Theta^*_n) > 0$. For example, for any feasible allocation such that $\mu_n((\theta^*_{n-2}, \theta^*_n)) = c > 0$, the social planner could do strictly better by setting $\mu_n((\theta^*_{n-1}, \theta^*_n)) = \mu_{n-1}((\theta^*_{n-2}, \theta^*_{n-1})) = c$, and the constraints would still be satisfied. Intuitively, the social planner's constraints allow assets (θ, θ') to be split up into any arbitrary collection of sub-assets, so the social planner always chooses to allocate assets (θ, θ') in such a way that each type n holds assets with $[\theta', \theta] \subset [\theta^*_n, \theta^*_{n-1}]$. It is precisely those assets that type n values most relative to other agents.

The social optimum is achieved by the allocation $\mu_n((\theta^*_{n-1}, \theta^*_n)) = 1, \mu_n(\theta, \theta') = 0$ for all other pairs (θ, θ'). Up to trivial changes in the measures μ_n, there is no other feasible allocation such that $\mu_n(\hat{A} \backslash \Theta^*_n) = 0$ for all n. In such an allocation, agents with $\theta^*_n = \theta^*_{n-1}$ receive nothing, so it is convenient to enumerate the agents who receive nonzero asset holdings by setting $k_1 = \operatorname{argmax}_n(\theta^*_n)(\mathbf{1}\{\theta^*_n < \overline{s}\})$, $k_j = \operatorname{argmax}_n(\theta^*_n)(\mathbf{1}\{\theta^*_n < \theta^*_{k_{j-1}}\})$. This allocation resembles the payoff structure of a tranched security. Agents of type k_1 receive payoffs that start falling in value immediately as s decreases below $\overline{s}$ and reach zero at $\theta^*_{k_1}$. Type k_j's payoffs are constant over $[\theta^*_{k_{j-1}}, \overline{s}]$ and fall to zero as s goes from $\theta^*_{k_{j-1}}$ to $\theta^*_{k_j}$.

Why should this be the constrained social optimum? At the margin, the planner decides whether to reduce the threshold $\underline{\theta}_n$ at which agent n's payoffs reach zero. An reduction in $\underline{\theta}_n$ reduces type n's perceived payoff by $1 - F_n(\underline{\theta}_n)$ since this reduction effectively gives type n an extra dollar in each state $s \geq \underline{\theta}_n$. Similarly, this reduction decreases type $n+1$'s payoff by $1 - F_{n+1}(\underline{\theta}_n)$, so the planner's decision depends on (1) the probabilities assigned to the tail event $s \geq \underline{\theta}_n$ by each type, and (2) the weights

λ_n, λ_{n+1}. Given that n is more optimistic than $n+1$, even if $\lambda_{n+1} > \lambda_n$ there will be some $\theta^*_{n,n+1}$ such that the relative weight assigned to the tail event by type n outweighs the larger weight placed by the social planner on type $n+1$.

In general, this tranched allocation will not be an unconstrained social optimum in the following sense: if the planner were given access to a complete contingent set of securities rather than assets with payoffs $\min\{\theta, s\} - \min\{\theta', s\}$, the allocation would in general differ from the one obtained above. In fact, when agents' subjective probabilities satisfy the monotone likelihood ratio property (rather than the weaker hazard rate inequality), it can be shown that the planner chooses to give each type n the entire payoff of the asset over some range $s \in [\hat{\theta}_n, \hat{\theta}_{n-1}]$ and nothing in all other states. The optimality of tranching is therefore specific to an incomplete markets setting in which assets' payoffs resemble those that arise with collateralized lending.

2.3.2 Tranched Payoffs in Equilibrium

I now show that (1) the allocation that arises in any equilibrium corresponds to the allocation in some social optimum and (2) the equilibrium in this market exists and is unique.

Proposition 3.2. *Suppose $q(\theta), \{\mu_n\}_{n=1}^N, \{c_n\}_{n=1}^N$ is a competitive equilibrium. Then there exist weights $\{\lambda_n\}_{n=1}^N$ such that the payoffs of the constrained social optimum are achieved by $\{\mu_n\}_{n=1}^N$. Furthermore, there exist $\{\theta^*_n\}_{n=1}^N$ such that $q(\theta) - q(\theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ whenever $[\theta', \theta] \subset [\theta^*_n, \theta^*_{n-1}]$ (where $\theta^*_0 \equiv \overline{s}$).*

Conversely, there is a unique mapping from the set $\Lambda = \{\{\lambda_n\}_{n=1}^N : \lambda_1 \leq \cdots \leq \lambda_N\}$ to the set of equilibrium allocations such that the resulting allocation coincides with the solution to the social planner's problem with weights λ.

Given that every equilibrium corresponds to a constrained social optimum, the previous characterization of constrained social optima carries through as a characterization of equilibrium allocations. That is, in any equilibrium, there are values $\{\theta^*_n\}_{n=1}^N$ such that $0 = \theta^*_N \leq \theta^*_{N-1} \leq \cdots \leq \theta^*_1 \leq \overline{s}$ and each type n owns assets with a total payoff of $\min\{\theta^*_{n-1}, s\} - \min\{\theta^*_n, s\}$. Tranched payoffs therefore always emerge in equilibrium: the most optimistic agent suffers losses whenever the asset's value falls below its maximum possible value, but the most pessimistic agent does not take losses until the asset's value falls below θ^*_{N-1}. These payoffs resemble those of CDO tranches. An example of the payoff structure is illustrated in Figure 1.

The mechanism that causes these payoffs to arise, however, is conceptually different. In equilibrium, type 1 buys the asset and uses it as collateral to borrow from

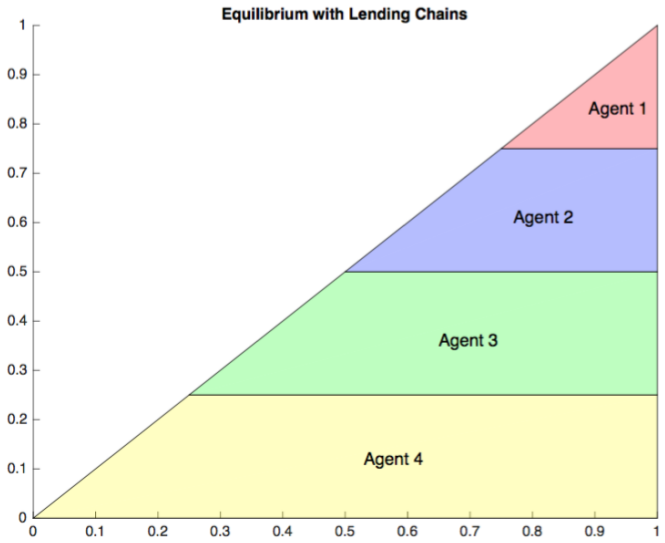


Figure 2.1: An example of the tranched payoff structure described in Propositions 3.2 and 3.3. Each agent's share of the asset's total payoff is shaded in a different color. An agent is wiped out when the realized state falls below a certain cutoff, at which point the next-most pessimistic type starts to take losses.

type 2, promising to repay θ_1^* dollars at $t = 1$. Type 2 uses this debt contract as collateral to borrow from agent 3 and promises to repay θ_2^* dollars at $t = 1$. This continues all the way down to type N, who lends to type $N-1$ (without borrowing) in exchange for a promise to repay θ_{N-1}^* at t=1, taking $N-1$'s debt contract with $N-2$ as collateral.

Collateral in this type of arrangement is a contract backed by another contract, which in turn is backed by a third contract, and so forth, until the last contract, which is backed by a physical asset. This type of arrangement is another salient feature of housing markets: a CDO is a set of debt contracts backed by subprime mortgage-backed securities, which in turn are backed by pools of mortgages, which are contracts backed by houses. Chains of lending are also common in loan markets. Banks often lend to firms directly and then repo those loans to outside investors.

Agents find this behavior optimal because of the structure of their beliefs. The hazard-rate ordering of Assumption 1 ensures not only that low types are more optimistic than high types, but also that low types are relatively more optimistic conditional on sufficiently high realizations of s. To see this, note that the hazard rate assumption implies that $1 - F_i(s)$ decreases at a slower rate than $1 - F_j(s)$ whenever $i < j$. Consider the events $B_s = \{s' : s' \geq s\}$. The probability assigned to event B_s by a type k is $1 - F_k(s)$. If $i < j$, the ratio of the probability of B_s in type

i's measure to that in type j's measure is $\frac{1-F_i(s)}{1-F_j(s)}$, which is increasing in s. In this sense, more optimistic types value risky payoffs relatively more than safe payoffs.

Intermediation emerges in this model precisely because intermediary types are most optimistic about intermediate states relative to optimists and pessimists. In order to construct a portfolio that pays off above some intermediate state, they lend to optimists, bear risk when optimists are wiped out, and borrow from pessimists. A common theme in the theoretical literature on intermediation is that types with preferences that are in some sense between those of other types will tend to take offsetting positions. For example, in Farboodi, Jarosch, and Shimer (2017) fast traders are less sensitive to the alignment of their preferences and asset holdings and act as intermediaries for slower traders. Atkeson, Eisfeldt, and Weill (2014) present a framework in which banks with intermediate risk exposures act as dealers for "customer" banks with extreme risk exposures. The emergence of lending chains in the presence of belief heterogeneity is a novel implication of this model.

Why is it that in this setup tranched payoffs arise in chains of lending? After all, in reality tranching arises when a single asset is used as collateral for many contracts, whereas in chains of lending contracts are used to back other contracts, with the physical asset serving as the collateral for the original contract. The answer lies in the fact that by borrowing and lending, agents in this model are able to synthesize assets whose payoffs are constant over some range of states and then decline to zero. When agents' opinions differ, they bet with each other on the tail probabilities of events. The most optimistic agents will naturally make the riskiest bets. At the interest rate that they are willing to borrow, the second most optimistic agents are happy to lend to them and take a safe payoff for draws in the extreme upper tail of the distribution of s. The social planner's solution indicates that whenever the set of available (synthetic) assets is $\hat{A}$, absent other frictions, agents will always make the exact same bets they do in this model.

Indeed, in this model there is an equivalence between chains of lending and the direct sale of tranched securities. Instead of assuming external agents sell an asset to agents who may borrow and lend with collateral, it could instead be assumed that these external agents offer a menu of assets with tranched payoffs $\min\{\theta, s\} - \min\{\theta', s\}$ in a competitive market at prices $\tilde{q}(\theta, \theta') = q(\theta) - q(\theta')$. In this case, clearly, agents would choose the same asset holdings, and equilibrium prices would be exactly the same. An alternative assumption is that some type n has the ability to create tranched securities and sell them to types $n+1, n+2, \ldots, N$. Types 1 through n would then take part in a chain of lending, and types $n+1$ through N

would buy the tranches of the security sold by n, but again the equilibrium would be equivalent to the one obtained above. All results regarding chains of collateralized lending therefore also carry over to settings in which the asset is directly tranched.

Although Proposition 3.2 characterizes the equilibrium when it exists, it shows neither that an equilibrium always exists nor that it is unique. Fortunately, as Proposition 3.3 demonstrates, there always exists a unique equilibrium.

Proposition 3.3. *There exists a unique competitive equilibrium.*

2.3.3 Equilibrium Prices, Interest Rates, and Leverage

The equivalence between equilibrium allocations and constrained social optima yields interesting implications for prices and margins in the model. Consider an equilibrium with $0 = \theta_N^* < \theta_{N-1}^* < \cdots < \theta_1^* < \overline{s}$, and let $\{\lambda_n\}_{n=1}^N$ be the weights in the corresponding social planner's problem (assuming $\lambda_N = 1$). The weight placed on type n is the inverse of that type's return on wealth r_n (defined in Section 2.2). For the values θ_n^* to satisfy this property, it must be that $\theta_n^* = \theta_{n,n+1}^*$ (recalling that θ_{nm}^* is defined so that $\frac{1-F_n(\theta_{nm}^*)}{1-F_m(\theta_{nm}^*)} = \frac{\lambda_m}{\lambda_n}$). These assumptions may seem restrictive, but this is the only type of equilibrium in which all agents are involved in the chain of lending and the most pessimistic agent holds cash. That is, as long as each agent lends to the next most optimistic agent and borrows from the next most pessimistic agent, the equilibrium values of θ_n^* always take this form. Proposition 3.2 shows that the equilibrium always takes this form, so there is in fact no loss of generality. Whenever the most pessimistic agent holds cash, it must be that returns on wealth for type N are $r_N = 1$, since type N agents are content to receive zero net return on investment.

As shown in Proposition 3.2, for any $[\theta', \theta] \subset [\theta_{n,n+1}^*, \theta_{n-1,n}^*]$, the price function $q(\theta)$ satisfies

$$q(\theta) - q(\theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] = \frac{1}{r_n} E_n[\min\{\theta, s\} - \min\{\theta', s\}] \quad (2.1)$$

For any such pair (θ, θ'), $q(\theta) - q(\theta')$ is determined by the beliefs of type n. Relative to perceived returns on wealth r_n, type n has the highest valuation of the payoff $\min\{\theta, s\} - \min\{\theta', s\}$. Since $q(\theta) - q(\theta') > \frac{1}{r_n} E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ for any $[\theta', \theta] \not\subset [\theta_{n,n+1}^*, \theta_{n-1,n}^*]$, type n will only borrow or lend quantities in the range $[\theta_{n,n+1}^*, \theta_{n-1,n}^*]$. All agents $n < N$ therefore use leverage when lending. In order to increase their returns on wealth, they borrow from more pessimistic agents who are willing to lend at favorable interest rates.

Given $q(0) = 0$ and $p = q(\overline{s})$, it follows that

$$p = \frac{1}{r_1} E_1[s - \min\{\theta^*_{1,2}, s\}] + \sum_{n>1} \frac{1}{r_n} E_n[\min\{\theta^*_{n-1,n}, s\} - \min\{\theta^*_{n,n+1}, s\}] \tag{2.2}$$

The price of the asset is determined by the expectations of all agents. Even though only the most optimistic agent purchases the asset directly, the expectations of the other agents are important in determining the price because expectations determine each agent's willingness to lend. More lending means the most optimistic agent has more funds to use in purchasing the asset, so the market clearing price must be higher.

In particular, since

$$E_n[\min\{\theta, s\} - \min\{\theta', s\}] = (1 - F_n(\theta^*_{n-1,n}))\theta^*_{n-1,n} + \int_{\theta^*_{n,n+1}}^{\theta^*_{n-1,n}} s f_n(s) ds$$

the asset's price is determined by two features of agent n's beliefs: the upper tail probability $1 - F_n(\theta^*_{n-1,n})$ and the perceived risk $\int_{\theta^*_{n,n+1}}^{\theta^*_{n-1,n}} s f_n(s) ds$ (which is also determined by the lower tail probability). When deciding how much to lend, it makes no difference to type n how payoffs are distributed above $\theta^*_{n-1,n}$. All that matters is the perceived probability that the loan will be repaid in full. Below $\theta^*_{n-1,n}$, the distribution of payoffs does affect how much type n chooses to lend, as type n bears payoff risk in that region. These properties will be crucial in the analysis of comparative statics in Section 4.

The values r_n, which represent perceived return on wealth for each type, merit some explanation as well. Recall that $\frac{r_n}{r_{n+1}} = \frac{1-F_n(\theta^*_{n,n+1})}{1-F_{n+1}(\theta^*_{n,n+1})}$. This formula can be interpreted as follows: at the margin, type n agents decide whether to promise an extra unit of repayment θ at $t = 1$ to type $n+1$ agents. A small change in loan riskiness represents one extra dollar of repayment whenever $s \geq \theta$, since the agent will walk away from the loan regardless whenever $s < \theta$. The amount type $n + 1$ is willing to lend at $t = 0$ for that repayment is $\frac{1-F_{n+1}(\theta)}{r_{n+1}}$, and the cost of that repayment to type n is $\frac{1-F_n(\theta)}{r_n}$. Therefore type n's repayment choice must satisfy $\frac{r_n}{r_{n+1}} = \frac{1-F_n(\theta^*_{n,n+1})}{1-F_{n+1}(\theta^*_{n,n+1})}$.

Observe that $\frac{1}{r_n} = \prod_{k=n}^{N-1} \frac{r_{k+1}}{r_k} = \prod_{k=n}^{N-1} \frac{1-F_{k+1}(\theta^*_{k,k+1})}{1-F_k(\theta^*_{k,k+1})}$. The beliefs of type N then appear to be important in determining the return on wealth for all agents, since r_n has a factor of $1 - F_N(\theta^*_{N-1,N})$ for all n. This is a leverage effect: the returns on wealth of type n agents are determined in part by their ability to borrow, which is in turn

related to type $n+1$'s ability to borrow, and so forth all the way down to type N. If type N is sufficiently pessimistic, the pyramid of lending can effectively collapse, leaving type n agents to fund asset purchases using little more than their endowments. Indeed, type n's leverage can be written as

$$L_n = 1 + \frac{q(\theta^*_{n,n+1})}{w_j} = 1 + \frac{\sum_{m>n} \frac{1}{r_m} E_m[\min\{\theta^*_{m-1,m}, s\} - \min\{\theta^*_{m,m+1}, s\}]}{w_j} \tag{2.3}$$

When agents of type N are very pessimistic, the term $1 - F_N(\theta^*_{N-1,N})$, which appears in all terms of the sum in the numerator, is small. Leverage should thus be lower when agents of type N become more pessimistic.

Beliefs are also an important determinant of the marginal interest rates perceived by agents (that is, the interest rate on each extra dollar of $t = 0$ borrowing). In equilibrium, the marginal perceived interest rate R_n must equal the perceived return on a dollar r_n for each type n. The marginal interest rate can therefore be written as

$$R_n = \frac{1}{1 - F_N(\theta^*_{N-1,N})} \left(\prod_{m=n+1}^{N-1} \frac{1 - F_m(\theta^*_{m,m+1})}{1 - F_m(\theta^*_{m-1,m})} \right) (1 - F_n(\theta^*_{n,n+1})) \tag{2.4}$$

The first two terms are the inverse probabilities that types $m > n$ in the lending chain assign to being repaid conditional on not defaulting themselves. The third term is the inverse of type n's subjective probability of default. When pessimistic types think they are not likely to be repaid, they charge high interest rates on loans. On the other hand, when type n agents believe they will likely default (i.e., $F_n(\theta^*_{n,n+1})$ is high), they perceive a low interest rate because they believe that there is a good chance they will never repay the loan.

Of course, all of the preceding analysis relies on partial equilibrium logic, since the values $\theta^*_{n,n+1}$ are endogenous objects. However, these partial equilibrium effects provide important intuition for the results of Section 4, and for the most part, these intuitions carry through in the full general equilibrium of the economy.

2.3.4 Equilibrium Characterization

In this subsection, I briefly list and explain the equations defining an equilibrium. The proof of Proposition 3.3 shows that there exists some K such that the following

equations hold for $n \in \{1, \ldots, K-1\}$:

$$w_n = \left(\prod_{k=n}^{N-1} \frac{1 - F_{k+1}(\theta^*_{k,k+1})}{1 - F_k(\theta^*_{k,k+1})} \right) E_n[\min\{\theta^*_{n-1,n}, s\} - \min\{\theta^*_{n,n+1}, s\}] \tag{2.5}$$

where $\theta^*_{0,1} \equiv \overline{s}$. The values $r_n = \prod_{k=n}^{N-1} \frac{1-F_k(\theta^*_{k,k+1})}{1-F_{k+1}(\theta^*_{k,k+1})}$ must be greater than or equal to 1 for $n < K-1$ and equal to 1 for all $n \geq K$ if $K < N$. Any vector $(\theta^*_{1,2}, \ldots, \theta^*_{K-1,K})$ satisfying this system of equations yields an equilibrium. This equation says that each type's wealth is equal to its expected payoff in equilibrium divided by its expected return on wealth. Type n's expected return on wealth, as mentioned in Section 3.3, is a function of each higher type's subjective probability of default.

Observe that equation (2.5) can be rewritten as

$$w_n = (1 - F_N(\theta^*_{N-1,N})) \left(\prod_{m=n+1}^{N-1} \frac{1 - F_m(\theta^*_{m-1,m})}{1 - F_m(\theta^*_{m,m+1})} \right) E_n[\min\{\theta^*_{n-1,n}, s\} - \theta^*_{n,n+1} | s \geq \theta^*_{n,n+1}] \tag{2.6}$$

Equation (2.5) can therefore be rewritten in terms that only depend on each type's beliefs conditional on not defaulting. This observation is interesting in light of other observations in the literature, such as the claim in Simsek (2013) that it is the *skewness* of beliefs that determines asset prices. This formula shows that it is really each type's belief about the upside of its portfolio that pins down the asset price.

2.4 What Does Intermediation Do?

This model raises an obvious question of both theoretical and practical interest: how does intermediation affect equilibrium allocations? The main result of this section will show that the increase in the capitalization of an intermediary (or, similarly, a lengthening of the intermediation chain) tends to drive up the riskiness of loans taken by more optimistic agents and lower the riskiness of the loans made by more pessimistic agents. In Section 5's discussion of applications this characterization will play a key role.

In order to prove the results discussed above, the following definition will be helpful:

Definition 4.1. *Aggregate wealth is abundant if* $\sum_{i=1}^{N} w_i > E_1[s]$. *The wealth of a subset of types* $S \subset \{1, \ldots, N\}$ *is scarce if* $\sum_{i \in S} w_i < E_N[s]$.

Intuitively, when aggregate wealth is abundant, some type must hold cash because the price of the asset can never exceed the expectation of the most optimistic agent. As a consequence, some type is indifferent between lending and holding cash, meaning that this type always breaks even in expectation when lending. When the wealth of a subset of types is scarce, that group of types must borrow from wealthier agents if they are net purchasers of the asset, since the price of the asset can never be lower than the expectation of the least optimistic agent when aggregate wealth is abundant.

Suppose there are N types with priors $F_1(s), \ldots, F_N(s)$ satisfying Assumption 1. If K of these types have initial wealth equal to 0, the maximum possible length of a lending chain is $N - K$. The length of the lending chain can increase if the wealth of some type with zero wealth is increased. Proposition 4.2 is motivated by this notion of chain length.

Proposition 4.2. *When total wealth is abundant,*

1. *The face value of loans made to type n agents is increasing in the wealth of types $m > n$;*

2. *The riskiness of and the perceived marginal interest rate on the debt contract purchased by type n agents is decreasing in the wealth of types $m < n$.*

The logic behind this proposition is straightforward. When type n intermediaries have more wealth, they bid up the amount they are willing to lend to type $n-1$ agents, who in turn lend more to type $n-2$, and so forth. The increase in type n's demand for debt contracts propagates all the way up the chain and allows more optimistic types to take riskier loans at lower interest rates. Intermediaries also provide a cash buffer for the most pessimistic lenders, meaning that those agents are willing to lend at lower marginal interest rates. When type n agents are wealthier, they are able to bear more of the risk associated with holding the collateral. Then type $n+1$ agents make safer loans to type n agents, thereby reducing their borrowing needs and allowing them to take safer loans from type $n+2$ agents. As intermediary capitalization increases, the riskiness of the end-lenders' loans approaches zero. Intermediation effectively tranches the risky asset's payoffs to such an extent that it produces almost perfectly safe assets for pessimists.

An increase in the length of the intermediation chain can be interpreted as follows: if some intermediate type starts with wealth equal to zero, increasing that type's wealth to some positive value will cause a discrete change in the length of the chain. If the length of the chain is increased in this way, then, the leverage of borrowers

will increase and the interest rate charged by lenders will decrease. The following corollary follows naturally from the fact that increasing the length of the lending chain is equivalent to increasing the wealth of some intermediary in this setting.

Corollary 4.3. *The leverage of the most optimistic type is increasing in the number of intermediaries. The marginal interest rate charged by the most pessimistic lender and paid by the most optimistic borrower is decreasing in the number of intermediaries.*

Financial intermediation chains and the reuse of collateral are key to the simultaneous increases in leverage and the supply of safe assets. Indeed, in this model intermediary capitalization, leverage, and the safety of end-lenders' portfolios all move together. This relationship is in accordance with some recent stylized facts. Lenel (2017) documents that the supply of safe assets produced by the private sector underwent an unprecedented expansion prior to the financial crisis of 2007-2009. Geanakoplos (2009), on the other hand, shows that margins on the contracts backing those safe assets were countercyclical. Tranched securities are the archetypical example of this phenomenon: if the riskiness of the underlying assets is held fixed, a greater number of tranches must mean that the lower tranches are safer and the upper tranches are riskier.

Proposition 4.2 highlights a fundamental property of intermediation in heterogeneous belief models. In any such model, intermediation will focus underlying risks on agents who are most willing to bear them. When optimists are poor and there are no intermediaries, pessimists are forced to bear some risk because optimists can only afford to purchase the asset if they take risky loans. When intermediaries enter, they can act as optimists' agents by providing them with funding while protecting more pessimistic agents from default. Intermediation focuses borrowers' risks into the region of the state space where they are most optimistic relative to other agents.

It is also useful to characterize intermediation in terms of agents' perceived returns on wealth r_n, which are endogenous quantities. The following result is similar to Proposition 4.2 in spirit and will provide insight during the discussion of applications in Section 5:

Proposition 4.4. *Fix some type $n \in \{2, \ldots, N-1\}$. Holding the returns on wealth of types $m \neq n$ constant, a decrease in r_n raises the riskiness of loans taken by type $n-1$ agents and decreases the riskiness of loans made by type $n+1$ agents.*

Recall that there is a mapping from returns on wealth to the distribution of wealth– in particular, the returns on wealth of a given type are inversely related to

that type's wealth. This proposition is then akin to a local rephrasing of Proposition 4.2, since it relates intermediation to the positions taken by a given type rather than those taken by all types.

2.5 Applications of the model

In this section I present several applications of the model. I show that the model generates contagion in margin spikes both along the chain of lending and across asset classes and that leverage cycles coincide with cycles of safe asset production by intermediaries. I also analyze the consequences of financial innovations by comparing the model to a benchmark in which it is not permitted to use debt as collateral.

2.5.1 Contagion in Margins

While contagion is often discussed in the context of asset prices, during the financial crisis of 2007-2009 spikes in margins for mortgage-backed CDOs propagated to other markets as well. As Gorton and Metrick (2012) document, in bilateral repo markets margins rose in essentially all risky asset classes. Furthermore, the increase in CDO margins coincided with a decrease in loan-to-value ratios in the housing market, which is a phenomenon that Geanakoplos (2010) terms the *double leverage cycle*: increases in the capital requirements for purchasing asset-backed securities translate into increases in margins on the securities themselves, which leads to instability both in the market where intermediaries lend and in the market in which they borrow. I show that my model explains both of these phenomena in a straightforward way.

There are two forces that determine the margin put down by agents of type n in this model: the beliefs of other agents and the wealth of other wealth-constrained agents. Broadly speaking, the margin paid by a given type of agent rises when either other agents become more pessimistic or other agents are more wealth-constrained. For now, I focus on a shock to the pessimism of the end-lenders who hold the safe tranche (the most pessimistic type) as the driver of margins, since this type of shock is the one most often discussed in the literature on asset pricing with endogenous collateral constraints.

I first show that a shock that "scares" pessimists into becoming even more pessimistic leads to all types $n = 1, \ldots, N-1$ paying higher margins on the securities they purchase. Proposition 5.1 provides a formal statement:

Proposition 5.1. *Suppose that in the equilibrium with initial endowments $w_1, \ldots, w_N$ and beliefs $F_1, \ldots, F_N$, the margin paid by type i agents is*

$$m_n = \frac{\frac{1}{r_n} E_n[\min\{\theta^*_{n-1}, s\} - \min\{\theta^*_n, s\}]}{\sum_{k=n}^{N} \frac{1}{r_k} E_k[\min\{\theta^*_{k-1}, s\} - \min\{\theta^*_k, s\}]}$$

. Then if the belief of type N is changed to $\tilde{F}_N$ such that $\tilde{F}_N(s) \geq F_N(s)$ for all s and all other beliefs and endowments are held fixed, the new equilibrium margins $\{\tilde{m}_n\}_{n=1}^{N-1}$ satisfy $\tilde{m}_n \geq m_n$ for all n.

This model therefore features not only double leverage cycles in which the margins on two different securities are linked, but rather leverage cycles of degree N where the margins on securities at all levels of the lending chain move together. Empirically, this corresponds to comovement in margins on all asset-backed debt of all different ratings, which did occur during the collapse of the securitized products market in 2008.

Of course, increased pessimism by those holding the safest tranche of debt will lead pessimists to lend less against any collateral. This feature is shared by essentially all models with endogenous collateral constraints. The feature that is particular to a model with endogenous *chains* of lending is that once the lenders at the end of the chain become less willing to lend, intermediaries increase their margin requirements as well. The key insight behind why an increase in margins at the bottom level of the chain leads to a spike in margins all along the chain is that intermediary lenders are wealth-constrained. When cash-rich pessimists suddenly stop lending against the debt contracts purchased by these intermediaries, they are simply unable to lend as much to end-borrowers (optimists) because their leverage has decreased. Put differently, the spike in margins at the bottom of the chain effectively increases the cost of capital for intermediaries because pessimists demand higher interest payments, so intermediaries are less willing to pay optimists for a given promised future cash flow, which decreases margins at the top of the chain as well.

This double leverage cycle result is directly linked to the results of Section 4 on the consequences of intermediation. Intermediaries must always hold a cash buffer to protect pessimists from negative shocks to the asset's payoff. When pessimists feel sufficiently protected, they lend large quantities to intermediaries, which in turn enables them to turn around and lend to optimists, which generates high leverage at the top of the chain. However, when pessimists perceive additional tail risk on the asset, they require a larger cash buffer to make a loan of a given size. Barring a positive

shock to intermediary capitalization, intermediaries are unable to provide a larger cash buffer than previously, so the size of loans made by pessimists to intermediaries (and thus the size of loans made by intermediaries to optimists) must decrease.

I now show that in an extension of the model, a shock to pessimism of the type in Proposition 5.1 can cause an increase in margins across asset classes even when payoffs are uncorrelated across classes. Suppose that there are two assets A and B (each in unit supply) that yield payoffs s_A, s_B at $t = 1$. There are four types of investors: type 1_A investors, type 1_B investors, type 2 investors, and type 3 investors. Type 1_A investors have the most optimistic assessment $F_1^A(s_A)$ about asset A's payoffs, type 2 and type 1_B investors share the second-most optimistic beliefs $F_2^A(s_A)$, and type 3 investors are the most pessimistic about A's payoff, with beliefs $F_3^A(s_A)$. In market B, agents' beliefs are analogous to those in market A (with the roles of 1_A and 1_B flipped). The contracts are as in the baseline model. A debt contract may be collateralized by either a unit of asset A, a unit of asset B, or another debt contract. Hence in market A, 1_A will borrow from 2 who borrows from 3, whereas in market B, 1_B borrows from 2 who borrows from 3. Both 2 and 3 are "crossover investors" in the sense that they invest in both markets, but the critical feature of this model will be that wealth-constrained type 2 agents lend to optimists and determine margins in both markets. Type 1_A and 1_B investors are relatively more optimistic about one of the two assets, so they put their entire endowment into their respective markets.

The main result in this extension of the baseline model is that a pessimistic shock to type 3's belief about asset A causes margins paid by both type 1_A agents and type 1_B agents to rise. This result is summarized in Proposition 5.2.

Proposition 5.2. *Suppose that in the equilibrium with initial endowments w_{1A}, w_{1B}, w_2, w_3 and beliefs $\{F_1^k, F_2^k, F_3^k\}$ for $k \in \{A, B\}$, margins paid by end-borrowers are m_A and m_B. Then in an alternative equilibrium where only 3's beliefs about asset A change to $\tilde{F}_3^A$ such that $\tilde{F}_3^A(s) \geq F_3^A(s)$ for all s, margins increase in both markets; i.e., $\tilde{m}_A \geq m_A$ and $\tilde{m}_B \geq m_B$.*

Proposition 5.2 shows that a disruption to the collateral capacity of asset A also reduces the collateral capacity of asset B regardless of how the payoffs of asset B are related to those of asset A. Although contagion in asset prices across markets is a common feature of models with wealth-constrained investors who trade multiple asset classes, contagion in margins is not. For example, in the setting of Simsek (2013) in which there is no scope for intermediation, a shock to pessimism about asset A would cause end-borrowers to move their wealth out of asset B and into asset A to

cushion the crash in prices, which would decrease margins in market B rather than increasing them. In a model with information frictions like that in Dang, Gorton, and Hölmstrom (2013), a shock increases margins across several asset classes only if that shock affects the value of information about a fundamental common to the payoffs in all of those asset classes. Finally, in Brunnermeier and Pedersen (2009), margins increase across asset classes only because lenders misinterpret the wealth effect that leads to departures of prices from fundamentals across markets as a shock to the fundamentals of assets in all those markets.

By comparison, this explanation seems more parsimonious than those offered in the literature, as it achieves contagion in margins without any assumption about the correlation of assets' payoffs or the bounded rationality of lenders. The mechanism of Proposition 5.2 relies on the simple fact that *lenders* may be wealth constrained, and when those lenders participate in several markets, they will be able to lend less against collateral in all markets. This explanation certainly seems plausible in the case of the bilateral repo market during the "run on repo" of 2008 studied by Gorton and Metrick (2012). In that market, leveraged investors such as banks and hedge funds obtain funding from each other rather than from institutions with safer portfolios such as MMMFs and securities lending firms. In the context of this model, the observed rise in margins on all risky securities during the run on repo is consistent with financial stress in the intermediary sector.

2.5.2 Leverage and Safe Asset Production Cycles

During the boom in securitization leading up to 2007, both leverage and private-label safe asset production exploded along with real estate prices and investment. After the collapse of the securitized products market, private production of safe assets and loan-to-value ratios collapsed. Figures 2 and 3, taken from Fostel and Geanakoplos (2016) and Lenel (2017), respectively, depict these patterns. I now show that the baseline model augmented to allow for investment explains this episode as a consequence of the rise and subsequent fall of net worth in the intermediary sector, which in this case can be thought of as the shadow banking system.

In order to isolate the effect of intermediary net worth on leverage and safe asset production, I modify the model by assuming that there are only three types ordered by their optimism and that type 1 agents and type 3 agents have large endowments. However, type 1 agents are relatively more impatient than type 3 agents: their utility is $u_1(c_0, c_1) = r_1 c_0 + c_1$ with $r_1 > 1$, whereas type 2 and 3 agents have utility

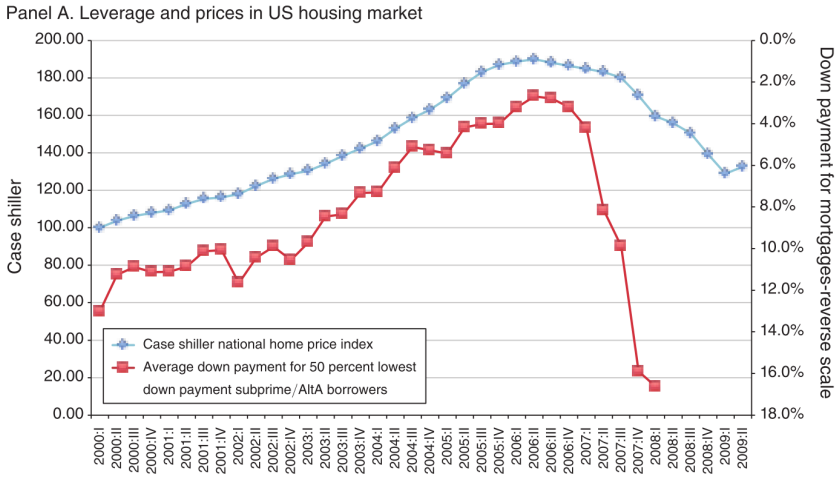


Figure 2.2: House prices (in blue) and loan-to-value ratios (in red) during the period 2000-2009. Loan-to-value ratios are plotted as margins on an inverse scale.

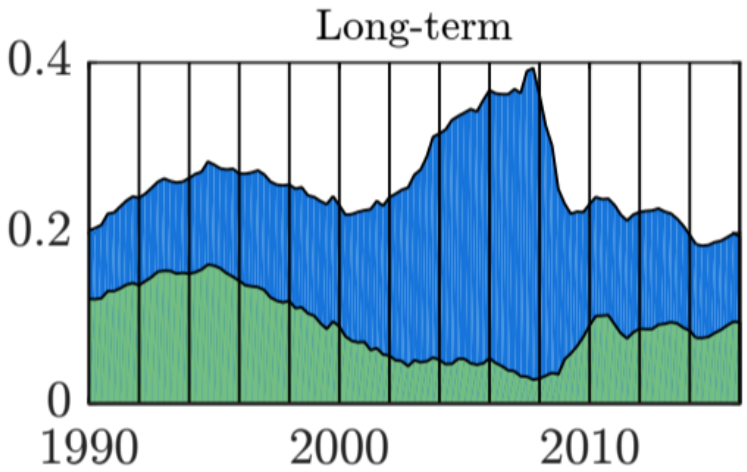


Figure 2.3: Private-label long-maturity safe asset production (in blue) and long-maturity Treasury debt (in green) during the period 1990-2017.

$u_2(c_0, c_1) = u_3(c_0, c_1) = c_0 + c_1$. The assumption that the initial endowments of types 1 and 3 are large implies that their discount factors are essentially exogenous: type 1 will price any asset with payoff $\phi(s)$ as $\frac{1}{r_1}E_1[\phi(s)]$, and type 3 agents price assets as $E_3[\phi(s)]$ as in the baseline model. The assumption that type 1 is impatient can alternatively be seen as an assumption that type 1 agents have a superior outside investment option that yields expected returns r_1. Ultimately, this assumption just serves to capture the idea that type 1 will invest in the asset only if it is possible to lever up and obtain high expected returns.

I also introduce investment into the baseline model to capture the boom and bust in investment that coincided with the rise and fall of leverage and private safe asset production. Formally, there is an investment technology that allows any agent to produce a unit of the asset by paying a price $p(q)$, where q is the quantity of the asset that has been produced in aggregate and p is a continuous increasing function. This assumption does not change any of the conclusions of the baseline model– it is still the case that type 1 will borrow from type 2, who uses that debt contract as collateral to borrow from type 3. The only differences between this model and the baseline model are that (1) the returns on wealth perceived by type 1 are exogenous, and (2) the supply of the asset is endogenous.

The equations that pin down equilibrium in this extension of the model are

$$r_1 = \frac{1 - F_1(\theta_1^*)}{1 - F_2(\theta_1^*)} \frac{1 - F_2(\theta_2^*)}{1 - F_3(\theta_2^*)}$$

$$w_2 = \frac{1 - F_3(\theta_2^*)}{1 - F_2(\theta_2^*)} q(p) E_2[\min\{\theta_1^*, s\} - \min\{\theta_2^*, s\}]$$

$$p = \frac{1}{r_1} E_1[\max\{s - \theta_1^*, 0\}] + \frac{1 - F_3(\theta_2^*)}{1 - F_2(\theta_2^*)} q(p) E_2[\min\{\theta_1^*, s\} - \min\{\theta_2^*, s\}] + E_3[\min\{\theta_2^*, s\}]$$

where $q(p)$ is the inverse of $p(q)$. It can easily be shown that these equations yield a unique solution for $(\theta_1^*, \theta_2^*, p)$.

The main result regarding the model with investment links intermediary wealth v_2, leverage, safe asset creation, investment, and asset prices:

Proposition 5.3. *The margins m_1 paid by type 1 agents on a unit of the asset, the riskiness θ_1^* of loans taken by type 1 agents, investment q, and prices p are all increasing in intermediary net worth w_2. The riskiness of loans made by type 3 agents θ_2^* is decreasing in w_2.*

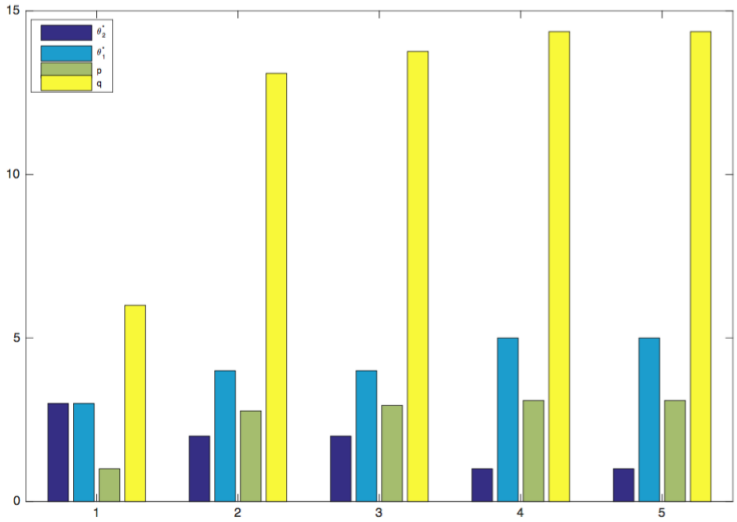


Figure 2.4: Riskiness of loans taken by type 2 θ_2^*, riskiness of loans made by type 2 θ_1^*, asset prices p, and investment quantities q in Example 5.4 as a function of the realized state $s \in \{1, 2, 3, 4, 5\}$ at $t = 1$.

The logic underlying this result is similar to the intuition behind Proposition 4.2. Intermediation by agents who are willing to bear risk boosts the collateral capacity of the asset, which raises its price, and allows those least willing to bear risk to make safer loans. When investment is introduced, the boom in asset prices additionally leads to increased investment, since it becomes more profitable to invest than to purchase the asset directly.

The other main benefit of this alteration of the model, however, is that it is numerically solvable in a dynamic setting. I present an example that suggests the static result of Proposition 5.3 goes through to a dynamic setting even when intermediaries might be incentivized to "keep their powder dry" by making safer loans and not exposing themselves to risk early on. Intermediary net worth is correlated the shocks to the asset's payoff, which yields a dynamic version of the proposition's result.

Example 5.4: There are three periods, $t = 0, 1, 2$. There are three types of agents with the preferences specified above. In each period $t \in \{1, 2\}$, a unit of the asset pays off $s \in \{1, 2, 3, 4, 5\}$ and depreciates completely. The asset's payoffs (as perceived by each type) are iid across periods. Agents' beliefs satisfy the hazard rate ordering.[1] Type 2 agents start at $t = 0$ with an endowment of $w_2 = 20$ dollars. The function that pins down the cost of investment is $q(p) = 2 + 4p$.

[1]I do not explicitly write down agents' beliefs in order to conserve space. It makes no difference that the agents believe the payoffs of the asset take values in a discrete set– all of the conclusions of the baseline model go through.

As shown in Figure 4, the results in the dynamic model are exactly those suggested by Proposition 5.3. When the state realized at $t = 1$ is good, the intermediary is wealthy at $t = 1$, investment booms, the asset price rises, optimists are able to take riskier loans, and pessimists hold only riskless debt. By contrast, when the realized asset payoff at $t = 1$ is low, the intermediary is wealth-constrained and leverage, prices, and investment collapse while pessimists are forced to hold risky debt. When the worst outcome is realized, in fact, the intermediary's net worth is completely wiped out and pessimists have to lend directly to optimists. These results broadly capture the trends in the run-up to and fallout of the crash displayed in Figures 2 and 3.

2.5.3 The Impact of Financial Innovation

In this section, I introduce an alternative economy in which it is not possible to use debt contracts as collateral and ask how the benchmark economy differs from this alternative one. I primarily compare risk-taking in the two economies and show that some agent always takes greater risk in the benchmark economy than in the alternative economy.

The model is identical to that described in Section 2.1 with the exception of the space of assets that are traded. In this version of the model, the only admissible collateral is the physical asset a. Therefore, there are two types of positions that agents may hold: they may borrow using the asset as collateral and promise to repay θ in the next period, or they may lend to a borrower and take possession of the asset in case of default. In other words, when a unit of the asset is purchased, the borrower receives a payoff $\max\{s - \theta, 0\}$ and the lender receives $\min\{\theta, s\}$ when the loan has riskiness θ.

The optimization problem of a type-n agent is

$$\max_{a,\mu_+,\mu_-,c} aE_n[s] + c + \int E_n[\min\{\theta, s\}]d\mu_+(\theta) - \int E_n[\min\{\theta, s\}]d\mu_-(\theta)$$
$$s.t.\ w_n = pa + c + \int q(\theta)d\mu_+(\theta) - \int q(\theta)d\mu_-(\theta)$$
$$a \geq \int d\mu_-(\theta),\ \mu_+(\theta) \geq 0,\ \mu_-(\theta) \geq 0$$

This differs from the optimization problem in Section 2.2 only in that debt contracts cannot be used as collateral: in this case, the total number of borrowing contracts

must be less than or equal to the number of units of the asset a purchased by the agent.

An alternative formulation of this problem is

$$\max_{\mu_b,\mu_l,c} \int E_n[\max\{s-\theta,0\}]d\mu_b(\theta) + \int E_n[\min\{\theta,s\}]d\mu_l(\theta) + c$$
$$\text{s.t.} w_n = \int (p - q(\theta))d\mu_b(\theta) + \int q(\theta)d\mu_l(\theta) + c,\ \mu_b(\theta) \geq 0, \mu_l(\theta) \geq 0$$

In this formulation, $\mu_b(\theta)$ represents the number of borrowing contracts of riskiness θ collateralized by a unit of the asset and $\mu_l(\theta)$ represents the number of lending contracts of riskiness θ held by the agent.

An equilibrium is defined in the obvious way. From now on, for ease of exposition I assume that there are only three belief types, although the model can easily be extended to include more. In this setting, as long as types 1 and 2 are wealth-constrained there are only two possible types of equilibria: ones in which both types 1 and 2 purchase the asset and borrow from type 3, and ones in which only type 1 buys the asset and borrows from types 2 and 3. Proposition 5.5 formalizes the main result.

Proposition 5.5. *If, given fixed initial endowments and beliefs, type 2 agents buy the asset directly in the equilibrium of the alternative economy, type 1 agents take riskier loans in the benchmark economy than in the alternative economy. If type 2 agents lend to type 1 agents in the equilibrium of the alternative economy, type 2 agents take greater risk in the benchmark economy than in the alternative economy. The asset price is always higher in the benchmark economy*

This result, while almost obvious, will have important implications for the dynamics of asset prices. When natural intermediaries are lenders in the setting without debt contracts as collateral, they take advantage of financial innovation by levering up to lend to optimists. On the other hand, when these natural intermediaries are buyers of the asset before financial innovation, the introduction of debt contracts as collateral actually makes them take a slightly safer position: they borrow to make risky loans to optimists rather than buying the asset directly, which in turn puts optimists in a much riskier position.

As in the previous subsection, I construct a dynamic numerical example to illustrate this effect.

Example 5.6: There are three periods, $t = 0, 1, 2$. Agents with a desire to consume immediately sell a single unit of an asset in periods 0 and 1. Agents of types 1, 2,

and 3 buy this asset and trade collateralized debt contracts in the first two periods. Type 1 and 3 agents are short-lived, meaning a new generation of type 1 and 3 agents enters the market at time 1, but type 2 agents are long-lived. This assumption focuses the example on the consequences of risk-taking by intermediaries rather than end-borrowers. The asset's payoff in each period falls in the set $s \in \{u, m, d\}$, where $u = 1$, $m = 0.5$, and $d = 0.2$. At $t = 0$, agents have the following beliefs:

$$Pr_1(s_0 = u) = 0.8,\ Pr_1(s_0 = m) = 0.1,\ Pr_1(s_0 = d) = 0.1$$

$$Pr_2(s_0 = u) = 0.1,\ Pr_2(s_0 = m) = 0.6,\ Pr_2(s_0 = d) = 0.3$$

$$Pr_3(s_0 = u) = 0.02,\ Pr_2(s_0 = m) = 0.18,\ Pr_2(s_0 = d) = 0.8$$

If the realization of s_0 is either u or d, all agents agree that at $t = 1$ the asset will pay u. On the other hand, if $s_0 = d$, agents' beliefs about s_1 are identical to their $t = 0$ beliefs about s_0. Type 1 agents are endowed with 0.16 dollars in each period, and type 2 agents start with an endowment of 0.2 dollars at $t = 0$.

At $t = 1$, if the realization of s_0 was u or m, the asset's price will be equal to 1. In order to solve the example, it is necessary to find the allocation and prices at $t = 1$ after d is realized and then work backwards to determine the allocation and prices at $t = 0$. When debt cannot be used as collateral, in this example types 2 and 3 lend to type 1. At $t = 1$ after d is realized, type 2 still has some remaining wealth in the alternative economy without debt as collateral but is completely wiped out in the benchmark model. The asset price at $t = 0$ is 0.56 in the benchmark economy, whereas in the economy without debt as collateral, the asset price at $t = 0$ is 0.42. In the initial period, the asset's price is higher in the benchmark economy because intermediaries sell off the safe tranche of debt that they hold to pessimists and then use that additional capital to lever up and lend to optimists, at which point they sell off the lower tranche again, and so forth. This cycle of lending and levering up greatly increases the asset's collateral capacity. However, once d is realized, the asset's price crashes to 0.36 in the benchmark economy, but in the alternative economy it falls to only 0.40. The price in the benchmark economy is both higher in the initial period and lower in the second period, and the volatility exceeds that in the alternative economy by an order of magnitude. Volatility is higher in the alternative economy simply because allowing lenders to lever up may mean that they go bankrupt and are absent when the market crashes, which incidentally is when they are most needed to prop up asset prices. This example thus illustrates how the ability of the financial

sector to use leverage and stretch collateral among several parties may lead to large run-ups and precipitous falls in asset prices.

2.6 Discussion

In this section I briefly address issues regarding the plausibility of the modeling choices and how they affect the results. I consider two issues: the introduction of alternative sets of securities and the use of differences in beliefs as a motive for trade.

2.6.1 Alternative Asset Markets

It is not immediately clear from the discussion in previous sections how the results would change if the set of feasible portfolios did not consist solely of leveraged positions in collateralized debt contracts. For example, if agents were permitted to trade Arrow-Debreu securities, would chains of intermediation emerge? Would information aggregation be impaired in situations with long chains of intermediation?

The results presented above are quite general. The fundamental ingredients of the model are the following: a function $g_n(\theta) = E_n[\min\{\theta, s\}]$ of a parameter θ corresponding to each agent's expected asset payoffs and an assumption on the ordering of beliefs, which in this case is that

$$-\frac{d}{d\theta}\log(g_n'(\theta)) = -\frac{g_n''(\theta)}{g_n'(\theta)} = \frac{f_n(\theta)}{1-F_n(\theta)}$$

is ordered by n. As long as agents are permitted to take matched long-short positions in contracts with payoffs $g_n(\theta)$ (as in this model) they can take positions with payoffs $g_n'(\theta)$ (which in this model is $1-F_n(\theta)$). Any feasible payoff can then be written as an integral $\int\limits_{\theta\in\Theta} g_n'(\theta)d\theta$ for some set Θ. When agents have perceived returns on wealth $\{r_n\}_{n=1}^N$, type n will price positions such that $\frac{1}{r_n}g_n'(\theta) = \max\limits_m \frac{1}{r_m}g_m'(\theta)$. Note that in this discussion limited commitment is not explicitly mentioned– it is only necessary in order to motivate the specific form of the payoff function.

A simple way to construct such a model with complete markets is to take $g(\theta) = E_n[\mathbf{1}\{s \leq \theta\}]$. Then $g_n'(\theta) = f_n(\theta)$, so agents are able to synthesize Arrow-Debreu securities. In this context, the necessary assumption on beliefs is that $\frac{g_n''(\theta)}{g_n'(\theta)} = \frac{f_n'(\theta)}{f_n(\theta)}$ is a decreasing function of n, which is equivalent to assuming the monotone likelihood ratio property. Even in an environment with complete markets, then, the ordering of beliefs has a natural interpretation. When this restriction on beliefs is imposed,

agents form a chain: the most pessimistic agent holds claims on the worst states, the next-most pessimistic agent holds claims on the next interval of states, and so forth.

I use collateralized debt contracts in my analysis because they are the predominant financial instrument used to fund investments in several markets, such as the repo market and the housing market. This set of instruments thus allows for the analysis of empirically relevant issues. On the other hand, it is unclear to what an Arrow-Debreu contract corresponds in the data. There are also other reasons outside the scope of this model that one might want to restrict the analysis to simple collateralized debt contracts. For example, Dang, Gorton, and Holmström (2013) show that debt contracts optimally minimize the loss of liquidity due to asymmetric information.

2.6.2 Beliefs as a Motive for Trade

In the literature, heterogeneous beliefs often serve in as an ad hoc method of introducing a motive for borrowing and lending. However, the assumptions on beliefs in this model may raise some concerns. First, they are somewhat stronger than those in a binomial model, for example, since a binomial model requires only that agents disagree about one state, whereas here a hazard rate ordering must be satisfied over the entire state space. Second, it may not seem natural to equate financial intermediaries with agents who have intermediate beliefs. I discuss two alternative modeling choices – risk limits and ambiguity aversion – that take a step towards addressing these concerns yet result in an isomorphic model with exactly the same predictions.

Suppose all agents have a common prior $F(s)$ but are subject to risk limits. An extreme example of risk limits is a case in which agents of type n are not allowed to take any position that takes a loss with probability greater than $F(\hat{\theta}_n)$. In this case, agents act as if they had beliefs satisfying $1 - F_n(s) = \lambda_n(s)(1 - F(s))$ where

$$\lambda_n(s) = \begin{cases} 1 & s \leq \hat{\theta}_n \\ 0 & s > \hat{\theta}_n \end{cases}$$

That is, they act as if they believe states higher than $\hat{\theta}_n$ are impossible. More generally, one could assume that type n agents are subject to compensation schemes and restrictions that incentivizes them to act as if they had beliefs satisfying $1 - F_n(s) = \lambda_n(s)(1 - F(s))$. In this case, the hazard rate ordering reduces to the assumption that $\frac{\lambda_n'(s)}{\lambda_n(s)}$ is a decreasing function of n.

Such risk limits are not uncommon in the financial industry, and it is natural to think of lenders such as money market funds as having tight risk limits and

borrowers such as hedge funds as having loose limits. In practice, intermediaries such as investment banks are able to take greater risks than money market funds and can therefore serve as a link between hedge funds and sources of funding. In a model with this type of constraint, intermediation would be essential in facilitating trade between cash-constrained borrowers and lenders who are unable to bear the risk associated with the volatility of the underlying asset's payoffs. The interpretation of borrowing constraints would be different, but the implications would be exactly the same as those of the baseline model.

One could also assume that investors are ambiguity averse and evaluate payoffs according to their worst-case belief. If some investors face significant model uncertainty whereas others have a better grasp of the statistical properties of the asset's payoffs, it is natural to assume that each type n considers a set of possible priors μ_n such that $\mu_1 \subset \mu_2 \subset \cdots \subset \mu_N$. Under this assumption, type n is more certain of the distribution than type $n+1$ for all n. The worst-case distribution of type n is therefore always more optimistic than the worst-case distribution of type $n+1$. If the distributions within each μ_n can be ordered by their hazard rates, then once again this model is isomorphic to the baseline model. The assumption that people are ambiguity averse has experimental support, and it is not unreasonable to assume that natural lenders like money market funds would be uncertain about the payoffs of risky assets relative to hedge funds that employ traders who specialize in trading those assets.

Thus there are several equivalent ways of arriving at the same model. The assumption of heterogeneity of beliefs serves to simplify the exposition, but it really amounts to assuming nothing more than the existence of some agents who are natural borrowers and others who are natural intermediaries or lenders. Additionally, the model with heterogeneous beliefs is extremely tractable and allows for extensive theoretical analysis. Even though a model with more realistic institutional frictions could provide additional insight, such a model would likely be more difficult to solve.

2.7 Conclusion

How are assets' payoffs allocated in markets with reusable collateral? In an environment where agents are ordered by their optimism, chains of lending emerge endogenously. As a consequence of belief heterogeneity, payoffs are tranched: agents begin to take losses only when their direct debtors are wiped out. Belief heterogeneity also leads to intermediation. Types with intermediate optimism act as intermediaries

by borrowing from pessimistic agents and lending to optimists. These chains of collateralized lending resemble those that arise in financial markets through contracts such as CDOs and CLOs.

Intermediation creates heterogeneity in risk profiles and allows agents to take positions that match their beliefs. When an intermediary enters, more optimistic borrowers are able to take greater risks while more pessimistic lenders make safer loans. Intermediation is critical in bridging the gap between agents who want to take risky loans and the pessimistic agents who hold the majority of the wealth.

In extensions of my model, I study several applications of empirical relevance. This framework sheds light on the fact that increases in margins may be contagious both across markets and along the chain of lending, both of which are documented empirical phenomena. The observation that levered lenders as well as borrowers may be wealth-constrained allows my model to capture these phenomena in a parsimonious way. Second, endogenous intermediation produces an environment in which leverage, intermediation capacity, and the quantity of safe assets all move together. Finally, I show that financial innovation leads to greater risk-taking either on behalf of end-borrowers or intermediaries who lend to those borrowers, both of which can have serious consequences for asset markets after bad news arrives.

A challenge for future empirical work is to disentangle the causes of crises in collateralized debt markets. In my model, the source of essentially all crises in these markets are increases in pessimism that cause intermediaries to become wealth-constrained and cut back on risky lending. In other theories, haircuts rise because of asymmetric information (such as in Dang, Gorton, and Hölmstrom, 2013) or because volatility spooks cash lenders such as money market funds (as in Brunnermeier and Pedersen, 2009).

References

Atkeson, Andrew, Andrea Eisfeldt, and Pierre-Olivier Weill, "Entry and Exit in OTC Derivatives Markets," *Econometrica*, 2015, *83* (6), 2231–2292.

Bottazzi, Jean-Marc, Jaime Luque, and Mário R. Páscoa, "Securities market theory: Possession, repo and rehypothecation," *Journal of Economic Theory*, 2012, *147*, 477–500.

Broer, Tobias, "Securitization Bubbles," *Journal of Financial Economics*, 2018, *127* (3), 505–518.

Brunnermeier, Markus and Lasse Heje Pedersen, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 2009, *22* (6), 2201–2238.

Copeland, Adam, Antoine Martin, and Michael Walker, "Repo Runs: Evidence from the Tri-Party Repo Market," *Journal of Finance*, 2014, *69* (6), 2343–2380.

Dang, Tri Vi, Gary Gorton, and Bengt Hölmstrom, "Opacity and the Optimality of Debt for Liquidity Provision," *Unpublished Manuscript*, 2012.

_ , _ , and _ , "Haircuts and Repo Chains," *Unpublished Manuscript*, 2013.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer, "The Emergence of Market Structure," *Unpublished Manuscript*, 2017.

Fostel, Ana and John Geanakoplos, "Leverage Cycles and the Anxious Economy," *American Economic Review*, 2008, *98* (4), 1211–1244.

_ and _ , "Tranching, CDS, and asset prices: How financial innovation can cause bubbles and crashes," *American Economic Journal: Macroeconomics*, 2012, *4* (1), 190–225.

_ and _ , "Financial Innovation, Collateral, and Investment," *American Economic Journal: Macroeconomics*, 2016, *8* (1), 242–284.

Garleanu, Nicolae and Lasse Pedersen, "Margin-based Asset Pricing and Deviations from the Law of One Price," *Review of Financial Studies*, 2011, *24* (6), 1980–2022.

Geanakoplos, John, "Promises, Promises," *Cowles Foundation Working Paper*, 1997.

_ , "Liquidity, Default, and Crashes," *Advances in Economics and Econometrics: Theory and Applications II*, 2003, pp. 170–205.

_ , "The Leverage Cycle," *NBER Macoreconomics Annual*, 2009, pp. 1–65.

Geerolf, Francois, "Leverage and Disagreement," 2017. Working Paper.

Gorton, Gary and Andrew Metrick, "Securitized banking and the run on repo," *Journal of Financial Economics*, 2012, *104* (3), 425–451.

Infante, Sebastian, "Liquidity Windfalls: The Consequences of Repo Rehypothecation," *Unpublished Manuscript*, 2015.

Kahn, Charles M. and Hye Jin Park, "Collateral, Rehypothecation, and Efficiency," *Unpublished Manuscript*, 2015.

Krishnamurthy, Arvind, Stefan Nagel, and Dmitry Orlov, "Sizing Up Repo," *Journal of Finance*, 2014, *69* (6), 2381–2417.

Lenel, Moritz, "Safe Assets, Collateralized Lending and Monetary Policy." Working Paper.

Maggio, Marco Di and Alireza Tahbaz-Salehi, "Collateral Shortages and Intermediation Networks," *Unpublished Manuscript*, 2015.

Martin, Antoine, David Skeie, and Ernst-Ludwig von Thadden, "Repo Runs," *Review of Financial Studies*, 2014, *27* (4), 957–989.

Muley, Ameya, "Collateral Reuse in Shadow Banking and Monetary Policy," *Unpublished Manuscript*, 2016.

Reinhart, Carmen and Kenneth Rogoff, *This Time is Different: Eight Centuries of Financial Folly*, Princeton University Press, 2009.

Rytchkov, Oleg, "Asset Pricing with Dynamic Margin Constraints," *Journal of Finance*, 2014, *69* (1), 405–452.

Shin, Hyun Song, "Financial Intermediation and the Post-Crisis Financial System," *BIS Annual Conference*, 2009, *8* (304), 1–32.

Simsek, Alp, "Belief Disagreements and Collateral Constraints," *Econometrica*, 2013, *81* (1), 1–53.

Singh, Manmohan, "Velocity of Pledged Collateral: Analysis and Implications," *IMF Working Papers*, 2011, *11* (256).

Appendix

Proof of Proposition 2.1:

Proof. Observe that $A^1 = \{\min\{\theta, s\} : \theta \in S\}$. The payoff of an asset in A^2 is then $\min\{\theta_1, \min\{\theta_2, s\}\} = \{\min\{\theta_1, \theta_2\}, s\}$, but this asset is in A^1. □

Proof of claim in Section 2.1:

Proof. Suppose first that the constraint $a + \int\limits_{\theta \geq \hat{\theta}} d\mu_+(\theta) \geq \int\limits_{\theta \geq \hat{\theta}} d\mu_-(\theta)$ is satisfied for all $\hat{\theta}$. Define $h(\hat{\theta}) = \inf\limits_{\theta}(a + \int\limits_{\theta' \geq \theta} d\mu_+(\theta') \leq \int\limits_{\theta' \geq \hat{\theta}} d\mu_-(\theta'))$. Then observe that whenever $\mu_-([\theta_1, \theta_2]) > 0$, it must be that $\mu_+([h(\theta_1), h(\theta_2)]) > 0$ as well. Equivalently, defining a new measure $\tilde{\mu}_+([\theta_1, \theta_2]) = \mu_+([h(\theta_1), h(\theta_2)])$, μ_- is absolutely continuous with respect to $\tilde{\mu}_+$. By the Radon-Nikodym theorem, there exists a function $\beta(\theta)$ such that $\int\limits_U d\mu_-(\theta) = \int\limits_U \beta(\theta) d\tilde{\mu}_+(\theta)$ for any measurable U. Let $\alpha(\theta, \theta') = \mathbf{1}\{\theta = h(\theta')\}\beta(\theta')$. Then

$$\int\limits_{\theta, \theta' \in [\theta_1, \theta_2]} \alpha(\theta, \theta') d\mu_+(\theta) = \int\limits_{\theta' \in [\theta_1, \theta_2]} \beta(\theta') d\mu_+(h(\theta')) = \int\limits_{\theta' \in [\theta_1, \theta_2]} \beta(\theta') d\tilde{\mu}_+(\theta') = \mu_-([\theta_1, \theta_2])$$

Define $\mu(U) = \int\limits_U \alpha(\theta, \theta') d\mu_+(\theta)$. Then $\mu \geq 0$, as in the second optimization problem.

Now suppose $\mu \geq 0$, and let $\mu_+(U) = \int\limits_{\theta \in U, \theta'} d\mu(\theta, \theta')$, $\mu_-(U) = \int\limits_{\theta \in U, \theta'} d\mu(\theta', \theta)$. Then

$$\int\limits_{\theta \geq \hat{\theta}} d\mu_+(\theta) = \int\limits_{\theta \geq \hat{\theta}, \theta'} d\mu(\theta, \theta') \geq \int\limits_{\theta, \theta' \geq \hat{\theta}} d\mu(\theta, \theta') = \int\limits_{\theta \geq \hat{\theta}} d\mu_-(\theta)$$

using the fact that μ puts zero weight on sets $[\theta_1, \theta_2] \times [\theta_1', \theta_2']$ with $\theta_1' > \theta_2$. □

Proof of Lemma 3.1:

Proof. Define

$$f_{nm}(\theta, \theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] - \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$$

Observe that $f_{nm}(\theta, \theta) = 0$ for all θ. The derivatives of f_{nm} are

$$f_{nm,1}(\theta, \theta') = \lambda_n(1 - F_n(\theta)) - \lambda_m(1 - F_m(\theta))$$

$$f_{nm,2}(\theta, \theta') = \lambda_m(1 - F_m(\theta')) - \lambda_n(1 - F_n(\theta'))$$

The hazard rate inequality implies $\frac{1-F_n(\theta)}{1-F_m(\theta)}$ is an increasing function of θ, so if $f_{nm,1}(0,0) < 0$ and $f_{nm,1}(\overline{s}, \overline{s}) > 0$, there exists θ^*_{nm} such that

- $\frac{1-F_n(\theta^*_{nm})}{1-F_m(\theta^*_{nm})} = \frac{\lambda_m}{\lambda_n}$,
  - $f_{nm,1}(\theta, \theta') > 0$ iff $\theta > \theta^*_{nm}$,
  - $f_{nm,2}(\theta, \theta') > 0$ iff $\theta < \theta^*_{nm}$.

For $\theta < \theta^*_{nm}$, define $\overline{g}_{nm}(\theta)$ to be the $\hat{\theta}$ that satisfies

$$\int_{\theta^*_{nm}}^{\hat{\theta}} f_{nm,1}(\theta', \theta) d\theta' = -\int_{\theta}^{\theta^*_{nm}} f_{nm,1}(\theta', \theta) d\theta'$$

if it exists and $\overline{g}_{nm}(\theta) = \overline{s}$ otherwise. Note that the integral on the left-hand side is positive and the one on the right is negative. The value $\hat{\theta} = \overline{g}_{nm}(\theta)$ is weakly decreasing in θ because the integral on the right-hand side is increasing in θ. By the fundamental theorem of calculus, $f_{nm}(\hat{\theta}, \theta) = 0$, and for $\theta' > \overline{g}_{nm}(\theta)$, $f_{nm}(\theta', \theta) > 0$ (because the derivative $f_{nm,1}$ is positive to the right of θ^*_{nm}). Hence statement (1) in the lemma has been proved. Statement (2) follows in an analogous fashion.

Statement (3) in the lemma follows from the observation that $f_{nm}(\theta, \theta) = 0$ and $f_{nm,1}(\theta', \theta) > 0$ for all $\theta' \geq \theta$ implies $f_{nm}(\theta', \theta) > 0$ for all $\theta' \geq \theta$. Given that $f_{nm,2} > 0$ for $\theta' < \theta^*_{nm}$, it is always the case that $f_{nm}(\theta', \theta) < 0$ for $\theta < \theta'$. □

Proof of Proposition 3.2:

Proof. Let $\{r_n\}_{n=1}^N$ be the equilibrium values of

$$\max\left\{1, \max_{(\theta,\theta')} \frac{E_n[\min\{\theta, s\} - \min\{\theta', s\}]}{q(\theta) - q(\theta')}\right\}$$

and set $\lambda_n = \frac{1}{r_n}$. Then for each (θ, θ'),

$$q(\theta) - q(\theta') \geq \max_m \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$$

Define $\{\theta_n^*\}_{n=1}^N, \{\Theta_n^*\}_{n=1}^N$ in the same way as in Section 3.1, and let

$$\Theta_n^{opt} = \{(\theta, \theta') : q(\theta) - q(\theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}]\}$$

Note that the results of Section 3.1 imply $\Theta_n^{opt} \subset \Theta_n^*$. Optimality implies $\mu_n(\hat{A} \backslash \Theta_n^{opt}) = 0$ for all n, so it must be that $\mu_n(\hat{A} \backslash \Theta_n^*) = 0$. Then for any agent n, given $s < \theta_n^*$,

$$\int (\min\{\theta, s\} - \min\{\theta', s\}) d\mu_n(\theta, \theta') = \int (\min\{\theta, s\} - \min\{\theta', s\}) \mathbf{1}(\theta' < \theta_n^*) d\mu_n(\theta, \theta') = 0$$

Furthermore, for any $s > \theta_{n-1}^*$,

$$\int (\min\{\theta, s\} - \min\{\theta', s\}) d\mu_n(\theta, \theta') = \int (\theta - \theta') d\mu_n(\theta, \theta') \equiv c_n$$

That is, if $s > \theta_{n-1}^*$, agent n's payoff is equal to a constant that is independent of s. Agent n bears risk only in the region $\theta_n^* \leq s \leq \theta_{n-1}^*$.

Agent n's payoff is 0 for $s \leq \theta_n^*$, some linear function $a_n s + b_n$ for $\theta_n^* \leq s \leq \theta_{n-1}^*$, and a constant c_n for $s \geq \theta_{n-1}^*$. The conditions that the payoff should be piecewise linear and continuous come from the fact that the payoffs of all assets (θ, θ') are piecewise linear and continuous. Continuity implies $b_n = -a_n \theta_n^*$ and $c_n = a_n(\theta_{n-1}^* - \theta_n^*)$. Market clearing will pin down the constant a_n.

For $\theta_n^* < s < \theta_{n-1}^*$, the market clearing condition is

$$\begin{aligned} s &= \sum_{m=1}^{N} \int (\min\{\theta, s\} - \min\{\theta', s\}) d\mu_m(\theta, \theta') \\ &= a_n(s - \theta_n^*) + \sum_{m>n} a_m(\theta_{m-1}^* - \theta_m^*) \\ &= a_n s + C \end{aligned}$$

Equating coefficients on the left- and right-hand sides, it immediately follows that $a_n = 1$ for all n. Thus agent n's payoff is zero for $s \leq \theta_n^*$, $s - \theta_n^*$ for $\theta_n^* \leq s \leq \theta_{n-1}^*$, and $\theta_{n-1}^* - \theta_n^*$ for $s \geq \theta_{n-1}^*$, just as in the social planner's problem.

In this setting, the law of one price holds: assets with equal payoffs have equal prices. The total price agent n pays to acquire assets is then

$$q(\theta_{n-1}^*) - q(\theta_n^*) = \lambda_n E_n[\min\{\theta_n^*, s\} - \min\{\theta_{n-1}^*, s\}]$$

If there is $[\theta', \theta] \subset [\theta_n^*, \theta_{n-1}^*]$ such that

$$q(\theta) - q(\theta') > \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}]$$

it follows that

$$\begin{aligned} q(\theta_{n-1}^*) - q(\theta_n^*) &= q(\theta_{n-1}^*) - q(\theta) + q(\theta) - q(\theta') + q(\theta') - q(\theta_n^*) \\ &> \lambda_n \Big(E_n[\min\{\theta_n^*, s\} - \min\{\theta, s\}] + E_n[\min\{\theta, s\} - \min\{\theta', s\}] \\ &+ E_n[\min\{\theta', s\} - \min\{\theta_{n-1}^*, s\}] \Big) \\ &= \lambda_n E_n[\min\{\theta_n^*, s\} - \min\{\theta_{n-1}^*, s\}] \end{aligned}$$

yielding a contradiction. Hence $q(\theta) - q(\theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ for all $[\theta', \theta] \subset [\theta_n^*, \theta_{n-1}^*]$, as desired. □

Proof of Proposition 3.3: Existence

Proof. The proof proceeds by an application of Brouwer's Fixed Point Theorem. Let $S = \{(\theta_1, \ldots, \theta_{N-1}) : 0 \leq \theta_{N-1} \leq \cdots \leq \theta_1 \leq \overline{s}\}$. Define a map $f : S \to S$ as follows:

1. Set

$$u_1(\theta_1) = E_1[\max\{s - \theta_1, 0\}]$$

$$u_n(\theta_{n-1}, \theta_n) = E_n[\min\{\theta_{n-1}, s\} - \min\{\theta_n, s\}],\ 1 < n < N$$

$$u_N(\theta_{N-1}) = E_N[\min\{\theta_{N-1}, s\}]$$

   (a) Define $\lambda_n = (\max\{\frac{u_n}{w_n}, 1\})^{-1}$ for $1 \leq n \leq N$.

   (b) Solve the problem

$$\max_{\overline{s} \geq \tilde{\theta}_1 \geq \cdots \geq \tilde{\theta}_N \geq 0} \lambda_1 E_1[\max\{s - \tilde{\theta}_1, 0\}] + \sum_{n=2}^{N-1} \lambda_n E_m[\min\{\tilde{\theta}_{n-1}, s\} - \min\{\tilde{\theta}_n, s\}] + \lambda_N E_N[\min\{\theta_{N-1}$$

   in order to obtain the optimum $(\theta_1', \ldots, \theta_{N-1}')$.

   (c) Define $f(\theta_1, \ldots, \theta_{N-1}) = (\theta_1', \ldots, \theta_{N-1}')$.

The map f is continuous and the values $\theta_1, \ldots, \theta_{N-1}$ correspond to an equilibrium in the sense of Proposition 3.2, as will be shown below.

To see that f is continuous, note that f is a composition of the function g that maps $(\theta_1, \ldots, \theta_{N-1})$ to $(\lambda_1, \ldots, \lambda_N)$ with the function h that maps $(\lambda_1, \ldots, \lambda_N)$

to $(\theta'_1, \ldots, \theta'_{N-1})$. Note that g is itself a composition of a continuous function of $(\theta_1, \ldots, \theta_{N-1})$ and the max function, so g is continuous.

The proof that h is continuous is slightly more involved. For $n > m$, define θ^*_{nm} to be the unique solution to the equation

$$\frac{\lambda_m}{\lambda_n} = \frac{1 - F_n(\theta)}{1 - F_m(\theta)} \equiv \psi(\theta)$$

if such a value θ^*_{nm} exists. Otherwise, if $\frac{\lambda_m}{\lambda_n} > \psi(\theta)$ for all θ, set $\theta^*_{nm} = \overline{s}$. Finally, if $\frac{\lambda_m}{\lambda_n} < \psi(\theta)$ for all θ, set $\theta^*_{nm} = 0$. As long as $\psi(\theta)$ is continuous, this mapping from λ_n, λ_m to θ^*_{nm} is also continuous. As shown in Section 3.2, the solution $(\theta'_1, \ldots, \theta'_{N-1})$ can be expressed recursively as $\theta'_{N-1} = \theta^*_{N-1,N}$, $\theta'_n = \max\{\theta_{n-1}, \max_{m<n} \theta^*_{nm}\}$. This defines a continuous function from $\{\theta^*_{nm}\}_{n>m}$ to $(\theta'_1, \ldots, \theta'_{N-1})$, so h is continuous as well, meaning f is continuous.

The function $f : S \to S$ is therefore a continuous function on a compact, convex set in $\mathbb{R}^{N-1}$, so f has a fixed point $\theta^* = (\theta^*_1, \ldots, \theta^*_{N-1})$ by Brouwer's Fixed Point Theorem. It remains to show that this fixed point corresponds to an equilibrium.

With some abuse of notation, let $\lambda_n = (\max\{\frac{u_n(\theta^*)}{w_n}, 1\})^{-1}$, and set cash holdings $c_n = w_n - \lambda_n u_n(\theta^*)$. Furthermore, let $\mu_n(\theta^*_{n-1}, \theta^*_n) = 1$ (with $\theta^*_0 = \overline{s}$, $\theta^*_N = 0$). Recall that

$$\lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}] = \max_m \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$$

as long as $[\theta', \theta] \subset [\theta^*_n, \theta^*_{n-1}]$. Then set $q(\theta) - q(\theta') = \lambda_n E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ for all such pairs (θ, θ'). Clearly, then, $q(\theta) - q(\theta') \geq \max_m \lambda_m E_m[\min\{\theta, s\} - \min\{\theta', s\}]$, as required in equilibrium. As shown in Section 3.2, (μ_n, c_n) solves type n's optimization problem taking $q(\theta)$ as given, since for all $(\theta', \theta) \not\subset [\theta^*_n, \theta^*_{n-1}]$, $\frac{u_n(\theta, \theta')}{q(\theta) - q(\theta')} < \frac{1}{\lambda_n} = \frac{u_n(\theta^*_{n-1}, \theta^*_n)}{q(\theta^*_{n-1}) - q(\theta^*_n)}$ (as long as $\lambda_n > 1$). When $\lambda_n = 1$, the same result holds.

For the converse result, first fix some $\lambda = \{\lambda_n\}_{n=1}^N \in \Lambda$ and let $\Theta^*_n = \{\theta : \lambda_n(1 - F_n(\theta)) = \max_m \lambda_m(1 - F_m(\theta))\}$. Then define $\mu_n(\theta, \theta') = \mathbf{1}\{\theta = \sum \theta^*_n, \theta' = \inf \Theta^*_n\}$. This allocation is identical to the one obtained from the social planner's problem with weights λ. The uniqueness of this mapping follows from the uniqueness of the solution to the planner's problem. □

Proof of Proposition 3.3: Uniqueness

Proof. Consider two equilibria (1 and 2) such that n_1 types invest all of their wealth in the underlying asset or debt contracts in equilibrium 1 and $n_2 \geq n_1$ types do so in equilibrium 2. Suppose that, as in Proposition 3.2, the cutoff values $\{\theta^i_j\}$ correspond

to equilibrium i. That is, in equilibrium i, the price function $q_i(\theta)$ satisfies

$$q_i(\theta) - q_i(\theta') = \lambda_j E_j[\min\{\theta, s\} - \min\{\theta', s\}]$$

whenever $[\theta', \theta] \subset [\theta^i_j, \theta^i_{j-1}]$. Let k be the lowest integer such that $\theta^1_k \neq \theta^2_k$.

First consider the case in which $\theta^1_k < \theta^2_k$. Define

$$v^i_j = \prod_{l=1}^{n_i} \frac{1 - F_l(\theta^i_l)}{1 - F_{l+1}(\theta^i_l)}$$

The k-th budget constraint in equilibrium i is

$$w_k = \frac{1}{v^i_{k+1}} \frac{1 - F_{k+1}(\theta^i_k)}{1 - F_k(\theta^i_k)} E_k[\min\{\theta^i_{k-1}, s\} - \min\{\theta^2_k, i\}]$$

Since $\theta^1_k < \theta^2_k$, $\theta^1_{k-1} = \theta^2_{k-1}$, it must be that $v^1_{k+1} > v^2_{k+1}$. Now assume $\theta^1_j < \theta^2_j$, $v^1_{j+1} > v^2_{j+1}$ for some $j > k$. Then

$$w_{j+1} = \frac{1}{v^i_{j+2}} \frac{1 - F_{j+2}(\theta^i_{j+1})}{1 - F_{j+1}(\theta^i_{j+1})} E_{j+1}[\min\{\theta^i_j, s\} - \min\{\theta^i_{j+1}, s\}]$$

This equation implies $\theta^1_{j+1} < \theta^2_{j+1}$, since the right-hand side is decreasing in v^i_{j+1} and increasing in θ^i_j. It also implies $v^1_{j+2} > v^2_{j+2}$ once we have $\theta^1_{j+1} < \theta^2_{j+1}$. By induction, then, $v^1_{n_1+1} > v^2_{n_1+1}$, but this is impossible because $v^1_{n_1+1} = 1$ (since in equilibrium 1, type $n+1$ holds cash).

Now suppose $\theta^1_k > \theta^2_k$. By the same line of reasoning as above, $v^1_{n_1+1} < v^2_{n_1+1}$ and $\theta^1_{n_1+1} > \theta^2_{n_1+1}$. Given that type $n_1 + 1$ invests all wealth in the underlying asset or debt contracts in equilibrium 2 but not in equilibrium 1, it must be that

$$w_{n_1+1} = \frac{1}{v^2_{n_1+1}} E_{n_1+1}[\min\{\theta^2_{n_1}, s\} - \min\{\theta^2_{n_1+1}, s\}] > E_{n_1+1}[\min\{\theta^1_{n_1+1}, s\}]$$

Given that $\theta^2_{n_1+1} > 0$, this is clearly impossible if $v^1_{n_1+1} < v^2_{n_1+1}$ and $\theta^1_{n_1+1} > \theta^2_{n_1+1}$. Therefore $\theta^1_k > \theta^2_k$ is also impossible, so two distinct equilibria cannot exist.

I have omitted the proof of uniqueness in some edge cases, but those cases follow using exactly the same reasoning as in the above arguments. □

Proof of Proposition 4.2

Proof. Suppose that when endowments are $\{w_i\}_{i=1}^N$, the equilibrium default cutoffs are $\{\theta^*_i\}_{i=1}^{N-1}$ and the most pessimistic type who invests is K. Now assume that when

the wealth of type $m > 1$ is $w'_m > w_m$, the new equilibrium default cutoffs are $\{\theta_i^{*'}\}_{i=1}^N$ and that $\theta_1^{*'} < \theta_1^*$.

Note that in the new equilibrium, all agents $n < m$ must participate, and $\frac{w_1}{w'_n} \leq \frac{w_1}{w_n}$ for all n. Then, denoting $E_n[\min\{\theta, s\} - \min\{\theta', s\}]$ by $E_n(\theta, \theta')$,

$$\begin{aligned}\frac{w_1}{w_2} &= \frac{1 - F_2(\theta_1^{*'})}{1 - F_1(\theta_1^{*'})} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_2(\theta_1^{*'}, \theta_2^{*'})} \\ &\geq \frac{1 - F_2(\theta_1^*)}{1 - F_1(\theta_1^*)} \frac{E_1(\overline{s}, \theta_1^*)}{E_2(\theta_1^*, \theta_2^{*'})}\end{aligned}$$

so it must be that $\theta_2^{*'} < \theta_2^*$. Assume by way of induction that $\theta_k^{*'} < \theta_k^*$ for all $k \leq n$, and observe that

$$\begin{aligned}\frac{w_1}{w'_n} &= \prod_{i=1}^{n} \frac{1 - F_i(\theta_{i-1}^{*'})}{1 - F_{i-1}(\theta_{i-1}^{*'})} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_m(\theta_{n-1}^{*'}, \theta_n^{*'})} \\ &> \prod_{i=1}^{n} \frac{1 - F_i(\theta_{i-1}^*)}{1 - F_{i-1}(\theta_{i-1}^*)} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_m(\theta_{n-1}^*, \theta_n^{*'})}\end{aligned}$$

where $\theta_n^{*'} \equiv 0$ if n holds only cash in the new equilibrium. On the other hand,

$$\frac{w_1}{w'_n} \leq \frac{w_1}{w_n} = \prod_{i=1}^{n} \frac{1 - F_i(\theta_{i-1}^*)}{1 - F_{i-1}(\theta_{i-1}^*)} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_m(\theta_{n-1}^*, \theta_n^*)}$$

so $\theta_n^{*'} < \theta_n^*$. This is true for all $n \leq K$, so

$$w_1 = \prod_{i=1}^{K} \frac{1 - F_i(\theta_{i-1}^{*'})}{1 - F_{i-1}(\theta_{i-1}^{*'})} E_1(\overline{s}, \theta_1^{*'}) < \prod_{i=1}^{K} \frac{1 - F_i(\theta_{i-1}^*)}{1 - F_{i-1}(\theta_{i-1}^*)} E_1(\overline{s}, \theta_1^*)$$

a contradiction. Thus $\theta_1^{*'} \geq \theta_1^*$. The same argument by contradiction then applies if it is assumed that $\theta_2^{*'} < \theta_2^*$, and so on up to $\theta_{m-1}^{*'}$.

Now suppose $\theta_m^{*'} > \theta_m^*$. Given that $w'_m > w_m$, $\frac{w'_m}{w_n} > \frac{w_m}{w_n}$ for all $n > m$. Then

$$\begin{aligned}\frac{w_1}{w_{m+1}} &= \prod_{i=1}^{m+1} \frac{1 - F_i(\theta_{i-1}^{*'})}{1 - F_{i-1}(\theta_{i-1}^{*'})} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_{m+1}(\theta_m^{*'}, \theta_{m+1}^{*'})} \\ &> \prod_{i=1}^{m+1} \frac{1 - F_i(\theta_{i-1}^*)}{1 - F_{i-1}(\theta_{i-1}^*)} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_{m+1}(\theta_m^*, \theta_{m+1}^{*'})}\end{aligned}$$

but

$$\frac{w_1}{w_{m+1}} = \prod_{i=1}^{m+1} \frac{1 - F_i(\theta^*_{i-1})}{1 - F_{i-1}(\theta^*_{i-1})} \frac{E_1(\overline{s}, \theta_1^{*'})}{E_{m+1}(\theta^*_m, \theta^*_{m+1})}$$

so it must be that $\theta^{*'}_{m+1} > \theta^*_{m+1}$. By induction, it is possible to show in exactly the same way that $\theta^{*'}_n > \theta^*_n$ for all $n > m$. Again, this yields a contradiction because

$$w_1 = \prod_{i=1}^{K} \frac{1 - F_i(\theta^{*'}_{i-1})}{1 - F_{i-1}(\theta^{*'}_{i-1})} E_1(\overline{s}, \theta_1^{*'}) < \prod_{i=1}^{K} \frac{1 - F_i(\theta^*_{i-1})}{1 - F_{i-1}(\theta^*_{i-1})} E_1(\overline{s}, \theta^*_1) = w_1$$

so it must be that $\theta^{*'}_{m+1} \leq \theta^*_{m+1}$. This argument can then be repeated to show that $\theta^{*'}_{m+2} \leq \theta^*_{m+2}$ once it is known that $\theta^{*'}_{m+1} \leq \theta^*_{m+1}$ and so forth until concluding that $\theta^{*'}_n \leq \theta^*_n$ for all $n > m$. □

Proof of Proposition 4.4

Proof. Recall that $\theta^*_{i-1}, \theta^*_i$ satisfy $\frac{r_{i-1}}{r_i} = \frac{1-F_{i-1}(\theta^*_{i-1})}{1-F_i(\theta^*_{i-1})}$, $\frac{r_i}{r_{i+1}} = \frac{1-F_i(\theta^*_i)}{1-F_{i+1}(\theta^*_i)}$. The proposition then follows immediately from the observation that $\frac{1-F_j(s)}{1-F_k(s)}$ is increasing in s for $j < k$ (which, in turn, follows from Assumption 1). □

Proof of Proposition 5.1

Proof. First, observe that given θ^*_{N-1}, it is possible to derive the default thresholds θ^*_n of all other agents using the equilibrium conditions

$$w_n = \left(\prod_{k=n}^{N-1} \frac{1 - F_{k+1}(\theta^*_k)}{1 - F_k(\theta^*_k)} \right) E_n[\min\{\theta^*_{n-1}, s\} - \min\{\theta^*_n, s\}].$$

Clearly, given θ^*_k for $k > n$, it is possible to derive θ^*_n from this equation. The other boundary condition is that $\theta^*_0 = \overline{s}$. Hence, finding an equilibrium reduces to finding a θ^*_{N-1} that yields $\theta^*_0 = \overline{s}$.

Second, observe that for agent $N-1$, the equilibrium condition is

$$w_{N-1} = \frac{1 - F_N(\theta^*_{N-1})}{1 - F_{N-1}(\theta^*_{N-1})} E_{N-1}[\min\{\theta^*_{N-2}, s\} - \min\{\theta^*_{N-1}, s\}].$$

If $\tilde{F}_N(s) \geq F_N(s)$ for all s, then given a value of θ^*_{N-1}, the value of θ^*_{N-2} implied by this equation must be larger. Then, a short inductive argument implies that for any n, a larger value of θ^*_{N-1} implies a larger value of θ^*_n. Hence, once beliefs are changed

from F_N to $\tilde{F}_N \geq F_N$, it must be that the equilibrium value of θ^*_{N-1} is lower than the original value in order to ensure $\theta^*_0 = \overline{s}$.

Now note that the expression for margins can be rewritten as

$$m_n = \frac{w_n}{E_N[\min\{\theta^*_{N-1}, s\}] + \sum_{k=N}^{n+1} w_k}$$

since

$$w_n = \frac{1}{r_n} E_n[\min\{\theta^*_{n-1}, s\} - \min\{\theta^*_n, s\}]$$

for $n < N$. We have shown that the equilibrium value of θ^*_{N-1} decreases after beliefs shift to $\tilde{F}_N$, and endowments w_n are exogenous, so in fact $\tilde{m}_n \geq m_n$ for all n, as desired. □

Proof of Proposition 5.2

Proof. Let θ^*_{nA} be the riskiness of the debt contract taken by type $n \in \{2, 1A\}$ in the market for asset A, and define θ^*_{nB} analogously. The equilibrium conditions in this economy are

$$w_2 = \frac{1 - F_3(\theta^*_{2A})}{1 - F_2(\theta^*_{2A})} E_2[\min\{\theta^*_{1A}, s\} - \min\{\theta^*_{2A}, s\}] + \frac{1 - F_3(\theta^*_{2B})}{1 - F_2(\theta^*_{2B})} E_2[\min\{\theta^*_{1A}, s\} - \min\{\theta^*_{2A}, s\}],$$

$$w_{1A} = \frac{1 - F_3(\theta^*_{2A})}{1 - F_2(\theta^*_{2A})} \frac{1 - F_2(\theta^*_{1A})}{1 - F_1(\theta^*_{1A})} E_1[s - \min\{\theta^*_{1A}, s\}],$$

$$w_{1B} = \frac{1 - F_3(\theta^*_{2B})}{1 - F_2(\theta^*_{2B})} \frac{1 - F_2(\theta^*_{1B})}{1 - F_1(\theta^*_{1B})} E_1[s - \min\{\theta^*_{1B}, s\}],$$

and

$$\frac{1 - F_3(\theta^*_{2A})}{1 - F_2(\theta^*_{2A})} = \frac{1 - F_3(\theta^*_{2B})}{1 - F_2(\theta^*_{2B})}.$$

This final condition is key in the proof of the proposition– it equates the intermediary's returns across the different markets. Note that a choice of θ^*_{2A} pins down θ^*_{1A} through the budget constraint of type $1A$, but then this indifference condition pins down θ^*_{2B} as well (and thus θ^*_{1B}). Hence, a choice of θ^*_{2A} pins down the entire equilibrium. Importantly, a lower θ^*_{2A} implies a higher θ^*_{1A}, a higher θ^*_{1B}, and a lower θ^*_{2B}.

If the beliefs of type 3 change such that $\tilde{F}_{3A} \geq F_{3A}$, the values of θ^*_{1A} and θ^*_{1B} implied by a particular choice of θ^*_{2A} decrease. The value of θ^*_{2B} increases. Thus, for a particular choice of θ^*_{2A}, the right-hand side of the budget constraint of type 2 is lower under this change of beliefs. Therefore, θ^*_{2A} (and thus θ^*_{2B}) must be lower in

the new equilibrium. The same technique used in the previous proof can be used to show that this implies higher margins for both the intermediary type 2 as well as both end-borrowers, types $1A$ and $1B$. In short, this is because the prices of both assets decrease, so the wealth of each end-borrower comprises a larger share of the corresponding asset's price. □

Proof of Proposition 5.3

Proof. First note that the condition

$$r_1 = \frac{1 - F_1(\theta_1^*)}{1 - F_2(\theta_1^*)} \frac{1 - F_2(\theta_2^*)}{1 - F_3(\theta_2^*)}$$

immediately implies that the equilibrium value of θ_1^* is decreasing in the value of θ_2^*. Next, note that since r_1 must always be greater than r_2, the price of the asset must be decreasing in θ_1^* and θ_2^*.

When type 2 becomes wealthier, the budget constraint of type 2 implies that for a given value of θ_2^*, it must be that θ_1^* is larger. That is, the curve defined by type 2's budget constraint shifts. The curve defined by type 1's indifference condition does not shift, nor does the price function, so in equilibrium it must be that θ_2^* decreases and θ_1^* increases. This implies that the most pessimistic agent makes a safer loan while the most optimistic agent takes riskier loans, as desired.

□

Proof of Proposition 5.5

Proof. Suppose that in the alternative economy, type 2 agents buy the asset directly. The equilibrium conditions are then

$$w_1 = \alpha \frac{1 - F_3(\theta_1^*)}{1 - F_1(\theta_1^*)} E_1[s - \min\{\theta_1^*, s\}],$$

$$w_2 = (1 - \alpha) \frac{1 - F_3(\theta_2^*)}{1 - F_2(\theta_2^*)} E_2[s - \min\{\theta_2^*, s\}],$$

where $\alpha \in (0, 1)$ represents the share of the asset purchased by type 1, the most optimistic agent.

If in the benchmark economy θ_1^* is smaller than in the alternative economy, $\theta_{1B}^* < \theta_{1A}^*$, then

$$\begin{aligned} w_1 &= \frac{1-F_3(\theta_{2B}^*)}{1-F_2(\theta_{2B}^*)}\frac{1-F_2(\theta_{1B}^*)}{1-F_1(\theta_{1B}^*)}E_1[s-\min\{\theta_{1B}^*,s\}] \\ &> \alpha\frac{1-F_3(\theta_{1B}^*)}{1-F_2(\theta_{1B}^*)}\frac{1-F_2(\theta_{1B}^*)}{1-F_1(\theta_{1B}^*)}E_1[s-\min\{\theta_{1B}^*,s\}] \\ &= \alpha\frac{1-F_1(\theta_{1B}^*)}{1-F_1(\theta_{1B}^*)}E_1[s-\min\{\theta_{1B}^*,s\}] \\ &> \alpha\frac{1-F_3(\theta_{1A}^*)}{1-F_1(\theta_{1A}^*)}E_1[s-\min\{\theta_{1A}^*,s\}] = w_1, \end{aligned}$$

which is a contradiction. This proves that $\theta_{1A}^* \leq \theta_{1B}^*$, so type 1 takes greater risk in the benchmark economy.

If in the alternative economy type 2 agents lend to type 1, then it is automatically true that they take riskier loans in the benchmark economy. In the alternative economy, they do not take loans at all.

□

Chapter 3

Blockchain Economics

3.1 Introduction

Traditionally, records have been maintained by centralized entities. Blockchain has provided us with a radical decentralized alternative to record information. It has the potential to be as groundbreaking as the invention of double-entry bookkeeping in fourteenth-century Italy. Blockchain could revolutionize record-keeping of financial transactions and ownership data.

The central problem in digital record-keeping is how to ensure agents come to a *consensus* on the true history of events. Consensus, in turn, requires that a ledger's record-keepers have the incentive to report honestly, i.e., the ledger should be devoid of fraud. These incentives can be provided in three ways. First, agents may face external punishments for dishonest behavior, which may be social, commercial, or legal in nature. In this case, the ledger is not self-sufficient: a failure of the external punishment mechanism results in a failure of the ledger. Second, record-keepers may be punished through a loss of rents if users of the system abandon it upon discovering fraudulent activity. Third, record-keepers may face physical resource costs to write on the ledger, as in most decentralized blockchains, rendering dishonesty unprofitable from an ex-ante perspective.

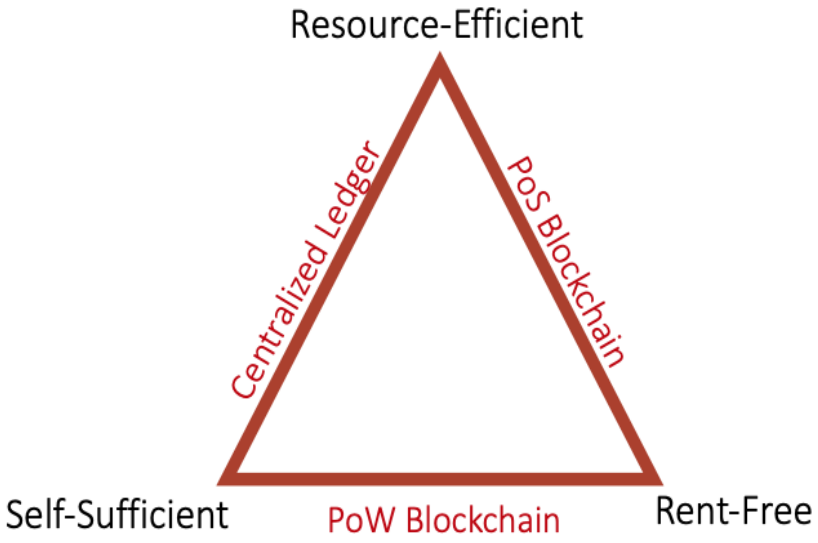


Figure 3.1: The Blockchain Trilemma.

In this paper we prove a "Blockchain Trilemma": consensus requires external punishments, rents, or ex-ante resource costs to write on the ledger. Hence, it is impossible for any digital ledger to simultaneously be (i) self-sufficient, (ii) rent-free, and (iii) resource-efficient.

In centralized record-keeping systems, consensus is achieved through *trust* in the record-keeper. Agents simply ask the record-keeper to report the history of events to them, giving the record-keeper ample opportunity to benefit by misreporting. Trust can emerge from the record-keeping system itself if the record-keeper earns sufficient rents to keep it honest. Trust can also stem from external mechanisms: there may be legal authorities to punish fraud or commercial relationships that incentivize the record-keeper to keep its reputation intact.

Blockchains seek to minimize the role of trust in achieving consensus. A blockchain is a type of *distributed ledger* written by decentralized and usually anonymous groups of agents rather than known centralized parties. In a decentralized setting, users of the system may be faced with multiple internally consistent but mutually conflicting ledgers, making consensus all the more important. Record-keepers effectively "vote" on the history of events that they believe to be correct, and users of the system then aggregate those votes to determine the current state using a *consensus algorithm*.

The consensus algorithms used by some blockchains are completely objective in nature: they permit any two agents looking at the same set of ledgers to come to the same conclusion regarding the current state. The most popular blockchain consensus

algorithm, called proof-of-work (PoW), has this property. Anonymous record-keepers (known as "miners") effectively vote on the true state (i.e., a chain of blocks) by extending that chain, which in turn requires an expenditure of computational power. When deciding the true state, agents simply look for the chain of blocks to which the greatest amount of computational power has been contributed. A new user of the system with no prior knowledge of the state, therefore, would come to the same conclusion as all others. Furthermore, this consensus algorithm disincentivizes misbehavior by making it costly for any agent to alter the state, so there is no need for trust in any particular entity.

Some blockchain consensus algorithms have sought to eliminate the resource costs entailed by the PoW algorithm. The most popular such consensus algorithm, known as proof-of-stake (PoS), instead allocates voting power based on the number of tokens held in each account. It is not costly to vote, so voting is secure only to the extent that the record-keepers can be punished for casting conflicting votes. Record-keepers are identified only by the pseudonyms attached to their accounts. Once the accounts that held a majority of the votes at some point in the past are empty, there is no way to guarantee that the owners of those accounts will not misbehave. They may conspire to use their historical majority to produce an alternate history that, to a new user, looks indistinguishable from the true history. To protect against such "long-range" attacks, new users need access to a trusted source of information in order to begin using the system, although no trust is required thereafter. Therefore, PoS blockchains require some external social trust and are not self-sufficient.

We formalize these intuitions underlying our Trilemma by building a general model of record-keeping. Agents in our model keep track of transfers of tokens using a digital record-keeping system. In order to focus on the issue of achieving consensus, we keep the environment simple. There are two periods and three agents. Initially, only two of the agents are present– there is a buyer and seller who can transact, but a record of the transaction must be kept in order to incentivize the seller to participate. In the second period, the third agent (a "new user") arrives and needs to be updated on the state of the ledger. Agents who agree on the state of the ledger in the second period are able to generate a surplus, representing continued interactions using the digital record-keeping system. The key property of this environment is *asynchronous* interaction between agents. The new user is not present when the buyer and seller interact in the first period, raising the possibility that the buyer and seller agree to a transaction, but that when the new user arrives, she is presented with a conflicting

version of the ledger in which the buyer never spent tokens in the first place. This is the *double-spending* problem that complicates digital record-keeping.

While our environment is simple, our model of communication is general enough to allow for essentially arbitrary forms of digital record-keeping. Our model can capture record-keeping by a single centralized entity, a group of known entities, or anonymous entities. Different systems of record-keeping correspond to different ways of allocating voting power on the ledger and forming a consensus on the current state. In a fully centralized record-keeping system, a single entity votes by signing the ledger with a digital signature. By contrast, in a PoW system, any agent may vote by paying a computational cost, and in a PoS system, voting power is given to token holders. Resource costs correspond to proof-of-work, rents are associated with surplus accruing to record-keepers, and external punishment (trust) comes from external social interactions in which agents have mutually beneficial relationships that may break down.

Importantly, in our model, nothing links agents directly to digital identities (e.g., accounts or digital signatures). That is, the digital record-keeping system is *pseudonymous*, creating the possibility that agents may generate alternate, internally consistent ledgers by using pseudonyms to simulate the behavior of honest record-keepers who observed the corresponding alternate history. In a centralized system, the record-keeper could, for example, circulate different ledgers with different groups, always maintaining with each group that any other ledgers it receives are forgeries. In a PoW system, any agent may create an alternate history by acquiring enough computing power (simulating the behavior of a large group of "miners"), and in a PoS system, agents who held a majority of votes in the past may engage in a long-range attack. Formally, we prove that in the class of models we consider, a *double-spending lemma* obtains: the buyer in the initial period may, in principle, create a ledger that appears indistinguishable from the true consensus ledger. In this alternate ledger, the tokens transacted in the first period are sent to one of the buyer's pseudonyms (i.e., accounts) rather than to the seller, effectively allowing the buyer to spend those tokens twice.

The Trilemma is a consequence of the double-spending lemma. We consider the situation faced by a new user of the system who sees conflicting ledgers reflecting two different histories: the true history and an alternate history created by an attacker (the buyer). If the two ledgers contain identical sets of votes, the new user will have no a priori way to distinguish between the two ledgers. If the user has an external trusted source of information (e.g. a social connection), she may ask which one is the true ledger, but in order to rely on this information, she must have some

capacity to punish that source if the information turns out to be false. In this case, the ledger is not self-sufficient because honest record-keeping relies on punishment mechanisms external to it. Otherwise, there is the possibility that the new user may be fooled by the attacker's alternate history. Attacks can then be dissuaded in two ways: via an ex-ante cost to create an alternate history or via an ex-post punishment. Ex-ante costs can be imposed if the record-keeping system requires costly computations (proof-of-work) to achieve consensus, in which case the system is not resource-efficient. It is possible to inflict ex-post punishments if record-keeping is performed by an entity that extracts rents from users of the system. In this case, a double-spending attack could cause the attacker to lose those rents, since such an attack would produce evidence that the record-keeper deviated. Ex-post punishments are possible only if the record-keeping system is not rent-free.

After proving the Trilemma, we discuss important potential extensions of our model. While our benchmark model has only three agents and two periods, this discussion addresses how our results naturally extend to an environment with an arbitrary number of agents. We also informally discuss an infinite-horizon environment that generates the same reduced-form payoff structure as our benchmark model.

We further make the important point that while blockchains guarantee transfers of *ownership*, some sort of enforcement is required to ensure transfers of *possession*. For example, in a housing market, the owner of the house is the person whose name is on the deed, but the possessor of the house is the person who resides in it. The buyer of the deed needs to be certain that once she holds the deed, her ownership of the house will be enforced. In the stock market, the purchaser of a share has ownership of future dividends but not necessarily possession, since the delivery of dividends needs to be enforced. Broadly, blockchains can record obligations. Currency, for example, is special because its value derives only from the fact that it can be passed on, so no obligations need to be enforced. Punishing those who default on their obligations is another matter: it typically requires a trusted legal enforcement entity. We propose that it may be beneficial to bundle record-keeping and enforcement duties given that the punishment mechanisms that create trust in the enforcer also allow for trust in its role as record-keeper.

Related Literature. Our paper is related to the emergent literature on the economic properties and implications of blockchains. The paper most closely related to ours is Budish (2018), which studies the costs of incentivizing honesty for cryptocurrency blockchains in isolation, whereas our work compares the cost and incentive schemes required to secure both centralized and decentralized record-keeping systems.

Biais et al. (2019) study coordination among miners in a blockchain-based system. They show that while the strategy of mining the longest chain proposed by Nakamoto (2008) is in fact an equilibrium, there are other equilibria in which the blockchain forks, as observed empirically. We study forks as well, in the sense that agents in our model may attempt to defraud others by forking the blockchain. Cong and He (2019) focus mostly on the issue of how ledger transparency leads to a greater scope for collusion between users of the system, placing more emphasis on users' perspective than on record-keepers', which is where our interest lies.

Some of the recent literature on blockchains in economics focuses on the security and the costs of the system. Huberman, Leshno, and Moallemi (2017) study transaction fees in Bitcoin and compare that environment to one with a monopolistic intermediary. They emphasize the role of free entry and conclude that the blockchain market structure completely eliminates the rents that a monopolist would extract in an identical market, whereas we focus on the role of rents as a disciplining device for centralized record-keepers and compare that to the incentive provision mechanisms of decentralized blockchains. Easley, O'Hara, and Basu (2017) use a game-theoretic framework to analyze the emergence of transaction fees in Bitcoin and the implications of these fees for mining costs. The R&D race between Bitcoin mining pools is described in Gans, Ma, and Tourky (2018), who argue that regulation of Bitcoin mining would reduce the overall costs of the system and improve welfare. We focus on the key economic role of mining costs, which is to dissuade agents from creating alternate histories of events for their own benefit.

We also relate to the literature on cryptocurrencies. Chiu and Koeppl (2017) develop a macroeconomic model in which the sizes of cryptocurrency transactions are capped by the possibility of an attack on the blockchain and derive optimal compensation schemes for record-keepers. Schilling and Uhlig (2018) study cryptocurrency pricing in a monetary model and derive necessary conditions for speculation to occur in equilibrium. Pagnotta and Buraschi (2018) derive a pricing framework for cryptocurrencies that explicitly accounts for the interplay between demand for the currency and the cryptographic security provided by miners.

The seminal paper in the computer science literature on consensus algorithms is Lamport, Shostak, and Pease (1980), who showed that consensus is impossible when one-third of agents behave dishonestly. Our paper builds on their argument in an environment in which economic agents act strategically (as opposed to their environment, in which communication devices may simply be faulty). Recent computer science literature has also studied blockchain security extensively. Most papers in

computer science, such as that by Gervais et al. (2016), study how to defend against "double-spend" attacks or other types of attacks that could be undertaken by a single individual who holds control over a large portion of the network's computing power. The conclusion of studies in the computer science literature is that a large fraction of the blockchain record-keepers must always play honestly in order for the network to be secure. In contrast, we do not assume any record-keepers are compelled to play honestly. Rather, agents are permitted to act and collude in arbitrary ways, and they have well-defined incentives to behave as the system's protocol dictates they should. Our model shows that the real cost of operating a PoW blockchain is intrinsically linked to the benefit of a successful attack. It also outlines the types of incentives required for a PoS blockchain to operate correctly.

The rest of the paper is structured as follows. Section 3.2 discusses the basics of blockchain technology and introduces some concepts and notation used in the model. Section 3.3 develops intuition for the trilemma by providing examples of the opportunities for fraud and the incentive schemes used in centralized record-keeping systems as well as PoW and PoS blockchains. Section 3.4 presents our general model of record-keeping. Section 3.5 and proves the double-spending lemma and the Blockchain Trilemma. Section 3.6 discusses other issues related to blockchain from a less formal perspective as well as extensions to our model. Section 3.7 concludes.

3.2 Digital Ledger Technology

In this section, we provide an introduction to digital ledger technology. We give an overview of common aspects of digital ledgers, including blockchains, in order to motivate our model's formulation of these record-keeping systems and the underlying problems that make digital record-keeping challenging. Throughout, we will work with a simple example to build intuition for the concepts introduced. In our example, the ledger will keep track of digital token holdings, but the concepts we discuss should be taken to apply to any ledger that tracks holdings of any arbitrary set of digital assets.

3.2.1 The digital record-keeping problem

The motivation behind blockchains, and digital record-keeping algorithms in general, is that it is often necessary to maintain ledgers that keep track of sequences of events. In particular, it is often important to determine whether one event occurred before

or after another. A canonical example of the necessity of sequential record-keeping arises in the context of digital assets (often referred to as coins, or tokens).

In account-based digital token systems, users typically have secure accounts. These are often protected by *public key encryption*. With public key encryption, it is possible for users to sign messages with a signature that publicly proves they own an account without revealing the secret "password" (private key) that grants access to the account. Tokens are transferred by messages: the owner of an account can send a signed message indicating a transfer from her own account to another. Thus, it is not possible to spend others' tokens.

Unlike physical assets, however, the transfer of digital assets does not inherently preclude their reuse by the original owner. That is, there is no fundamental scarcity of digital assets. To see this point clearly, we consider a simple example. There are three agents, Alice (A), Bob (B), and Carol (C). Alice wants to purchase goods from Bob using digital tokens, keeping a record of the transaction so that all three agents can agree on the transfer of tokens that took place.

The digital ledger keeps track of the quantity of tokens held by each *digital identity* j (account) in the system. In a centralized record-keeping system, digital identities usually correspond to account numbers. In public blockchain-based systems (like Bitcoin), digital identities correspond to public keys. Initially, Alice has two accounts, j_A and $j_{A'}$, and Bob has one account j_B. Alice's digital identity j_A holds a single token, whereas the other accounts are empty. We will denote the quantity of tokens held by a digital identity j by $a(j)$ (where a stands for "assets"), so $a(j_A) = 1$ and $a(j_B) = a(j_{A'}) = 0$.

If Alice may transfer the token held by j_A to Bob by sending a message $m =$ "j_A sends one token to j_B", what prevents her from sending the same token again to her other identity $j_{A'}$ by sending another message $m' =$ "j_A sends one token to $j_{A'}$"? Alice may want to do this, for instance, because she wants Carol to believe that she is still in possession of the token. One might say that as long as it is commonly known that m preceded m', all users of the digital token system would simply ignore m' because they consider the tokens to be in Bob's account, concluding that $a(j_B) = 1$ and $a(j_A) = a(j_{A'}) = 0$.

There are two issues with this reasoning. First, latency in the arrival of digital messages may result in Carol seeing m' *before* m if the two messages were sent in close succession. Second, if Carol joins the system long after both messages were sent, she would be unable to discern which message came first. She would see two messages

attempting to send the same token but would be uncertain as to whether it currently resides in account j_B or account $j_{A'}$.

The problem of sequencing the two messages is known as the *double-spend* problem, which highlights the need for a stable consensus on the ordering of messages. The double-spend problem arises due to two frictions commonly associated with digital ledgers. First, digital ledgers are often *pseudonymous*: one agent can hold multiple digital identities without being directly linked to those identities. If Carol could tell that Alice's actually owned both j_A and $j_{A'}$, she may be able to reasonably conclude that Alice actually sent the token to Bob. Second, communication in digital record-keeping systems is *asynchronous*: users receive messages in different orders, so it is impossible to agree on a system in which everyone simply accepts only the first transaction of a particular token and rejects all others. Pseudonymity and asynchronicity will be key elements of our formal model of digital record-keeping.

3.2.2 What is a blockchain?

A blockchain is a data structure that can be used for digital record-keeping.[1] Information is recorded sequentially in structures known as *blocks*. A block consists of (1) a set of messages and (2) a *pointer* to another block. In the context of a cryptocurrency (or other digital asset) ledger, the messages contained in blocks correspond to token transactions as well as messages meant to validate the block, which play into the consensus algorithms we discuss below.[2] The blockchain is effectively a device to keep track of an assignment of digital assets to accounts. Pointers serve to order blocks. The first block in a blockchain, known as the *genesis block*, has no pointer, and pointers on other blocks indicate their immediate predecessors in the blockchain. The state of the system (an allocation of tokens to accounts) can be derived from a blockchain by reading off the transactions in each block.

Simply adding pointers to collections of messages does not solve the double-spend problem, however. In the context of the previous example, suppose that the token is in Alice's account in the genesis block B_0, and there is a block B_1 pointing to B_0 that contains message m. Then the blockchain $\omega = \{B_0, B_1\}$ results in an allocation of tokens $a(j_A) = a(j_{A'}) = 0$, $a(j_B) = 1$. However, the existence of chain ω does

[1] Other data structures can also be used for digital record-keeping, such as Directed Acyclic Graphs (DAGs). The principles we outline in our model apply equally to those systems.

[2] For instance, in Bitcoin there are two types of messages contained in a block: transactions and a set of messages known as a *header*. The header contains an integer known as a *nonce*, and the block is judged to be valid if, when the nonce is passed through a hash function, it produces a number with sufficiently many leading zeros.

not preclude the possibility that someone could create a block B_1' that points to B_0 and contains message m'. This would create another chain $\omega' = \{B_0, B_1'\}$ that conflicts with ω and implies instead that the allocation of tokens is $a(j_A) = a(j_B) = 0$, $a(j_{A'}) = 1$, so Alice still owns the token. A situation in which there are two conflicting ledgers (chains of blocks) with a common genesis is known as a *fork*, and blocks on different branches of a fork are called *conflicting blocks*.

When faced with multiple valid yet conflicting chains, users need a *consensus algorithm* to dictates which chain to consider as the current state. The form of the consensus algorithm depends on whether the digital record-keeping system is centralized (or permissioned), proof-of-stake, or proof-of-work, so we discuss each case separately below.

3.2.3 Consensus algorithms

In our setting, the purpose of a consensus algorithm is to permit agents to agree on a *consensus state* (or chain). Consensus algorithms are essentially voting systems: when faced with multiple valid chains, agents can look at the "votes" cast on each chain in order to come to a consensus. We examine three distinct classes of blockchains that differ in their consensus algorithms: permissioned blockchains, proof-of-stake (PoS) blockchains, and proof-of-work (PoW) blockchains. These types of systems differ mainly in how they allocate voting power, and, as we will show later, in how they incentivize correct record-keeping.

A permissioned blockchain is one in which a known entity (or consortium of entities) has full, unimpeachable power to update the ledger. When the blockchain is operated by a single known entity, it is called a private blockchain. A simple example of a private blockchain is one in which a monopolist maintains and updates the ledger of a transaction system it operates. The monopolist may add transactions to blocks and approve blocks by signing them with a digital signature (e.g., by proving that he has access to a secret password). As long as no other agent has access to this digital signature, the monopolist will be the only one able to alter the ledger. In principle, the monopolist can even communicate different chains of blocks to different agents, breaking consensus (perhaps to his own advantage).

PoS and PoW blockchains are types of *public blockchains*, in which anonymous agents can become eligible to write on the ledger. The question, then, is how voting power to approve blocks should be allocated. In an ideal world, one might imagine a "one-person-one-vote" system in which all users of the ledger have equal voting

power. However, this type of voting system is infeasible, because it is not costly for one person to acquire multiple digital identities (e.g., IP addresses) and pretend to be a larger group of people. This type of attack, which will play an important role in our analysis, is known as a *Sybil attack*. The possibility of Sybil attacks creates an *identity management problem*, which is what the PoS and PoW algorithms attempt to solve. Our model will explicitly permit for the possibility that one agent creates multiple identities in order to conduct such an attack.

The PoS and PoW algorithms differ in how they allocate voting power. In PoS blockchains, voting power is allocated to accounts in proportion to the number of tokens held in those accounts. PoW blockchains, by contrast, allocate voting power to agents who prove they have solved intrinsically useless computational problems. Effectively, then, PoW allocates voting power in proportion to expenditures on computational resources. The act of solving computational problems is typically known as *mining*, and miners are compensated for their expenses through *block rewards*. Mining expenditures and block rewards are both sizable in reality. For example, as of this writing, the Bitcoin block reward was worth over $350,000 and was distributed every ten minutes on average.

In our model, we prove an impossibility result about digital record-keeping systems. As such, we will wish to keep the communication protocol as general as possible, and we will not explicitly distinguish between messages that contain information about votes and those that contain information about transactions. The consensus algorithm will contain an *updating rule* $\omega(M)$ that dictates which chain of blocks each user should consider to be valid after receiving a history of messages M. This history of messages should be understood to contain blocks as well as votes on each block.

3.2.4 Blockchain security

In communication across anonymous digital channels, nothing intrinsically links votes to individuals. At best, votes are linked to accounts (as in PoS and private blockchains), but additional information is needed to determine which accounts belong to specific individuals. This feature of digital communication creates opportunities to deviate that are at the heart of our analysis.

In the context of a completely centralized private blockchain, a monopolist can cheat by creating two internally consistent ledgers with different sets of transactions. It can then send these different ledgers to different groups. If one large group sends messages to the other saying it has received a different ledger, the monopolist may

simply act as if that large group is simply one attacker conducting a Sybil attack, causing the other group to continue to believe it has the true ledger.

In PoS systems, the problem is similar, as voting power is linked to accounts holding monetary tokens. As in a centralized system, a group of attackers who held a large fraction of tokens in the genesis block may want to build out a fraudulent ledger of transactions. The attackers would have all of the voting power necessary to update that new ledger as they chose, making it indistinguishable from a ledger used for transactions in the economy. The attackers could then proliferate this seemingly valid ledger to unsuspecting new users, who would need additional information to distinguish it from the true ledger used by others. More realistically, attackers can attempt to acquire the private keys corresponding to accounts that held a supermajority of votes long in the past (e,g., on the black market). It is plausible that the accounts that held a supermajority in a block B_{past} are eventually emptied, so it could be cheap to acquire these keys and start a fork from block B_{past} that looks just as valid as the true consensus chain to new users. This attack, known as a "long-range attack," is of principal importance in the study of PoS security.

Finally, in PoW systems, attackers can mimic the behavior of agents reporting honestly simply by paying a high enough computational cost to create blocks at the same pace as the rest of the network. Nothing about the attackers' blocks indicates that they were fraudulently created, as the votes on those blocks look identical to those on any other block. The consensus algorithm that PoW blockchains use is typically a "longest chain rule": all users converge on the chain with the greatest amount of work committed to it, which typically coincides with the longest chain of blocks. Hence, an attacker can generate a consensus on a chain he privately created, so long as he controls a majority of the network's computational power. This type of "51% attack" is the primary concern in the security of PoW systems.

3.3 Examples of consensus algorithms

In this section, we discuss consensus algorithms in more detail in order to illustrate how the different types of algorithms used in reality work and how they provide incentives for honest record-keeping. A consensus algorithm determines how a user of the record-keeping system should update her view of the ledger and communicate with others based on the history of messages she has seen. Formally, if the history of messages seen by an agent with account j is M_j, a consensus algorithm specifies (1) an *updating rule* $\hat{\omega}(M_j)$ whose output is a chain of blocks j should consider to be

valid, and (2) a *messaging strategy* $\sigma_m(M_j)$ dictating which messages the owner of j should send to others.

We now give a rough description of the three broad classes of consensus algorithms and outline the problems that may arise in each case. We will continue to work with our example in Section 3.2.1. For concreteness, we assume there are two periods, $t = 1, 2$. Alice and Bob are both present at $t = 1$, but Carol arrives only at $t = 2$. Everyone knows the true genesis block B_0. Carol, however, will not observe communication between Alice and Bob in real time, so she may not be able to distinguish between multiple potentially valid blocks B_1 when she arrives at $t = 2$.

We will argue that a consensus algorithm always requires incentives to ensure the history of events is recorded honestly– this will be the essence of the Blockchain Trilemma. Intuitively, incentives are needed because a form of double-spending is always possible: Alice can create a new identity $j_{A'}$ and communicate with that identity in precisely the same way as Bob's identity j_B, regardless of the consensus algorithm used. The result will be two conflicting ledgers that appear to be equally valid, potentially deceiving Carol. To model agents' incentives, we will assume a simple payoff structure. We assume that at $t = 1$, there are gains from trade between Alice and Bob. Bob can give Alice a good that he values at 1 but that Alice values at $\theta > 1$. We view $t = 2$ as representing the future. Going forward, each pair of agents who agree on the state of the ledger can generate a surplus s, which is allocated among those agents in proportion to the quantity of tokens they hold. For instance, if Alice, Bob, and Carol all agree on a state in which Bob holds a token, then Bob receives future payoff $3s$. Consensus may break down, however. If Alice and Bob privately agree that Bob holds one token, but Alice tricks Carol into thinking that she still holds the token, then Bob receives the entire surplus s generated by his interactions with Alice, but Alice receives the surplus s generated by her interactions with Carol going forward. Bob and Carol do not agree on the state of the ledger, so they are not able to generate any surplus.

3.3.1 Private blockchain

For simplicity, here we consider a situation in which Alice is a monopolist operating the record-keeping system. She may send signed messages from her identity j_A to Bob and Carol of the form $m =$ "*The current state is summarized by the blockchain* ω." That is, she simply sends a complete ledger to other agents to inform them of the state. If an agent has observed a history of messages $M = \{\omega_0, \omega_1, \ldots, \omega_K\}$ from Alice,

each one containing a blockchain, the consensus algorithm specifies the updating rule

$$\omega(M) = \omega_K.$$

Each agent's view of the ledger is simply the most recent blockchain provided by Alice's digital identity j_A.[3]

One way to provide Alice with incentives to keep records honestly is to permit her to extract rents from the other agents. We assume that at $t = 0$, Alice may purchase goods from Bob for 1 token, but that she is able to extract a fee f for processing the transaction. Hence, any agent who views the transaction between Alice's account j_A and Bob's account j_B as valid will conclude that j_B has $1 - f$ tokens while j_A retains f tokens, $a(j_B) = 1 - f$ and $a(j_A) = f$. If Alice keeps records honestly, then, she receives a payoff of θ at $t = 1$. All agents will come to a consensus at $t = 2$, so going forward they will generate a surplus of $3s$, of which $f \cdot 3s$ is allocated to Alice.

Alice may deviate, however, by providing Bob and Carol with conflicting ledgers. She may send Bob a ledger confirming that j_A sent a token to j_B at $t = 1$, and then send Carol a ledger at $t = 2$ confirming that j_A sent the token to $j_{A'}$. That is, Bob's ledger contains message m while Carol's contains m'. In this case, Alice and Bob privately agree on one ledger, while Alice and Carol privately agree on another. Hence, Alice receives a payoff $f \cdot s$ from her future interactions with Bob, but she receives the full surplus s from her interactions with Carol (because Carol still believes Alice's accounts hold all tokens).

How could Alice be incentivized to provide Carol with the correct state? Note that after Alice deviates, Bob and Carol disagree about the state of the ledger, though, so they are not able to interact further– and Alice is not able to extract any rents from the interactions between them. In fact, when Alice's fees are large enough ($f \geq \frac{1}{2}$), the costs she incurs by failing to extract future rents from Bob and Carol exceed the benefit of making Carol believe she never transacted with Bob.

3.3.2 Proof-of-Stake blockchain

In a proof-of-stake (PoS) system, voting power is typically assigned to token holders. When determining which message is ultimately included in the blockchain, voting power is assigned based on token holdings before transfers are executed rather than

[3]Other monopolistic consensus algorithms are possible as well. For instance, it may be that each agent accepts the *first* blockchain provided by Alice and refuses to accept any chain that conflicts with that one thereafter.

after. That is, following block B_0, Alice has all of the voting shares, no matter which message is chosen. Proof-of-stake blockchains typically operate using a supermajority rule. Accounts that hold tokens are permitted to vote on the next block to be included in the blockchain. We say a fraction α of tokens vote on a block if the fraction of tokens held in accounts voting on that block is α. The consensus algorithm is as follows: a user who has received a history of messages M containing blockchains $\{\omega_0, \ldots, \omega_K\}$, the consensus algorithm chooses $\omega(M)$ to be the longest chain of blocks in M, ω^*, such that each block in ω^* has at least a fraction $\frac{2}{3}$ of votes.

Note that this consensus algorithm does not necessarily generate a consensus if token holders misbehave. In particular, if agents holding $\frac{2}{3}$ of tokens vote on conflicting chains of blocks, there may be several chains of equal length with a supermajority of votes. Alice can deviate precisely in this way: when she communicates with Bob at $t = 1$, she may send Bob a message saying that her account (which holds all tokens) confirms the transaction in which j_A sends one token to j_B. This causes Bob to conclude that the transaction has a supermajority of votes. However, at $t = 2$, Alice may send Carol a message saying that her account confirmed the transaction in which j_A sent one token to $j_{A'}$, but Bob may pass on the initial message confirming the transaction in which j_A sent a token to j_B. Then, Carol has two conflicting ledgers, each having a supermajority of votes. The Proof-of-Stake algorithm, by itself, is not able to resolve this type of conflict.[4]

This type of attack is simply the "long-range attack" at the center of discussions about proof-of-stake security. Alice waits until she has spent all of her tokens and then creates a conflicting ledger using her previous supermajority of voting power. The punishments used to deter this behavior in the short run involve taking away tokens from those who vote on conflicting chains. These are useless once Alice has spent her tokens. Therefore, Carol still needs some way to decide on the true history of events, and Alice must be deterred in another way.

In particular, Carol needs access to a trusted source of information to tell her that the transaction between j_A and j_B actually occurred first. If Carol has some sort of interaction with Alice outside of the digital record-keeping system, she may be able to rely on Alice to externally communicate the true state to her. In reality, these types of messages are known as "checkpoints": in order to ensure security, users are encouraged to get a recent block's header from an external source. Trust between

[4]It is generally agreed upon that there is no secure algorithm to establish consensus when a sufficiently large fraction of the voting stake belongs to an agent who has equivocated. For instance, in a white paper for the Casper TFG proof-of-stake algorithm, Vitalik Buterin instead proposes that these instances of equivocation be resolved by market forces or social consensus instead.

Alice and Carol may stem from a mutually beneficial business or personal relationship that may be broken off if Alice lies (which could be represented by the loss of some benefit v). As long as $v \geq s$, Alice will have an incentive to report the true state honestly to Carol.

3.3.3 Proof-of-Work blockchain

Proof-of-work (PoW) blockchains allocate voting power in proportion to the computational resources spent by agents (typically known as "miners"). Miners are then compensated for their efforts with tokens later on. In a PoW system, each block B can be thought of as being associated with an amount of work $w(B)$ done to "mine" it.[5] The updating rule in PoW systems is extremely simple: fix some $\alpha > 0$, and let the set of valid blocks be all blocks with $w(B) \geq \alpha$. The consensus algorithm picks $\omega(M)$ to be the longest chain of blocks ω^* consisting only of valid blocks.

This rule is the analogue of the "longest chain rule" used by most PoW blockchains. In practice, the consensus state in a PoW blockchain is given by the chain with the greatest aggregate difficulty of PoW problems solved, and adding an additional block requires a fixed amount of work (on average).[6] Observe that under this definition, all users seeing the same set of blocks will agree on the current state. There is no question of how messages were ordered in the past because, no matter the order in which previous blocks were added, all users agree on which is the longest chain at any given time.

In a PoW system, Alice can deviate by conducting what is known as a "51% attack" in which she performs enough computational work to change the consensus achieved by all agents. Suppose that at $t = 1$, Alice's account j_A sends a token to j_B, and then Bob performs work α to produce a block containing the transaction. Alice may double-spend and create a new chain that is two blocks long, paying 2α to create a longer chain that instead contains the transaction sending tokens from j_A to $j_{A'}$. All agents then agree that account $j_{A'}$ has one token, so Alice's payoff going forward will be $3s$. She receives θ in the first period from transacting with Bob but has to pay 2α to execute the deviation. On the other hand, if she plays honestly, she will simply

[5] In reality, matters are slightly more complex. Each block must contain a nonce that hashes to a small enough number. If it does so, one can compute the expected quantity of work required to generate such a nonce.

[6] One element we omit from our analysis is the endogenous adjustment of the difficulty of PoW problems. Typically, the difficulty of the problems to be solved adjusts to target a fixed rate at which blocks should be added to the chain. If blocks are added at a higher rate than desired, the difficulty adjusts upwards, and the opposite occurs when blocks are added at a lower rate than desired.

receive θ. Therefore, it must be that $\alpha \geq \frac{3}{2}s$ to incentivize honesty. That is, Alice will not double-spend only if the computational work she needs to do to produce a new block is large enough. Whereas Alice faced ex-post costs of deceiving Carol in the private blockchain and PoS models, in the PoW model the cost of deceiving Carol is incurred ex-ante. The only thing preventing Alice from reversing her payment to Bob (in everyone's eyes) is the cost of computing power required to override the initial payment.[7]

3.4 Model

The model has two periods, $t = 1, 2$, and three agents: Alice (A), Bob (B), and Carol (C). The agents keep records of actions at $t = 1$ using a digital ledger, which we describe below. We use this minimal setting for simplicity so that we may focus on the details of how consensus is reached. Later on, we discuss how the structure of payoffs in our two-period setting can be derived from those in an infinite-horizon setting. We view $t = 1$ as being analogous to a stage of an infinite-horizon game, whereas $t = 2$ will represent the continuation of the game into the infinite future. Further, we discuss how all of our results extend to the case in which there are more than three agents. In this respect, our analysis parallels the classical work on consensus algorithms in the computer science literature: the seminal paper of Lamport, Shostak, and Pease (1980) first proves a consensus impossibility result with three nodes and then proves that the result extends to any number of nodes.

There is a digital ledger tracks a *state*, which corresponds to an allocation of tokens to digital identities, which are numbers j in a set $\mathcal{J}$. We normalize the quantity of tokens in all states to one.[8] We denote the quantity of tokens held by digital identity j in state ω by $a_\omega(j) \geq 0$ (where a stands for "assets"). The state space is therefore $\Omega = \{\omega \in \mathbb{R}_+^{\mathcal{J}} : \sum_{j \in \mathcal{J}} a_\omega(j) = 1\}$. Because the ledger assigns tokens to digital identities rather than directly to agents, it is *pseudonymous*. The initial state ω_0 endows a single identity j^* with one token and all others with nothing. This state is common knowledge.

Initially, one of the three agents is assigned the digital identity j^*. Agents may also create as many new identities as they would like. When an agent creates a

[7]For simplicity, here we assume that Bob will accept the payment after it is included in a single block. In reality, vendors usually deliver goods only after several blocks have followed the initial block including the payment, so Alice would have to overwrite multiple blocks rather than just one.

[8]This assumption is without loss of generality. In reality, the stock of many cryptocurrencies grows over time.

new identity, she draws a new $j \in \mathcal{J}$ randomly. Since these digital identities are pseudonyms, agents know only the set of identities that they personally own, but they do not know which identities correspond to the other agents.

Agents will communicate with one another via private messages from a set $\mathcal{M}$ in order to come to an agreement on the state and update the ledger. When two agents come to an agreement on an update to the ledger, they may take corresponding actions to reflect the transaction that was entered in the ledger. We will later specify the transactions that can occur at $t = 1$. As in the examples we discussed in the previous section, agents will have incentives to transact with an agent and then report to the other that the transaction never occurred (i.e., double-spend).

Interactions in this model are *asynchronous*. At $t = 1$, only two of the agents are present, and the other is absent. The two agents who are present will transact, but they will not be able to report their transaction to the absent agent until her arrival at $t = 2$. The lack of synchronicity will create the possibility that the absent agent is fooled by double-spending attempts.

The *digital record-keeping system* used by agents to facilitate trade consists of three elements, listed below.

1. **Consensus algorithm**: A pseudonymous communication protocol that permits agents to come to a consensus on the state (i.e., the history of token transactions). This consensus algorithm may or may not require agents to perform computational work.

2. **Token-based mechanism**: A mechanism that specifies how agents should transact and how tokens should change hands at $t = 1$. This mechanism determines agents' payoffs at $t = 1$ as well as their continuation values at $t = 2$ along the path of play.

3. **External environment**: Communications and interactions between agents that occur outside of the digital ledger but that may help to sustain honest record-keeping. We will sometimes refer to these as social interactions.

We will explain each of these elements in turn. Each of our desired properties of a record-keeping system relates to one of these elements: rent-freeness will be a property of the token-based mechanism related to allocative efficiency, resource-efficiency will be a property of the consensus algorithm related to computational expenditures, and self-sufficiency will depend on whether punishments coming from the external environment are necessary to sustain honest record-keeping.

3.4.1 Communication and the consensus algorithm

We first describe communication among agents and the consensus algorithm, which determines how they attempt to reach agreement on the state of the ledger. It is possible to describe the process by which the state of the ledger is updated without reference to the types of physical transactions made by agents, which we introduce in the next section.

Each period begins with K rounds of communication among agents. Only agents who are present may send and receive messages. Agents send each other private messages from an abstract set $\mathcal{M}$, known as the message vocabulary. In particular, within a round, each agent $n \in \{A, B, C\}$ will choose a message $m_{n,n'} \in \mathcal{M}$ to send to each other (present) agent n'. We keep the set of messages abstract because our goal is to prove an impossibility result, so we must allow for essentially arbitrary forms of communication. Each message $m \in \mathcal{M}$ is associated with a computational cost $w(m)$, which may be positive (if the message requires proof-of-work) or zero (if it does not). The *communication protocol* used by agents consists of two objects: a *messaging rule* σ and a *updating rule* $\hat{\omega}$, which we describe below.

Messaging rule: As communication proceeds, agents will accumulate private information. In particular, an agent's information set consists of (1) the set of identities owned by that agent $\mathcal{J}_n \subset \mathcal{J}$, and (2) the (ordered) set of messages observed by the agent in previous rounds of communication, denoted M_n. This private information will determine how the agent communicates with others. We will sometimes refer to an observed history of messages M_n as an *input*. The messaging rule dictates how each agent n should communicate with other agents n' based on the agent's digital identity j and observed history of messages M_n.[9] The output of the messaging rule is a message $m_{n,n'} = \sigma_{n,n'}(j, M_n)$. An agent is said to be *honest* if she follows her messaging strategy.

The messaging strategy can be viewed as a revelation mechanism. Agents report a digital identity j and a history $\tilde{M}$, and the mechanism outputs the messages to be sent to each other agent. However, agents may deviate from the communication strategy given by the mechanism. An agent can choose to send any message m' such that $m' = \sigma(j, \tilde{M})$, where j is *any* digital identity owned by the agent and $\tilde{M}$ is any *subset* of the agent's observed history of messages M_n. Agents cannot produce evidence that they to own digital identities that do not belong to them, and they cannot produce evidence that they have seen messages that were never sent to them.

[9] Agents are not supposed to hold more than one digital identity, so it is not necessary for the rule to specify behavior for an agent who has multiple identities.

Nevertheless, they can pretend that they have never received messages that were sent to them, or they can pretend to have received those messages in a different order. Below, we discuss why we restrict the set of messages that agents can send in this way, and we argue that these restrictions reflect the types of proofs embedded in digital record-keeping systems in practice.

Updating rule: The updating rule is a function $\hat{\omega}(M)$ specifying the state of the ledger as a function of an input M (i.e., a history of messages). Intuitively, the updating rule tells agents how to read the ledger given a (possibly conflicting) set of messages received in the past.

The updating rule can also be used to determine whether agents are in agreement about the state of the ledger. A message $m = \sigma(j, M)$ is said to acknowledge state ω if $\hat{\omega}(M) = \omega$. That is, if $\hat{\omega}(M) = \omega$, an agent who sends message m has observed a history of messages that indicates the current state is ω, so the message implicitly acknowledges ω as the current state. If, in the final round of communication of a given period, two agents send each other messages that acknowledge state ω, they are said to *agree* on state ω in that period. This is because the messages sent by each agent acknowledge that ω is the current state of the ledger, so the agents themselves should be willing to act as if token holdings are as specified by that state. We use this idea to define our notion of a consensus algorithm.

Definition 4. *A communication protocol $(\sigma, \hat{\omega})$ is said to* ***achieve consensus*** *if any agent communicating with honest agents only eventually reaches agreement with some subset of those agents on the state. Such a communication protocol is called a* ***consensus algorithm****.*

We will restrict attention to consensus algorithms that are symmetric in the sense that they treat all digital identities equally. This simply means that there is nothing special about any of the digital identities $j \in \mathcal{J}$, so the consensus algorithm generates the same result no matter how these identities are permuted. We formally define this notion of symmetry in the Appendix.

Properties of the message space: The assumption that agents are restricted in the set of messages they can send actually makes our results more general rather than less. It allows agents' messages to embed several types of proofs commonly used in digital record-keeping systems, such as the following.

1. **Proof of identity**: A proof that an agent owning a particular digital identity j sent the message, which comes in the form of an unforgeable signature;

2. **Proof of succession**: A proof that the message was sent after a previous message $m' \in \mathcal{M}$;

3. **Proof of work**: A proof that a certain quantity $w \in \mathbb{R}_+$ of physical resources were expended to send the message.

In reality, an unforgeable signature associated with a proof of identity corresponds to signing a message using a secret password known only to the sender. This type of proof can be useful in attaining consensus because it allows others to track messages to a particular digital identity, allowing them to discern whether the owner of that identity is misbehaving and trying to disrupt or alter consensus in some way. A proof of succession links a message back to an earlier one observed by the sender. This is precisely how blockchains are formed: each block "points" to the previous one. A proof of succession prevents the sender of a message from claiming to have seen a history it did not. Finally, proofs of work are common in blockchain-based consensus algorithms because they make it costly to send messages and alter consensus.

3.4.2 Payoff structure and token-based mechanisms

We now introduce the actions and payoff structure in the game played by agents and token-based mechanisms that can be used to implement particular action profiles. The environment we present is a simple one meant to capture a role for record-keeping. We interpret $t = 1$ as the time at which records must be kept and $t = 2$ as the future, when agents receive continuation payoffs.

Actions and payoffs: At $t = 1$, two agents are present. One of the two agents is a *buyer* and the other is a *seller*. For simplicity, we assume the buyer is always the agent who owns the digital identity j^* that holds all tokens initially. The seller is able to produce goods for the buyer at a linear cost: if the seller produces $x \in \mathbb{R}_+$ goods, she suffers disutility x. The buyer, on the other hand, enjoys utility $u(x)$ when consuming x goods, where u is increasing, concave, and differentiable. The buyer and seller communicate throughout $t = 1$, and they transact if they agree on the state of the ledger at the end of the period. The absent agent does not observe this communication, providing a role for record-keeping: the seller must be assured that a record of the transaction will be kept in order to receive a payoff at $t = 2$. The quantity that the seller produces for the buyer will be specified by the token-based mechanism we introduce below.

At $t = 2$, the previously absent agent arrives. This agent must be updated on the state of the ledger. Throughout $t = 2$, all three agents communicate with each other.

In our setting, agents will be able to coordinate when they agree on a state of the digital ledger, enabling them to continue interacting using the digital record-keeping system. Agents are able to generate some value by coordinating with each other. In particular, at the end of the period, any pair of agents who agree on the state generate a surplus $\frac{1}{3}s > 0$. So then, for example, if all three agents coordinate, they generate a total surplus of s. If, on the other hand, A and B coordinate and A and C coordinate, but B and C do not coordinate, then the total surplus generated is $\frac{2}{3}s$. Utility is transferable at $t = 2$, so surplus can be split arbitrarily among agents.

Token-based mechanisms: The token-based mechanism specifies how many tokens τ^* should be transferred from the buyer to the seller. Equivalently, the mechanism specifies a future state of the ledger $\omega^* = (a_B(\omega^*), a_S(\omega^*)) = (1 - \tau^*, \tau^*)$. The buyer starts with one token, so his final holdings will be $a_B(\omega^*) = 1-\tau^*$, whereas the seller's final holdings will be $a_S(\omega^*) = \tau^*$.

The mechanism also defines agents' payoffs conditional on the state of the ledger. At $t = 1$, if the buyer and the seller agree on a state of the ledger ω, the quantity that the seller should produce for the buyer is given by $x(\omega)$. At $t = 2$, the mechanism dictates how surplus is to be split between agents: the fraction of the future surplus allocated to the buyer and seller in state ω are given by $\mathbf{v}(\omega) = (v_B(\omega), v_S(\omega))$ with $v_B(\omega) + v_S(\omega) = 1$. Note that the token-based mechanism does not directly specify actions and payoffs. Instead, it specifies payoffs for all possible states ω. This is necessary because dishonest record-keeping by agents may cause the realized state to differ from ω^*, so the mechanism must provide an assignment of surplus in all contingencies. Finally, the mechanism determines how much computational work the buyer and seller should do, $\mathbf{w} = (w_{B,}w_S)$. These computations may be necessary if the consensus algorithm requires proof-of-work.

A token-based mechanism can then be written as $\mathcal{T} = \{\omega^*, x(\omega), \mathbf{v}(\omega), \mathbf{w}\}$. There is no need to check incentive-compatibility for this mechanism– the buyer cannot produce and the seller cannot consume. Instead, it is necessary to check that a participation constraint binds for both the buyer and seller. We define these constraints below.

Definition 5. *A token-based mechanism $\mathcal{T}$ is **feasible** if*

- *The buyer's participation constraint holds,*

$$u(x(\omega^*)) + v_B(\omega^*)s - w_B \geq s;$$

- *The seller's participation constraint holds,*

$$v_S(\omega^*)s - x(\omega^*) - w_S \geq 0.$$

The buyer must not prefer to hold tokens until the final period, in which case he would receive the full surplus s (because he would hold all tokens). The seller, on the other hand, must not prefer to do nothing and sit out of the mechanism.

It is useful to characterize the optimal mechanism that could be implemented if honest record-keeping were not a concern. When a utilitarian social welfare function is used, the optimal quantity of production is x^* such that $u'(x^*) = 1$. It is feasible to set $x = x^*$ as long as $s \geq x$. Otherwise, the seller is never willing to produce more than s. Thus, the optimal mechanism would set

$$x^{FB} = \min\{s, x^*\}.$$

When it is necessary to incentivize honest record-keeping, it will be more difficult to reach the optimal level of production x^*. There are two reasons that this might be the case. First, it may be necessary to provide the buyer with some continuation payoffs to incentivize good behavior. This implies that the full surplus s generated at $t = 2$ cannot be transferred to the seller. Second, the consensus algorithm may involve costly computations that disincentivize double-spending. This also tightens the participation constraint if the seller is forced to bear some of these costs.

We now define our notion of rents in a token-based mechanism, which will be important in our statement of the Blockchain Trilemma.

Definition 6. *The buyer **extracts rents** in the token-based mechanism $\mathcal{T}$ if $v_B(\omega^*) > 0$. Otherwise, the mechanism is **rent-free**.*

Allowing the buyer to keep tokens in the second period can only reduce the allocative efficiency of the mechanism. The need to provide the buyer with continuation payoffs (to discipline bad behavior) means there is less surplus to be transferred to the seller, discouraging production. This is why rent-freeness is an ideal property of a token-based mechanism.

In reality, there are several ways that the buyer could extract rents. Under the most literal interpretation of the mechanism described here, rent extraction corresponds most closely to putting down collateral tokens that can be taken away to punish misbehavior. This type of rent extraction corresponds closely to the short-run incentive mechanisms provided to "validators" in PoS systems, who are required to

stake tokens during the period in which they vote on updates to the blockchain. However, another interpretation is possible– the buyer could remain in possession of tokens at $t = 2$ by charging a transaction fee. Indeed, if the buyer spends one token to purchase x goods but charges a fee f to process the transaction, the seller would be left with $1 - f$ tokens and the buyer would have f tokens at $t = 2$. Furthermore, the tokens held by the buyer at $t = 2$ could also represent the expectation of *future* fees to be extracted from users of the system.

3.4.3 Social interactions

Outside of the token-based mechanism, agents have external channels through which they can interact, which we refer to as social interactions. These interactions can help to sustain honest record-keeping in the token-based mechanism.

We assume that through their social interactions, agents have some capacity to directly punish each other. In particular, at $t = 2$, an agent n can costlessly inflict a punishment (i.e., a negative payoff) $v_{n,n'} \in [0, \overline{v}_{n,n'}]$ on agent n', where $\overline{v}_{n,n'}$ represents the capacity of n to punish n'. These punishments are meant to reflect a break-down of some type of cooperation outside the token-based mechanism. For instance, they could reflect the abandonment of a business partnership or a social relationship. We assume punishments are bilateral for simplicity, but nothing precludes other types of collective punishments from being used to sustain honest behavior within the record-keeping system.

3.4.4 Equilibrium concept

We now summarize the actions, information, and payoffs of each agent and define our equilibrium concept. The choices made by each agent n who is present at $t = 1$ are

1. A number $J_n \in \mathbb{N}$ of digital identities to create;
2. Messages $m_{n,n',k}$ to send to the other agent n' in each round k of communication.[10]

The transaction between the buyer and the seller is derived from the token-based mechanism $\mathcal{T}$. When the buyer and seller agree on a state ω, the seller produces $x(\omega)$ for the buyer. The computational work done by an agent is derived from the properties

[10]Recall that if a message m is to be feasible for agent n, it must be that $m = \sigma_{n,n'}(j, \tilde{M})$ for some digital identity j owned by n and some observed history of messages $\tilde{M}$ that is a subset of the true set of messages M_n observed by the agent.

of the message space and the set of messages sent by the agent. Communication at $t = 2$ is identical to that at $t = 1$, but if there are external social interactions, agents additionally choose a punishment $v_{n,n'}$ to inflict on each other agent n'.

Recall that when two agents n and n' agree on the state at the end of $t = 2$, then they are able to coordinate and generate the surplus s. This surplus is allocated to token holders as specified by the token-based mechanism $\mathcal{T}$. The utility obtained by an agent who produces x_n^S, consumes x_n^B, does computational work $\{w_{n1}, w_{n2}\}$ in each period, is allocated surplus s_n, and suffers social punishments $\{v_{n',n}\}_{n' \neq n}$ is

$$U_n = u(x_n^B) - x_n^S - \sum_{t=1}^{2} w_{nt} - \sum_{n' \neq n} v_{n',n} + s_n. \tag{3.1}$$

The first two terms relate to the allocation induced by the mechanism. The third represents proof-of-work costs, the fourth is related to social punishments, and the fifth is related to rents. The elements of the Blockchain Trilemma can then be seen from this expression for an agent's utility.

This is a game of incomplete information, so agents have different information sets. An agent's information set I_n consists of the identities she owns and the messages she observes. Agents form expectations over the information sets of others based on their own information and equilibrium strategy profiles. We will focus on pure-strategy Perfect Bayesian Equilibria (PBE) in which agents take a unique action given their information set, defined below.

Definition 7. *A **pure-strategy Perfect Bayesian Equilibrium (PBE)** of the token-based mechanism $\mathcal{T}$ consists of the following objects for each agent $n \in \mathcal{N}$: a node creation strategy $J_{nt}(I_n)$ for $t \in \{1, 2\}$, a feasible communication strategy $m_{n,n',k,t}(I_n)$ with each other agent n' for $t \in \{1, 2\}$, a social punishment strategy $v_{n,n'}(I_n)$ for each other agent n, and beliefs $\mathbb{P}(I_{-n}|I_n)$ over other agents' information sets such that*

1. *At each information set I_n, an agent's strategy constitutes a best response to the expected strategies of others;*

2. *Beliefs are consistent with Bayes' rule whenever possible.*

We do not impose any restrictions on agents' beliefs about the information sets of others off the path of play. This is because our focus here is not on understanding signaling in the game we study. Rather, we prove an impossibility result about the types of behavior that can emerge in equilibrium (the Blockchain Trilemma). As such,

putting fewer restrictions on the beliefs agents may hold makes our impossibility result more robust.

We define a digital record-keeping system as a token-based mechanism, a consensus algorithm, and an external environment such that in equilibrium, the token-based mechanism is implemented through honest communication.

Definition 8. *A **digital record-keeping system** consists of a a consensus algorithm $(\sigma, \hat{\omega})$, a token-based mechanism $\mathcal{T}$, and a pure-strategy PBE of $\mathcal{T}$ such that*

1. *The consensus algorithm induces the evolution of the state implied by the token-based mechanism $\mathcal{T}$;*

2. *In the PBE of $\mathcal{T}$, agents never create new identities at any information set and always communicate according to the consensus algorithm's message strategy σ.*

*The digital record-keeping system is **rent-free** if the associated token-based mechanism is rent-free, **resource-efficient** if agents do not pay computational costs to keep records on the path of play, and **self-sufficient** if external punishments are not used even off the path of play.*

3.5 The Blockchain Trilemma

With the model set up, we are now in a position to prove our main result: the Blockchain Trilemma. We will begin with an intermediate result showing that it will always be possible for the buyer to effectively double-spend, tricking the agent who arrives at $t = 2$ into believing that the buyer is actually the one who *received* tokens at $t = 1$. Then we will show that this result implies there must always be incentives for the buyer not to double-spend. In this environment, these incentives can come only from rents, a costly communication protocol, or the possibility of external punishment.

3.5.1 Double-spending

A double-spend occurs when the seller believes that he has received tokens at $t = 1$, but then at $t = 2$, the newly arriving agent believes that the transaction at $t = 1$ did not send tokens to the seller. Formally, under this definition, a double-spend occurs when the newly arriving agent agrees on a state ω with other agents that is different from the state ω^* prescribed by the token-based mechanism.

Our key intermediate result is that no matter how the consensus algorithm works, there always exists a deviation for the buyer that can cause a double-spend to occur. The proof of the proposition is somewhat notation-intensive, so we relegate it to the Appendix. However, the main intuition is quite simple, so we provide a heuristic sketch of the proof.

Lemma 3. *In any digital record-keeping system, there exists a communication strategy for the buyer such that (1) the buyer and seller agree on the state ω^* prescribed by the token-based mechanism at the end of $t = 1$, and (2) the buyer and the agent who was previously absent agree on a different state $\omega^{*'}$ at $t = 2$ in which the buyer holds all tokens.*

Proof Sketch. The buyer can double-spend as follows. First, at $t = 1$, the buyer communicates honestly with the seller, whose digital identity is some $j_S \in \mathcal{J}$. This communication convinces the seller that he has received tokens (i.e., the buyer and seller agree on the state ω^* prescribed by the token-based mechanism). In the meantime, the buyer creates a new digital identity j'_S to mimic the seller and generates messages as if j'_S had been the one to receive tokens at $t = 1$ instead of j_S. Figure 3.2 depicts this part of the deviation.

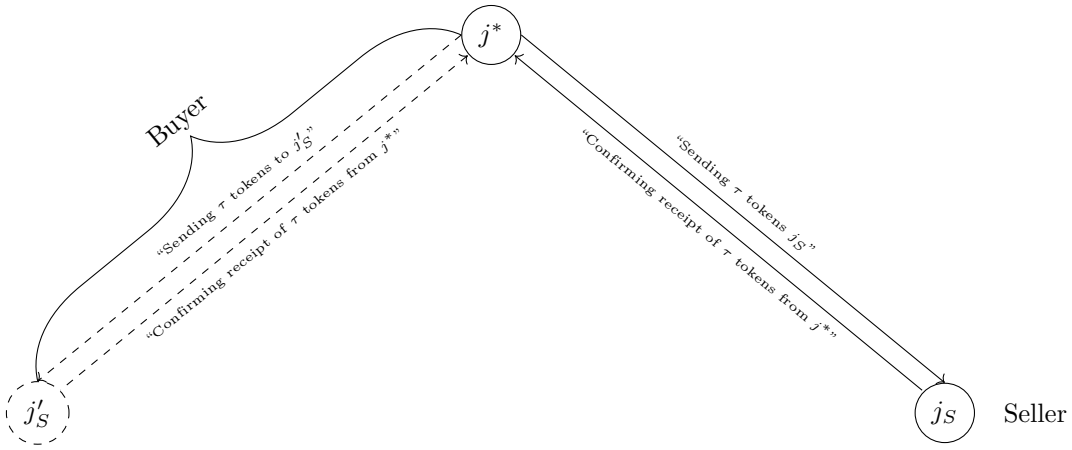


Figure 3.2: The double-spend deviation at $t = 1$. The buyer creates a new identity j'_S and secretly generates precisely the messages that would have been sent between the buyer and seller if tokens had been sent to j'_S instead of the seller's account j_S.

At $t = 2$, the buyer simply communicates with the agent who was previously absent as if (1) he is actually the seller, and (2) his identity is j'_S. If the seller is communicating honestly, he will follow an identical strategy with the exception of the fact that he will say his identity is j_S rather than j'_S. The evidence that j'_S (the

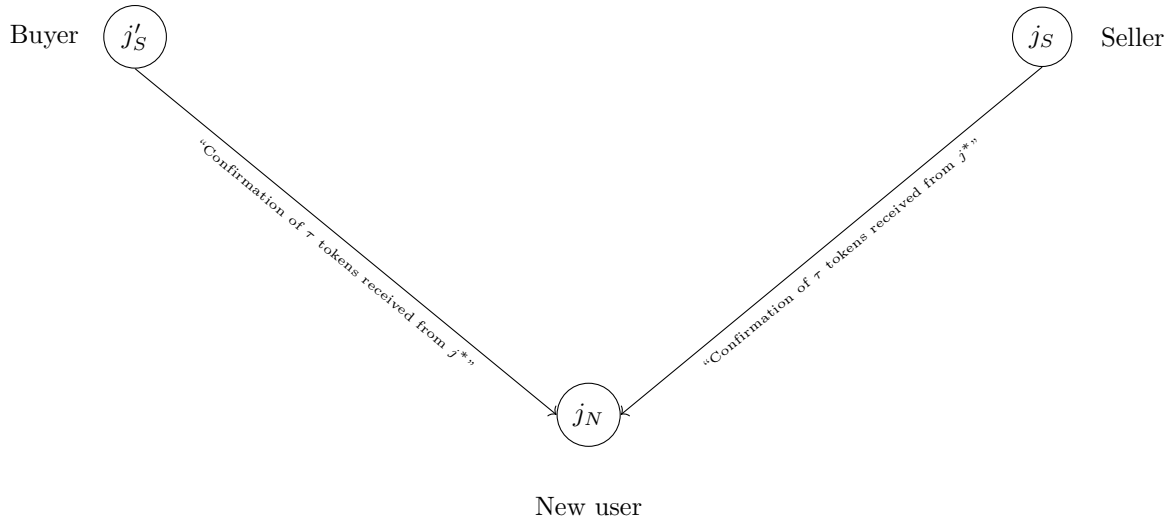


Figure 3.3: The double-spend deviation at $t = 2$. The seller sends confirmation of the $t = 1$ transaction using identity j_S to the new user. Using identity j'_S, the buyer relays an identical confirmation that tokens were sent to j'_S instead.

buyer) received tokens is then the same as the evidence that j_S received tokens. The agent who was absent is therefore just as likely to ultimately come to an agreement with the buyer as with the seller. If this agent ends up agreeing with the buyer, then the double-spend was successful. Figure 3.3 illustrates this part of the deviation.

□

This proof illustrates a basic point that makes digital record-keeping difficult: nothing prevents the emergence of two valid-looking ledgers, and there is no a priori way to distinguish between such ledgers. This is why consensus requires not only a suitably designed communication protocol, but also incentives. Broadly, this is how our analysis differs from the typical approach to consensus in computer science. The literature in computer science considers non-strategic agents and concludes that with sufficiently many agents who misbehave, consensus is impossible. We instead consider strategic agents who may all want to misbehave. Under these assumptions, consensus is in general impossible unless they have incentives not to do so.

We conclude this section by briefly mentioning how the logic of our proof relates to the security issues that face digital ledgers used in practice. One of the key elements of our proof is that an agent can create a new identity and act as if they originally received tokens from the identity that sent tokens. This is possible in essentially all digital record-keeping systems: it is only necessary for the original owner of a token to create a new account and send a message attempting to pass it from the original account to the new one.

Where digital record-keeping systems differ is in how the attacker can make this new message credible. In a monopolistic system in which all agents get the consensus state from a single entity, the monopolist can make anyone believe the double-spend. In a PoS system, a group of agents that had a supermajority of tokens at some point in the past can go back and create a chain of blocks as if the double-spend transaction had been approved originally. In a PoW system, all the attacker would need to do is create a long enough chain of blocks (i.e., do enough computational work) to get all others to come to a consensus on the double-spend transaction. The common thread in these examples is that there always exists some group of attackers who can create exactly the ledger that would have emerged if the double-spend had actually come first. A new user of the system will not be able to tell the difference between the truth and the attackers' ledger.

3.5.2 Proof of the Trilemma

We can now state and prove the Blockchain Trilemma.

Proposition 11. *There does not exist a digital record-keeping system that is (1) rent-free, (2) resource-efficient, and (3) self-sufficient.*

Proof. We proceed by contradiction. Assume that there is a digital record-keeping system that is rent-free, resource-efficient, and self-sufficient. Rent-freeness implies that the buyer does not receive any of the surplus generated at $t = 2$. Resource-efficiency implies that there is no use of costly messages in the consensus algorithm, both on and off the equilibrium path. Self-sufficiency implies that external punishment is never used as a threat off the equilibrium path. From Equation 3.1, then, the buyer's payoff in equilibrium will be

$$U^B = u(x^*),$$

where x^* is the quantity of production specified by the token-based mechanism $\mathcal{T}$.

By Lemma 3, there exists a deviation for the buyer that permits double-spending with positive probability. Under this deviation, the buyer still consumes x^* at $t = 1$, but there is a probability $\pi > 0$ that the newly arriving agent ultimately concludes that the state is one in which the buyer holds all tokens. Furthermore, given that the consensus algorithm is resource-efficient, the buyer will not incur a computational cost for pursuing this deviation. Self-sufficiency implies that no agent can externally punish the buyer ex-post for deviating. Then, the buyer's expected utility $\tilde{U}^B$ under

the double-spend deviation satisfies

$$\mathbb{E}[\tilde{U}^B] \geq u(x^*) + \pi \cdot \frac{1}{3}s,$$

where the term $\frac{1}{3}$ enters because the buyer and the newly arriving agent can generate a surplus of $\frac{1}{3}s$ together that will be allocated to the buyer if he double-spends.

Clearly, the buyer's expected utility from double-spending is greater than his utility when playing honestly, since $s > 0$. This result contradicts the assumption that there exists a rent-free, resource-efficient, and self-sufficient record-keeping system. □

The logic behind our proof is straightforward. Proposition 3 says the buyer always has a deviation that permits double-spending with positive probability. If this deviation is to be deterred, the buyer needs to be provided with incentives. Otherwise, attempting the deviation is always profitable: if he plays honestly, he gets to consume only the amount purchased at $t = 1$, whereas if he double-spends, he gets to use those tokens again.

Incentives can be provided in three ways. The simplest way is to make the buyer pay an up-front cost to engage in the deviation. In order to deviate, the buyer needs to create a valid ledger in which the double-spend transaction was recorded, so deviation can be deterred only by creating a consensus algorithm that requires proof-of-work to send messages. In this case, the digital record-keeping system is not resource-efficient.

The buyer can also be punished ex-post for engaging in the deviation. There are two ways to punish the buyer. First, the buyer can be punished within the system. If the buyer is promised some continuation value in the future through token holdings, these tokens can be taken away after a deviation. For instance, if the buyer will receive some transaction fees in the future and others prove that a deviation occurred, the protocol can cancel those future fees. When the system is rent-free, however, it must be able to operate without the buyer being promised any continuation payoffs. Second, the buyer can be punished outside the system. If others observe a deviation from the equilibrium path, they may simply punish the buyer through the external environment. This could involve cutting off relationships with that agent, taking him to court, and so on. In a self-sufficient system, though, such punishments are not available.

Taken together, these arguments imply that there is no way to achieve a stable consensus in a digital record-keeping system that is rent-free, resource-efficient, and

self-sufficient. In such a system, it is impossible to make double-spending costly up-front, and it is also impossible to punish double-spending ex-post.

The need for rents, costly computations in communication, or external trust has implications for the efficiency of the token-based mechanism as well as, potentially, for the digital record-keeping system's stability. Rent extraction and a costly computation requirement both make it more difficult to incentivize the seller to produce. In the case of rents, some agent other than the seller is promised payoffs in the future, reducing the value that can be transferred to the seller at $t = 2$ in return for production at $t = 1$. In the case of a proof-of-work cost in achieving consensus, the deadweight cost of resources required to achieve consensus reduces the surplus produced by the system, also capping the value that the seller can ultimately receive.

A lack of self-sufficiency does not imply any inefficiency along the path of play. In fact, external punishments are used only off the path in order to incentivize honest behavior. Nevertheless, self-sufficiency may be desirable for several reasons. For one, if the trust between agents stemming from the external environment disappears for some reason, a record-keeping system that relies purely on that trust will collapse. In a sense, then, a system that is not self-sufficient is less robust than one that is. Additionally, if agents' ability to monitor the deviations of others in the record-keeping system is imperfect, external punishment may be triggered unnecessarily. For example, an external punishment may be triggered when an agent makes a mistake in record-keeping rather than a true attempt to deviate.

It is worth noting, though, that the external trust required to run a record-keeping system is not the type of centralized trust that proponents of decentralized ledgers seek to avoid. That is, a system that is not self-sufficient does not require all agents to trust the same centralized entity (e.g. the legal system). Instead, it is possible for the record-keeping system to be sustained by local trust, in which each agent trusts only a separate set of agents or entities. In fact, this is precisely how some decentralized ledgers without proof-of-work establish consensus. The blockchain used by Ripple requires that each user of the system choose a set of "validators" whose updates to the ledger they wish to trust, but there is no requirement that all users choose the same set of validators. As long as there is overlap between the sets chosen by different users, the system can achieve consensus.

3.6 Discussion

In this section, we discuss possible extensions to the model, the generality of the results, and other issues related to record-keeping with blockchains that are outside the scope of our formal analysis. Specifically, we comment on the viability of blockchains for which enforcement of ownership rights is necessary.

3.6.1 Extension to arbitrary set of agents

One apparent limitation of our benchmark model is that there are only three agents. One may wonder if the double-spending deviations that we derive in our analysis would still be possible if there were more than three agents. Intuitively, a setting with more agents provides honest record-keepers with more opportunities to communicate amongst themselves to deduce which agent is lying.

Fortunately, our results would continue to go through with more than three agents. For one, note that the existence of unilateral double-spending deviations in a setting with three agents implies that when there are more than three agents, there always exists a *coalition* that can double-spend. To see this, note that we can always split the set of agents into three groups: one group A consisting of more than one-third of agents, one group B consisting of the remaining agents who are present at $t = 1$, and a third group C consisting of all agents who are absent at $t = 1$. The coalition A can then double-spend in exactly the same way as a single agent in the benchmark model: it can communicate honestly with B while creating new identities B' and generating an identical set of messages. Then, at $t = 2$, the deviating coalition can communicate with C with those new identities B'. Group C will not be able to tell whether B or B' actually received tokens at $t = 1$, precisely as in the benchmark.

Since it is always possible for a coalition to deviate and double-spend, it is in fact always possible for a single agent to do so as well. The reason why a unilateral double-spending deviation is possible is that, since agents may always create as many digital identities as they would like, a single agent can communicate with others in precisely the same way as a larger coalition. Thus, even with more than three agents, there always exists a unilateral deviation that permits an agent to double-spend, providing a need for incentives to maintain honest record-keeping.

3.6.2 Extension to an infinite-horizon environment

We interpret our model as consisting of a mechanism that takes place in one period and then an infinite future in which continuation payoffs are realized, and we represent these payoffs by a reduced-form surplus that is generated between agents. We do not explicitly model continuation play, so here we outline how our model could map to a full infinite-horizon setting.

A simple way to map our model to an infinite-horizon setting is as follows. Consider a model with periods $t = 0, 1, \ldots,$ in which each period has two subperiods. In the first subperiod, each agent has a possibility of matching with another. When a match is formed, one of the two agents will be able to produce for the other. If the seller produces x, the buyer gets utility $u(x)$ and the seller suffers disutility x. Just as in our benchmark model, this interaction generates a role for record-keeping. In the second subperiod, agents can simply transfer utility to one another. They also learn if they will be buyers or sellers in the next period. Agents who will be buyers will then transfer utility to others in order to acquire tokens. This setup actually shares similarities with the framework introduced by Lagos and Wright (2005) in monetary economics.

Suppose that the infinite-horizon token-based mechanism specifies that a seller should produce x^* for a buyer. If any pair of agents match with probability $\frac{\lambda}{\overline{N}^2}$, where $\overline{N}$ is the total number of agents, then the total surplus generated by a subset of N agents who agree on a ledger is (approximately) proportional to $s(N) = \left(\frac{N}{\overline{N}}\right)^2 \lambda(u(x^*) - x^*)$. If the quantity of tokens is normalized to one and all tokens are used to purchase goods, then in the second sub-period of any t, future buyers will be willing to transfer utility to sellers in proportion to $s(N)$. Then if N agents agree on a ledger, an agent who holds a tokens in the second sub-period receives utility proportional to $a \cdot s(N)$. This extended model thus generates a payoff structure quite similar to that in our benchmark model.

3.6.3 Enforcement

In our model, the tokens traded on the blockchain are useful only to the extent that others accept them in exchange. They do not represent claims to physical or financial assets. In reality, there have been several proposals for blockchains on which physical assets such as houses or automobiles can be traded, or on which companies could issue securities that constitute promises to deliver certain cash flows to the owner. The difference between these types of blockchains and the ones we consider is that when

tokens represent legally binding claims, there must be some entity willing to enforce those claims. There is an important distinction between *ownership* and *possession*. Blockchains can record transfers of ownership of a digital asset, but an enforcing entity must ensure that the owner of a digital claim can come to be in possession of the financial or physical asset underlying that claim. In most situations, the enforcer would be the relevant legal authority.

Power ultimately lies with the enforcer. To see this point clearly, consider a simple example of a blockchain on which each token represents ownership of a particular house. Suppose that the blockchain forks into two chains, ω and ω'. On ω, Alice owns house H in one of her accounts, whereas on ω', Bob owns H. Then who is actually able to live in the house? The enforcer is endowed with the power to decide that one of the two has the legal right to live in the house and may evict the other. Then the enforcer effectively decides which branch of the fork should be considered the true state. There may be rules governing how the enforcer should act, but these rules must be enforced by a legal authority as well. Then, to ensure that the enforcer does not act with impunity, there must be some way to punish the enforcer through the blockchain mechanism (e.g., by decreasing the benefits it derives from the mechanism) or externally (e.g., through elections).

These considerations could be made explicit by extending our model to incorporate physical assets with *digital identities* in the form of tokens traded on the blockchain. There would be an additional player, called an enforcer, who would be tasked with allocating the physical assets. Without the enforcer's assistance, players would have no way of ensuring that their assets are not stolen by others. The mechanism would dictate how the enforcer should allocate assets based on the state of the blockchain. Importantly, in case of a fork, the enforcer would have to make a decision as to which branch of the fork to honor. Of course, the enforcer would need the correct incentives to make the right decision. The Trilemma indicates that there are two options: the enforcer may directly extract rents from the mechanism in some way, or there may be external arrangements that permit others to punish the enforcer (what we call "social trust" in our benchmark model). Realistically, a legal authority may enjoy private benefits from its position, so any mechanism that allows for the removal of that authority would constitute an appropriate punishment option. As long as the enforcer has the correct incentives to follow the rules, all other players may behave as in the benchmark model.

In this modification of our model, the enforcer has full power to decide the current state. That is, the enforcer may ultimately decide the allocation of physical assets in

each period. If the enforcer behaves, it is only because it has the correct incentives to do so. This observation has several important implications. First, if a new user needs to acquire information about the current state from social connections, it is enough for that new user to ask the enforcer only. The enforcer decides which physical assets will be allocated to the new user, so any reports that conflict with the enforcer's are irrelevant for that user. Second, it may be efficient to bundle the responsibility of record-keeping (i.e., updating the ledger) with enforcement. The enforcer must enjoy some rents or social trust, so it may be that its incentives are sufficient to ensure it keeps records honestly as well. This type of bundling could be beneficial, for example, if allowing another record-keeper to extract rents would create inefficiencies. Nevertheless, there may still be some benefits of decentralized record-keeping even when there is a need for centralized enforcement. As discussed in the previous subsection, decentralization may be an effective tool to lower the transaction fees paid by participants in the mechanism.

3.7 Conclusion

The fundamental question that inspired the invention of public blockchains is whether consensus can be achieved without trust in a single entity. The Blockchain Trilemma answers this question: it is possible to replace trust in a single entity (generated by rents) either by imposing resource costs or by relying on external sources of trust among individuals. Proof-of-work blockchains take the former approach. It is possible for a new user who knows nothing other than the blockchain protocol to read the ledger, deduce the current state, and use the blockchain-based mechanism with the confidence that records will be kept honestly. Proof-of-stake blockchains take a less radical approach, achieving decentralization by changing the *structure* of trust required for the system to operate correctly. A new user must rely on external trust with an existing user in order to acquire the state, but there is no need for all new users to share a common trusted source. Instead, users may choose who they trust to provide that basic information. The Trilemma implies that these are, in fact, essentially the only possibilities.

We leave to future research the question of which types of ledgers should be centralized and which should be distributed. The benefits of decentralization come from the elimination of rents, potentially above and beyond the minimum level of rents required to ensure honest reporting. Our model is too general to make specific predictions relating the record-keeping protocol to rent extraction, but it seems

natural that the free entry of record-keepers permitted by decentralized blockchains should prevent the type of collusion required to extract large fees from users. We also clarify the distinction between ownership, which can be recorded on a blockchain, and possession, which must be enforced by a trusted entity. When there is a need for a trusted entity regardless of whether record-keeping is centralized or not, it may be more efficient to bundle record-keeping with enforcement.

References

Biais, Bruno, Christophe Bivière, Matthieu Bouvard, and Catherine Casamatta, "The Blockchain Folk Theorem," *The Review of Financial Studies*, May 2019, *32* (5), 1662–1715.

Budish, Eric, "The Economic Limits of Bitcoin and the Blockchain," 2018. Working paper.

Chiu, Jonathan and Thorsten Koeppl, "The Economics of Cryptocurrencies–Bitcoin and Beyond," 2017. Working paper.

Cong, William Lin and Zhiguo He, "Blockchain Disruption and Smart Contracts," *The Review of Financial Studies*, May 2019, *32* (5), 1754–1797.

_ , Ye Li, and Neng Wang, "Tokenomics: Dynamic Adoption and Valuation," 2019. Working Paper.

Easley, David, Maureen O'Hara, and Soumya Basu, "From Mining to Markets: The Evolution of Bitcoin Transaction Fees," *Journal of Financial Economics*, October 2019, *134* (1), 91–109.

Gans, Joshua, June Ma, and Rabee Tourky, "Market Structure in Bitcoin Mining," 2018. NBER Working Paper.

Gervais, Arthur, Ghassan Kharame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf, and Srdjan Capkun, "On the Security and Performance of Proof of Work Blockchains," in "Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security" 2016.

Huberman, Gur, Jacob Leshno, and Ciamac Moallemi, "An Economic Analysis of the Bitcoin Payment System," 2019. Working Paper.

Lamport, Leslie, Robert Shostak, and Marshall Pease, "Reaching Agreement in the Presence of Faults," *Journal of the ACM*, April 1980, *27* (2).

Nakamoto, Satoshi, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. Whitepaper.

Pagnotta, Emiliano and Andrea Buraschi, "An Equilibrium Valuation of Bitcoin and Decentralized Network Assets," 2018. Working paper.

Schilling, Linda and Harald Uhlig, "Some Simple Bitcoin Economics," 2019. Forthcoming, Journal of Monetary Economics.

Sockin, Michael and Wei Xiong, "A Model of Cryptocurrencies," 2018. Working paper.

Wüst, Karl and Arthur Gervais, "Do you need a Blockchain?," *IACR Cryptology ePrint Archive*, 2017.

Appendix

Definition of a symmetric consensus algorithm: Formally, define $\mathcal{M}_0$ to be messages of the form $m = \sigma(j, \emptyset)$ for some $j \in \mathcal{J}$. Recursively define $\mathcal{M}_k$ to be messages of the form $m = \sigma(j, M_{k-1})$ for some history M_{k-1} containing only messages in $\mathcal{M}_{k-1}$. Note that $\cup_k \mathcal{M}_k = \mathcal{M}$. Let $\pi : \mathcal{J} \to \mathcal{J}$ be a permutation. We define the permutation of a history $M \in \mathcal{M}_k$ by π as follows:

- For $m \in \mathcal{M}_0$ such that $m = \sigma(j, \emptyset)$, $\pi(m) = \sigma(\pi(j), \emptyset)$, yielding a definition of $\pi(M)$ for $M \subset \mathcal{M}_0$;
- For $m \in \mathcal{M}_k$ such that $m = \sigma(j, M)$ with $M \subset \mathcal{M}_{k-1}$, $\pi(m) = \sigma(\pi(j), \pi(M))$, yielding a definition of $\pi(M)$ for $M \subset \mathcal{M}_k$.

Then a consensus algorithm $(\sigma, \hat{\omega})$ is symmetric if $a_{\hat{\omega}(\pi(M))}(\pi(j)) = a_{\hat{\omega}(M)}(j)$ for all $j \in \mathcal{J}$. That is, in the state $\hat{\omega}(\pi(M))$ implied by the permuted history, the token holdings of $\pi(j)$ are the same as those of the original identity j in the original state $\hat{\omega}(M)$.

Proof of the double-spending lemma: Let the seller's identity be j_S. Denote the equilibrium communication strategy of the seller in round k of period t by $\sigma(j_S, M_{j_S,k,t})$. Similarly, let $\sigma(j^*, M_{j^*,k,t})$ be the buyer's equilibrium communication strategy. Here, $M_{j_S,k,t}$ and $M_{j^*,k,t}$ denote the message histories that emerge under honest play. Suppose the buyer plays the following strategy:

- At $t = 1$, communicate honestly with the seller according to $\sigma(j^*, M_{j^*,k,1})$, i.e., according to the consensus algorithm. Secretly create a new identity j'_S and generate all messages between j^* and j'_S that were sent during honest communication between j^* and j_S, creating a set of messages $M_{j'_S}$ that is equivalent to the set of messages M_{j_S} in all respects. Formally, $M_{j'_S}$ is equal to $\pi(M_{j_S})$ under the permutation $\pi(j_S) = j'_S$, $\pi(j'_S) = j_S$, and $\pi(j) = j$ for all other j.

– At $t = 2$, stop communicating with the seller. Let $\tilde{M}_{j'_S,k,2}$ be an altered history of messages observed by j'_S that omits all messages sent by j_S at $t = 2$. That is, the buyer pretends that the seller is no longer communicating. Communicate with the agent who was previously absent according to $\sigma(j_{S'}, \tilde{M}_{j'_S})$.

Under these strategies, at $t = 2$ the buyer and seller effectively pursue identical communication strategies with the previously absent agent. Each acts as if (1) they are the seller and (2) they no longer receive messages from the other agent. The only difference is that the seller communicates as identity j_S while the buyer communicates as j'_S. Formally, for each message m sent by the seller to the absent agent, the buyer sends the message $\pi(m)$.

These are both honest communication strategies. Therefore, the previously absent agent must come to a consensus with either the buyer or the seller (from the definition of a consensus algorithm). Given the symmetry of the consensus algorithm, in fact, consensus is reached with the buyer just as often as it is reached with the seller. The buyer pursues a communication strategy as if only j^* and j'_S had been present at $t = 1$. Hence, if the absent agent reaches agreement with the buyer on a state $\omega^{*'}$, all tokens in that state must either be allocated to j^* or j'_S, proving the proposition.