

**Cooperation in Infinite Games:
Applications to Finance and Public
Economics**

by

James Richard Lowery

Dissertation

Submitted to the **Tepper School of Business**

at **Carnegie Mellon University**

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

March 23, 2009

Abstract

Essay 1: Financial Intermediation, Trust, and Asset Values

In thin financial markets where intermediation is necessary to facilitate exchange, the intermediary may have an informational advantage in addition to his cost advantage for acquiring the security. If all information is eventually revealed, the intermediary may have an incentive to truthfully reveal the value of the security to a customer as soon as he learns it rather than attempting to profit from the asymmetric information. Whether this incentive is sufficiently strong depends on the patience of the intermediary and the probability that he will have future interactions with the same client. This probability depends on the value of the security because declining values lead to increased risk that the intermediary will fail. I study a model of repeated interaction between clients and intermediaries that takes into account the correlation between value and continuation probabilities. The model captures the unresponsiveness of thinly traded securities to bad news and explains the breakdown in liquidity following declines in asset values. This fact can help explain why relatively illiquid securities, such as those based on subprime mortgages, can experience apparent bubbles and crashes.

Essay 2: Imperfect Monitoring and Fixed Spreads in the Market for IPOs

Characteristics of the investment banking industry, particularly the extreme concentration of spreads at exactly 7%, seem consistent with some form of collusion through which underwriters can extract surplus from the IPO. I present a model of investment banking that, under the assumption of optimal collusion, generates a distribution of spreads qualitatively similar to that observed. The model is extended to show that underpricing and spread rigidity may arise together, each one reinforcing incentives to engage in the other.

Essay 3: Social Capital as Economic Overlap

This paper presents a model of endogenous social capital where location decisions can generate the necessary means to sustain cooperative behavior in the absence of legal institutions or social conventions. By choosing to locate close to each other, agents create public goods that facilitate cooperative behavior on other endeavors. The model can serve to explain both initial agglomeration decisions and cooperation in extra-legal environments, even in the absence of frequent repetition.

Committee:
Richard C. Green (Chair)
Burton Hollifield
Bryan Routledge
Juan Dubra

Acknowledgments

Contents

Abstract	ii
Acknowledgments	iv
Introduction	4
1 Financial Intermediation, Trust, and Asset Values	5
1.1 Introduction	6
1.2 The Game	10
1.3 Example: Higher Frequency Information	11
1.3.1 Discussion	16
1.4 Continuous Values	18
1.5 Bubbles, crashes, and crises: How bad can they be?	24
1.6 Policy Implications	29
1.7 Conclusion	31
1.8 Appendix	33
1.8.1 Description of Equilibrium in section 1.3	33
2 Imperfect Monitoring and Fixed Spreads in the Market for IPOs	41
2.1 Introduction	42
2.2 Model	44
2.2.1 Public History and Equilibrium	47
2.3 Spreads under Competition and Perfect Monitoring	49
2.4 Imperfect Monitoring	54
2.4.1 Optimal Rigid Spread	56
2.4.2 Partially Rigid Spread	57
2.4.3 Flexible Spreads	60

2.4.4	Comparing Values: An Example	63
2.4.5	Optimal Spreads	66
2.4.6	Impatient Firms	69
2.5	Underpricing	71
2.6	Seasoned Equity Offerings	75
2.7	International Comparisons	76
2.8	Policy Implications	77
2.9	Conclusion	78
2.10	Appendix	78
2.10.1	Finite Approximation	78
2.10.2	Competitive and Monopoly Spreads: General Model	80
2.10.3	Competitive Spreads	81
2.10.4	Oligopolistic Competition	82
2.10.5	Monopoly	83
2.10.6	Perfectly Observable Value and Preference	86
2.10.7	Partial Collusion with Impatient Firms	88
2.10.8	Perfectly Observable Value and Unobservable Preference	90
2.10.9	Proof of Proposition 5	91
2.10.10	Flexible Spread Upper Bound	91
3	Social Capital as Economic Overlap	96
3.1	Introduction	97
3.2	Model	100
3.3	Example	102
3.3.1	Costs of Crowding	105
3.3.2	Maintaining Cooperation	106
3.4	General Characteristics of Equilibria	112
3.4.1	Uniqueness	115
3.5	Applications and Implications	119
3.5.1	Financial Services	119
3.5.2	Formal Institutions and Social Capital	120
3.6	Three Agent Sequential Arrival	121
3.7	Conclusion	124
3.8	Appendix	126
3.8.1	proof of lemma 2	126

3.8.2	Proof of lemma 3	127
3.8.3	Proof of lemma 4	127
3.8.4	Proof of proposition 17	128
3.8.5	Proof of lemma 5	130
3.8.6	Proof of proposition 16	131
3.8.7	Proof of proposition 19	131
3.8.8	Proof of proposition 22	132
3.8.9	Tragedy of the Commons and Area under Cultivation: Parametric Example	132

Bibliography		135
---------------------	--	------------

Introduction

...

Chapter 1

Financial Intermediation, Trust, and Asset Values

1.1 Introduction

For many assets that trade in relatively thin markets, a significant decline in perceived value appears to lead markets in that asset to freeze up. Liquidity within a market seems to disappear after a downward price movement. This phenomenon is puzzling because it is not obvious that a downward movement in perceived value should adversely affect the gains from trade associated with an asset. Furthermore, markets appear to respond less quickly to information that a security is overpriced than to information that it is underpriced, introducing downward price stickiness and sudden crashes. These phenomena are often seen as evidence of uncertainty aversion or irrational behavioral biases in asset markets (see, for example, Routledge and Zin [2004]). I show that both of these phenomena can appear in a model where customers repeatedly purchase securities from an intermediary who has superior information about the value of the asset. Asymmetric price responsiveness and periodic liquidity breakdowns arise because the inventory risk borne by the intermediary affects the incentives for the intermediary to behave honestly and for the customer to trust her intermediary.

In many financial transactions, like those taking place in public equity markets, the role of a financial intermediary is limited to matching buyers and sellers, and the intermediary has little economic interest in the asset transacted and little information advantage. In contrast, here I focus on markets with trade that is less frequent and where the price and value of the asset are less transparent. Examples include markets for municipal bonds, high-yield corporate debt, and mortgage-backed securities. The model here will rely on three characteristics of financial intermediation in opaque markets. First, the financial intermediary holds an inventory of the traded asset that can be sizable for securities initially being placed.¹ This is important here since fluctuations in the value of the inventory change the likelihood that an intermediary will survive. Second, the financial intermediary is better informed about the value of the asset. This assumption is natural in a market without transparent pricing and where the final purchasers are generally retail investors or relatively unsophisticated delegated portfolio managers. In this situation, an intermediary will have an opportunity to attempt to exploit this private information to effectively cheat the customer, although eventually the customer will learn the true value of the asset. Third, financial intermediaries

¹These inventories are generally obtained directly from issuers or created through securitization of loans. Brunnermeier and Pedersen [2007] notes the tendency for market makers to be net long in the assets in which they specialize and cites the relevant direct evidence (Ibbotson [1976] and Hameed et al. [2005]).

transact repeatedly and a large portion of their value is represented by profits on future trades. This last characteristic will be crucial in mitigating the adverse selection problem arising from the intermediary's better access to information.

The requirement that the intermediary hold an inventory in the asset which he trades exposes the intermediary to risks associated with changes in the value of that asset. Since financial intermediaries make extensive use of leverage and short-term borrowing in general, they face the constant risk that they will not be able to meet some obligation if they face unexpected adverse conditions. In this paper, we treat this risk as the risk that the business fails; specifically, sufficient adverse events will cause an intermediary to cease to do business with clients, at least for a significant period of time. The crucial link between this risk of collapse, liquidity, and asset prices arises because perhaps the most significant risk faced by an intermediary is that the value of his inventory will decline to the point where he is effectively insolvent. Specifically, we assume that the probability that the interaction between the intermediary and the customer comes to an end is a function of the value of the asset that the intermediary attempts to sell the customer; when the value of the asset falls, the intermediary is more likely to face financial distress. This in turn causes the intermediary to discount the potential gains from maintaining a cooperative trading relationship more after he has learned that the value of his inventory has fallen.

This is effectively a reduced form approach to modeling the incentives facing a levered financial intermediary. Lowery and Routledge [2008] explore in more detail but in a simpler strategic interaction the effect of capital structure on the incentives of an intermediary. They find that indeed the increased risk of bankruptcy resulting from adverse changes in the financial position of an intermediary will decrease the relative weight the intermediary gives to future profits relative to current gains.²

The link between liquidity breakdowns, price stickiness, and bankruptcy risk is relatively straightforward once the role of repetition in inducing the intermediary to behave honestly is considered. Since the customer eventually learns the true value of the asset, he can change his behavior depending on whether the intermediary has behaved honestly in the past.³ Naturally, when the intermediary is sufficiently patient

²That paper considers several other mechanisms through which leverage affects the ability of intermediaries and customers to sustain cooperative trade and find an ambiguous relationship between leverage and the ability to sustain cooperative trade. Ultimately, however, leverage will cause sufficiently severe negative shocks to the worth of the intermediary to undermine cooperation in a manner similar to what is described here.

³For simplicity, here we consider a perfect monitoring game where the intermediary is able to

relative to the available short-term benefits to cheating the customer, he can be induced to reveal his private information about the security. When the customer knows that the intermediary cares relatively more about future trading opportunities versus short-term gains, the customer can then safely trust that the intermediary will accurately reveal his private information. If public information arrives that the value of the asset has fallen and thus the intermediary is at risk for bankruptcy, the customer will be less willing to trust the further information provided by the intermediary and trade will become less frequent. Information about the value of the asset, however, is most likely available to the intermediary before it is available to the customer. Thus, when bad news about the value of the asset first develops, the intermediary is likely to learn first that he is at risk of bankruptcy. He therefore has no incentive to honestly report the true value of the asset to the customer since he can profitably cheat the customer by refusing to reveal the bad news and because the customer, thinking the value is still high, will be prepared to trust the intermediary. This second effect is what leads to asymmetric price responses; the intermediary will hide bad news for as long as possible, perhaps even falsely reporting good news and generating an asset price “bubble” where prices rise while intrinsic value falls. Eventually, however, the customer will learn both that he has been cheated in the past and that the intermediary has no incentive to behave honestly in the future. Thus, bubbles will be followed not only by crashes but by periods of illiquidity.

In the model, the asymmetry in price response arises because the intermediary is selling assets and thus prefers the price to be high. If, however, we instead consider an intermediary who also purchases securities from customers, we would not then observe symmetric stickiness or reverse bubbles. Bubbles arise because of the coincidence of two events following the intermediary’s discovery of surprisingly bad news about the asset. First, the intermediary no longer has an incentive to invest in maintaining a trusting relationship with the customer. Just as importantly, however, the customer must be prepared to believe the lie the intermediary tells. Since asset values are at least somewhat persistent, the customer’s prior beliefs will make him susceptible to a lie reporting a small change in value. Following surprisingly good news, on the other hand, an intermediary would have an opportunity to cheat customers who need to sell their assets by pretending that the value has not changed. Surprisingly good news, however, will make the intermediary care more about the future because his risk of

observe *exactly* the final payoff of the asset and, consequently, the customer will know for certain if the intermediary reported the true value honestly.

failure has declined; an intermediary making a secondary market as well as a primary market will still maintain a long position in the asset, if for no other reason than that short positions in such assets are not feasible.

By assuming that the rate at which players discount the future changes as a function of a random state variable, this paper adds to the very small literature on repeated games with random discount rates. Baye and Jansen [1996] proves anti-folk theorems in the context of such games, while Dal Bo [2007] is the only application of such games that I am aware of. These types of games can be viewed as special cases of games with random payoffs where efficient cooperation breaks down, most notably Rotemberg and Saloner [1986]. The paper also relates to the theory of asymptotically finite games (for example, Bernheim and Dasgupta [1995], Jones [1998], and Jones [1999]) which is concerned with games where the likelihood that the game ends changes over time in a *deterministic* fashion.

The implications of intermediation for the pricing of thinly traded assets explored in this paper are complementary to those explored by Duffie et al. [2005]. I abstract from the effects on bargaining in the market by assuming a fixed “fair” division of surplus between customers and intermediaries, while they abstract from asymmetric information and inventory concerns, focusing on the role of bargaining in the price setting process. Another paper considering similar observations about asymmetric price responsiveness, Hong et al. [2000], documents that “bad news” appears to enter stock prices slowly. The authors argue that this is consistent with the behavioral model presented in Hong and Stein [1999] if informed insiders, specifically managers, prefer for the asset price to be high. While my model is designed to capture a very different trading environment in which prices are not publicly observable, it is related in the sense that the downward price stickiness arises from the intermediary’s preference for a higher price. My model, however, admits the possibility that that information provided by the informed insider is not verifiable and therefore can apply to environments where rumor and misinformation can exist.

Carlin et al. [2008] also consider the role of repetition in maintaining cooperation, and therefore liquidity, in financial markets; they focus is on interdealer markets with symmetric information. More generally, the importance of repetition to maintaining profitable trade has been explored in, for example, Greif et al. [1994], Routledge and von Amsberg [2003], Greif [1993], Greif [2006], and Dixit [2004].

This paper is organized as follows. First, I present a general model of repeated trade in an over the counter market. I then consider a very simple example of the model where

the asset value follows a finite Markov chain. This example introduces the concept of downward price stickiness (or bubbles), liquidity crises, and market unraveling. I then introduce a model with an infinite state space and solve (numerically) for the equilibrium of the stage game induced by a simple strategy of punishing any lie by refusing all future trade. This equilibrium presents more clearly the effects identified in the finite case. Finally, I present a limit result showing that market unraveling can lead to bubbles, crashes, and liquidity crises of arbitrary severity.

1.2 The Game

This section presents the financial intermediation game in a general, indefinitely repeated setting. This general model nests the numerical examples treated in the next two sections and the more abstract model treated later in the paper.

Two players, an intermediary I and a customer C , meet in every period $T = \{1, 2, \dots\}$. There is a security with a value process $\{v_t\}_{t=0}^\infty$, where $v_t \in V_t \subset V$. Each player receives a signal at the beginning of each period, x_t^I for the intermediary and x_t^C for the customer.

The customer's signal, x_t^C , is common knowledge, while x_t^I is the private information of the intermediary. The value process v_t is correlated with x_t^I and weakly correlated with x_t^C . That is, the customer's signal may be uninformative in some periods. Let $\mathcal{F}^I = \{\mathcal{F}_t^I\}_{t=0}^\infty$ and $\mathcal{F}^C = \{\mathcal{F}_t^C\}_{t=0}^\infty$ be the filtrations of the intermediary and the customer, respectively, with $\mathcal{F}_t^C \subseteq \mathcal{F}_t^I$. That is, the intermediary's information is finer than the customer's. To avoid issues associated with imperfect monitoring, however, we assume that $\mathcal{F}_t^C = \mathcal{F}_t^I$ infinitely often.

The stage game proceeds as follows: Each agent observes his signal. Then, simultaneously, the intermediary chooses a price $r_t \in V_t$ while the customer chooses the set of prices that he will accept, $A_t \subseteq V_t$. In equilibrium, we will see that this is equivalent to assuming that the customer chooses a threshold above which he will not purchase the security. We do not impose thresholds in the general model and thus permit the customer to potentially learn precise information about the value of the asset from the intermediary's offer. In the finite example we do restrict attention to threshold strategies for simplicity.

If $r_t \in A_t$, payoffs are

Intermediary: $c_I + (r_t - v_t)$

Customer: $c_c + (v_t - r_t)$

Otherwise, both players receive zero. Here, $\{c_I, c_c\}$ represent the gains to trade between the intermediary and the customer, reflecting different preferences for holding the security and the cost advantage for the intermediary in obtaining the security. The “fair” division of surplus between the intermediary and the customer is also given exogenously. We assume in the initial examples that there are infinitesimal costs associated with switching intermediaries and an infinitesimal probability that the intermediary is a commitment type who always chooses $r_t = v_t$. By appropriately choosing these parameters, we can guarantee that the best response of the customer after learning that his intermediary has charged a price above v_t is to switch intermediaries. As is standard, we ignore these infinitesimal quantities in the analysis.⁴

The game is repeated indefinitely, continuing with probability $\delta_t(v_t)$ in each period, where $\delta'_t(v_t) \geq 0$. Note that the continuation probability can depend both on the value and the period.

1.3 Example: Higher Frequency Information

I first analyze a simple four state version of the model that can demonstrate the basic relationship between continuation probabilities, liquidity, and the response of prices to information.

Here we assume that the uninformed agent learns the truth in every other period but has to make an uninformed decision about whether to buy at a given price in the intervening period. In this case, we can investigate how price responds to bad news when such news is only available to the intermediary.

⁴This assumption serves two purposes, one primarily expositional and one more fundamental. In the absence of the chance to switch intermediaries, the worst available punishment would be reversion to the static Nash equilibrium, which would provide small but positive continuation values to the intermediary. By assuming that the customer would change intermediaries, this continuation value can simply be set at zero. More fundamentally, if we seek the equilibrium that provides the greatest trade or the highest payoffs, grim trigger strategies are not optimal. A customer may learn that the intermediary has lied, but if the intermediary lied in a period where he could not have been expected to tell the truth (because continuation values are too low) the optimal equilibrium would require forgiveness. Assuming the ability to switch intermediaries allows us to focus on simple grim trigger strategies. It is also arguably more realistic, as customers are unlikely to continue to engage in business with someone they know is opportunistically seeking to cheat them. Finally, we assume that the intermediary will “tremble” and tell the truth with infinitesimal, positive probability, and that this probability is very large relative to the probability of facing a commitment type. This eliminates strategies where the intermediary seeks to influence the posterior beliefs about him being a commitment type in any way except avoiding revealing himself with certainty.

Let $V_t = V = \{1, 2, 3, 4\}$ and $x_t^I = v_t$ for all t , while $x_t^C = v_t$ for all odd t and \emptyset for even t . $\{V\}_t^\infty$ is a homogenous Markov chain with transition matrix

$$\begin{pmatrix} 0.5 & 0.3 & 0.15 & 0.05 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.05 & 0.15 & 0.3 & 0.5 \end{pmatrix}.$$

Let $c_I = 0.5$ and $c_c = 0.2$. Let $\delta_t = 1$ if t is odd. That is, the game never ends after a period with symmetric information. This assumption is strictly to simplify algebra. For even number periods, let $\delta_t(v_t) \equiv \delta_{v_t}$, where $\delta_1 < \delta_2 < \delta_3 < \delta_4$. That is, the security continues to exist with higher probability when the value is high. Note that a proper subgame begins in each odd numbered period, since both intermediary and customer know the full history and the current state at this point. I refer to each two period unit of time starting with an odd period as a stage. See Figure 2.1 for a summary of play in each stage.

In this example, the intermediary learns the true value of the security before trading, while the customer receives this information only every other period. In periods where v_t is not common knowledge before trading occurs, v_t is revealed immediately before the beginning of period $t+1$. For example, a firm issuing junk bonds will disclose information at regular intervals as required by accounting regulations, but intermediaries specializing in the industry will collect information relevant to the value between these disclosures.

We will consider simple strategies of the following form: The strategy has two phases. In the initial phase, the buyer and intermediary engage in the “fair” transaction in the odd periods.⁵ In the even period, the buyer chooses a threshold that depends on the previous odd period (note that the threshold will be mixed), and the intermediary reports a value. If the intermediary’s report turns out to be true, the game remains in this phase. Otherwise, the customer switches intermediaries. The customer then enters into a new relationship with an intermediary starting in the original phase, thus continuing to obtain the same payoffs. The intermediary, however, no longer receives any payoffs from interactions with this particular customer.

Given the assumption that strategies take this form, the game can be summarized

⁵Recall that the “fair” transaction refers to the transaction that a customer would complete with a commitment type intermediary.

as a finite game in a manner reminiscent of Abreu et al. [1990] and Spear and Srivastava [1987]. That is, the payoff to the intermediary of a particular strategy can be decomposed into the payoff for the current stage and the promised continuation values. In this case, the promised continuation value is exactly zero after every false report. For the continuation value following a truthful report, define γ_i as the continuation value for telling the truth in an even numbered period where the true value is i . Note that this is sufficient to cover all possible continuations since the value process is a one-stage Markov chain and complete punishment is always triggered by any misreport. Also, define U_i as the expected present value of play starting in an odd period where the value observed by the intermediary is i . Furthermore, define the stage-game payoff for the intermediary in an even period where the previous odd period had value i and the even period has value j as u_{ij} . We can now write U_i as follows:

$$U_i = \frac{1}{2} + \sum_{k=1}^4 P(k|i)(u_{ik} + \mathbf{1}_{truth,ik}\gamma_k),$$

where $\mathbf{1}_{truth,ik}$ is an indicator of whether a truthful report ever occurs when value k follows an odd period value of i . This expression for U_i is something of an abuse of notation since u_{ik} will not be well defined when the intermediary is playing a mixed strategy in which he sometimes tells the truth and sometimes overstates the value of the security. When the intermediary does play a mixed strategy, u_{ik} will represent the stage game payoff when he is honest.

The simplest way to analyze this game is to assume continuation values that can be promised to the intermediary after each even period (which will depend *only* on the current state in said even period). Assuming these continuations allows for the calculation of the possibly mixed strategies in each stage, taking into consideration the future benefits available to the intermediary for telling the truth. From these continuation values and the implied stage game payoffs, it is then possible to back out the probability of the game ending following a realization of each value. This approach is simpler than assuming continuation probabilities and then calculating continuation values because the continuation values are determined by the strategies that are played after each of the eight possible two-period price realizations, and these strategies are again determined by the available continuation values.

From the assumed continuation values, it is straightforward to calculate the implied

		Even State Value			
		1	2	3	4
Odd State Value	1	1.11	2	3	4
	2	1.7	2	3	4
	3	2.43	2.09	3	4
	4	2.54	2.38	3	4

Table 1.1: Average Price following each price path realization

		Even State Value				
		1	2	3	4	overall
Odd State Value	1	0.90	0.50	0.30	0.21	0.66
	2	0.57	0.50	0.30	0.21	0.45
	3	0.78	0.93	0.60	0.40	0.64
	4	0.73	0.78	0.60	0.40	0.53

Table 1.2: Liquidity following each price path realization

discount factor in terms of the U_i :

$$\delta_i = \frac{\gamma_i}{\sum_{k=1}^4 P(k|i)U_k}$$

Since it is straightforward to calculate u_{ik} (the payoff in a particular stage) given the assumed continuation values,⁶ the above expression gives the continuation probability associated with each value that is consistent with the assumed continuation values (γ_i).

For example, if we assume the customer can “reward” the intermediary with continuations of $\{0.25, 0.5, 1, 1.5\}$ following values in an even period of $\{1, 2, 3, 4\}$, respectively, the implied continuation probabilities are $\delta = \{0.169, 0.324, 0.578, 0.827\}$. High continuation values are indeed associated with high probabilities of the game continuing. That is, telling the truth is more valuable when the security value is high because the probability that the game ends is lower.

Continuing with this example, we can look at the strategies played by the customer and the intermediary for each realization of the two-period price path and show how the breakdown in trust leads to downward price stickiness and liquidity breakdowns.

Table ?? shows the average price of a transaction observed in the market following an odd period of *row* and an even period of *column*. That is, with a commitment to truth telling, all rows would read 1,2,3,4. This table shows that after “good news”

⁶For certain parameterizations, this game may have multiple equilibria. I focus on the Pareto-dominant equilibria when possible. I don’t know yet the extent or importance of multiple equilibria.

Value	1	2	3	4	overall
1	0.90	0.50	0.30	0.21	0.66
2	0.57	0.50	0.30	0.21	0.45
3	0.78	0.93	0.60	0.40	0.64
4	0.73	0.78	0.60	0.40	0.53

Table 1.3: Liquidity following each price path realization

about the asset value becomes available to the intermediary, it is immediately reflected in price; an increase in the value of the security between the common knowledge, odd numbered period and the even period where only the intermediary knows the value is immediately reflected in the price. Bad news does not fully enter the price. Instead, a large drop in value is hidden by the intermediary since he now knows (1) the game will probably end soon so he should attempt to squeeze as many profits as possible from the customer and (2) the customer is expecting a high value (since the price process is persistent) and will be willing to believe a false, high report. Thus, an econometrician who had *ex post* access to the time series of news available to the intermediary would conclude that prices respond more slowly to bad news. This sluggishness will effectively result from deception on the part of the intermediary. This deception takes one of two forms; for certain large price drops (from 3 to 1 or 4 to 1) the intermediary mixes over lies, sometimes reporting 2 and sometimes reporting 3. This mixing makes detecting the lie more difficult. For other price drops, the intermediary will mix between reporting the truth and reporting higher values. In principle, it is possible that an intermediary would lie after an increase from a low value to another still low value (say, report 3 or 4 after an increase from 1 to 2), but this is ruled out by the decreasing probability of the game stopping and the persistence in the underlying value process.

Table ?? shows the related liquidity story. Each cell is now the probability that the transaction goes through at any price, given a price path of *row* then *column*. The most revealing comparison here is between stages that begin with $v_t = 3$ and stages that begin with $v_t = 2$. The market is more liquid when starting from 3 than 2 regardless of the realized value in $t + 1$. This comparison follows from the fact that a customer must use the period t information about the value to estimate the probability that the game will continue after period $t + 1$, which will determine whether the intermediary has an incentive to report the truth. Liquidity is also high when the stage starts in the very lowest state, but this is an artifact of the assumption that the security cannot fall below this level. The intermediary knows that the customer will not trust any

reports since the customer's belief about the probability that the game continues is pessimistic, but the intermediary can credibly report that the value of the security is 1 since this is the lower bound of possible values. Liquidity also breaks down slightly at the upper bound (that is when the initial value is 4) simply because 4 is the most likely value to arise but is also the most profitable cheat given any other realization. A more symmetric problem where, for example, the information about value came from discrete time observations of Brownian motion would exhibit more monotonicity in liquidity but would also be far less simple to solve.

Two relevant simulated price processes are included in Figures 1.2 and 1.3 to demonstrate downward price stickiness and relative liquidity. Recall that there is asymmetric information only in even-numbered periods. The solid red line (which is always above the solid blue line) represents the average price in the market, while the solid blue line represents the value as observed by the intermediary. The dashed line, for comparison, is the price process in a one-shot version of the game. Liquidity in each even period is denoted by an asterisk (with the scale along the y-axis) and, for comparison, x 's denote the liquidity in the one-shot version.

1.3.1 Discussion

This section considers how the characteristics of the equilibrium of the stage game relate to price responsiveness and liquidity in a repeated asset market. We first consider the question of liquidity. In this model, liquidity in the over the counter market is measured by the probability that trade will occur. When available continuation values are very high, this probability is clearly one since the intermediary will never prefer to cheat. These continuation values can, in turn, be supported by high continuation probabilities and, crucially, a high probability that the asset value remains high. When the value is sufficiently persistent, high continuation probabilities are sufficient to guarantee high continuation values. When the asset price is not persistent or the continuation probability is very sensitive to small changes in value, it will be more difficult to maintain cooperation.

When cooperation can be sustained for high values, liquidity will break down as the value declines. As the value of lying increases relative to the value of honest reporting, the customer will have to decrease the value of lying by lowering the probability that he will accept a report of a given increase in the value in order to keep the intermediary mixing between honest reports and lies. As continuation probabilities get very low,

no cooperative equilibrium can be sustained and the game will resemble a standard lemons market, where trade occurs only after the lowest realizations. Furthermore, the asset price will cease to reveal the intermediary's information even if trade occurs.

When the asset value is relatively high but not very high, on the other hand, trade will occur frequently. Following the worst possible realizations the intermediary will lie about the asset value with positive probability. The customer will of course anticipate this possibility. The customer will still accept the trade with positive probability, but now the price does not perfectly convey the underlying information of the intermediary. Trust has effectively partially broken down, and the customer's beliefs following a trading period now place positive probability on two values. This decrease in the informativeness of price introduces the possibility of large price swings. Specifically, if very bad news arrives, intermediaries lie and report good news instead, and then more bad news arrives in the following period, we can see either a breakdown in liquidity or a large swing in price where the price change from one period to the next exceeds the feasible change in value of the asset. Once public information has confirmed that the value is low customers will know that they cannot trust the intermediaries and liquidity will dry up. For simplicity, we have assumed that public information from the previous period is revealed before each trading period, but in principle several periods could lapse between public signals. In such a case prices could deviate even more from fundamental values. The customer would continue to place positive probability on the true path of the asset value but observed prices might diverge significantly from that value, particularly since once an intermediary starts to lie he would continue to do so until the next public signal. Solving the model in this case of persistent private information is quite challenging technically and computationally⁷ and will be pursued in future research.

Finally, we have focused exclusively on the sell side. For the class of securities under consideration, this is a reasonable simplification since we are concerned with OTC securities without an active secondary market and thus without observable market prices. If we were to consider an intermediary who both places initial offerings and makes a secondary market the results would be substantially the same. The intermediary would still be long in the security for institutional and legal reasons and thus more exposed to bankruptcy following surprisingly bad news rather than surprisingly good news. When facing someone who wanted to sell the security, he would prefer to claim that the true

⁷See, for example, Williams [2008a], Williams [2008b], and Fernandes and Phelan [2000].

value was in fact low. The intermediary could effectively only get away with such a lie if the value had actually increased; the customer will tend not to believe very bad reports when he is trying to sell the security because refusing to sell after such reports will seldom prevent him from making a fair transaction. So, the intermediary will only be in a position to successfully cheat the selling customer after a surprising increase in the value of the asset. This is exactly when the intermediary has the *least* incentive to cheat since he is most likely to remain in business. Thus, the microstructure of the market generates asymmetries between the incentives to try to cheat buyers and the incentives to try to cheat sellers and explains the apparently asymmetric behavior of price deviations from fundamentals.

1.4 Continuous Values

The previous section demonstrates why a dealer facing inventory risk will prefer to honestly report good news while hiding bad news. It also shows that trade may become less frequent after bad news becomes common knowledge. The discreteness of the space of possible values, however, limits the analysis and introduces the contrary finding that assets may trade very frequently when public information indicates that the value is very low. Untangling the relative importance of these effects requires considering a more general, or at least more realistic, model of the asset value. In this section, I present such a model and describe behavior in a stage game that would be induced by the dynamic game described. Significant issues arise in attempting to guarantee the existence and optimality of such an equilibrium in the fully dynamic game, but even this simplified framework can provide intuition for how prices and liquidity will respond to the private information available to intermediaries. Analysis of the fully dynamic version of the game is the subject of ongoing research.

In this section, we assume that the customer knows the distribution of the value of the asset while the intermediary knows the value exactly. In an explicit dynamic setting equivalent to that investigated above, the parameters of this distribution would arise from the previous public signals received by the customer and the information that he could infer from the reports of the intermediary. We assume that the customers can reward honest reports with a continuation payoff $\gamma(v)$, which is increasing in v .⁸ The solution to this stage game can be characterized in a straightforward manner given

⁸For technical reasons, we allow the customer to choose to prevent the intermediary from receiving $\gamma(v)$ even if he responds truthfully.

the restrictions on strategies imposed by the presence of a commitment type and the exogenously given division of fair surplus.

Several characteristics of the equilibrium are immediate and will be useful for maintaining notational simplicity.

1. The customer must play a mixed strategy above some threshold.
2. The probability that the customer accepts the report is weakly decreasing in the report and strictly decreasing if said probability is less than 1.
3. The intermediary will always report the truth above some threshold and always lie below some threshold.⁹
4. The lie told will be increasing in the true value for all values that generate lies.

The proofs of each of these characteristics are almost immediate so we will simply state the basic logic. For characteristic (1), observe that if the customer were to believe all reports, the intermediary would prefer to report arbitrarily high values regardless of the true value. But, if the customer were to always refuse to buy at some value, the intermediary would never find it profitable to lie and report such a value. Then, the customer must believe, and therefore accept, any report of this value. So, the customer must mix between accepting and rejecting above some threshold. The probability of accepting must decline in the reported value because otherwise the intermediary would strictly prefer to make the higher report rather than the lower report conditional on lying since the continuation value available does not depend on what lie was told but only on whether a lie was told. Given such a decreasing probability of accepting a report, characteristics 3 and 4 arise. The lie is increasing in the true value because the intermediary is relatively more concerned about getting rid of securities with low values than securities with high values. Effectively, the costs of failing to sell a high value security at a given price are lower than the costs of failing to sell a low value security at that same price. The threshold arises from this fact and the fact that $\gamma(v)$ is increasing in v .

From the above characteristics, we can characterize an equilibrium with the following objects. Let v^* be the smallest value which the intermediary reports truthfully and let $\lambda : (0, v^*) \rightarrow (\lambda^*, \infty)$ give the lie told by the intermediary for any value less than v^* , where λ^* is the infimum of the lies. Finally, let $\pi^A : (\lambda^*, \infty) \rightarrow (0, 1)$ be the probability

⁹We will confine attention to problems where these thresholds are identical for ease of exposition.

that a customer accepts a report that might be associated with a lie. It is immediate that λ and π^A will be continuous in equilibrium.¹⁰

We can now write the payoffs to the intermediary and customer given a true value v and a report r as:

- If $v = r$ and customer accepts:

customer: c_c

intermediary: $c_I + \gamma(v)$

- If $v \neq r$ and customer accepts:

customer: $c_c + v - r$

intermediary: $c_I + r - v$

- If $v = r$ and the customer rejects:

customer: 0

intermediary: $\gamma(v)$

- If $v \neq r$ and the customer rejects:

customer: 0

intermediary: 0

An equilibrium of the game will then be defined as a function λ that leads the customer to be indifferent between accepting and rejecting any offer greater than λ^* and a function π^A that makes $\lambda(v)$ the optimal lie when the intermediary observes that the true value is $v < v^*$. It must also be the case that $\lim_{v \rightarrow v^*-} \pi^A(\lambda(v))(c_I + \lambda(v) - v) = \pi^A(v)c_I + \gamma(v)$, that is, the intermediary strictly prefers to tell the truth at the threshold but strictly prefers to tell lies below the threshold.

We can first find conditions for λ by imposing the requirement that the customer be indifferent between purchasing and not purchasing securities at any report greater than λ^* . In order to do this, we must calculate the expected value of accepting a

¹⁰The function λ will necessarily be continuous, but the domain may not be connected. To simplify exposition we assume that γ induces a connected set for lies.

report, which must be equal to zero. Letting f be the distribution of the asset value, that expectation is given as¹¹

$$\frac{f(v)}{f(v) + \lambda'(v)f(\lambda(v))}(c_c + v - \lambda(v)) + \frac{\lambda'(v)f(\lambda(v))}{f(v) + \lambda'(v)f(\lambda(v))}c_c.$$

Thus, we have the condition that, for all $v < v^*$,

$$f(v)(c_I + v - \lambda(v)) + \lambda'(v)f(\lambda(v))c_c = 0.$$

This identifies λ up to a constant as the solution to a nonlinear differential equation. The λ that solves this equation will necessarily be unbounded; very large values are associated with very low values of the density since the density is single peaked but has support over $(0, \infty)$, and thus $\lambda'(v)$ must go to ∞ for the indifference condition to hold. Since v^* must be finite (since $\pi^A(v) \rightarrow 0$ as $v \rightarrow \infty$), we know $v^* = \lim_{r \rightarrow \infty} \lambda^{-1}(r)$. This will serve, effectively, as the boundary condition.

We can now find conditions for π^A . This probability must be chosen by the customer to make the intermediary prefer to report $\lambda(v)$ when the true value is v . As such, the solution to the problem:

$$\max_r \pi^A(r)(c_I + r - v)$$

must be $\lambda(v)$ for all v . The first order condition is then

$$\dot{\pi}^A(r)(r - v - c_I) + \pi^A(r) = 0$$

which in equilibrium gives the differential equation

$$\dot{\pi}^A(r)(r - \lambda^{-1}(r) - c_I) + \pi^A(r) = 0$$

which has a solution of the form:

$$\pi^A(r) = k \exp \left\{ - \int^r \frac{1}{s - \lambda^{-1}(s) + c_I} ds \right\}$$

with initial condition

$$\pi^A(\lambda^*) = 1.$$

¹¹We arrive at this expression by considering the expected value of a report of $v \in (\lambda(v) - \varepsilon, \lambda(v) + \varepsilon)$ as $\varepsilon \rightarrow 0$. This is the only calculation for the regular conditional distribution consistent with the Borel σ -algebra.

Thus,

$$\pi^A(r) = \frac{1}{\exp \left\{ \int_{\lambda^*}^r \frac{1}{s - \lambda^{-1}(s) + c_I} ds \right\}}.$$

We have thus solved for the equilibrium, up to finding λ^* such that

$$\lim_{r \rightarrow \infty} \lambda^{-1}(r) = v^* \quad (1.1)$$

$$\lim_{v \rightarrow v^*} \pi^A(\lambda(v))(\lambda(v) - v + c_I) = c_I + \gamma(v). \quad (1.2)$$

Note that the probability of accepting a report of v^* must be one by the assumption that the intermediary plays a pure strategy since, were $\lambda^* < v^*$, any report $r < v^*$ would always be associated with a lie and would therefore be rejected.¹²

If it is possible to find a λ^* such that the above holds, then that defines a perfect Bayesian Nash equilibrium, where we define off-equilibrium path beliefs to induce the customer to reject with probability 1 and the intermediary to expect not to receive the reward or have the deal accepted even if he reports the truth. Since, however, the payoff to following the optimal lie function is not monotonic, it is not immediate that there will not be some region entirely below the posited threshold where telling the truth is slightly preferred to lying if the customer always rewards the intermediary for honest behavior, even off the equilibrium path. Furthermore, the customer must accept with probability one any report of a value less than c_c . Whether these complications are important will depend on the shape of the γ function. In the dynamic game, there will be significant freedom in the choice of γ , so an equilibrium with these basic characteristics will exist. This equilibrium would not necessarily be optimal in the sense of producing the most frequent honest trade if the maximum available continuations grow slowly in the value.¹³

Figure 1.4 shows an example of the strategy of the intermediary in an equilibrium of the form described here. The asset value is distributed lognormally. When the true value of the asset turns out to be relatively high, the continuation values available are sufficient to induce the intermediary to reveal his private information. When the true value is low, and in particular when the true value is low relative to the prior of the customer, the intermediary will find cheating tempting. The lies the intermediary tells

¹²The customer will only accept reports that he knows are lies when the report is less than c_c . We ignore this for the time being and treat those cases separately.

¹³In this case, λ will be continuous but over unconnected support and π^A would exhibit a kink, significantly complicating the analysis.

are not only increasing in the value, but are increasing particularly steeply. The slope of the lie function is determined by the need to keep the customer indifferent between buying and refusing to buy the security. The very highest values of the asset, however, must occur relatively infrequently since the density function must decrease toward zero. In order to induce the customer to be just willing to purchase following such a high report, the lie function must become increasingly and, in the limit, infinitely steep. When the lie function is very steep, any small interval of reports of a very high value is associated with an extremely small interval of low values that would generate lies of those values. This is the only way in pure strategies for the intermediary to induce the customer to be indifferent between purchasing and not purchasing following very high reports.¹⁴

A crucial point of this example is that when the possible values of the security are not bounded from above, the possible divergence between the true value and the reported value is also not bounded. Securities with low values will be reported as having higher values, and for some low values the report will be extremely high. In this sense, a bubble can arise; bad news about the security can generate reports of very good news. This is the exact behavior that was observed by certain sectors of Bear Sterns immediately before its collapse(Goldstein and Henry [2007]). In the credit default swap market, another type of exotic and thinly traded security, bond insurers remained “upbeat” even after learning about the extent of their exposure to losses.¹⁵

Another point to take away from this example is that the posterior beliefs of a customer about the true value of the asset will be markedly “bimodal.” That is, reports of values in some high region will be associated with true values either in that region or in a much lower region. If information gleaned from intermediary reports is used to make irreversible decisions about real investments, this has potentially serious consequences. While the customer will form the correct beliefs given the observed report, the ability to adjust activities to take into account the possibility that the underlying state is vastly different may be limited and adjustments costs may be high following the eventual revelation of the true state. In this sense, the breakdown in information transmission associated with financial intermediation in opaque markets may end up having spillover effects in the real economy. Only for certain moderate

¹⁴Were we to confine attention to finite but extremely fine price reports, this same effect would have to be achieved through mixed strategies, complicating the analysis significantly.

¹⁵“ACA Financial Guarantee ‘has never been in better financial condition’ said...[the] chief financial officer.” Pulliam and Ng [2008]

levels of the value will the customer be able to certainly trust the intermediary since the smallest lie told will, in general, exceed the lowest value for which the intermediary tells the truth. These periods where the customer learns that truth with certainty following the report of the intermediary are, of course, the only states in which the customer expects positive profit. As the available continuation values increase, the smallest lie will also increase, making prices more informative and providing the customer with greater expected profits.

1.5 Bubbles, crashes, and crises: How bad can they be?

This section addresses the question of how severe a bubble or liquidity crises could be in the environment described here. A partial answer to this question is given by a limit result which indicates that infinitesimal information that agents learn about the value process can lead to changes in behavior in equilibrium that are easily interpreted as liquidity crises or bubbles, and these events can be as severe as possible given the specification of the game. To understand this result, we first describe in more detail the basic mechanisms that lead to the behavior described in this paper.

There are two basic dynamics governing the extent to which cooperation can be maintained in the types of games considered here. First, there is a direct effect associated with the probability that the game ends. Other things equal, a higher probability of continuing the game leads to (weakly) more cooperation being sustainable, and thus to more liquidity. The second dynamic is slightly more subtle. The ability to sustain cooperation depends not just on the likelihood that the game continues but also on the likelihood that the game continues along a path where cooperation is maintained.

While the primary effect of the reputations concerns addressed here can be understood largely through the first channel, the second channel will greatly influence the qualitative and quantitative characteristics of the equilibria that may arise. Specifically, the sensitivity of the degree of cooperation to small informational events will be determined primarily through the second channel. This section discusses the role of this second channel and presents a result showing that this potential for cooperation to unravel along certain paths can lead to arbitrarily sudden and severe liquidity crises and arbitrarily explosive price bubbles.

First, we note that it is trivial to construct equilibria in settings similar to those

treated in previous sections where the second channel is of primary importance. Basically, if values are very persistent, cooperation can be maintained even when the probability of continuing is relatively moderate since future periods are likely to also be associated with such moderate probabilities of continuing. If, however, values (and therefore continuation probabilities) are less persistent, the value may increase or decrease. Increases would lead to more likely continuations, while decreases would lead to less likely continuations in future periods. Decreases, however, will have the secondary effect of causing the game to enter an “uncooperative phase” where liquidity breaks down and the gains to trade vanish, at least until the value rises again. The risk of entering such phases decreases the available punishments since the threat of reverting to a punishment phase is far less potent when there is a significant chance that the payoffs even for cooperating are low. This effect has the potential to completely unravel cooperation, even in periods where the game continues almost surely.

As mentioned, constructing such an example is a simple task but does not provide much insight into what characterizes situations in which this unraveling takes place. Such insights are important because this potential unraveling can lead to catastrophes in the sense that seemingly small changes in beliefs can induce both liquidity crises and bubbles. Fortunately, a reasonably straightforward application of the theory of asymptotically finite games developed in Bernheim and Dasgupta [1995] can provide some insights into the scenarios in which liquidity and price informativeness may disappear quickly. In the remainder of this section, we develop a limit result for a subclass of intermediation games. The goal of this exercise is to show that the unraveling of cooperation associated with the inability to sustain cooperation when the game is very likely to end can lead very small changes in the beliefs of the players about the exogenous state of the game to generate catastrophic changes in the behavior of the players. To understand why this is the case, it is necessary to first introduce the concept of an asymptotically finite game and describe the fundamental results on this important class of indefinitely repeated games.

The basic idea of asymptotically finite games is that when the probability of the game continuing declines deterministically toward zero, equilibria providing payoffs higher than perpetual play of the stage game Nash equilibrium¹⁶ can be sustained

¹⁶This theory is concerned exclusively with games with a unique stage game Nash equilibrium to avoid overlapping with Benoit and Krishna [1985]. We will follow this tradition by focusing on intermediation games with unique Markov strategies. Given the importance of inventory in the model, this focus is without loss of excessive generality.

only when the probability of continuing falls sufficiently slowly to permit an orderly unwinding of cooperation. If the probability changes too quickly or the action space is discrete, such a gradual unwinding is impossible. Specifically, Bernheim and Dasgupta [1995] are able to precisely characterize the set of games with asymptotically finite horizons in which payoffs strictly superior to the stage game Nash equilibrium can be obtained. Somewhat remarkably, this characterization depends exclusively on the rate at which the continuation probability descends toward zero; if $\lim_{\tau \rightarrow \infty} \sum_{k=1}^{\tau} \frac{1}{2^k} \ln \delta_k$ is summable, some cooperation, in the sense of actions that lead to payoffs that Pareto dominate the payoffs in the stage game Nash equilibrium, can be sustained in equilibrium for a very general class of games.

Such a clean characterization is, unfortunately, unavailable in the setting we consider. To see why, note that in general neither the intermediary nor the customer will know what “path” to the end of the game they will follow. In a given period, punishments sufficient to enforce better than Markov¹⁷ equilibrium payoffs will be available only if the game is sufficiently likely to remain on a path on which cooperation is maintained. But, of course, which paths involve cooperation will depend in part on the extent to which cooperation can be maintained at the value in question, and so forth. Consequently, the problem is fundamentally recursive.¹⁸ It is, however, fairly simple to answer the question of how important this unraveling effect may be. Specifically, by looking at a game that is close in the appropriate sense to the Bernheim and Dasgupta [1995] setting, we can see that a very small change in beliefs about the process describing the value of the asset can lead to an arbitrarily severe drop in liquidity and arbitrarily large bubbles.

To see this, we will construct a highly stylized version of the financial intermediation game. This version of the game focuses on the importance of information about the path the game is likely to follow. This information would generally be revealed in the information about the current value and possibly also in the information about value changes. That is, a large drop in value may signal not only that the future distribution of values is lower (due to the persistence of value), but may also signal

¹⁷We are, of course, considering a dynamic, rather than simply repeated, game. This concern is of little importance as long as the strategy space and payoffs of the game are stationary up to the most recent period where all information is common knowledge. This clearly conflicts with the specific examples described in previous sections but is consistent with the general approach where customers care directly only about changes in value rather than the value itself.

¹⁸It is natural to ask whether an algorithm similar to that developed in ? can be constructed to compute equilibria for this type of game. This is likely to be the case, but remains for future research.

that the likelihood of large drops is higher than previously believed. Here, we assume simply that the intermediary and customer may observe some “bad news” event B that alters their beliefs about the path that the values during the common knowledge periods follow. We also assume that the game can only end immediately *before* some common knowledge period, there are m periods where the intermediary learns more information about the value than the customer following every common knowledge period, and the probability of the game ending depends exclusively on the most recent *common knowledge* value. The game is stationary in the sense that $v_{t+1} - v_t$ is distributed identically to $v_{t+1+(m+1)} - v_{t+(m+1)}$ whenever v_{t+1} is not common knowledge, for all t . That is, if we call each set of periods starting with a common knowledge period and ending the period before the next common knowledge period a “stage” the game is effectively a repeated (rather than a more general dynamic) game since the strategy space is identical up to a constant and the payoffs are identical in all periods. This environment and additional technical assumptions¹⁹ guarantee that when the path of the value in common knowledge periods is deterministic and eventually declines such that the probability of the game continuing reaches zero at an infinite horizon, the game is strategically equivalent to a repeated game with an asymptotically finite horizon.

To simplify exposition, we will index stages with k to distinguish them from periods, indexed with t . The notation v_k will refer to the value of the asset at the beginning of the stage, and $\delta(v_k)$ will designate the probability that the game ends after a stage where the value begins at v_k . We consider the following game. For the first K^* stages, v_k follows a deterministic process where v_k remains sufficiently high such that $\delta(v_k)$ remains in some small neighborhood of 1. After stage K^* , if even B did not occur in an period, v_k declines *deterministically* at a rate such that $\delta(v_{k-K^*}) = a\lambda^{(2-\varepsilon)^{k-K^*}}$. If, however, event B did occur, $\delta(v_{k-K^*}) = a\lambda^{(2+\varepsilon)^{k-K^*}}$. As long as δ is a continuous function of v_k , the event B contains, in a sense, a trivial amount of information about the value process as $\varepsilon \rightarrow 0$.

Our goal is now to show that the trivial information event B can completely unravel cooperation in the following sense. There exist games of the form described such that, if B arrives in a common knowledge period, behavior in any equilibrium that is not Pareto dominated switches from perfect liquidity and informative prices to repeated play of

¹⁹Intermediaries observe and arbitrarily fine but finite coarsening of the true value but report prices from an infinite, compact interval of the real line in each period, while customers choose the percentage of the offering to purchase following a report within some small exogenously given interval. Assumptions A-C of Bernheim and Dasgupta [1995] are satisfied.

the stage game Nash equilibrium, which we interpret as a liquidity crisis. If B should arrive in a period that is not common knowledge, then the crisis will not begin until the next stage begins (and the customer learns B); in the intervening periods we will have a bubble, where the intermediary reports almost the highest possible value and the customer purchases nearly all²⁰ of the offered security with arbitrarily high probability. This result is a limit result; as the probability of the event B occurring decreases toward zero, behavior is as described. The basic point is that, when B is unlikely, players who have not observed B will play as if they are playing an asymptotically finite game where the likelihood that the game ends increases slowly enough to permit payoffs that exceed the stage game Nash equilibrium. When the end of the game is sufficiently far off, a folk theorem holds and for an arbitrarily large number of initial stages the game proceeds as if it were infinitely repeated with almost no discounting. Should B occur, however, the agents would start playing as if they were playing an asymptotically finite game with a rapid decrease in the probability of continuing. Since the equilibrium value set for asymptotically finite games is discontinuous in the rate at which the value declines in the distant future, this switching can produce arbitrarily sudden changes in behavior in the dynamic game we consider, even when agents only very slightly change their beliefs about the process governing the exogenous variables.

Clearly, as B becomes more likely or the learning process about the future trajectory of values becomes more smooth the severity of the crises and bubbles will wane in the sense that the drop in liquidity and the deviation of prices from fundamentals will be lower. The point of the exercise here is to show that the contribution of the unraveling of cooperation can, in principle, generate crises and bubbles of effectively unbounded severity, and thus it should not be surprising that we do, in fact, observe bubbles and crashes of otherwise surprising magnitude in environments similar to those described here.

This section concludes by formalizing the above argument by explicitly invoking the appropriate results from Bernheim and Dasgupta [1995]. The first proposition establishes that all cooperation will cease after event B :

Proposition 1. *If event B occurs in stage $k < K^*$, no subgame starting at $k' > k$ involves play of any strategy other than the stage game (or, equivalently, Markov) Nash equilibrium.*

Proof. Note first that the concept of a subgame does, in fact, impose restrictions in this

²⁰Not almost all, in the formal sense

game since all uncertainty across stages in the game is resolved after event B occurs. All subgames following event B must involve no cooperation by theorem 2 of Bernheim and Dasgupta [1995]. \square

The above proposition is sufficient to demonstrate that event B , which contains a trivial amount of information about the value process, completely undoes any cooperation that was being sustained up to event B . It remains only to show that, in some games, the degree of cooperation that can be sustained before event B is nearly complete:

Proposition 2. *Assume event B occurs in every period with probability η , conditional on B not having occurred before. Then, for every $K \in \mathbb{N}$ and $\varepsilon > 0$, there exists an η and K^* such that any individually rational payoff of the stage game can be achieved in the first K rounds.*

Proof. It is immediate from proposition 4 of Bernheim and Dasgupta [1995] that any individually rational payoff can be sustained in the first K periods for sufficiently large K^* when $\eta = 0$ and $\delta(v_{k-K^*}) = a\lambda^{(\alpha-\frac{\varepsilon}{2})^{k-K^*}}$.

Now, suppose $\delta(v_{k-K^*}) = a\lambda^{(\alpha-\varepsilon)^{k-K^*}}$ when $\eta > 0$ but event B happens not to occur before K^* . We can now construct an equilibrium where, for small enough η , agents achieve any individually rational payoff for the first K periods. This equilibrium simply calls for exactly the same strategies to be played as in the equilibrium for $\eta = 0$. To see that no agent has an incentive to deviate from such a strategy for η sufficiently small, observe that at stage $K^* - 1$ both players would strictly prefer to adhere to the strategy assigned in the game with $\eta = 0$ mentioned immediately above when η is close enough to zero. This is immediate because the difference between the payoff to adhering to the strategy and the payoff to reverting to Nash equilibrium is strictly greater than in the $\eta = 0$ case because the game is more likely to continue in each period conditional on B not occurring, but we can make the probability that B occurs between $K^* - 1$ and K^* arbitrarily small. Now, backwards induction shows that no player has an incentive to deviate at any point in the game, unless they observe B . \square

1.6 Policy Implications

This paper argues that the proximate cause of both liquidity breakdowns and the failure of prices to respond smoothly to bad news is the vanishing incentives for financial

intermediaries to provide truthful information about a security when the value of the security falls. The incentives disappear because the intermediary's specialization in and exposure to the asset in question imply that the intermediary will be less likely to be in a position to complete profitable transactions in the future. When the intermediary's customers are aware that the intermediary will behave opportunistically in the short run, they hesitate to buy from him except at the lowest possible prices. If, on the other hand, the intermediary's customers are unaware of the decline, then the intermediary will take advantage of their continued trust and will not inform them of the adverse news. Customers, in turn, will have to take into consideration this risk when deciding how frequently to buy securities and at what price, thus making the market less liquid even in good states.

This situation presents two possible concerns for policymakers. First, from a traditional social-planner perspective, agents do not reach an unconstrained Pareto-optimal allocation. Gains to trade that would be realized if information were symmetric are lost. Second, customer's effectively face being cheated by their intermediary. From a strictly economic perspective, it is hard to argue that this is a problem since the customer fully anticipates this possibility when deciding to engage in trade. The customer can in fact guarantee that he is not cheated by only agreeing to buy the security when the intermediary offers it at a very low price, but this will be neither best for him nor an optimal equilibrium. From a policy perspective, however, such "cheating" may be viewed as undesirable. Furthermore, if there are externalities from having information flow from intermediaries to customers then the lack of consistent truthful reporting of private information has additional costs. It seems quite plausible that a well functioning, liquid market would have positive spillover effects on other markets since liquidity problems appear to have a tendency to spread among seemingly unrelated sectors.

Taking as given that the policy maker does not have access to the private information of the intermediary and cannot simply take over the trading business to exploit its own symmetric ignorance with the customer, a natural policy to consider is one of bailing out intermediaries in financial distress. This policy is particularly relevant in the case where the game stops because of bankruptcy caused by overexposure rather than by the actual disappearance of a market. If an intermediary can always count on transfers or easy loans from the government or central bank when he faces bankruptcy, he will continue to provide accurate information about the security since his business will not be in danger following a collapse in the intrinsic value of his holdings.

This conclusion, however, should not be taken as a policy recommendation. Such

a policy could prove quite costly. Eliminating the possibility of large losses on an intermediary's portfolio might encourage the intermediary to take on more correlated risk than necessary for its core business, requiring that the social planner engage in frequent and large bailouts. A policy based on making collateralized loans to intermediaries in distress would also present a problem. Ideally, a lender of last resort makes collateralized loans at a penalty rate to balance-sheet solvent but illiquid financial intermediaries. Here, the illiquidity of the market is a result of the insolvency of the intermediary, meaning that the loans made by a central bank would lead to losses with high probability. In particular, if the social planner intervenes in a period of illiquidity by purchasing securities that are not traded or making loans backed by those securities, he may face large losses if he bases his valuation of the security on the recent prices in the market since the prices that prevail before a market becomes illiquid may be based on false reports of good news. Furthermore, if the intermediary expects that he may be able to sell his inventory to the social planner at prices closer to those that prevailed *before* the liquidity event, he will no longer have an incentive to sell even the worst securities at a pooling price. Thus, the increase in liquidity that might be observed in the very worst states can be undermined by even the suggestion that the government will purchase "toxic" assets. The social planner then must carefully balance the benefits of eliminating bankruptcy risks with the costs of interfering with the incentives for prudent behavior by the intermediary.

Finally, a policy of subsidizing the takeover of a distressed intermediary, as suggested in Acharya and Yorulmazer [2008] in the context of banks and implemented by the Federal Reserve in response to the Bear Stearns debacle, would be counterproductive in the context of this model. Maintaining a policy of forcing a firm in distress to close and hand over its assets to another firm rather than permitting it to attempt to save itself for as long as possible will lead employees to act as if they have an even shorter time horizon and may lead to unraveling of liquid trading in even less distressed states.

1.7 Conclusion

This paper has proposed a general framework for considering asymmetric information between intermediaries and customers in thin asset markets. The model takes into consideration the potential gains from trade and the role of reputation in maintaining

trust. The key feature of the model is that the degree of trust between the customer and the intermediary is a function of the underlying value of the asset since the value of the asset determines the probability that profitable trading opportunities will arise in the future. Examples that demonstrate the role of this link in liquidity breakdowns and price stickiness show that certain qualitative characteristics of asset prices can be understood as resulting from the breakdown in trust that occurs when values fall. Further work is needed to apply the model to a richer environment that will allow the full implications of this interaction to be developed.

1.8 Appendix

1.8.1 Description of Equilibrium in section 1.3

The strategy of the retail customer is measurable with respect to the common knowledge (odd numbered) period value. That strategy is summarized in the following table:

		Threshold			
		1	2	3	4
Odd State Value	1	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{3}{35}$	$\frac{3}{14}$
	2	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{3}{35}$	$\frac{3}{14}$
	3	0	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
	4	0	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

The intermediary's strategy is, of course, more complex as it is measurable with respect to both the odd and even period states. The following four tables summarize the strategy:

Odd State:		1				Odd State:		2			
		1	2	3	4			1	2	3	4
Truth	1	$\frac{17}{21}$	$\frac{3}{20}$	$\frac{1}{30}$	$\frac{1}{140}$	Truth	1	$\frac{11}{504}$	$\frac{5}{8}$	$\frac{1}{9}$	$\frac{1}{28}$
	2	0	1	0	0		2	0	1	0	0
	3	0	0	1	0		3	0	0	1	0
	4	0	0	0	1		4	0	0	0	1

Odd State:		3				Odd State:		4			
		1	2	3	4			1	2	3	4
Truth	1	0	$\frac{4}{9}$	$\frac{5}{9}$	0	Truth	1	0	$\frac{1}{3}$	$\frac{2}{3}$	0
	2	0	$\frac{8}{9}$	0	$\frac{1}{9}$		2	0	$\frac{17}{27}$	0	$\frac{10}{27}$
	3	0	0	1	0		3	0	0	1	0
	4	0	0	0	1		4	0	0	0	1

Here, each table represents the strategy of the intermediary following the realization of the value in the common knowledge period given in the table heading. Then, the row represents the true value in the next period (which is not observable to the customer) and the column represents the probability that the intermediary will give the report associated with that column.

The above equilibrium implies the following payoffs to the intermediary following each pair of realizations. These payoffs are the u_{ik} from the main text. Recall that if an intermediary plays a mixed strategy that sometimes calls for truth telling and

sometimes calls for lying, the associated u_{ik} is when the intermediary tells the truth.

		Even			
		1	2	3	4
Odd	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{20}$	$\frac{3}{28}$
	2	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{20}$	$\frac{3}{28}$
	3	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{5}$
	4	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{5}$

It is now straightforward to verify that the continuation values posited in the main text, $\gamma = \{0.25, 0.5, 1, 1.5\}$ are in fact the continuations implied by the equilibrium here. It is of course necessary to note that states 31 and 41 involve always lying when making said calculation.

We can also calculate the price transition matrix:

	1	2	3	4
1	0.452	0.342	0.156	0.051
2	0.077	0.609	0.212	0.103
3	0	0.248	0.543	0.210
4	0	0.144	0.327	0.529

The asymmetry of the above matrix compared to the state transition matrix highlights the asymmetry of price response to news. It is also informative to summarize the posteriors of the customer following the reports of the intermediary in order to see how well the customer can estimate the state following a transaction. It is clear that such price information is not fully informative as it would be in a frictionless economy; furthermore, even the information that is revealed to the customer is not generally available to non-participants in the trade, so the informativeness of trade is limited to those engaged in the trade.

Odd State:		1				Odd State:		2			
Report	1	1	2	3	4	Report	1	1	2	3	4
	2	1	0	0	0		2	1	0	0	0
	3	0.2	0.8	0	0		3	0.2	0.8	0	0
	4	0.1	0	0.9	0		4	0.1	0	0.9	0
	4	0.067	0	0	0.93		4	0.067	0	0	0.93

Odd State:		3				Odd State:		4			
Report	1	Undef.	Undef.	Undef.	Undef.	Report	1	Undef.	Undef.	Undef.	Undef.
	2	0.2	0.8	0	0		2	0.15	0.85	0	0
	3	0.1	0	0.9	0		3	0.1	0	0.9	0
	4	0	0.1	0	0.9		4	0	0.1	0	0.9

The guess-and-verify method of finding an equilibrium used above does not immediately guarantee uniqueness. That is, it remains possible that, for the posited δ 's, there

is an equilibrium that involves more trade. This possibility arises because continuation values higher than those derived here may imply that trade occurs more frequently, which will in turn support the higher continuation values. As the goal of this paper is to show that the correlation between continuation probabilities and asset values leads to liquidity breakdowns and slow transmission of bad news, it is necessary to show that equilibria of the game that support trade generally exhibit these characteristics. To show this, we will posit the existence of equilibria where prices and value move together more and more trade occurs than in the equilibrium derived and show that such equilibria do not exist.

We can first show that there is no equilibrium in which the intermediary reports truthfully in every period. This follows from the fact that the continuation promises required to sustain such truthful reporting in every period are inconsistent with the continuation payoffs available.

That is, when the intermediary reports truthfully in every period, the only equilibrium possible will be the one in which the customer sets his threshold at 4 in every period. This implies that the intermediary will receive a stage game payoff of $\frac{1}{2}$ in each period, his fair share of the surplus. Consequently, we can express the continuation value available following a realization in an even numbered period of i as

$$\gamma_i = \delta_i \sum_{k=1}^4 P(k|i)U_k$$

where

$$U_k = \frac{1}{2} + \sum_{j=1}^4 P(j|k) \left(\frac{1}{2} + \gamma_j \right)$$

$$U_k = 1 + \sum_{j=1}^4 P(j|k)\gamma_j$$

since the stage game payoff is always $\frac{1}{2}$ and the game always remains in the cooperative phase. Thus, we have

$$\gamma_i = \delta_i \left(1 + \sum_{k=1}^4 P(k|i) \sum_{j=1}^4 P(j,k)\gamma_j \right).$$

This defines a set of 4 independent linear equations which determine γ uniquely. In

this example, we have $\gamma = \{0.30, 0.60, 1.17, 1.77\}$. But, in order to sustain truthful reporting in every period, it must be the case that

$$\frac{1}{2} + \gamma_1 \geq 3\frac{1}{2}$$

since otherwise the intermediary will have an incentive to report 4 following a realization in the even period of 1. This clearly does not hold, so truthful reporting in every period is not an equilibrium.

In the posited equilibrium, state 3 and 4 are always truthfully reported, while state 2 is truthfully reported following the lower states but not always truthfully reported following the higher states. This characteristic of the equilibrium is central to the idea that prices are sticky downward. We therefore now show that there is not an equilibrium where states 2,3, and 4 are always truthfully reported.

Assume there is an equilibrium where states 2,3, and 4 are always truthfully reported. This requires that, following a realization of 2 in an even period, the intermediary expects an opportunity to cheat at some point in the future and that opportunity has present value in excess of 1.4 (since it is impossible to achieve a continuation above 0.60 starting from state 2 and engaging in only fair transactions, as shown in the discussion of non-existence of a fully truthful equilibrium).²¹ The value of the cheat is maximized if it becomes available in the next stage. But the value of the cheat is bounded above by 3 the intermediary will have at most one opportunity to cheat. Furthermore, the cheat can only occur when the value in the even period of the stage is 1 since, by assumption, truthful reports follow all other realizations. So, the present value of the cheat is bounded from above by $3 \times 0.324 \times P(s1|2)$, where $P(sx|y)$ is the probability that the next stage ends in state x conditional on the current state ending in state y . Note that this is not a very tight upper bound because it assumes that the retail customer continues to purchase in every period, which would not in general occur in an equilibrium with lying. The value of the cheat is clearly below 1.4 and therefore truthful reporting in state 2 cannot always be sustained.

²¹It is possible for the customer to reduce the value of cheating by refusing to purchase the security even when said customer believes that the report is true with probability 1. We will assume that the customer is sufficiently impatient to guarantee that he cannot commit to such a strategy. In any case, such a strategy will also reduce continuations available and will not increase the set of equilibria.

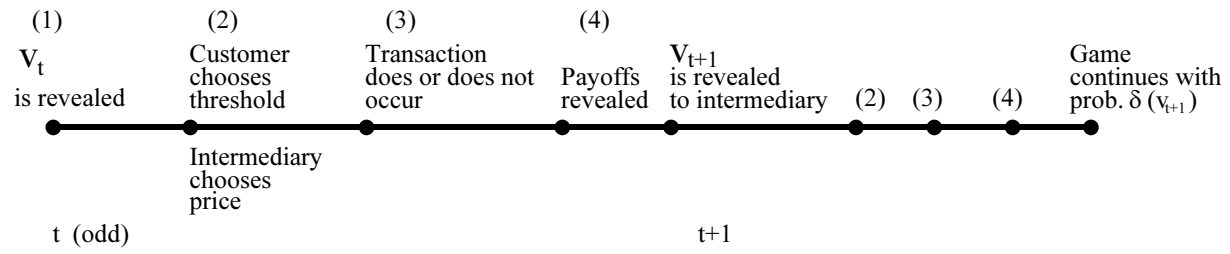


Figure 1.1: Single Stage

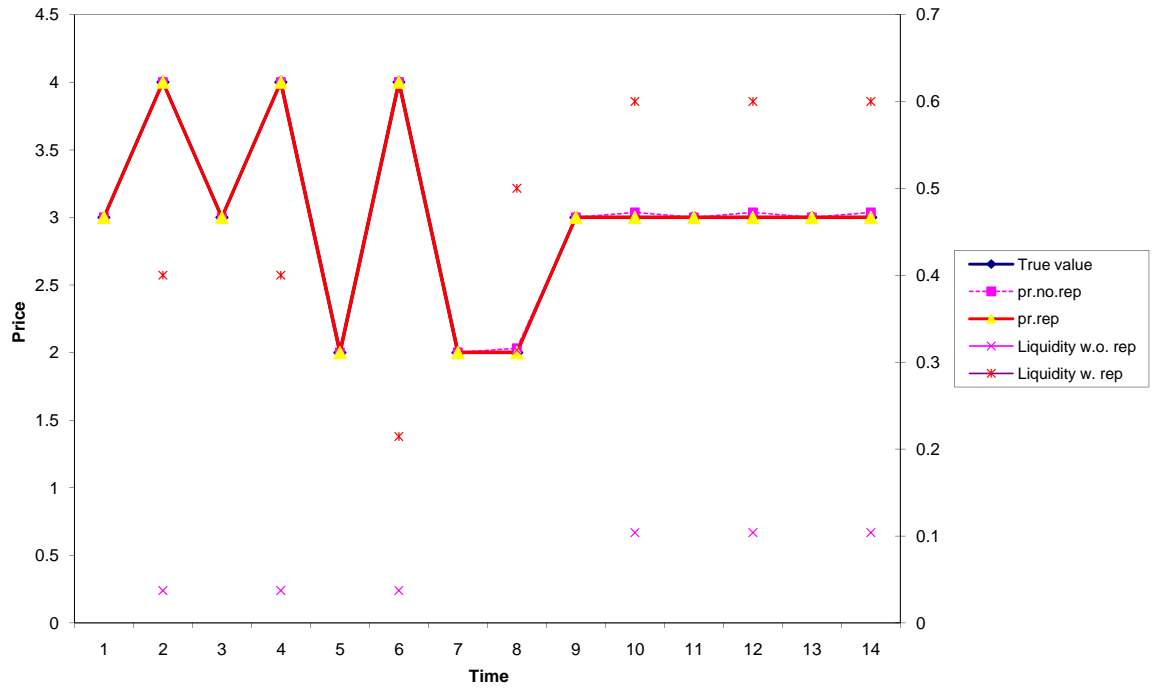


Figure 1.2: Liquidity Effect of Repetition

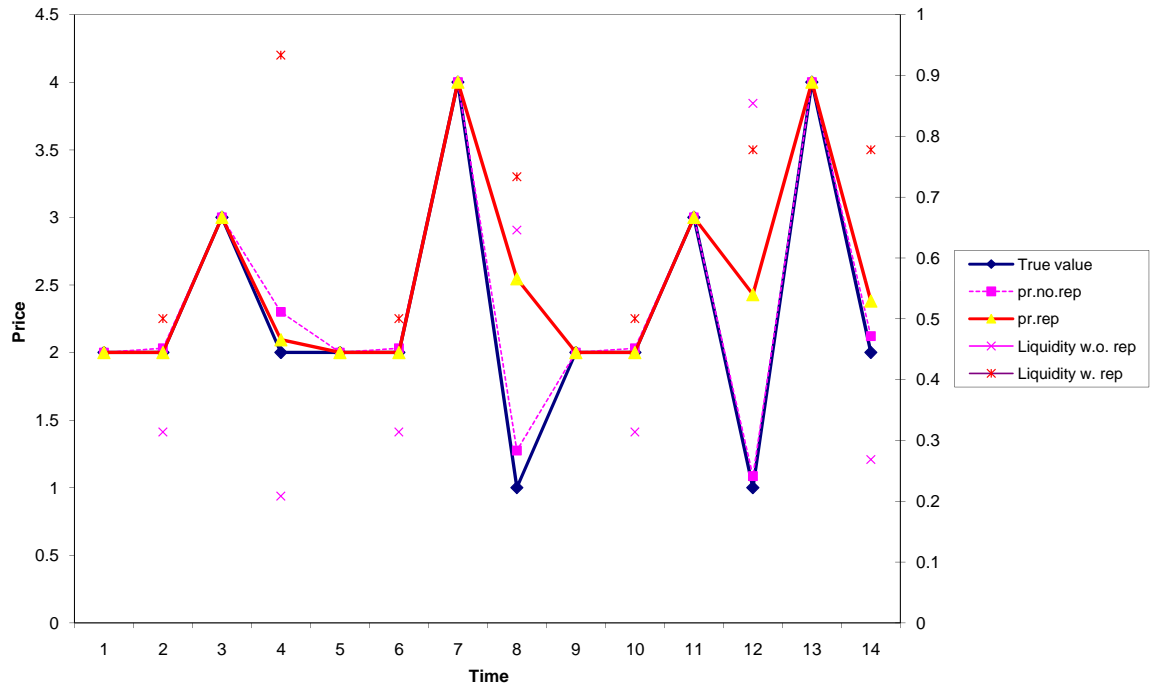


Figure 1.3: Example of Downward Price Stickiness

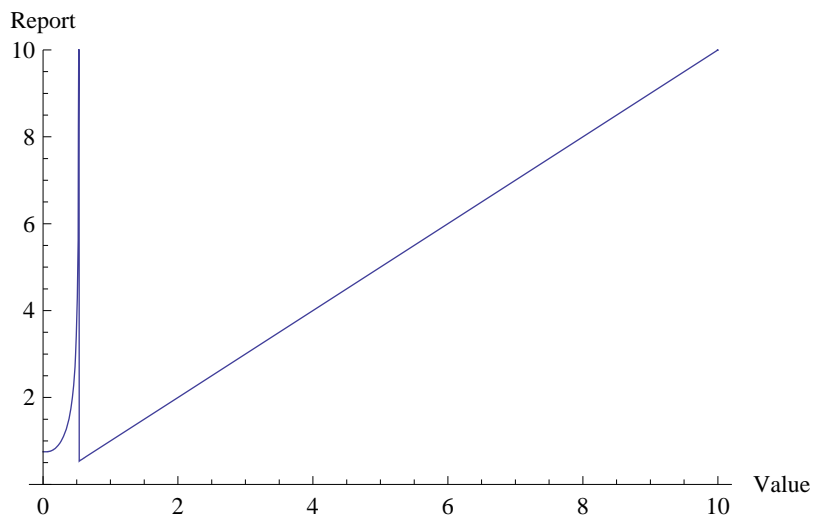


Figure 1.4: A Bubble

Chapter 2

Imperfect Monitoring and Fixed Spreads in the Market for IPOs

2.1 Introduction

In the U.S., over 95% of privately held firms of intermediate value pay a spread of exactly 7% to an underwriter when they decide to go public (Chen and Ritter [2000]¹). Very large firms pay lower spreads and very small firms higher. Since the total payment for investment banking services is determined by the spread and the proceeds of the offering, this rigidity is surprising. Costs to an underwriter almost certainly contain a fixed component. Either competition or efficient collusion between underwriters should thus lead to spreads that fall with the size of the offering.

A debate has developed in the corporate finance literature as to whether this extreme spread rigidity is evidence of collusive behavior by underwriters. Chen and Ritter [2000], who discovered the pervasiveness of the seven percent spread, argue informally that it is evidence of collusion, while Hansen [2001] argues that coordination on 7% naturally arises in an efficient contract. The question of whether underwriters collude on spread offers is important. Standard concerns about inefficient provision of goods and services in monopolistic or collusive industries apply, but, more importantly, if underwriters can collude to extract profits that would otherwise have accrued to entrepreneurs and venture capitalists, the incentives to engage in risky but positive expected value projects are diminished. If, on the other hand, underwriters are competitive in pricing underwriting services, then incentives for entrepreneurial activity will be appropriately aligned.

I present a formal model of the IPO process as an infinitely repeated game of imperfect information with public monitoring. In this context, I show that optimal collusion by underwriters will lead to spreads qualitatively similar to those observed, while competition or monopoly provision will imply that spreads depend on the size of the firm over the entire distribution of firm values.

I assume firms have two incentives to go public. Going public increases the present value of the firms expected future earnings. This “common value” element of the IPO process likely follows from improved access to capital markets.² Going public also provides private benefits (or implies private costs) to the managers of the firms, who are

¹See also Jay Ritter’s website for more current data.

²Chemmanur and Fulghieri [1999] treat this as the primary motivation for going public, and Hale and Santos [2006] present direct evidence that having held an IPO for *debt* reduces the interest rate charged on bank loans and private bond issues for a firm. Pagano et al. [1998] and Bharath and Dittmar [2007] also provide useful discussions of the motivations for a firm to go public from the “common value” perspective.

ultimately responsible for choosing whether or not to go public. This private element would follow primarily from preferences for personal liquidity,³ which will vary across the owners or managers of different firms.

In the model, underwriters collude to extract the most profits possible from issuing firms. Underwriters exploit their repeated interactions in the IPO market to maintain a pricing strategy that provides greater profits than one-shot competition. They cannot, however, collude on the spread schedule that maximizes total profits, even conditional on the information that they receive about the value of the firm and the preferences of the managers. Because underwriters receive different signals about the value of the firm, colluding on a fully efficient spread schedule becomes a game of imperfect monitoring and consequently bears costs associated with on-equilibrium-path punishments. I show that underwriters can, under certain circumstances, improve profits by using a partially rigid spread that calls for high spreads for small firms and a lower, uniform spread for intermediate and large sized firms. I also argue that the results on cyclical pricing in Rotemberg and Saloner [1986] explain why very large firms are charged lower spreads.

In addition to demonstrating that collusion is a likely explanation for spread rigidity, the model also suggests links between unobserved variables in the investment banking industry, such as the costs of providing underwriting services and the effect of IPOs on the value of firms, and the observed characteristics of the distribution of spreads. While estimating the parameters of the model is beyond the scope of the present work, such an exercise could prove useful in quantifying the costs of collusion and investigating the welfare effects of anti-collusive policies.

The idea that rigidity in pricing may make collusion easier has been understood informally in the industrial organization literature for some time. More recently, Athey et al. [2004] developed a model of collusion in Bertrand competition and show that under asymmetric and imperfect signals about firm costs, rigid pricing schemes may be optimal. The techniques applied rely crucially on the assumption that players receive private information about their own costs, a *private* value component of payoffs. Their approach does not readily generalize to cases where participants in the market are differentially informed but symmetric in all other respects. The model of Hanazono and Yang [2007] introduces private information over a common value aspect of payoffs, specifically the state of demand, and shows that in a simple two-state environment price rigidity can also arise under optimal symmetric collusion. While the tradeoffs

³See Ritter [2003].

between price rigidity and price flexibility in this model are similar to those that arise in the auction setting that I consider, the restriction to two states prevents their model from addressing the extent of rigidity, and in particular it cannot be used to investigate the partially rigid schedule observed in the cross-section of IPO spreads.

The paper proceeds as follows. I first present a model of the underwriting process. I then explore the likelihood that spreads will exhibit rigidity when underwriters behave competitively, as a competitive oligopoly, or as a monopoly. Under each of these assumptions, price rigidity will only be observed in very special cases that I argue are unlikely to characterize the underwriting industry.

I then show that partially rigid spreads similar to those observed in the data can arise when underwriters receive different but informative signals about firm value and play symmetric strategies in each period. Partially rigid spreads are *strictly* preferred to any fully rigid or fully flexible (i.e. everywhere downward sloping) spread and so the conclusion that partially rigid spreads can arise under collusion with imperfect information is robust to small perturbations of any element of the environment considered.

Next, I discuss an extension of the model to permit underpricing and argue that spread rigidity and underpricing will reinforce each other. When underwriters collude on a rigid spread to reduce the costs of monitoring collusive agreements they will be unable to extract all possible rents from those firms that choose to go public. By introducing an “offer price” stage *after* the firm has committed to one underwriter, I show that underpricing has a natural role as a means to capture more of these rents.

I also address the marked difference between IPO spreads, where rigidity is pervasive, and SEO spreads, where rigidity is nonexistent. In the SEO market the restriction to symmetric equilibria is shown to be inappropriate, which in turn permits efficient, and therefore non-rigid, spread pricing.

The paper concludes with a brief discussion of international differences in the market for underwriting services and the implications for regulatory policy that follow from the analysis here.

2.2 Model

This section presents a model of IPO underwriting. The model admits heterogeneity in firm value and manager preferences for going public. Underwriters receive (potentially noisy and private) signals about these quantities before bidding for the right to take

a firm public. The model therefore allows for the possibility of a relationship between firm value and IPO spreads and can thus provide insight into the source of spread rigidity over a wide range of firm values.

Firms have two incentives to go public. First, there is a “common value” element of going public; the net present value of a firm increases when it goes public. Second, managers of firms have idiosyncratic private incentives to take their firms public (or to keep their firms private). A detailed examination of the source of the “common value” or “private value” benefits of taking a firm public is beyond the scope of this paper. The common value benefits most likely relate to improved access to capital markets, while private benefits can be thought of as representing a tradeoff between preferences for control and preferences for personal liquidity.

Institutional constraints prevent a firm from going public without the assistance of one of the underwriters. In taking a firm public, an underwriter incurs a fixed cost, which implies that returns to scale are increasing. Certain costs associated with underwriting, particularly those involving accounting tasks, might increase with the value of the firm going public. Many costs, however, will have a fixed component; for example, the costs of certain legal and regulatory tasks would not vary much with firm value. Other costs, such as those associated with advertising the issue, might actually fall in firm value as larger firms would tend to be more well known and therefore require less effort to elicit interest during the book-building process. Post-issue stabilization, a potentially significant cost to underwriters, also has an ambiguous relationship to firm value since the expected magnitude of such activity most likely depends both on the size of the issue and on the riskiness of the price from the perspective of the underwriter (Prabhala and Puri [1998]), which may be decreasing in the value of the firm.⁴

Underwriters compete for a sequence of opportunities to take firms public by making simultaneous spread offers after they receive their signal about the firms, where the spread determines the proportion of the common value of the firm that the underwriter receives as fees for facilitating the IPO; the IPO market is thus modeled as a repeated procurement auction with security bids (see DeMarzo et al. [2005] for a general treatment of auctions with security bids). I assume that the offer price of the issue is exactly the true value; that is, book building perfectly reveals the value of the firm and underwriters have no incentive to underprice offerings. Section 2.5 addresses

⁴The direction of the change in the costs of stabilization is an empirical question that, to my knowledge, has not been addressed. Aggarwal [2000] provides a method for determining the extent of after-market activities by underwriters, and could likely be extended to address this question.

the role of underpricing in the model considered here. The repeated setting captures the presence of the established investment banks which dominate the market for IPO services. The entire structure of the auction is exogenous. This abstracts away from considerations about optimal security design for IPOs, which has been explored, for example, in Chakraborty et al. [2008]. Optimal design in the presence of collusion is left for future research.

Following the bidding for each firm, underwriters learn the true common value of the firm and the winning bid is made public. This additional information, which becomes common knowledge among firms, captures a crucial element of the IPO process; before going public, information about a firm is disaggregate and held privately, whereas after the IPO there is a perfectly observable summary statistic aggregating such information, the trading price. This public signal allows the repeated auction to be modeled as a game of imperfect public monitoring. By confining attention to strategies conditioned on current private information and the history of the public signal, the repeated interaction can be recast in a recursive structure and analyzed using standard techniques developed in Abreu et al. [1990].

Formally, there is a sequence of short-lived firms and two long lived investment banks who discount the future at rate δ . Each firm has type $\{x_t, \varepsilon_t\}$, where x_t is value and ε_t summarizes the idiosyncratic preferences of the owners and managers of the firm. Each underwriter, indexed by i , observes a potentially two dimensional signal $\{\xi_t^i, \eta_t^i\}$, where ξ_t^i is informative about x_t and η_t^i is informative about ε_t . Underwriters simultaneously submit spread offers $\alpha_t^i \in [0, \infty]$.⁵⁶ The firm then chooses between the two underwriters, or decides to remain private.

The payoffs in the stage game can then be summarized as follows:

- If the firm goes public using underwriter i :

firm: $(1 - \alpha_t^i)\beta x_t + \varepsilon_t$

⁵For expositional clarity, throughout the paper I treat the action space as a continuum. Because I will use the bang-bang result from Abreu et al. [1986] and Abreu et al. [1990], formally we must consider arbitrarily fine approximations to the continuous stage game as bang-bang has not, to my knowledge, been shown to apply to games with continuous action spaces. All of the arguments showing that the results for the continuous case also apply to the limit of an appropriate discretization of the game are collected in appendix 2.10.1.

⁶It is obviously unlikely to observe a bid greater than 1 as that would require payments in excess of the total value of the firm. Since some of the benefits of going public are not directly related to the increase in the value of the firm, however, it is possible that managers of very small firms would pay more than the entire value of the firm in order to realize the private benefits of going public. For simplicity, I do not rule out even such obviously pathological examples.

underwriter i : $\alpha_t^i \beta x_t - \kappa$

underwriter j : 0

- If the firm does not go public:

firm: x_t

underwriter i : 0

underwriter j : 0

Here, β summarizes the common value benefits of going public. Since the IPO provides a proportional increase in value, going public is more valuable to large firms from an absolute perspective.

2.2.1 Public History and Equilibrium

After each stage of the game, the true value x_t and a public signal $a_t = \{\min_{i \in \{1,2\}} \alpha_t^i\}$ are revealed publicly. For simplicity, I assume that the lowest spread offer is always reported publicly, regardless of whether the firm decides to go public. See Figure 2.1 for a summary of the timing of the stage game.

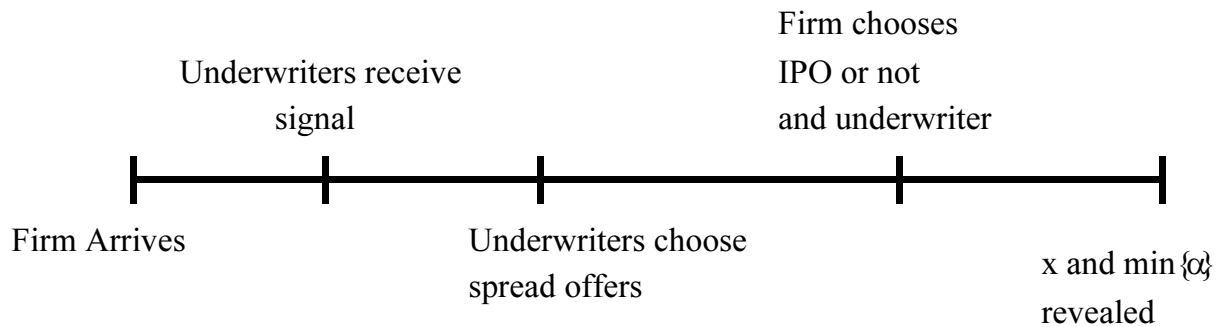


Figure 2.1: Timing of the Game

I will analyze the game by restricting attention to “almost pure strategy symmetric perfect public equilibria” where this set of strategies is defined as follows:

Definition 1. *A perfect public equilibrium is a profile of public strategies that, for any public history, specifies a Nash equilibrium for the repeated game starting at that history. A perfect public equilibrium is symmetric if all players use the same strategy profile following every history.*

The restriction to symmetric equilibria is not innocuous and provides significant restrictions on the strategies available to the underwriters. Fully efficient collusion in Bertrand competition (Athey and Bagwell [2001] and Kandori and Matsushima [1998]) and in auctions (Aoyagi [2003] and Aoyagi [2007]) generally relies on exploiting asymmetric bid rotation schemes and communication. The restriction to strongly symmetric strategies renders the techniques exploited in these papers out of bounds in my application. In fact, permitting even simple bid rotation schemes with pre-play communication will immediately permit underwriters to collude efficiently in a pure common values setting, which will imply spreads that do depend on firm value. I interpret the restriction to strongly symmetric equilibria as a restriction to collusive strategies that are not easily detectable by regulators. Since firms are effectively anonymous to underwriters, asymmetric collusion would require bid rotation. Anti-trust regulators are experienced in identifying bidding rings of this type, and maintaining such coordination in a more complex setting than described here would likely require either an extremely sophisticated initial agreement or regular cartel discussions to resolve questions and assure continued adherence. Such discussions are the primary means of identifying and prosecuting collusion, and thus the restriction to strongly symmetric equilibria makes sense within the legal framework faced by underwriters. The difference between symmetric and asymmetric collusion also provides a useful way to understand the differences between spreads charged on IPOs and spreads charged on SEOs, which do not cluster to any meaningful extent. This comparison is explored in section 2.6.

Definition 2. *An almost pure strategy symmetric perfect public equilibrium is a symmetric perfect public equilibrium in which in each period underwriters choose pure actions or choose mixed actions that are consistent with an equilibrium of the one-shot version of the game.*

That is, I consider only strategies that generally call for both long-lived players to play the same pure strategy action profile following a given public history, with the action chosen by any underwriter being measurable with respect to the public history and the current signal received by that underwriter. Only when long-lived players are

playing as if the game were not repeated can mixed strategies be included.⁷ And, the continuation strategy following each history must be a Nash equilibrium of the repeated game. To avoid cumbersome repetition of this phrase, I will simply refer to the perfect public equilibria of the game with the understanding that I am referring to this particular class of perfect public equilibria.

Note that this equilibrium concept implies that all short-lived players must play their static best response to the long-lived players' equilibrium actions. For a full treatment of the concept of symmetric perfect public equilibrium see Mailath and Samuelson [2006], and for related applications to price rigidity see Hanazono and Yang [2007] and Athey et al. [2004]. Finally, I introduce a public correlation device to simplify optimal punishments. This assumption is without loss of substantial generality as such devices are readily available, particularly in financial markets.

2.3 Spreads under Competition and Perfect Monitoring

The primary goal of this study is to show that the concentration of spreads at exactly one number and the behavior of spreads for the largest and the smallest firms are best explained by a combination of collusion and imperfect information among underwriters. To reach this conclusion, it is necessary to consider the likelihood that spreads would have the documented characteristics in the absence of collusive arrangements of the type considered. The model presented here will not generate rigid spreads under either perfect competition or oligopolistic competition, except in certain pathological cases. As long as signals about firm value are informative, the bid that leads to zero profit in expectation and the bid that maximizes profits in the one-shot auction will both depend on the signal. Since for a given spread a more valuable firm will generate a higher payoff to the underwriter, and since returns to scale are increasing, larger firms must receive lower spread offers, on average, than smaller firms. This will be true locally at every point, so contrary to the claims in Hansen [2001] and Torstila [2003] a competitive underwriting industry will not use a contract with a partially rigid spread.

⁷Permitting mixed strategies when long-lived players are simply playing stage-game equilibria permits a larger class of punishment strategies. Optimal collusion will generally take the form of pure strategies, while stage game play may involve mixed strategies, particularly if the distribution of signals is not atomless. These mixed strategies are admitted to avoid artificially and unnecessarily reducing the set of punishments available to long-lived players.

On the other hand, if underwriters can effectively act as a monopoly, signals about value will be informative about the reservation spread of the firm as long as manager preference is not identically zero. Since signals are informative about firm value and firm value is informative about the distribution of the reservation spread, underwriters will charge slightly different spreads in response to slightly different signals. If manager preference were identically zero for all firms and the common value benefit of going public took exactly the proportional form described, then monopoly spreads would be rigid, as in Chen [2001]. Any slight perturbation of either of these assumptions would undo rigidity, and such assumptions cannot account for the increase in spreads charged on the smallest issues.

The impossibility of observing spreads that are rigid even over a small region is not sensitive to the parametric assumptions on costs and benefits assumed in the main text of the paper. Appendix 2.10.2 presents a more general model and argues that almost all cost and benefit functions will necessarily imply spreads that are not rigid over any interval. Here, however, I present a parametric example that develops the intuition for why spreads will not be rigid under competition, oligopolistic competition, or collusion with symmetric information.

In the absence of potentially imperfect monitoring (that is, when both underwriters receive the same signal before each auction), the costs and benefit functions above generate a pattern of spreads that bears some resemblance to the observed spread distribution but does not imply rigid spreads. Figures 2.2 and 2.3 show examples of this, first in the case where underwriters both learn the true value of the firm and the true value of the preferences of the manager and then in the case where they only learn the value of the firm. Figure 2.4 shows the spreads that would arise under perfect information and competitive pricing. Details of the straightforward derivation of these spread schedules can be found in appendix 2.10.6 and 2.10.8.⁸

⁸There is no closed form expression for the spread function for unobservable ε ; in the appendix, I solve for the closed form expression when ε is unobserved but distributed uniformly rather than normally.

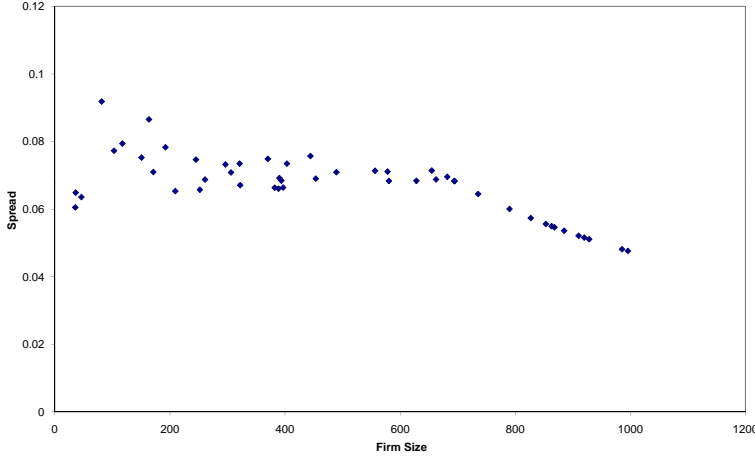


Figure 2.2: Perfect Information Spreads with Collusion: $x \sim U[0, 1000]$, $\varepsilon \sim N(0, 1)$, $\beta = 1.075$, and $\kappa = 1$

In both of these cases, the fact that the signals are identical makes collusion easy since any deviation from the pure strategy spread schedule prescribed in a collusive equilibrium is detectable and can therefore be deterred without costs; the perfect public equilibrium then coincides with collusion at the monopoly spread schedule. Under such a collusive outcome, information about firm value and manager preference will prove useful for choosing the spread offer. When both value and preferences are observed, underwriters will push firms to their participation constraints; that is, firms will be exactly indifferent between going public at the collusive spread offer and remaining private. This participation constraint will depend both on the preferences of the manager and the offer price. When preferences are not observed, signals about value effectively become signals about the elasticity of demand for holding an IPO.

Other things equal, a firm with a manager with strong preferences for going public will be willing to pay a higher spread and will consequently be charged the higher spread. The value of the firm will determine how sensitive the spread offer is to individual preferences. More valuable firms will in effect care more about the spread relative to individual preferences. Spreads above the “common value” benefit of the

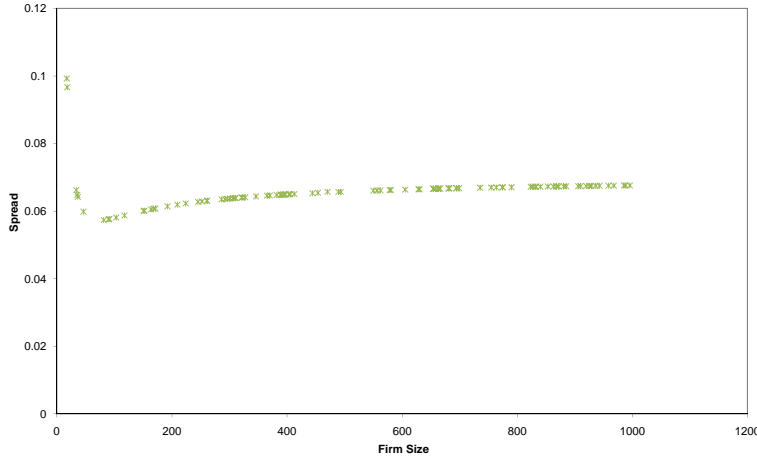


Figure 2.3: Collusive Spreads with Unobservable ε

IPO are costly in the sense that the post-IPO firm is less valuable than the pre-IPO firm since the payments to underwriters exceed the common value gains. For small firms, the individual manager benefits of going public can justify the costs, but for very valuable firms, even ones controlled by a manager with a strong preference for taking his firm public, the loss associated with holding an IPO at a very high spread will overwhelm the idiosyncratic private benefits. Symmetrically, firms with managers who prefer to keep their firms private will require spreads below the common value component of the benefit of going public. Small firms may demand very low spreads or, in extreme cases, refuse to go public at any non-negative spread. But, since underwriters must pay a fixed cost to take a firm public, they will refuse to offer a spread low enough to induce a small firm with low preference for going public to hold an IPO. The distribution of spreads for small firms is thus truncated from below, and consequently the average spread charged to small firms will be high relative to the average spread charged to larger firms.

Additionally, in the case of observable manager preference, underwriters are assumed to be patient but not arbitrarily patient. Since the value of going public increases with the value of the firm while the optimal spread remains approximately the

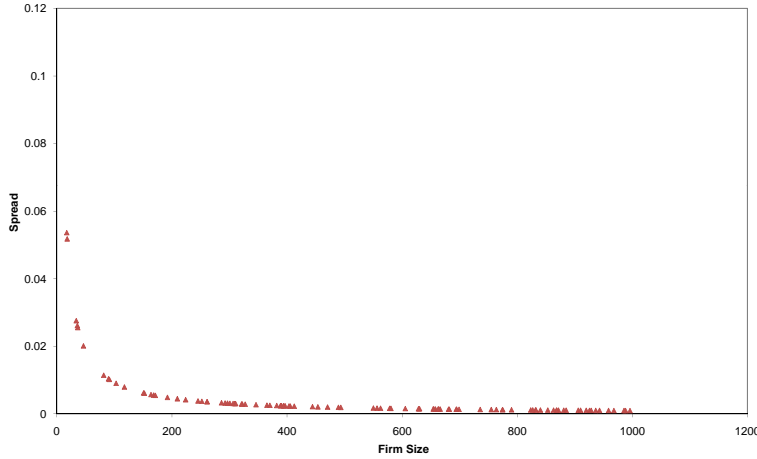


Figure 2.4: Competitive Spreads with Perfect Information

same, larger firms will, on average, prove more profitable for underwriters. Thus, to sustain an equilibrium, spreads must decrease for such firms so that a deviation will not be too profitable relative to expected future profits from maintaining collusion. This is, of course, an application of the result in Rotemberg and Saloner [1986].

Note that the spreads charged when underwriters collude and only value is observable is not necessarily monotone decreasing. For very small firms, the underwriters will charge a high spread to capture rents from those firms with strong preferences for going public. In intermediate ranges, charging relatively low spreads may be valuable because it can induce firms with idiosyncratic preferences for staying private to agree to go public. For the very largest firms spreads will be higher again since the relative importance of idiosyncratic preferences decreases. The participation constraints of very large firms will cluster very tightly around one particular spread, although the spread function will never be exactly flat. The fact that optimal collusive spreads asymptote to one particular spread strictly greater than zero will play an important role in explaining why spreads can be rigid over such a wide range of firm values while still responding to value for small and large firms.

2.4 Imperfect Monitoring

We have now seen that when underwriters receive identical signals about firm value and manager preference they optimally collude on spread offers that call on average for high spreads for small firms. Spreads for moderate sized firms decline sharply from these high levels, and the spread offers become more “flat” in the sense that firms of intermediate size are charged similar, but not identical, spreads.

This section shows that when signals are informative but asymmetric such that underwriters have private information about the characteristics of the firm optimal collusion may call for a partially rigid spread. For simplicity, we focus on a model where underwriters receive conditionally independent signals about the value of the firm and no signal about the preferences of the manager, but all conclusions will be robust to admitting minimally informative private signals about manager preference.

Colluding on a rigid or partially rigid spread implies that underwriters ignore information that would be useful for setting a spread that extracts the most possible surplus from issuing firms. The benefit of ignoring such information comes from the fact that deviations from a rigid spread schedule can be perfectly detected and prevented with punishments that do not occur along the equilibrium path. A spread schedule that uses all available information about firms would require a different spread offer for each signal.⁹ Since underwriters cannot observe the signal of the other underwriter, deviations from the prescribed spread schedule cannot be directly observed. The punishments necessary to enforce such a spread schedule are then triggered by *apparent* deviations, and as such will occur along the equilibrium path. This inefficiency will in certain cases prove more severe than that associated with ignoring private information.

Finally, in the non-identical-signals setting the model still can predict that spreads will be high for small firms even when rigidity is better than fully separating spreads. Such a partially rigid spread function can arise without requiring on-path punishments. An underwriter with a very low signal about firm value will not necessarily have an incentive to imitate a higher signal (and thus increase his chances of winning the IPO). If such an action requires him to bid far below the first-best spread offer, the efficiency loss may exceed the gain to himself of capturing more of the market. Thus, a spread schedule that calls for most firms to be charged the same spread but for very small firms to be charge a higher spread can provide greater profits than either a fully rigid

⁹In certain cases not considered in this section, the spread schedule may be non-monotonic and thus a measure zero set of signals could call for the same spread offer.

or fully separating spread function, all without requiring on-path punishments.

For simplicity, I consider a particular class of problems, summarized in the following assumptions:

Assumption 1.

1. $x \sim U[0, \bar{x}]$
2. $\xi_i = \begin{cases} x & \text{with probability } p \\ U[0, \bar{x}] & \text{with probability } 1 - p \end{cases}$
3. $\varepsilon \sim \exp(\lambda)$
4. $\delta \rightarrow 1$

It follows immediately that $\xi_i \sim U[0, \bar{x}]$ and that

$$x|\xi_i \sim \begin{cases} \xi_i & \text{w.p. } p \\ U[0, \bar{x}] & \text{w.p. } 1 - p \end{cases}$$

Assumption 1.3 indicates that we are in a special circumstance in which the manager-specific value of an IPO is always positive. This assumption is made only for analytical tractability, as it helps guarantee that the optimal spread schedule will be weakly decreasing everywhere.

I now solve for the optimal rigid spread, the optimal two-step self-enforcing spread, and an approximately optimal fully separating spread. Calculating these spreads makes it possible to compare the value of each type of equilibrium and to conclude that, in certain cases, the two-step spread schedule can provide profits higher than any fully rigid spread or any fully separating spread. From this finding, it is possible to demonstrate that the optimal collusive spread schedule when a two-step self-enforcing spread is preferred to either a rigid spread or a fully separating spread will exhibit characteristics similar to those observed in data. Spreads for intermediate and large sized firms will almost always be the same, while small firms will be charged spreads that will decrease on average as the firm gets larger (but remains “small”), but firms of the exact same value will be charged different spreads with positive probability. These characteristics are exactly those documented in Chen and Ritter [2000]. Finally, we demonstrate that if underwriters are less than completely patient and the upper bound on firm value is high enough, underwriters must demand lower spreads following

the very highest realizations of their signals about firm value. These elements of the optimal spread function combine to demonstrate that the “seven percent solution” almost certainly results from attempts by underwriters to collude to extract as much surplus as possible from issuing firms.

2.4.1 Optimal Rigid Spread

The optimal rigid spread is the solution to a relatively straightforward univariate maximization problem. The only complications are that certain parameter values imply that spreads should be set either so low as to guarantee that all firms hold an IPO or so high as to guarantee that no firms hold an IPO. Specifically, when both the costs of holding the IPO and the mean of the idiosyncratic manager preferences are sufficiently low, the underwriters will set a rigid spread of $\alpha^r = 1 - \frac{1}{\beta}$, which guarantees that even the firm with $\varepsilon = 0$ holds an IPO. When costs are sufficiently high and the mean of the manager preference is sufficiently low, no spread leads to positive profit. We will ignore such cases by imposing a technical restriction on the relationship between the costs of holding an IPO, κ , and the mean of the distribution of manager preference, λ :

¹⁰

Assumption 2.

$$\kappa\lambda < 1$$

The following proposition shows how to derive the optimal rigid spread. The derived expression is an immediate consequence of the necessary condition for an interior optimum for profit maximization and the proof is thus omitted.¹¹

Proposition 3. *The optimal rigid spread is given by $\alpha^{r*} = 1 - \frac{1}{\beta} + \gamma^*$, where*

$$\gamma^* = \begin{cases} \gamma' & \text{if } \gamma' > 0 \\ 0 & \text{otherwise} \end{cases},$$

¹⁰Note that this condition is not tight, in the sense that there are problems with $\kappa\lambda > 1$ where there is an interior optimum for α ; we ignore these cases for simplicity. In general, characterizing the set of parameter values where conditions hold is tedious and uninteresting; further details on these sets are available from the author.

¹¹The objective function in this problem fails to be quasiconcave for certain parameter values, but it is possible to show that there are no more than two local maxima and that identifying the global maximum is not difficult. Details are omitted to save space.

and γ' satisfies

$$\gamma' = \frac{1}{\beta\lambda\bar{x}} \log \left[1 + B + (\beta - 1 + \beta\gamma') \frac{B^2}{A} \right]$$

with

I. $A = 2(1 - \beta) - \gamma'\beta(1 - \kappa\lambda)$

II. $B = \gamma'\beta\lambda\bar{x}$.

Charging the optimal rigid spread would imply that firms are ignoring all information contained in their signals. This eliminates the difficulties associated with monitoring deviations from equilibrium since any deviation can be identified perfectly. It is, however, possible to increase the profits accruing to the underwriters without requiring on-path punishments. The following subsection demonstrates the procedure for finding just such an equilibrium.

2.4.2 Partially Rigid Spread

The benefits of a fully rigid spread come from the elimination of the need for punishments that occur along the equilibrium path. The costs are that spreads cannot be chosen optimally to extract as much surplus as possible from the firms given the information available to the underwriters. However, it is possible to choose a spread function that both uses information contained in the signals received by the intermediaries and does not require punishments along the equilibrium path. This section presents the form of such a “self-enforcing” spread function, proves the existence of such spread functions for the case where an optimal rigid spread exists, and describes the procedure for finding the optimal spread function within a restricted class of such self-enforcing equilibria. This two-step spread function will call for higher spreads for the smallest firms and a single, fixed spread for all other firms.

Definition 3. *A self-enforcing spread function is a function $\alpha^{se} : [0, \bar{x}] \rightarrow \mathbb{R}^+$ such that if players 1 and 2 play $\alpha(\xi_k), k \in \{1, 2\}$ in the pricing stage following any history where no deviation has been detected with probability one (that is, in period t , $a_s \Rightarrow P(\exists \xi \in [0, \bar{x}] \text{ s.t. } \alpha_s = \alpha^{se}(\xi)) > 0$ for all $s < t$) and play the stage-game equilibrium otherwise, then no arbitrarily patient underwriter has an incentive to deviate.*

Such spread functions must take on a very specific form over the range of signals for which an underwriter would anticipate positive profits. To slightly simplify the discussion, we restrict attention to the class of spread functions that are weakly decreasing in signals about firm value:

Proposition 4. *Let $X^p(\alpha)$ be the set of signals for which an underwriter expects positive profits in the stage game under spread function α . Then, if α is part of a decreasing self-enforcing equilibrium, α restricted to $X^p(\alpha)$ is a step function.*

Proof. Assume the contrary. This implies that, in some region where expected profits are positive, the spread function is continuous but not flat. Then, there is some signal ξ' such that an arbitrarily small deviation from $\alpha(\xi')$ is not detectable with probability one. But, such an arbitrarily small deviation will increase expected profits by $\frac{p^2}{2}(\alpha(\xi')\beta\xi - \kappa)e^{-\lambda x(\alpha\beta + 1 - \beta)}$ since the implied change in α is infinitesimal and the deviator now captures the entire market when both agents receive the same, correct signal. Thus, a profitable deviation exists for underwriter type ξ' . \square

Given this result, I can restrict attention to decreasing step functions without loss of substantial generality. Such step functions effectively generate partially separating equilibria. That is, all underwriter types effectively pool with those other types within their step. Such a self-enforcing step function is guaranteed to exist under general conditions:

Proposition 5. *For any configuration of parameter values for which there is a rigid spread that implies positive profits, there is some self-enforcing two-step spread.*

See the appendix for a proof.

Calculating an optimal two-step self-enforcing spread is a relatively simple multivariate optimization subject to a single incentive compatibility constraint since indifference at the threshold implies strict preference away from the threshold, a form of a single crossing property. The optimization problem can be written as follows:

Let

$$\begin{aligned}\pi(\alpha, x) &= \alpha\beta x - \kappa \\ h(\alpha, x) &= e^{-\lambda x(\alpha\beta + 1 - \beta)}.\end{aligned}$$

$$\begin{aligned}
& \max_{\alpha^h, \alpha^l, x^*} \int_0^{\bar{x}} \left(\mathbf{1}_{x < x^*} \left(p^2 \left(1 - \frac{x^*}{\bar{x}} \right)^2 + \frac{2px^* \left(1 - \frac{x^*}{\bar{x}} \right)}{\bar{x}} \right. \right. \\
& \left. \left. + \mathbf{1}_{x > x^*} (1-p)^2 \left(\frac{x^*}{\bar{x}} \right)^2 \right) \pi(\alpha^h, x) h(\alpha^h, x) \right. \\
& \left. + \left(\mathbf{1}_{x < x^*} \left(1 - \left(1 - \frac{x^*}{\bar{x}} \right) \left(p^2 \left(1 - \frac{x^*}{\bar{x}} \right) + \frac{2px^*}{\bar{x}} \right) - \frac{x^*}{\bar{x}} \right) \right. \right. \\
& \left. \left. + \mathbf{1}_{x > x^*} \left(1 - (1-p)^2 \left(\frac{x^*}{\bar{x}} \right)^2 \right) \right) \pi(\alpha^l, x) h(\alpha^l, x) \right) \frac{1}{\bar{x}} dx
\end{aligned}$$

subject to

$$\begin{aligned}
& \frac{1}{2} \left(\left(p^2 + \frac{x^* p (1-p)}{\bar{x}} \right) \pi(\alpha^h, x^*) h(\alpha^h, x^*) \right) \\
& + p(1-p) \int_0^{x^*} \pi(\alpha^h, x) h(\alpha^h, x) \frac{1}{\bar{x}} dx + (1-p)^2 \frac{x^*}{\bar{x}} \int_0^{\bar{x}} \pi(\alpha^h, x) h(\alpha^h, x) \frac{1}{\bar{x}} dx \\
\geq & \frac{1}{2} \left(p^2 \left(1 - \frac{x^*}{\bar{x}} \right) + p \left(1 + \frac{x^*}{\bar{x}} \right) \right) \pi(\alpha^l, x^*) h(\alpha^l, x^*) \\
& + p(1-p) \left(\int_0^{x^*} \pi(\alpha^l, x) h(\alpha^l, x) \frac{1}{\bar{x}} dx + \frac{1}{2} \int_{x^*}^{\bar{x}} \pi(\alpha^l, x) h(\alpha^l, x) \frac{1}{\bar{x}} dx \right) \\
& + \frac{1}{2} (1-p)^2 \left(1 + \frac{x^*}{\bar{x}} \right) \int_0^{\bar{x}} \pi(\alpha^l, x) h(\alpha^l, x) \frac{1}{\bar{x}} dx.
\end{aligned}$$

The existence of such a two-step self-enforcing spread arises because the optimal spread rises sharply as firms become small. A two-step spread with a relatively high threshold could be more profitable as it would allow more efficient pricing of the most valuable IPOs, but it would be impossible to enforce such a spread since the temptation of the firms near the threshold to bid as if they had a higher signal would be too strong. Self-enforcement requires that the step be large and that the higher spread offer be relatively close to the optimal spread offer for the threshold firm. In this case, when deviating the benefits of capturing more of the market are offset by the costs of charging a less efficient spread. The costs to the underwriter of holding an IPO play an important though not absolutely essential role in the nature of the self-enforcing spread. When underwriting costs (i.e. κ) are relatively high the costs of misreporting one's signal following a low signal can be substantial as the market share increase comes

largely from winning IPOs that are unprofitable at the lower spread. With very small or even zero underwriting costs, however, it is still possible to construct two-step self-enforcing spreads; these will generally call for very high spreads for the very smallest firms. In this case the self-enforcement is driven exclusively by the fact that signals about firm value are informative about the expected reservation price of the firm; very small firms should be charged very high spreads since there is a small chance that the firm has a high idiosyncratic preference for going public and the firm's small size effectively magnifies this preference. Introducing underwriting costs implies two-step spreads much closer to what is observed in data.

A two-step self-enforcing spread will not, in general, prove optimal, but we will not seek to precisely characterize the optimal self-enforcing spread. Instead, it is sufficient to show that the two-step spread dominates the optimal fully flexible spread in order to demonstrate that the optimal spread function does not respond to firm size everywhere. Since relying on the self-enforcing characteristics of step functions is the only way to decrease the costs associated with monitoring for a fully separating spread and since such step functions must have thresholds calling for increased spreads for the smallest firms, we can conclude that if the optimal self-enforcing two step spread provides more profit than the optimal fully separating spread then the optimal spread function must call for rigid spreads over a large region of intermediate valued firms and higher spreads for smaller firms. These smaller firms may face a more complicated spread schedule than a simple two-step function, but we will show that when a two-step spread is preferred to either a rigid spread or a fully flexible spread the optimal spread schedule will exhibit both rigidity over a large region and behavior generally consistent with observed data for small firms.

2.4.3 Flexible Spreads

We have so far confined attention to spread functions that do not require punishments along the equilibrium path. These are exactly the partially rigid spreads that appear to match the empirical data. We now consider instead those spread functions that serve to fully separate types and thus exploit the information about firm value contained in the signals to the underwriters. For the class of problems under consideration, this reduces to searching for the optimal strictly decreasing spread function.

In this section, we discuss a method for deriving an upper bound on the equilibrium payoffs of a strictly decreasing spread function. We follow a simplified version of

the dynamic programming approach developed in Abreu et al. [1990]. For details on implementing the full procedure, see Mailath and Samuelson [2006] and Judd et al. [2003]. We assume throughout that underwriters have access to a public randomization device.

The problem of finding the optimal fully separating symmetric perfect public equilibrium reduces to finding the everywhere downward sloping spread function that satisfies all incentive compatibility constraints, conditional on promised continuation values falling in the interval $[\underline{v}, v^*]$, where \underline{v} is the payoff associated with the worst symmetric perfect public equilibrium and v^* is the payoff associated with the best such equilibrium. Incentive constraints take the basic form of requiring that, for every possible misreport, the increased likelihood of reverting to the “punishment” continuation value following any deviation from the proposed spread schedule is sufficiently high to overcome any short-term improvement associated with capturing more market share. Continuation values must, in turn, be drawn from the set of feasible equilibrium payoffs, $[\underline{v}, v^*]$. Furthermore, we know from the bang-bang result of Abreu et al. [1986] and Abreu et al. [1990] that any equilibrium payoff can be achieved with strategies that call for continuation values drawn only from the extreme points of the set of equilibrium values.¹² In the symmetric case considered here, this reduces the set of strategies under consideration to those that call for play of the “optimal” spread schedule until punishment is triggered, followed by perpetual play of the stage game equilibrium following transition to the “punishment” phase.¹³ Note that the optimal spread schedule will *not* be the spread schedule that a monopolist observing the first order statistic of the signals would offer. The need to satisfy incentive compatibility constraints will introduce a tradeoff between choosing an efficient spread schedule and choosing a spread schedule that does not tempt underwriters to deviate. This will, in general, lead to a spread function that is distorted downward for larger values.

The problem then becomes to find the optimal spread function and reversion probabilities to maximize the value of the game. Reversion probabilities, of course, cannot

¹²Readers concerned about the application of the bang-bang theorem to a game with a continuous action space may consult appendix 2.10.1 for an explanation of the applicability of the results to games with a finite but arbitrarily fine action space.

¹³It is straightforward to show that the worst sustainable symmetric perfect public equilibrium payoff is the worst payoff associated with repeated play of a stage-game symmetric Nash equilibrium for arbitrarily patient underwriters. When underwriters are not arbitrarily patient, it may be possible to sustain a lower value as a SPPE payoff for impatient underwriters. However, this value will clearly be above 0 (a lower bound on the set of individually rational payoffs), and abusing notation slightly we will continue to refer to this worst value as the stage-game payoff, Π^{SG} .

depend directly on whether an underwriter deviated from his assigned bid. Reversion probabilities can only depend on the public history, specifically the history of the true value of the firms and the lowest spread offers. The bang-bang restriction, in turn, allows us to consider only those strategies where reversion probabilities depend only the most recent realization of the public signal. The value of the game, then, will depend on the probability of reversion given that everyone adheres to the equilibrium strategy. The dynamic programming approach to solving for this value consists of finding a superset of the set of payoffs sustainable in a perfect public equilibrium, maximizing the value of the game over all downward sloping spread schedules, reversion probabilities and continuation values that are, taken together, incentive compatible for every signal and every possible deviation and where the continuation values fall in the posited superset. The maximum value that can be achieved by solving this problem is also an upper bound on the set of sustainable perfect public equilibrium payoffs. The same procedure with minimization replacing maximization gives a new lower bound. This procedure can be repeated indefinitely to obtain smaller and smaller supersets of the equilibrium value set, and will iterate to convergence in sufficiently regular programs. We will be concerned only with calculating a sufficiently restrictive upper bound and will therefore rely only on the monotonicity of this operator.

To summarize the approach used, note first that it is straightforward to find a superset of the set of symmetric perfect public equilibria. Underwriters would certainly be able to guarantee themselves at least $-\varepsilon$, where ε can be made arbitrarily small, by demanding an extremely high “proportion” of the firm in exchange for underwriting services. If underwriters are allowed to bid $\alpha = \infty$, it is clear that no perfect public equilibrium can call for payoffs below 0. On the other hand, we know that it is impossible for underwriters to do better in equilibrium than would a monopolist who observes the first-order statistic from the two signals about firm size. Calculating the optimal spread function for such a monopolist is a relatively simple matter, and the payoffs associated with playing such a strategy in every period represent an upper bound of what underwriters can achieve in a collusive equilibrium. Designate this value \bar{v}^0 . We can now maximize the value of the game treating continuation values as choices from the set $[0, \bar{v}^0]$. The continuation values, spread schedule, and reversion probabilities must satisfy the incentive compatibility conditions that prevent any underwriter from effectively misreporting his signal about the value. The maximum value that can be achieved by solving this program is an upper bound for the value that can be achieved in a symmetric perfect public equilibrium. This procedure can now be repeated indef-

initely to obtain a decreasing upper bound on the equilibrium value set. This upper bound can then be compared directly to the maximum value that can be achieved with a rigid spread or a self-enforcing two-step spread function, both of which can be calculated directly without relying on a dynamic programming approach since neither require punishments along the equilibrium path.

Details of the procedure can be found in the appendix.

2.4.4 Comparing Values: An Example

We have now established a procedure to find an upper bound on the value of a fully separating spread schedule under optimal collusion. We can also find the optimal value that can be achieved with a fully rigid spread or a two-step self-enforcing spread. If either the fully rigid spread or the two-step self-enforcing spread imply greater value than the fully separating spread, we can conclude that the optimal symmetric perfect public equilibrium strategy implies partial price rigidity.

It is clear that, for sufficiently small values of p , a rigid spread will dominate a flexible spread since the optimal flexible spread without incentive constraints converges pointwise to the optimal rigid spread. That is, when the private signals contain almost no information about the value of the firm, even a monopolist would not alter his spread offer much in light of the information received from the private signals. Maintaining incentive compatibility, however, remains costly.

It is not as clear, however, that a two-step self-enforcing spread schedule will be optimal for intermediate values of p . As $p \rightarrow 1$, a fully flexible spread will dominate the two-step spread since the value from the fully flexible spread converges to perfect information monopoly profits.¹⁴ For $p \rightarrow 0$, the rigid spread dominates the two-step spread since the two-step spread requires a large gap between the high spread and the low spread in order to maintain self-enforcement. A two-step self-enforcing spread will be optimal over some intermediate range of p if the costs of enforcement of the fully flexible spread rise sufficiently fast as p decreases relative to the decline in value of segmenting the market based on spreads. It is difficult to characterize the region of p for which the two-step spread will dominate. The technique described above, however, enables us to find examples where exactly this occurs.

Specifically, consider the case where $\bar{x} = 20$, $\beta = 1.01$, $\lambda = 3$, $\kappa = .1$, and $p = 0.6$.

¹⁴This follows from the fact that, as $p \rightarrow 1$, the underwriters will have almost perfect information. Furthermore, the probability that reversion will be triggered is bounded above by $1 - p^2$.

Then, an upper bound on the value¹⁵ that can be achieved with a fully flexible spread is 0.0416, while the rigid spread gives 0.0432 and the two-step self-enforcing spread gives 0.0465. This self-enforcing rigid spread calls for a spread offer of 0.22 following signals below 2.87 and a spread offer of 0.035 for those above this threshold. See figure 2.6 for a plot of the optimal two-step self-enforcing spread and an approximation to the optimal fully separating spread schedule resulting from the iterative procedure above. Also, see Figure 2.5 for an approximation of the punishment function ρ necessary to sustain the optimal flexible spread. Note the shape; small deviations are often more profitable than large deviations, so the probability of reverting to the punishment phase are lower following signals that indicate the possibility of a large deviation. For larger signals, however, payoffs will increase in the size of the deviation, at least initially, as more and more valuable market share is captured. Thus, the punishment function is not in general monotonic.

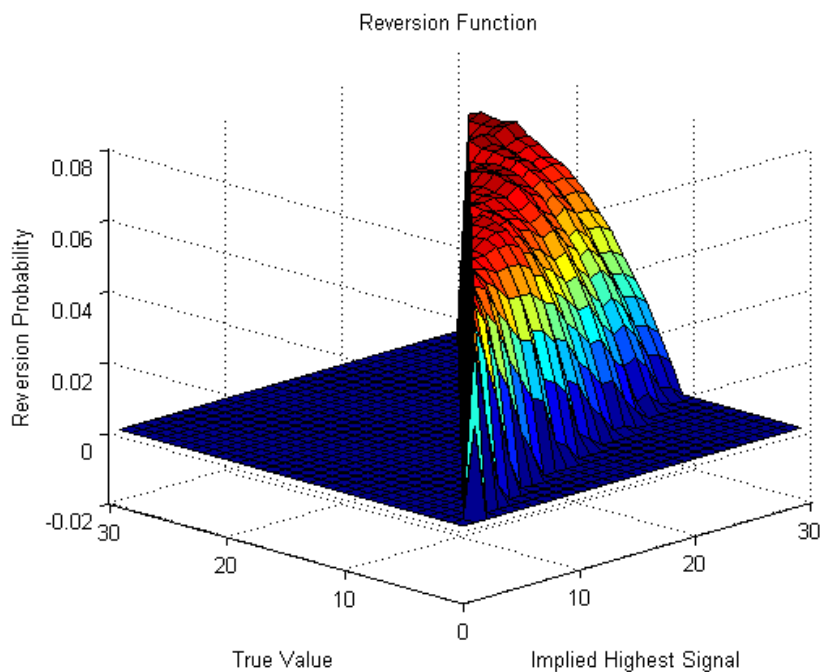


Figure 2.5: Approximate Reversion Function: $K = 30$

¹⁵All values are expressed as the expected normalized discounted sum of payoffs.

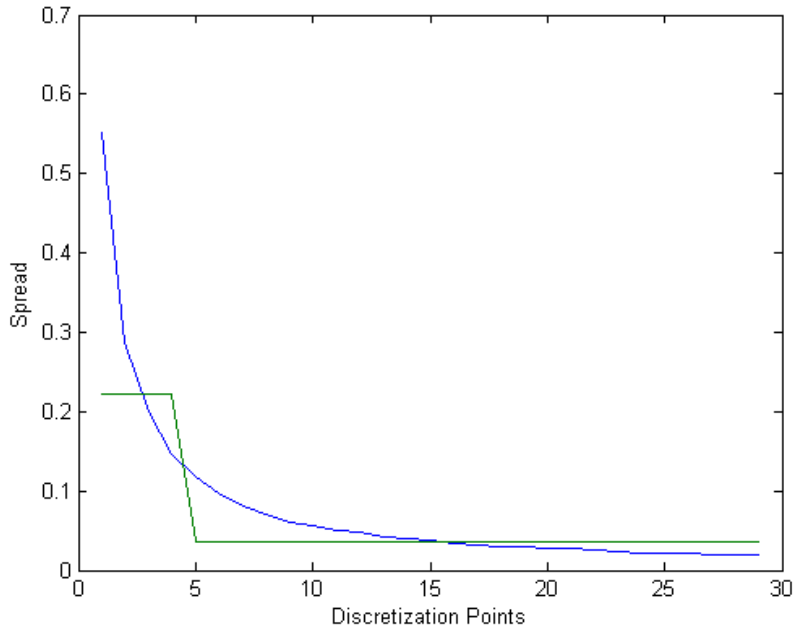


Figure 2.6: Optimal Two-Step Self-Enforcing Spread and Approximate Optimal Fully Separating Spread

Since the two-step spread is strictly preferred to either the fully rigid or the fully flexible spread, it is clear that minimal perturbations of the environment will not undermine the incentives to use a partially rigid spread. That is, small changes in the distribution of ε , x , or ξ would generate small changes in the value of the two-step spread and the value of the fully flexible spread, so the two-step spread would remain more profitable. Minimally informative private signals about manager preference would also not change the ordering of profitability; the optimal fully separating spread would now be a function both of the signal about firm value and the signal about manager preference, but the value of using such a schedule would converge to the value of ignoring manager preference as the informativeness of the signal decreases. So, in this sense, spread rigidity can arise for generic parameter values and functional forms for the relevant structural elements of the IPO industry when spread rigidity is motivated by the need to collude in an environment with private information. This contrasts with the scenarios where spreads can be partially rigid under competition, competitive oligopoly, or monopoly provision, which rely on knife-edge cases to generate spread rigidity.

The example considered in this section is particularly revealing as it represents a case where, arguably, colluding on a fully separating spread schedule should be relatively easy. With positive probability, both underwriters receive the same signal and therefore should even in a fully separating equilibrium frequently bid exactly the same when adhering to a collusive spread schedule. If signals were instead absolutely continuous, every single stage would result in different bids with probability 1. While the relative costs of monitoring will depend on the signal structure in a complex manner, the example presented here suggest that partially rigid spreads would be optimal for many signaling environments.

2.4.5 Optimal Spreads

The above discussion shows that optimal spreads may exhibit partial rigidity when underwriters collude but have private information about firms. Deriving the two-step spread function that provides higher payoffs than the optimal feasible fully separating spread function is sufficient to show that the optimal spread function must exhibit some degree of rigidity, but such a two-step spread is not necessarily the optimal feasible spread schedule. Indeed, data indicate that underwriters do not collude on a two-step spread schedule. Instead, for small firms, several different spreads are charged, with the spread generally, but not strictly, declining in firm value.

In this section, I demonstrate that in fact the optimal self-enforcing spread schedule in the example above must involve more than the single step up. I also argue that all of these steps must occur over the region of low value firms, where “low value” refers to those firms whose value is small enough that manager preference plays a very important role and, as a result, first-best spread offers decline steeply in the signal about firm value. The empirical implications of a spread function of this type provide a very close qualitative match to the distribution of spreads documented in Chen and Ritter [2000].

To see that the optimal self-enforcing spread function must call for more than two steps in the numerical example above, we rely on the following proposition:

Definition 4. *An n -step self-enforcing spread function is an $n+(n-1)$ -tuple $\left\{ \{\alpha^i\}_{i=1}^n, \{x^i\}_{i=1}^{n-1} \right\}$ where $\{\alpha^i\}$ is a decreasing sequence of spreads and $\{x^i\}$ is an increasing sequence of thresholds, and $x^0 = 0$ and $x^n = \bar{x}$, where underwriter j is assigned to demand spread α_j if it has a signal $\xi_j \in (x^{j-1}, x^j]$. Furthermore, the schedule can be enforced without recourse to on-path punishments.*

This is just a generalization of the two-step self-enforcing spread function.

Proposition 6. Consider an n -step self enforcing spread function $\{\{\alpha^i\}_{i=1}^n, \{x^i\}_{i=1}^{n-1}\}$.
If

$$\begin{aligned} \frac{p^2}{2}(-\kappa) &+ \frac{p(1-p)}{2} \frac{1}{\bar{x}} \int_0^{x^1} \pi(\alpha^1, x) h(\alpha^1, x) dx \\ &+ \frac{(1-p)^2}{2} \frac{x^1}{\bar{x}} \int_0^{\bar{x}} \pi(\alpha^1, x) h(\alpha^1, x) dx \\ &< 0 \end{aligned}$$

then there exists a self-enforcing spread schedule with $n + 1$ steps that provides higher payoffs by calling for higher spreads for a subset of signals $[0, x^1]$.

Proof. The expression in the proposition is simply the limit of the expected payoff to an underwriter for adhering to the proposed equilibrium as the signal goes to zero. If this expression is negative, that implies that underwriters with the smallest signals expect negative profit. A schedule calling for those underwriters with very small signals to charge higher spreads would thus be more profitable overall. By choosing the region for this higher spread to be arbitrarily small but with positive measure, the necessary changes to the existing steps in order to make the new schedule an equilibrium would be arbitrarily small. A standard continuity argument then shows that the $n + 1$ step spread is an equilibrium. \square

Applying this proposition to the numerical example shows that there must be at least 3 steps in the optimal self-enforcing step spread schedule. Note that it is not immediate that any self-enforcing step spread schedule can be improved upon with a spread schedule with one additional step. An underwriter with an arbitrarily small signal may still make positive profits if the profits accruing when he is wrong about the firm value exceed the losses when he is right. We can, however, conclude that all of the steps that do arise in an optimal or approximately optimal self-enforcing spread schedule will have the steps concentrated in the low value region.

To see this, note first that large steps are easier to enforce without on-path punishments. A small step cannot be easily enforced since total value accruing will not change much following a deviation from one step to another while the probability of winning the market will increase appreciably. Therefore, if underwriters want to enforce a step with a threshold at an intermediate value firm, the large step required to maintain self-enforcement will generate a large efficiency loss since the first-best spread schedule becomes relatively flat for intermediate and large sized firms. On the other

hand, first-best spreads rise sharply as firm value becomes small, so imposing a large enough step to maintain self enforcement in this region does not involve such a severe efficiency loss.

Exactly what constitutes the region of “small firms” is difficult to succinctly define, but the basic intuition is that spreads will be high where the decision to go public is driven primarily by the idiosyncratic private benefits to managers rather than by the opportunity to increase the common value of the firm.

The empirical implication of a spread schedule that calls for offering a relatively low, fixed spread following most signals and offering higher spreads following signals that the firm value is low, with the exact spread offer for low value firms depending on the signal in a coarse but non-trivial manner, will prove quite similar to the distribution of spreads in data. Since signals about firm value are imperfect, the relationship between true firm value (as measured by the closing price on the first day of trading or, in the model, the issue price) and realized spread will have the steps overlap. With positive probability, both the underwriters will receive and incorrect signal. If both of these incorrect signals are outside the (endogenously determined) step containing the true value, the realized spread offer will differ from the spread that would be charged if the signal were correct. If only one signal is incorrect but that signal leads to a spread offer below that of the correct signal, then again the realized spread will differ from the spread implied by the schedule. For example, if the true value of the firm were such that a correct signal would lead to the highest spread offer, but one firm receives an incorrect signal that the firm is actually of intermediate value, then that firm would be charged the lower spread.

Note that this argument is not symmetric. A firm of intermediate or high value will only be charged a high spread if *both* signals are wrong since the firm can choose the lowest spread offer. Furthermore, when both signals are wrong and both signals imply that the firm is small, an already unlikely event, the firm is very likely to decline to go public since its participation constraint is likely to be violated; the increasingly tight participation constraint is the primary reason why intermediate value firms are charged the lower spread in the first place, so clearly very few will agree to go public if underwriters demand the relatively high spread charged to small firms. Thus, while small firms will occasionally go public at lower spreads than most firms of their size (as observed in the data), few if any large or intermediate size firms will be observed going public at a high spread. This is consistent with the key finding of Chen and Ritter [2000] that price rigidity is nearly absolute over a significant range of firms while also

explaining the richer distribution of spreads for smaller firms.¹⁶

2.4.6 Impatient Firms

Up to this point, we have demonstrated that a model with private information and collusion by underwriters generates a relationship between firm value and spreads that closely resembles the observed data for small and intermediate value firms. In this section, we show that if underwriters are not arbitrarily patient, spreads may decline for the very largest firms while continuing to exhibit the same pattern described above for low and intermediate value firms. Underwriter impatience can thus account for the decline in spreads observed for the largest firms to hold IPOs.

The intuition for this decline is the same as described for the case where firm value and idiosyncratic preferences are perfectly observable. After receiving a signal that the value of the firm is very high, an underwriter expects profits from an optimally chosen spread to also be very high. If the underwriter is impatient and the signal indicates that the firm is sufficiently large, the underwriter will not adhere to the optimal *ex ante* collusive equilibrium but will instead make an (observable) deviation to undercut the other underwriter. This will trigger punishments, but the impatient underwriter will prefer the short-term benefit of the deviation to remaining in the collusive equilibrium.

Given that underwriters are too impatient to enforce optimal collusion, a collusive scheme that takes this constraint into consideration can provide greater profits than a naive attempt to collude on the optimal collusive equilibrium that arbitrarily patient underwriters would choose. The optimal spread schedule for impatient firms requires lower spread offers following high signals and on equilibrium path punishments, but, if impatience is only relevant for the very highest signals, the spread schedule will be effectively unchanged from the arbitrarily patient case except following the very highest signals.

Thus, collusion with private information and some degree of impatience can produce spread schedules that respond to firm value at the extremes of the distribution of firm value but also cause almost all firms of intermediate value to face the exact same spread. This model then explains the key qualitative characteristics of the “seven

¹⁶This point undermines the argument that the observation of 7% spreads in certain small offerings shows that collusion is not responsible for the price clustering (see Hansen [2001]). Once imperfect information about value is introduced, observing exactly 7% for certain small firms will not be surprising since only one underwriter need believe that the firm has intermediate, rather than low, value in order for 7% to arise.

percent solution.”

To demonstrate the above point formally, we will consider the effect of introducing impatience to those games with optimal collusive equilibria in the class described for the numerical example. The following assumption describes the set of problems under consideration and introduces minimal impatience:

Assumption 3.

1. *Parameter values are such that, if underwriters were arbitrarily patient ($\delta \rightarrow 1$), they would collude optimally on a spread schedule calling for a rigid spread for intermediate and large firms and higher spreads for the smallest firms, and would strictly prefer such a schedule to a fully separating or fully rigid schedule. This schedule will henceforth be referred to as the arbitrarily patient equilibrium*
2. *Underwriters are impatient such that, where δ^* is the minimal patience needed to sustain the equilibrium where underwriters are arbitrarily patient, $\delta < \delta^*$.*
3. $\delta^* - \delta \approx 0$

These assumptions guarantee that there is some $\hat{x} \in [0, \bar{x})$ such that underwriters have an incentive to deviate from the arbitrarily patient equilibrium if and only if the signal ξ is in $(\hat{x}, \bar{x}]$, and that this interval is arbitrarily small. With this δ , underwriters who attempt to enforce the arbitrarily patient equilibrium will eventually observe an “off path” deviation and revert to perpetual one-shot equilibrium play. A better schedule, and one that would remain an equilibrium, would call for underwriters to offer a lower spread following the signals above \hat{x} . This alternative equilibrium would not be self-enforcing when spreads are chosen optimally since underwriters receiving signals close to but below \hat{x} would have an incentive to claim a higher signal. Consequently, a public signal that the winning bid implied a signal $\xi \in (\hat{x}, \bar{x}]$ but that the true value x was less than \hat{x} must trigger reversion to stage game play with positive probability. Assumption 3 above guarantees that the probability that such an event will occur along the equilibrium path is arbitrarily small. Thus, by distorting the spreads down for the very highest signals, underwriters can avoid triggering the off-schedule deviations that would occur when attempting to collude on the arbitrarily patient equilibrium. Since the punishments required to enforce this nearly optimal equilibrium occur infrequently and the very large realizations of the signal that require downward distortion also occur

infrequently (since \hat{x} is arbitrarily close to \bar{x} and the probability of one of these large signals is $\frac{\bar{x}-\hat{x}}{\bar{x}}$), there is no incentive for underwriters to employ a schedule over signals in the range $[0, \hat{x}]$ that differs qualitatively from that called for in the arbitrarily patient equilibrium. That is, as $\delta \uparrow \delta^*$ the equilibrium that is identical to the arbitrarily patient equilibrium over $[0, \hat{x}]$ and exploits the most efficient combination of downward distortions for $\xi \in (\hat{x}, \bar{x}]$ and on-equilibrium-path punishments will provide payoffs converging to the payoffs to the arbitrarily patient equilibrium. Since a partially rigid spread schedule is strictly preferred to either a fully separating or fully rigid schedule in the case of arbitrarily patient underwriters, moving to a rigid or fully separating spread schedule cannot improve payoffs in the case of minimal impatience.

We will not attempt here to trace out the exact empirical implications of impatience. Since the distribution of large firms is unlikely to be well approximated by a uniform distribution, and a distribution with a long, thin tail would produce substantially different predictions about spreads charged by impatient underwriters, such an exercise would provide little in the way of validating or rejecting the model. We have, however, shown that in the context of the model with private information, impatience by underwriters will lead to a decrease in the spreads charged to the most valuable firms without undermining the incentives to apply the partially rigid spread described in the preceding subsection over the range of signals not near the maximum of the distribution.

2.5 Underpricing

While the great extent of spread rigidity in the market for IPOs has received significant attention in financial economics, the tendency for IPOs to be significantly underpriced has been the subject of far more research. Prominent examples include Loughran and Ritter [2004], Booth and Chua [1999], and Lowery and Shu [2002]. In this section, I show that spread rigidity and underpricing may be closely related.

Under the rigid or partially rigid equilibrium described above, underpricing has a natural role. Underwriters collude to extract as much surplus from firms as possible. However, given the need to collude on a rigid or partially rigid spread and the imperfect information about firm value and preferences, those firms that choose to hold an IPO will still benefit. That is, firms are not pushed to their participation constraint, and indeed underwriters are not even extracting all surplus that they could given their

information. This section shows that underwriters may have an incentive to underprice issues in order to extract this additional surplus and to exploit any additional information received during the book-building process. The opportunity to underprice, in turn, reinforces the incentives to collude on a rigid or partially rigid spread; part of the loss from failing to charge the most efficient, flexible spread is recouped through the underpricing stage without requiring costly punishments.

Admitting underpricing into the analysis requires explicitly considering the choice of the offer price. Above, I have implicitly assumed that the offer price for the shares perfectly reflects the true value of the firm. That is, while underwriters have imperfect information about the value of the firm when bidding to hold the IPO, the book building process functions perfectly to reveal the true value and there is no incentive to underprice the offer.

Given the high and highly variable underpricing observed in the industry, these simplifying assumptions clearly do not capture all important elements of the industry. Either underwriters remain uninformed about the true value of the firm relative to information available to the general investing public or they intentionally underprice issues. It is difficult to believe that the underwriter would be so poorly informed relative to the first pair of investors to engage in a transaction following the IPO,¹⁷ so we proceed under the assumption that by the end of the book building process the underwriter has access to all information necessary to accurately price the security. Additionally, the underwriter is likely to learn more about the preferences of the managers of the firm and will thus update the distribution of ε .

To introduce underpricing into the model, let the underwriter also have access to a costly technology for turning underpricing into profits. Specifically, assume that an underwriter is in a long term relationship with a class of institutional investors who will “kick back” part of their gains from receiving an underpriced issue. This sort of behavior is well documented, from the brazen (SEC [2004]) to more subtle *quid pro quo*. Underpricing will also have other less direct costs; a large degree of underpricing may attract scrutiny from regulators or lawsuits from issuers, and too much underpricing may hurt the reputation of the underwriter and make other firms less likely to choose that underwriter in the future.

We assume that underwriters are unable to compete on underpricing. Since the issue price is set very late in the IPO process, underwriters cannot contract on a

¹⁷Kirgman et al. [1999] show that an overwhelming proportion of the first day return, the standard measure of underpricing, is realized on the first trade of that day.

particular issue price. Thus, if underwriters wished to compete on underpricing, they would have to send credible signals to the firms that they will not underprice even when it is profitable to them. Any such signal would be just as observable to the other long-lived underwriter as it would be to the sequence of short-lived firms, so underwriters would be able to punish any attempt to use more favorable issue pricing to capture more of the market.

Formally, we consider the following environment:

- Once the underwriter is chosen to hold the IPO, it observes x perfectly
- With probability μ , the underwriter observes ε
- The underwriter chooses an offer price $x_o \leq x$, which implies a degree of underpricing $u = x - x_o$.
- The firm has an opportunity to withdraw from the issue and receive its reservation value x (in which case the underwriter receives $-\kappa$)
- If the issue proceeds, the payoffs are:

underwriter: $\alpha\beta x_o + \theta_t\beta u - \zeta_t\beta^2 u^2$, where $\theta_t \sim [0, 1]$ and $\zeta_t \sim [0, \infty)$

firm: $(1 - \alpha)\beta x_o + \varepsilon$

This environment guarantees that a firm's participation constraint in the underwriter-selection stage is unchanged; with positive probability the underwriter will have no incentive to underprice. That is, if $\theta < \alpha$ or ζ sufficiently large, a firm will prefer to choose $x_o \approx x$; when this occurs with positive probability, any firm that would choose to go public at a "fair" offer price will continue to agree to hold an IPO in the selection stage. Those firms who are almost indifferent between going public at the equilibrium α and remaining private will often be forced to their participation constraint when ε is revealed and will often cancel the IPO when ε is not revealed. They, however, still expect positive unconditional surplus, with the expected surplus decreasing in α .¹⁸

¹⁸It is not immediate that, in the presence of underpricing, firms would continue to choose the lowest spread. If underpricing is a costless way to generate additional profits, firms will be indifferent between a high spread with the expectation of low underpricing and a low spread with the expectation of high underpricing. The convex costs associated with underpricing guarantee that most firms will still prefer a low spread to a high spread. It is possible to construct examples even with convex costs where some small firms prefer larger spreads (since larger spreads serve, in part, to align the interest of the firm and the underwriter), but these cases are not of particular interest.

Underwriters will clearly benefit from the introduction of the underpricing stage. They have more information and another means to extract surplus. Furthermore, the value of coordinating on a fully flexible spread relative to a self-enforcing step function is reduced, thus preserving the spread-rigidity result above. That is, the offer-price phase will allow underwriters to extract some of the surplus lost by coordinating on a self-enforcing partially rigid spread, still without relying on on-equilibrium path punishments. To see this, consider the case where the mass of the distribution of θ becomes concentrated around 1 while the mass of the distribution of ζ becomes concentrated around 0, and $\mu \rightarrow 1$. Then, the underwriter extracts almost the full surplus from each firm, regardless of the spread offer. Taking a very small firm public at a relatively low α , however, will still imply an expected loss for the underwriter unless ε is very high. This situation implies that underwriters would still have an incentive to coordinate on a step-spread function rather than simply charging a flat, low spread regardless of the signal.

Under the assumptions presented above, it is straightforward to solve for optimal underpricing. When the underwriter does not observe ε , it will choose the optimal offer price taking into account the risk that too much underpricing will cause the firm to cancel the offer. The underwriter, of course, has effectively received a signal that the firm has $\varepsilon \geq x(\alpha\beta + 1 - \beta)$, and will therefore maximize conditional on an updated posterior distribution of ε . The optimal level of underpricing is then given by:

$$u^* = \frac{\left(\frac{2\zeta\beta}{\lambda(1-\frac{1}{x})} + \theta - \alpha\right) - \sqrt{\left(\frac{2\zeta\beta}{\lambda(1-\frac{1}{x})} + \theta - \alpha\right)^2 + 4\zeta\beta\frac{\theta-\alpha-\alpha x}{\lambda}}}{2\zeta\beta}$$

if said quantity falls in the interval $[0, x]$; a negative offer price or an offer price above the true value is infeasible.

In the cases where the underwriter does observe ε , the issue will never be withdrawn as the underwriter can now perfectly observe the participation constraint. In cases where the participation constraint does not bind, underpricing is given by

$$u^* = \frac{\theta - \alpha}{2\zeta\beta},$$

again with the *caveat* that underpricing must fall in the feasible interval. When the

participation constraint does bind, optimal underpricing is instead given by

$$w^* = \frac{\varepsilon}{(1-\alpha)\beta} - \left(\frac{\alpha\beta + 1 - \beta}{(1-\alpha)\beta} \right) x.$$

It is immediate that money left on the table will not depend on x when the participation constraint does not bind. It is less obvious, but nonetheless true, that the expected value of money left on the table conditional on firm value does not depend on the value of the firm when the participation constraint does bind. That is, the expected value of ε conditional on a firm choosing to go public at a given spread increases in x at a rate that exactly offsets the decrease in x of the degree of underpricing conditional on ε . This exact balancing is in part an artifact of the modeling structure, in particular the assumption that the costs of underpricing are convex in the absolute rather than relative degree of underpricing, but may help explain why the relationship between money left on the table and firm value is not obviously monotonic.

2.6 Seasoned Equity Offerings

While the spreads on IPOs exhibit remarkable price rigidity, spreads charged for seasoned equity offerings demonstrate dependence on issue size and also exhibit variance even conditional on issue size (See Chen and Ritter [2000], figures 4 and 5.) This suggests that the logic underlying the collusive solution for IPOs does not apply to SEOs. Since IPOs and SEOs are fundamentally different in many ways, this difference should not be taken as evidence that the market for SEOs is more competitive than that for IPOs, even though spreads on SEOs are on average much lower than those on IPOs. There are two crucial differences between the market for IPO services and the market for SEO services. First, there is no asymmetric information about firm value when a firm is considering holding an SEO since the stock price is already a publicly observable variable. And, firms holding SEOs, almost by definition, have a preexisting relationship with one investment bank. Krigman et al. [2001] document that 70% of firms completing an SEO within three years of their IPO use the same underwriter as they did for the IPO. This suggests that underwriters may use their preexisting relationships with various firms to coordinate on an asymmetric bid-rotation scheme, which would permit flexible spreads on SEOs that make the most use of all information available to the underwriter who held the firm's IPO. The lower spreads observed on

SEOs would then arise not from greater competition amongst underwriters, which is hard to reconcile with the 70% figure cited above, but from the presence of alternative sources of capital available to firms that are already public and the relative absence of idiosyncratic preferences over capital structure among managers of already public firms.

2.7 International Comparisons

The distribution of IPO spreads varies significantly across countries. In particular, both the average spread level and the degree of clustering at the modal spread differs markedly in cross country comparisons. Torstila [2003] presents evidence on spread rigidity in many countries. In Singapore and Germany the shape of the spread distribution is roughly similar to that in the United States, although spreads are on average much lower. The Netherlands, Italy, and the United Kingdom all exhibit very little rigidity, while Hong Kong, India, and Malaysia exhibit rigidity across a wide range of values without any apparent steps. This observation is broadly consistent with the model in this paper as collusion can produce any of these outcomes depending on, among other things, the quality of information available to the underwriting industry in a country. The exact implications for relative competitiveness in each country are ambiguous. The presence of rigidity is convincing evidence of collusion, but the absence of rigidity can imply either competitive pricing or efficient collusion exploiting bid rotation and communication. However, in each of the countries mentioned above as not exhibiting rigidity, there is significant variance in the spreads charged for a *given* level of offering proceeds. This is some evidence that underwriters are exploiting information about firm preferences to extract collusive rents. Conditional variance will arise under either myopic oligopolistic competition or perfect competition, largely as a result of imperfect signals about firm value. The wide variance relative to the slope of the average spread, however, is weak evidence for the hypothesis that underwriters are engaged in efficient collusion.

The relatively low spreads charged outside of the United States also has ambiguous implications for collusion. Lower average spreads in countries where collusion is apparent imply that the value of access to public capital markets is lower, which will drive the average reservation spread of private firms lower. Given the relative importance of public capital markets in the United States, it is not surprising that firms in

other countries are generally charged spreads lower than 7%. In particular, in countries where bank financing is an important source of capital for firms, it is unlikely that many firms would be willing to go public at such high spreads. If, on the other hand, spreads were determined competitively, the cross-country difference in spreads would primarily imply differences in costs to underwriters for running IPOs. It is not at all clear that this activity should be more expensive in the U.S., a country with relatively efficient and well functioning capital markets and regulatory infrastructure.

2.8 Policy Implications

Collusion in the pricing of IPO services is costly from a social perspective. Inefficiently high spreads result in fewer IPOs than would be optimal, and the incentives to start positive expected value firms will be suppressed if much of the value of taking the company public, which seems to represent a significant fraction of the total value of the firm (Ritter [1987]), is captured by the underwriters.

Regulatory intervention to reduce or eliminate collusion would therefore seem desirable. A natural if blunt approach to such regulation would involve imposing a binding cap on the spread charged for IPOs. Such a regulation might prove unwise if, as conjectured in Chen and Ritter [2000], small firms are charged high spreads in order to cover the fixed costs associated with taking a firm public. In such a circumstance, small firms would be effectively excluded from holding IPOs, even when going public is quite valuable to the firm. The model presented here indicates, however, that the higher spreads charged to firms with low value are driven in part, and perhaps for the most part, by “demand” effects rather than “supply” effects. Underwriters charge higher spreads to small firms in order to capture relatively large fees from those firms whose idiosyncratic preferences for going public are large. The small value of the firm effectively magnifies the importance of idiosyncratic preferences to the manager, relative to his concern about minimizing the fee. As long as underwriters charge higher spreads to small firms for this reason, a price cap will not drive small firms out of the market and will instead prove even more beneficial to those firms.

2.9 Conclusion

We have now seen that the empirically observed distribution of IPO spreads can best be explained by assuming that underwriters collude on price but receive noisy signals about firm value and preferences. Attempts to reach optimal collusion in a symmetric perfect public equilibrium are doomed to failure by the requirement that punishments occur along the equilibrium path, but underwriters will still capture significant rents. Collusion, price rigidity, and underpricing are closely linked, as underwriters exploit underpricing to extract additional surplus from firms. The model suggests links between unobservable variables relevant for underwriting and observed data, providing a blueprint for a structural analysis of the industry. Such analysis could guide regulatory policy designed to address the social costs associated with collusion in investment banking.

2.10 Appendix

2.10.1 Finite Approximation

In this appendix, I consider finite approximations to the continuous game described in the main text. Such finite games will exhibit the bang-bang property, whereas it is possible that there is some equilibrium of the continuous version of the game that provides better payoffs than the best bang-bang equilibrium. The arguments below presuppose familiarity with sections 2.3 and 2.4.

There are two complications that must be considered when treating the continuous stage game discussed in the main text as an approximation to a stage game with a finite but arbitrarily large action space. First, it is necessary to show that the payoffs to the optimal rigid and partially rigid spread schedules are approximated by the payoffs in the continuous version, and that the optimal strictly decreasing spread function is an appropriate approximation to the payoffs to a separating equilibrium in the finite game. The first two of these requirements are trivial. The third is less so because it is necessary to define the analog in the finite case of a strictly decreasing spread schedule. To address this, I consider a discretization of the following form. Spreads can be chosen from a finite subset of $[0, M]$ with cardinality N , and the spread schedule must be a decreasing step function. Furthermore, no step can have measure greater than ι . To approach the continuous case, let $M \rightarrow \infty$, $N \rightarrow \infty$, and $\iota \rightarrow 0$. As long

as these parameters change at an appropriate rate relative to each other, this approximation converges to the continuous game with a strictly decreasing spread schedule in the following sense. As $N \rightarrow \infty$ (at a rate fast relative to the increase in M), the underwriters cease to be constrained by the finiteness of the action space, while $\iota \rightarrow 0$ implies that they cannot exploit spread rigidity to reduce the probability of triggering punishment. Thus, in the limit, the underwriters will choose a spread that effectively uses all information available in the signals without exploiting rigidity to decrease the probability of on equilibrium path punishments. The maximum payoff to such an equilibrium employing bang-bang strategies converges to the maximum payoff to the optimal equilibrium supported with bang-bang strategies in the continuous game, and thus the upper bound on payoffs to an appropriate discretization of the game converges to the upper bound on the payoffs in bang-bang equilibria in the continuous case. Specifically, if there is some equilibrium of the continuous game in which payoffs are higher than the best bang-bang equilibrium of the continuous game, in an arbitrarily fine discretization of the game there is some profitable deviation from such an equilibrium since such an equilibrium cannot be supported with bang-bang strategies. This is *not* a contradiction of the claim that the discretization converges to the continuous game in the appropriate sense. Assuming there is some better-than-bang-bang equilibrium in the continuous game, there may be some deviation from the equilibrium that provides the same payoff to the deviator as adhering. In this case, the payoffs to the discrete analog of this deviation may give a higher payoff than adhering to the discrete analog of the better-than-bang-bang equilibrium. As the discretization becomes increasingly fine, the payoffs from the deviation must converge to the payoffs from adhering, but they are only equal in the limit and *not* for any actual finite game.

The second concern when evaluating a discrete version of the game is that the results from section 2.3 must be interpreted more carefully. These results depended effectively on showing that spreads will not be locally flat. In any finite approximation spreads will clearly have flat regions by construction; in fact, the spread schedule will be a step function. The appropriate approximation here is that as the cardinality of the finite action space increases, the number of steps will increase, and as the cardinality goes to infinity so will the number of steps. Furthermore, the measure of the set of signals generating any single spread will go to zero. This is the appropriate sense in which spreads will not be rigid in finite approximations to the continuous game under competition, competitive oligopoly, or monopoly.

That this statement is true follows from the fact that, for any finite strategy with

underwriters attempting to reach the imposed criteria (zero profits, one-shot equilibrium, or optimal monopoly), firms would always “deviate” from prescribed actions with finite support *if* they were not restricted to a finite action space. For example, if firms chose actions such that they would be as “competitive” as possible, then a restriction to a finite action space would imply that expected profit would be small but positive for some signals and small but negative for others.¹⁹ Thus, when the action space expands by a sufficiently rich set of additional spreads, the most competitive spread will have more steps than when the action space was smaller.

Further details on these approximations are available from the author on request.

2.10.2 Competitive and Monopoly Spreads: General Model

This section explores in a more general framework the spread schedules that would arise in the absence of collusion of the form posited. Specifically, we assume that the common value benefits of holding the IPO take the form of a potentially nonlinear function $\tilde{\beta}$, where $\tilde{\beta}'(x) > 0$ for all x . I also consider a general, weakly increasing cost function $C(x)$. Spreads are considered under three alternative assumptions about the process for bidding for an IPO. First, underwriters are assumed to bid in a “competitive” market for underwriting services such that they expect zero profit following every realization of the signal. Second, underwriters bid strategically as in an oligopoly without collusion. Finally, underwriters act effectively as a monopoly.²⁰ I consider the conditions necessary to generate intervals over which a fixed spread is charged. In each of these cases the conditions for flat spreads over some interval of firm value are shown to rely on very specific relationships among the costs to the underwriter of holding an IPO, the common value benefits, and the distribution of idiosyncratic preferences. As such, a slight perturbation of any one of these functions will eliminate the region of flat spreads. This requirement that the forms of the functions representing the structural elements of the underwriting industry exhibit remarkable coincidences²¹ in order to generate even small regions of fixed spreads will later be seen to contrast with the simple and robust setting in which collusion with imperfect information leads to

¹⁹Here we ignore the irrelevant fact that firms could guarantee themselves zero profit by charging ∞ , which is clearly not descriptive of the data and not in the spirit of the competitive assumption.

²⁰Note that of these three solution concepts, only competitive oligopoly represents an equilibrium of the game as defined above.

²¹The argument is that spreads will respond to information generically. Defining the appropriate sense of “generic” and formally proving the statement are tedious, uninteresting, and add little beyond the intuition presented here. Details are available from the author.

partially rigid spreads that are qualitatively consistent with observed spreads.

Throughout this section we confine attention to symmetric, pure strategies that call for weakly decreasing spreads. While these restrictions are in part for simplicity, the distribution of spreads generated by asymmetric, increasing, or mixed strategies would not be compatible with observed data.

With these restrictions, I can consider a general, weakly decreasing spread schedule $\alpha(\cdot)$ where spreads are rigid over some interval $[x^d, x^u] = \mathcal{I} \subset X$. That is, for all $\xi \in \mathcal{I}$, $\alpha(\xi) = \alpha \in \mathbb{R}^+$. The payoff to one underwriter following a signal in $\xi \in \mathcal{I}$ conditional on being among the lowest bidders can then be given in each case by

$$\mathbf{\Pi}(\xi, \alpha) = \mathcal{A}(\xi)E[\mathcal{H}(\alpha, x)(\alpha\tilde{\beta}(x) - C(x))|\xi, \xi' \in \mathcal{I}] + (1 - \mathcal{A}(\xi))E[\mathcal{H}(\alpha, x)(\alpha\tilde{\beta}(x) - C(x))|\xi, \xi' < x^d] \quad (2.1)$$

where ξ' is the signal observed by the other underwriter, $\mathcal{A}(\xi) = \frac{P(\xi' \in X|\xi)}{\frac{P(\xi' \in X|\xi)}{2} + P(\xi' < x^d|\xi)}$, and $\mathcal{H}(\alpha, x) = P(\varepsilon > x - (1 - \alpha)\tilde{\beta}(x))$. These expressions are, respectively, the probability of making the same spread offer as the other underwriter and being chosen to hold the IPO conditional on either making the lowest spread offer or making the same spread offer and being selected at random,²² and the probability that the firm will agree to go public given a spread α and firm value x . This expression for expected revenue can be decomposed into $\mathbf{\Pi}^T$, the part of the expected profit conditional on winning that accrues when the auction is a “tie,” and $\mathbf{\Pi}^S$, the part that accrues when the underwriter wins outright. The values $\mathbf{\Pi}$ and $\mathbf{\Pi}^T$ will provide the conditions necessary for spreads to be rigid under competition and competitive oligopoly; a related function will provide the conditions for monopoly.

2.10.3 Competitive Spreads

Here, we assume that following every signal about firm value the underwriter sets his spread such that he will earn, in expectation, zero profit. That is, $\mathbf{\Pi}(\xi, \alpha) = 0$ for all $\xi \in X$. Under the assumption that spreads are the same for every $\xi \in \mathcal{I}$, there is an additional restriction that $\frac{\partial \mathbf{\Pi}}{\partial \xi} = 0$ for all $\xi \in \text{int}(\mathcal{I})$. In fact, $\frac{\partial^n \mathbf{\Pi}}{\partial \xi^n} = 0$ for all n and all $\xi \in \text{int}(\mathcal{I})$. It is then clear that if signals are informative about value and value is informative about the probability that a firm will agree to the IPO at a given spread, flat regions can only occur when there is a remarkable coincidence among the costs of

²²Here, we use the convention that if both underwriters offer the same spread, the firm first chooses between the two underwriters and *then* consults its participation constraint. This is just a convenience.

holding an IPO, the common value benefits from an IPO, and distribution of manager preference.

The following proposition notes one such relationship that would generate flat spreads; the proof is an immediate consequence of equation 2.1 and is therefore omitted:

Proposition 7. *Spread offers are invariant in the signal about firm value over the interval $\mathcal{I} \subset X$ if*

$$C(x) = \gamma\tilde{\beta}(x).$$

for all x that occur with positive probability following signals in \mathcal{I} .

The proposition states that if the costs to the underwriter of arranging the IPO are exactly proportional to the common value benefits of the IPO, then flat spreads will be observed. This strict proportionality imposes strong restrictions on the relationship between the cost function of an underwriter, which will depend on the costs of providing analyst coverage and engaging in roadshows, and the benefits of being a public firm, which are related to improved access to capital markets. Relaxing this strict proportionality will then effectively require that $\frac{1-A}{A}$, which is determined entirely by the joint distribution of firm value and signals, and $\mathcal{H}(x, \alpha)$, which is determined by the distribution of the preference of the manager, to exactly adjust for the lack of proportionality in costs and benefits, an extremely unlikely event. This simple case thus demonstrates why it is so unlikely to observe flat spreads in a competitive market for IPO services.

2.10.4 Oligopolistic Competition

When the two firms bid for the right to hold the IPO and seek to maximize their current payoff (i.e. they do not attempt to collude), spreads are again unlikely to contain regions with flat areas. In this case, it is possible to derive some simple necessary conditions for the existence of an interval \mathcal{I} over which the spread schedule is flat. These conditions, however, are not sufficient for a the spread schedule to be flat in the interval.

Proposition 8. *A necessary condition for the spread schedule to be flat but greater than zero over the interval X is that $\mathbf{\Pi}^T(\xi, \alpha) = 0$.*

Proof. If spreads are flat over some interval in equilibrium, a deviation to an infinitesimally higher spread guarantees that the deviator never wins when both signals are in

said interval, and an infinitesimal deviation to a lower spread guarantees that the deviator always wins when both signals are in said interval. For both of these deviations to be unprofitable, it must be the case that the expected profit conditional on both signals being in the interval be zero at α (since the change in profit accruing in all events other than $\xi, \xi_i \in \mathcal{I}$ changes infinitesimally following an infinitesimal deviation). \square

The above condition can be satisfied in two ways; either costs are exactly proportional to benefits such that $C(x) = \gamma\tilde{\beta}(x)$ for all x with positive support following $\xi, \xi' \in X$, in which case $\alpha = \gamma$ satisfies this condition, or $\mathcal{H}, \tilde{\beta}$, and C have a particular dependence on α and x such that this condition is met for some α . Neither of these events are likely in the sense that small perturbations of any of the functions would undo the necessary relationship. Furthermore, this condition is effectively independent of the differential equation that will determine the symmetric equilibrium in the auction. Assuming that equilibrium spreads are in fact flat over \mathcal{I} , and thus the expected profit accruing conditional on both signals falling in \mathcal{I} is zero, the optimal bid for an underwriter with a signal in \mathcal{I} will be determined by maximizing expected profit conditional on the other signal not being in \mathcal{I} and on the equilibrium spread schedule outside of the interval. Not only must this happen to give the α that satisfies the zero profit condition for one particular signal in \mathcal{I} , it must do so for *all* signals in \mathcal{I} , an extremely unlikely event when signals are at all informative.

2.10.5 Monopoly

Several authors (for example Chen [2001]) have argued that fixed spreads can arise when underwriters collude on the optimal monopoly spread. In this section I argue that the optimal monopoly spread is unlikely to be characterized by fixed spreads over a significant region. It is important to draw a distinction between the argument that spreads are rigid because underwriters collude on a monopoly spread that happens to be rigid and the argument that rigid spreads arise in a second-best equilibrium where collusion at monopoly spreads is impossible. First, as I will show below, monopoly spreads will only be rigid for very specific environments that are unlikely to describe the IPO process. As such, explanations of spread rigidity based on monopoly pricing do not really provide theoretical support for the claim that spread rigidity is evidence of collusion rather than competition; both monopoly pricing and competition generate rigid spreads in effectively non-generic environments. Furthermore, optimal policy responses to collusion may differ depending on whether the market is effectively a

monopoly or only imperfectly collusive. Finally, the empirical implications for costs and benefits in the IPO industry of the data on spreads will depend crucially on the exact nature of the collusion in the industry. Attempting to recover empirical facts relating to the costs to underwriters and the benefits to firms of holding IPOs, while a daunting task, is worthwhile both because it would inform policy decisions and because few alternatives for measuring these quantities are available to financial economists.

In this section, I consider the case of a disaggregated monopoly, where the two underwriters each receive signals and set spreads to maximize joint profits but where underwriters cannot communicate their signals to each other before bidding.²³ This schedule must by construction satisfy the condition that, following a given signal, the underwriter receiving the signal cannot unilaterally “deviate” to an infinitesimally different spread and, by doing so, increase the expected *total* profits of the two underwriters. That is, defining $\tilde{\mathcal{A}}(\xi)$ as the probability of the event $\xi' \in \mathcal{I}$ conditional on ξ' either in \mathcal{I} or less than x^d , it must be the case that for all²⁴ $\xi \in \mathcal{I}$:

$$\begin{aligned} \frac{d}{d\alpha} \{ & (1 - \mathcal{A}(\xi))E[\mathcal{H}(\alpha, x)(\alpha\tilde{\beta}(x) - C(x))|\xi, \xi' < x^d] \\ & + \mathcal{A}(\xi)E[\mathcal{H}(\alpha, x)(\alpha\tilde{\beta}(x) - C(x))|\xi, \xi' \in \mathcal{I}] \} \\ & \geq 0 \end{aligned}$$

That is, an infinitesimal decrease (to below α , the rigid spread over \mathcal{I}) in the spread offer following a signal $\xi \in \mathcal{I}$ must decrease the total profit to both underwriters conditional on $\{\xi' \in \mathcal{I}\} \cup \{\xi' < x^d\}$. (The marginal decrease is irrelevant if the other underwriter would have won the IPO at a lower spread than α before the deviation.) The above expression follows from two facts: after a marginal decrease in the spread on the rigid interval, the total increase in profits conditional on winning must be negative, and a deviator to a marginally smaller spread following a signal that calls for the rigid spread now captures all of IPO’s where the other signal also calls for the rigid spread.²⁵

²³Permitting communication will not change the conclusions of this section in any meaningful way, although the exact environments where fixed spreads will arise will differ.

²⁴Technically, except on a set that arises with probability zero; we ignore such concerns.

²⁵The analogous condition to guarantee that the underwriter will not increase his spread is just

$$\frac{d}{d\alpha} \left\{ (1 - \tilde{\mathcal{A}}(\xi))E[\mathcal{H}(\alpha, x)(\alpha\tilde{\beta}(x) - C(x))|\xi, \xi' < x^d] \right\}$$

since, by raising the spread, the underwriter affects profits only when the other underwriter would have charged a spread higher than the rigid spread.

If the above inequality does not bind for almost every $\xi \in \mathcal{I}$, then there is some interval over which a marginal decrease in spreads can increase total profits, contradicting the optimality of $\alpha(\cdot)$. But, when the inequality does bind almost everywhere, we again have a condition that cannot hold in the generic sense described in the discussion of competition and non-collusive oligopoly.

Spreads will, however, be rigid in the case where firm value is not informative about the probability that a firm will accept a given spread offer. Specifically, if $\varepsilon = 0$ for all firms and $\tilde{\beta}(x) = \beta x$, a monopolist would charge exactly

$$\alpha = 1 - \frac{1}{\beta}$$

regardless of his signal, since this would guarantee that every firm exactly met its participation constraint, regardless of its true value.

Rigid spreads over some interval could also arise when manager preference is bounded from below. It is then possible to find cases where, for a large enough signal, underwriters will set spreads to guarantee that all firms go public, regardless of manager preference or firm value. That is, for signals that indicate that the distribution of firms is skewed toward high values, the monopolist underwriter might prefer to set a spread that guarantees that the largest possible firm goes public. The costs of failing to serve such firms may be so high as to make marginal differences in earnings elsewhere irrelevant, and the spread may thus be rigid exactly at the level that guarantees that the largest possible firm will agree to the IPO.

These two cases both involve a discontinuity in the distribution of ε so that the condition described above does not eliminate them. The second case also presents the most compelling alternative explanation for spread rigidity as it will involve decreasing spreads for small firms and flat spreads for large firms. Without extreme assumptions on the distribution of value and signals about value, however, it cannot generate the specific pattern observed in data, where spreads appear to step up for small firms rather than smoothly decreasing toward the flat spread. Furthermore, this explanation relies on the existence of a largest possible firm value and on significant mass being concentrated around that value.

2.10.6 Perfectly Observable Value and Preference

To analyze the game under perfect information, I will proceed under the following additional assumptions:

Assumption 4.

1. x and ε are common knowledge at the beginning of every period.
2. $E[\varepsilon] = 0$
3. Underwriters are sufficiently patient to sustain any degree of collusion.

The first assumption is just a restatement of the condition that both agents receive perfect signals at the beginning of the period. The second assumption is just for simplicity.

The spread that will be charged if the two underwriters can collude perfectly and set the monopoly price is then given by the following proposition:

Proposition 9. *When underwriters act as a monopoly, if*

$$\varepsilon \geq \kappa - (\beta - 1)x$$

the spread function is given by

$$\alpha^* = 1 - \frac{1}{\beta} + \frac{\varepsilon}{\beta x}$$

and the firm chooses to hold the IPO. Otherwise,

$$\alpha \geq 1 - \frac{1}{\beta} + \frac{\varepsilon}{\beta x}$$

and the firm does not hold the IPO.

Proof. Since the firm is short-lived, he will accept the lowest spread offer as long as his residual claim on the profits to the public firm plus his idiosyncratic private value for going public exceed the private value of the firm. Underwriters will force the firm to its participation constraint:

$$\begin{aligned} (1 - \alpha)\beta x - x + \varepsilon &= 0 \\ \alpha &= 1 - \frac{1}{\beta} + \frac{\varepsilon}{\beta x}. \end{aligned}$$

But, underwriters will only make such an offer if it is profitable to them. This condition is given by

$$\begin{aligned} \left(1 - \frac{1}{\beta} + \frac{\varepsilon}{\beta x}\right) \beta x - \kappa &> 0 \\ \varepsilon &> \kappa - (\beta - 1)x. \end{aligned}$$

□

I now consider how spreads behave for very small and very large firms. On average, very small firms will face arbitrarily high spreads, while larger firms face spreads that converge to the spread that would be charged to a firm with no idiosyncratic preferences for going public.

Proposition 10. *The expected spread as firms become infinitely valuable and as firms cease to have any value are given by:*

$$\lim_{x \rightarrow \infty} E[\alpha | x, \varepsilon > \kappa - (\beta - 1)x] = 1 - \frac{1}{\beta} \quad (2.2)$$

$$\lim_{x \rightarrow 0} E[\alpha | x, \varepsilon > \kappa - (\beta - 1)x] = \infty. \quad (2.3)$$

Proof.

$$\begin{aligned} E[\alpha | x, \varepsilon > \kappa - (\beta - 1)x] &= E\left[1 - \frac{1}{\beta} + \frac{\varepsilon}{\beta x} \mid x, \varepsilon > \kappa - (\beta - 1)x\right] \\ &= 1 - \frac{1}{\beta} + \frac{1}{\beta x} E[\varepsilon | \varepsilon > \kappa - (\beta - 1)x]. \end{aligned}$$

Since $\kappa - (\beta - 1)x$ is decreasing in x and covers the real line, we can conclude that $\lim_{x \rightarrow \infty} E[\varepsilon | \varepsilon > \kappa - (\beta - 1)x] = 0$, while $\lim_{x \rightarrow 0} \frac{1}{x} E[\varepsilon | \varepsilon > \kappa - (\beta - 1)x] = \infty$ since $E[\varepsilon | \varepsilon > \kappa]$ is positive, and, more generally, $\frac{d}{dx} E[\varepsilon | \varepsilon > \kappa - (\beta - 1)x] > 0$ since $\beta > 1$. Thus,

$$\begin{aligned} \lim_{x \rightarrow \infty} E[\alpha | x, \varepsilon > \kappa - (\beta - 1)x] &= 1 - \frac{1}{\beta} \\ \lim_{x \rightarrow 0} E[\alpha | x, \varepsilon > \kappa - (\beta - 1)x] &= 1 - \frac{1}{\beta} + \infty \\ &= \infty, \end{aligned}$$

establishing the proposition. □

Corollary 1. *The variance of the spread conditional on firm value disappears for large firm values.*

Proof. Direct calculation of the variance in spreads conditional on x gives

$$\frac{1}{\beta^2 x^2} E[\varepsilon^2].$$

□

The implications of the above results are as follows. First, without idiosyncratic preferences all firms would be charged exactly $1 - \frac{1}{\beta}$. This is exactly the spread that captures all of the “common value” of the IPO process, regardless of the size of the firm. However, this is not a complete explanation for the concentration of spreads at seven percent; rigidity would not be robust to the introduction of some small degree of idiosyncratic manager preference, and spreads would not rise at all for small firms, as they do in the data. With the introduction of manager preferences, spreads do depend on the value of the firm, but in such a way that the dependence disappears as firms grow large but matters a great deal for the smallest firms.

I will now address why spreads decrease for the most valuable firms.

2.10.7 Partial Collusion with Impatient Firms

When underwriters are impatient, it will not in general be possible to sustain optimal collusion for all firm values. When $\beta > 1$, more valuable firms provide, on average, more profitable opportunities for collusion. Consequently, when a very valuable firm enters and underwriters are insufficiently patient, they will have an incentive to deviate from an equilibrium that calls for optimal collusion. Thus, to sustain an equilibrium, spreads must decrease for such firms so that a deviation will not be too profitable relative to expected future profits from maintaining collusion. This is, of course, an application of the result in Rotemberg and Saloner [1986].

When ε has distribution F and associated density f , and x has distribution G with density g over support $[\underline{x}, \bar{x}]$,²⁶ the optimal spread can now be expressed as follows:

Proposition 11. *When underwriters are impatient and the upper bound on the value*

²⁶The existence of a density function for either x or ε is not necessary but is assumed to simplify notation.

of firms is sufficiently large, the optimal collusive duopoly spread is given by

$$\alpha = \begin{cases} \alpha^* & \text{if } \Pi^m(x, \varepsilon) \leq \bar{\Pi} \\ \frac{\bar{\Pi} + \kappa}{\beta x} & \text{otherwise,} \end{cases}$$

where α^* is again the collusive spread schedule when firms are perfectly patient, $\Pi(x, \varepsilon)$ is the equilibrium profit accruing to the underwriter in a period with a firm of type (x, ε) , $\Pi^m(x, \varepsilon)$ is the profit that would accrue to an underwriter if he forces the firm to its participation constraint, and $\bar{\Pi}$ is the largest value satisfying the condition

$$\bar{\Pi} = \sum_{t=1}^{\infty} \delta^t \left(\int_{\underline{x}}^{\bar{x}} \int_{-\infty}^{\infty} \Pi(x, \varepsilon) f(\varepsilon) g(x) d\varepsilon dx \right).$$

The proof of this proposition is standard and is omitted.

The following parametric example highlights the implications of the results in this section. Let $x \sim U[0, 1000]$ and $\varepsilon \sim N(0, 1)$ and set $\beta = 1.075$ and $\kappa = 1$. Note that this β implies $\lim_{x \rightarrow \infty} \alpha = 1 - \frac{1}{\beta} = 0.07$, the seven percent spread pervasive for IPOs. Finally, assume that δ is such that $\bar{\Pi}$, the highest level of profits that can be sustained through collusion, is 50. We can now, for specific draws of x and ε , calculate the equilibrium spread. Note that, despite the uniform distribution of x , fewer IPO's will be observed at low values of x since small firms will not find IPO's valuable unless they get an improbably high draw of ε .

This pattern contrasts with the spreads that would be observed if the underwriting industry were competitive. In this case, patience or impatience would be irrelevant, and spreads would be given by:

Proposition 12. *Competitive underwriters charge spreads*

$$\alpha = \frac{\kappa}{\beta x}.$$

The above is an immediate consequence of the zero profit condition for competition, since it must be the case that $\alpha\beta x = \kappa$. Note that the presence of idiosyncratic manager preference is irrelevant and spreads smoothly decline toward zero for the most valuable firms. The conditional variance of the spread charged will, of course, be zero over the entire distribution of values.

2.10.8 Perfectly Observable Value and Unobservable Preference

The specific environment considered in this section is described formally in the following assumption:

Assumption 5.

1. x is common knowledge at the beginning of every period.
2. Both underwriters observe F , the unconditional distribution of ε , but receive no other signal about ε .
3. $\varepsilon \sim U[-\eta, \eta]$
4. Underwriters are sufficiently patient to sustain any degree of collusion.
5. Both underwriters are risk-neutral.
6. Underwriters cannot demand a spread greater than 1.

Since there is no information about ε contained in the signals to the underwriters, a stationary, symmetric, public pure strategy will be a function mapping signals to spread offers. And, for any strategy, the public history is sufficient to identify deviations. Therefore, the underwriters can collude on the monopolist spread. I now derive the spread function implied by the above assumptions and then discuss the implications.

Proposition 13. *The optimal collusive equilibrium spread with symmetric imperfect information is $\alpha^* = 1$ if $x \leq \frac{\eta+\kappa}{1+\beta}$. Otherwise, the optimal spread is given by*

$$\alpha^*(x) = \begin{cases} 1 - \frac{1}{\beta} + \frac{\eta}{\beta x} & \text{if } x \leq \frac{\kappa-\eta}{\beta-1} \\ \frac{1}{2} \left(1 - \frac{1}{\beta} + \frac{\eta+\kappa}{\beta x} \right) & \text{if } x \in \left[\frac{\kappa-\eta}{\beta-1}, \frac{\kappa+3\eta}{\beta-1} \right] \\ 1 - \frac{1}{\beta} - \frac{\eta}{\beta x} & \text{if } x \geq \frac{\kappa+3\eta}{\beta-1} \end{cases}$$

The proof is by standard optimization and is omitted.

The first thing to observe is that, unsurprisingly, $\alpha^*(x) \rightarrow 1 - \frac{1}{\beta}$ as $x \rightarrow \infty$ since the constraint set collapses to $1 - \frac{1}{\beta}$. This occurs because, as x grows large, the idiosyncratic element of preferences for IPO's becomes relatively unimportant; a firm will accept an offer of a spread slightly below $1 - \frac{1}{\beta}$ with probability increasing toward 1. While the assumption of bounded support for ε makes this effect particularly dramatic, the

intuition will hold for virtually any distribution of ε independent of x . Even if mass is concentrated in the negative tails, indicating that most entrepreneurs prefer to keep their firms private, this mass will eventually constrict into a tight region around 0 as x grows and financial benefits become the overwhelming concern.

Furthermore, when costs are relatively small, α is declining in value for small firms. This result holds even if the costs of holding an IPO are zero. That is, the upward pressure on price as firms get very small does not result entirely, or necessarily at all, from the need to cover costs.

2.10.9 Proof of Proposition 5

Proof. The proof is by construction. Choose some $x^* \in (0, \bar{x})$, where there is some α such that the profits accruing to the underwriter of type x^* for pooling with all types $x < x^*$ are positive at α (such an x^*, α pair must exist by the assumption that there exists some profitable rigid spread). Now, choose α^h to maximize profits of type x^* conditional on pooling with the lower types. The profit accruing to type x^* for pooling with the types higher than x^* is clearly greater than the profit for pooling with low types at α^h , continuous in α , and reaches a minimum that is less than zero as $\alpha \rightarrow 0$. Thus, there is some α^l such that type x^* is indifferent between pooling with the low types at α^h and pooling with the high types at α^l .

Since α^h is chosen as the maximum for x^* , we know that the difference between the value of adhering and the value of deviating increases as ξ decreases away from x^* . So, types below x^* do not have an incentive to deviate. The same argument shows that types above x^* do not have an incentive to imitate a type below x^* .

□

2.10.10 Flexible Spread Upper Bound

The procedure used to find an upper bound on the set of symmetric perfect public equilibrium payoffs is described in greater detail here.

Define

$$R(\rho, \alpha) = \int_0^{\bar{x}} R(\rho, x, \alpha) \frac{1}{x} dx$$

$$R(\rho, x, \alpha) = \int_0^{\bar{x}} \rho(x, y) dF^{FOS}(y)$$

$$\rho(x, y) = P(\text{reversion} | x, y = \max \alpha^{-1}(\alpha_i)).$$

That is, $R(\rho, \alpha)$ is the unconditional reversion probability given than the game reverts to repeated play of the worst one-shot equilibrium strategy with probability $\rho(x, y)$ when the public signal reveals that the true value of the firm is x and the private signal implied by the lowest bid under the spread schedule α is y , assuming that both players use spread schedule α .

Let \mathfrak{I} be the set of closed intervals on the real line. And, let \mathfrak{C} be the family of incentive constraints:

$$\begin{aligned} & (1 - \delta)\pi(w, w) + \delta E_{z,y}[\rho(z, y)\Pi^{SG} + (1 - \rho(z, y))v | \xi_i = w, \alpha, \rho, m_i = \xi_i] \\ \geq & (1 - \delta)\pi(w, w') + \delta E_{z,y}[\rho(z, y)\Pi^{SG} + (1 - \rho(z, y))v | \xi_i = w, \alpha, \rho, m_i = w'] \end{aligned}$$

for all w and $w' > w$, where $m_i = \alpha^{-1}(\alpha_i)$; that is, m_i is the implicit report of the signal received by underwriter i when underwriters are both using the spread schedule α . Here,

$$\pi(w, w') = E_q[\mathbf{1}_{\{\xi_j < w'\}^*}(\alpha(w')\beta q - \kappa)e^{-\lambda q(\alpha(w')\beta + 1 - \beta)} | \xi_i = w, \alpha]$$

where

$$\mathbf{1}_{\{a < b\}^*} = \begin{cases} 1 & \text{if } a < b \\ \frac{1}{2} & \text{if } a = b \\ 0 & \text{if } a > b \end{cases}.$$

These incentive constraints are simply the requirement that an underwriter finds it optimal to truthfully reveal his signal through his bid rather than attempting to capture additional market share by decreasing the spread he demands.²⁷

Define $\mathfrak{B} : \mathfrak{I} \rightarrow \mathfrak{I}$ such that

$$\mathfrak{B}([a, b]) = [a, b'],$$

with

$$b' = \max_{\alpha, \rho, v^h, v^l} \pi(\alpha) - \frac{\delta}{1 - \delta} R(\rho, \alpha)[v^h - v^l]$$

subject to \mathfrak{C} and $v^h, v^l \in [a, b]$. The expression for b' is the maximized present value

²⁷Constraints preventing the underwriter from reporting a signal higher than that which he received will not bind in the problem of finding the maximum SPPE payoff and so are ignored.

of the game and is the direct analogy of the expression derived in Abreu et al. [1991]. Let Π^{SG} be the per-period unconditional expected payoff from repeated play of the symmetric stage game equilibrium with the lowest payoff, and let v^* be the maximum of the set of symmetric perfect public equilibrium payoffs. We will apply the following proposition to derive an upper bound on the set of symmetric perfect public equilibrium payoffs:

Proposition 14. \mathfrak{B} is a contraction over the domain $\{[a, b] | a \leq \Pi^{SG}, b \geq v^*\}$.

Proof. Recall

$$\begin{aligned} R(\rho) &= \int_0^{\bar{x}} R(\rho, x) \frac{1}{x} dx \\ R(x, \rho) &= \int_0^{\bar{x}} \rho(x, y) dF^{FOS}(y) \\ \rho(x, y) &= P(\text{reversion} | x, y = \max \alpha^{-1}(\alpha_i)). \end{aligned}$$

Let $R^*(\alpha)$ be the minimal unconditional reversion probability required to maintain incentive compatibility for all signals when both underwriters use spread schedule α .

Let α^* represent the spread schedule used initially in the optimal symmetric perfect public equilibrium.

By lemma 1 below, we know

$$v^* = (1 - \delta)\pi(\alpha^*) + \delta[R^*(\alpha^*)\Pi^{SG} + (1 - R^*(\alpha^*))v^*]$$

so

$$v^* = \pi(\alpha) - \frac{\delta}{1 - \delta} R^\alpha(\alpha^*)[v^* - \Pi^{SG}].$$

Thus if $a = \Pi^{SG}$ and $b = v^*$, a solution to

$$\max_{\alpha, \rho, v^h, v^l} \pi(\alpha) - \frac{\delta}{1 - \delta} R(\alpha)[v^h - v^l]$$

subject to \mathfrak{C} and $v^h, v^l \in [a, b]$, is α^*, ρ^*, v^* , and Π^{SG} .

If $a < \Pi^{SG}$ and/or $b > v^*$, the problem is exactly as above with one or two constraints relaxed.

Thus, $[a, v^*] \subseteq \mathfrak{B}([a, b])$. □

Lemma 1. Any symmetric perfect public equilibrium payoff can be supported with a strategy that uses only Π^{SG} and v^* as promised continuation values.

Proof. The existence of a public correlating device permits the construction of an absolutely continuous public signal for any discretized version of the game.²⁸ Therefore, any discrete approximation to the game will satisfy the requirements to apply the bang-bang result of Abreu et al. [1990]. Since we are considering arbitrarily fine discrete approximations to the continuous game, this is sufficient to apply the bang-bang result. \square

Having established that \mathfrak{B} is a monotone operator, it is straightforward to find an upper bound on the SPPE payoffs for a strictly decreasing spread function. By choosing a lower bound below Π^{SG} , in this case 0, and applying \mathfrak{B} to $[0, \bar{v}^0]$, we find a new upper bound \bar{v}^1 . This procedure can be repeated to find a (weakly) decreasing sequence of upper bounds. The maximization procedure is somewhat difficult because it is necessary to find α and ρ simultaneously. Since the game must be solved as the limit of a sequence of discrete approximations, this requires maximization over $3 * K + 2$ variables, where K is the number of grid points for the given discretization.²⁹ Fortunately, in the implementation studied a relatively coarse grid appears to give a good approximation for the continuous game, at least for the purposes of finding the upper bound. See figure 2.7 for a plot of the calculated upper bounds as a function of the fineness of the grid; the value appears to reach a maximum around 30 grid points, indicating that further increases in the number of grid points used would not lead to a significantly higher upper bound on the value of a fully separating spread.

²⁸The procedure is similar to the procedure when signals are drawn from a finite space. See, for example, Mailath and Samuelson [2006], remark 2.5.1.

²⁹The procedure used here takes into consideration that ties can occur when one or both signals are incorrect by chance. The fact that the spread schedule in the case where spreads are restricted to the rationals is a function mapping a continuum of signals into a countable set is then not a cause for concern.

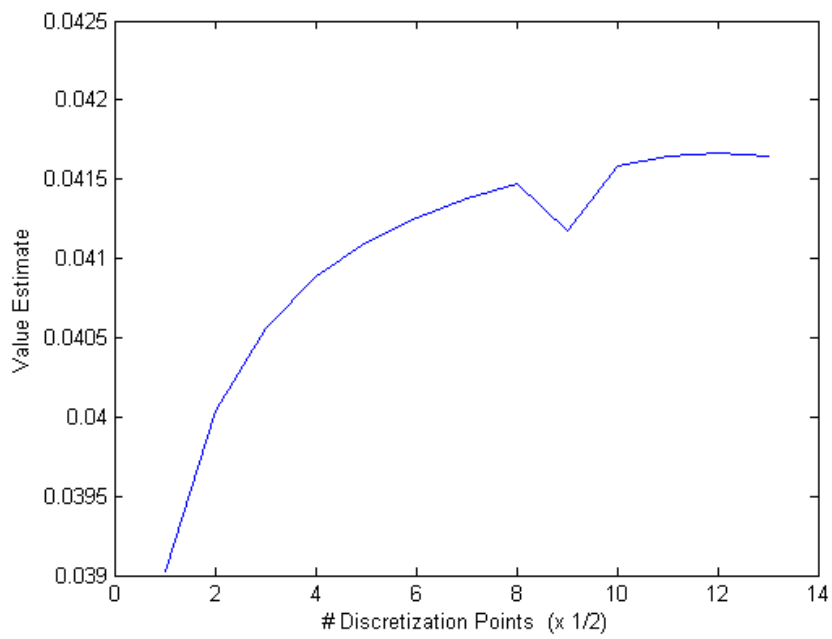


Figure 2.7: Calculated Upper Bound as a Function of Grid Fineness

Chapter 3

Social Capital as Economic Overlap

3.1 Introduction

People locate close to each other when engaging in economic activity. Cities form, and retailers congregate together. Such arrangements impose costs associated with crowding. The conventional explanation for the benefits that outweigh these costs focuses on some form of increasing returns to scale technology such as pure public goods. But, such a technology would in most cases require cooperative behavior or formal institutions to exploit. Of course, cooperation and formal institutions themselves may require pre-existing social and political structures, which would themselves have to arise from cooperative interactions among individuals. As such, a model in which agglomeration and cooperation arise simultaneously would prove useful for analyzing such situations.

I propose a model where two agents must choose a location and also must determine whether to behave cooperatively with the other agent. The model admits two types of costs associated with crowding. First, by crowding together agents leave resources and opportunities in other locations unexploited. Second, crowding together will inevitably lead to externalities between the agents, implying that each agent will no longer choose to act in the most efficient manner.

The model also includes an exogenously given opportunity to cooperate on an increasing returns to scale project that will, if successful, increase the value of the economic opportunities available. The project can be thought of as a mutually beneficial capital improvement that requires cooperation between both agents. But, this project takes the form of a work-shirk game summarized as a prisoner's dilemma, as is standard in the literature on governance in the absence of formal institutions (see Greif [2006], Dixit [2003], and Dixit [2004]). By locating close together and endogenously generating public goods, agents can effectively commit to cooperate in the work-shirk game by linking their welfare with that of their neighbor. This link leads to more efficient cooperation on projects where agents have incentives to cheat each other. In this sense, economic overlap generated by crowding serves as a form of social capital by facilitating cooperative behavior in situations where isolated individual agents would find such cooperation impossible. Location choice can be thought of as determining how "public" the economy is. More public economies lead to inefficient allocation of resources but can facilitate cooperation on joint projects by forcing agents to care about the incentives for other agents to contribute to the public good.

I show that location choice will be efficient in the sense that the agents will only crowd together when the costs of such crowding is outweighed by the benefit of in-

creased cooperation, and that when overlap is efficient the location decision leads to the least costly degree of overlap. This efficiency result is robust to relevant alternative solution concepts. I show by example, however, that in games with more than two agents efficient location decisions will not be stable and thus agents may end up locating in an inefficient manner. I also describe how the model can explain differences in geographic concentration and industry structure in financial services and explore the implications for the interaction between formal institutions and social capital.

By identifying an element of the social and economic structure that facilitates cooperation, this paper contributes to the literature on social capital. Similar attempts to define this concept and apply it to solving cooperation problems include Routledge and von Amsberg [2003], where social capital facilitates cooperation by generating repeated interactions, and Fryer [2006], where group specific investment signals the probability of repeated interaction. Both of these approaches rely on folk theorems to model cooperative behavior, whereas my model explains cooperation without relying on solution concepts that require repetition of the game and patient agents.

Since social capital here is defined as the degree of overlap between agents and crowding carries costs, social capital “investment” is costly as in Glaser et al. [2002]. But, social capital differs from other forms of capital, in particular human capital, in that it must arise from other elements of interaction rather than through direct investment. My model captures the fact that people cannot directly invest in social capital but can alter their behavior in a way that leads to optimal accumulation of this resource. The tradeoffs under consideration are consequently richer and more complicated than in standard capital accumulation models.

Other papers have explored the link between social capital and the provision of public goods. This research, however, focuses primarily on the role of social capital in facilitating the provision of public goods. Higher social capital is generally considered to lead to higher public good provision. See, for example Knack and Keefer [1997] and, for a more general discussion of social capital Putnam [2000]. The model of van Dijk and van Winden [1997] allows the level of public good to feed back into the level of social capital, but through a different mechanism and again in a repeated setting.

The role of public goods in providing social capital is a consequence of “transfer invariance” in public goods economies, as described first by Warr [1983] and given a more complete treatment in Bergstrom et al. [1986]. Small changes in the distribution of wealth in a public goods economy will not alter the level of provision or the distribution of utility when all agents choose a positive contribution to the public good. In my

model, agents exploit this property to locate so that they will not have an incentive to behave opportunistically when attempting to cooperate.

A more common approach to the interaction between public goods and location decisions posits the existence of local governments that serve as providers of public goods. Tiebout [1956] is the seminal paper in this field. Epple et al. [2001] is a current example of work in this framework. My paper, in contrast, focuses on the situations in which no government exists to provide public goods. It can thus apply to economic environments with underdeveloped political institutions or where cooperation is valuable along dimensions that governments do not encourage, such as collusive pricing.

The analysis in this paper focuses on the two agent case. I conclude, however, by introducing a three-agent example that suggests what characteristics of the equilibria might be robust to a more complicated environment. In this vein, the most closely related papers include Goyal [2005], who considers the role of strong and weak links in networks, and Bramouille and Kranton [2007], who consider the provision, though not the creation, of local public goods in networks and the stability of such networks. Additionally, Johnson and Gilles [2000] consider a spatial cost topology in in context of social networks and social capital that addresses similar questions to those raised when attempting to extend the model herein to a larger network.

The paper proceeds as follows. In the next section, I present and solve a parametric example of the model. This example captures the essential elements of the environment and demonstrates the conditions necessary for agents to reach a cooperative equilibrium. Then, I consider a general version of the two-agent model and show that the results in the example are robust to the dimension of the space and the functional form of the production function. I demonstrate the existence of an equilibrium in this two-agent game and provide properties of the equilibrium. Following this, I consider how the model can be applied to understand differences in the concentration of different types of financial services firms, and I explore how the model predicts that social capital will interact with formal institutions. Finally, I present an example that highlights important characteristics of the game with more than two agents.

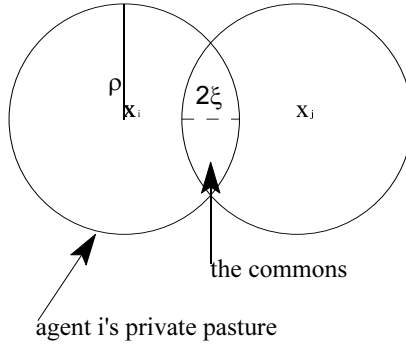


Figure 3.1: Pastures in \mathbb{R}^2

3.2 Model

There are two agents, i and j . They sequentially choose a location, $x_i \in \mathbb{R}^n$ and $x_j \in \mathbb{R}^n$. Each agent can engage in economic activity within a radius of ρ around his location, and each point in \mathbb{R}^n is identical for the purpose of economic activity. Agents may overlap in the sense that $|x_i - x_j| < 2\rho$ is permitted. Thus, the second agent's arrival decision can be summarized as choosing the degree of overlap with the first agent, $\xi \in [0, \rho]$.

For example, the agents may be shepherds choosing a location on a uniform plain around which to allow their sheep to graze. If $|x_i - x_j| < 2\rho$, where $|\cdot|$ is Euclidean distance, then the shepherds' plots overlap. This area can be thought of as an agricultural commons. We will exploit this example for expositional purposes to provide a more concrete sense of the meaning of the various elements of the model. Alternative interpretations include stores advertising in a local area around their business or investment firms searching for investment opportunities near their headquarters. This second example will be explored in more detail later.

Effort is required to make the land useful. The value of the flock of sheep is an increasing, concave function of the quality of its grazing land, while the quality of the grazing land is linear in the effort applied to it by the shepherd. Assume without loss of generality that the slope of this function is one and the intercept is zero. Each shepherd is endowed with an initial amount of effort E to spread over his land. Specifically, each shepherd $k \in \{i, j\}$ chooses a function

$$e_k : \bar{B}_\rho(x_k) \mapsto \mathbb{R}^+$$

where $\overline{B}_\rho(x_k)$ is the closed ball of radius ρ around agent k 's location, e_k is integrable with respect to the Lebesgue measure, and

$$\int_{\overline{B}_\rho(x_k)} e_k d\mu \leq E_k$$

where x represents coordinates on the plain, and μ is Lebesgue measure on \mathbb{R}^n . Thus, effort does not have to be applied uniformly over the area that an agent controls.

The value of the land when only one agent is on the plane can then be represented by the function

$$V(e_k) = \int_{\overline{B}_\rho(x_k)} v(e_k) d\mu$$

where $v(\cdot)$ is increasing and concave. Additionally, let $v(0) = 0$ so that a free disposal condition is satisfied; any subset of the area under control can be completely ignored without a direct cost to the agent. If this condition did not hold, less land might be preferred to more for a fixed level of effort. This assumption guarantees that total production will increase in the aggregate area under control of the agents.

If the shepherds choose non-overlapping plots, they will maximize $V(e_k)$ subject to the constraint

$$\int_{\overline{B}_\rho(x_k)} e_k d\mu \leq E_k.$$

Now, consider the case in which plots can overlap. Specifically, agents can locate sufficiently close to each other that part of their pastures are now used by both flocks (see Figure 3.1). Thus, the value of that land, which will be a function of the total effort expended on it by the two shepherds, will be divided evenly between them.¹ Considering only agent i , the value of his pasture is now

$$V_i(e) = \int_{\overline{B}_\rho(x_k) \setminus A} v(e_i) d\mu + \frac{1}{2} \int_A v(e_i + e_j) d\mu,$$

where A is the overlapping portion of the pastures and e_k is the effort function of each agent. Crowding on the shared area is complete; whatever benefit one agent

¹Adding an additional cost to overlap to represent fighting over shared product would not qualitatively change the predictions of the model. Also, by having the payoffs to the commons multiplied by something more than $\frac{1}{2}$ but less than 1, the model would admit the situation where crowding in the public good is less than complete. This will encourage greater provision of the public good and make generating cooperation less costly, but would again not fundamentally change the effects under consideration.

receives cannot then accrue to the other agent.² Recall that the effort function may take different values at different points on the plane. Specifically, $e_k(x) \neq e_k(y)$ in general when $x \in \overline{B}_\rho(x_k)$ and $y \in A$.

An exogenously specified opportunity for cooperation occurs after the location decision stage and before the effort allocation stage. This consists of a prisoners' dilemma that pays off in an effort bonus, *not* directly in utility. For concreteness, think of this as some sort of mutual capital investment, with the standard interpretation that cheating hurts the overall output of the project but is beneficial to the cheater. A work-shirk game with increasing returns to scale would fit into this paradigm. Thus, the shepherds play

	C	D
C	β, β	θ, η
D	η, θ	$0, 0$

where $\eta > \beta > 0 > \theta$, $\eta + \theta < 2\beta$.

To summarize, the game consists of three stages. In stage 1, agent 2 chooses $\xi \in [0, \rho]$ (with agent 1's decision irrelevant). In stage 2, each agent's action space is given by $\{C, D\}$. Finally, in stage 3, each agent chooses $e_k \in \{f \mid \int_{B_k} f d\mu \leq E_k\}$, where $E_k = \omega + \mathfrak{D}_k$ and \mathfrak{D}_k is determined by the actions in the second stage, with ω representing the initial effort endowment. Payoffs are $V_k = \int_{B_k \setminus A} v(e_k) d\mu + \frac{1}{2} \int_A v(e_k + e_{-k}) d\mu$.

3.3 Example

This section considers a parameterized version of the model on the real line, with $v(e) = e^\alpha$, where $\alpha \in (0, 1)$. The game is solved by backward induction. We find a minimal threshold ξ^* for overlap that can lead to cooperation in the prisoner's dilemma stage. We show how this threshold depends on the parameters η and α , which summarize the value of cheating in the prisoner's dilemma and the degree of decreasing returns to scale in the production function, respectively. We also describe how the two costs of crowding interact to determine the optimal degree of overlap.

²This type of good is sometimes referred to as an impure public good or a publicly provided private good (Epple and Romano [1996], for example). In the context of my model, there is no distinction between a public good with crowding and a pure public good that is half as productive as the private good.

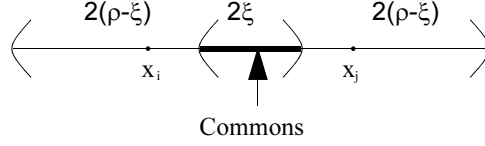


Figure 3.2: Pastures in \mathbb{R}

This example demonstrates the basic intuition of why agents will choose their location efficiently in equilibrium. By varying α , we can also see how efficient locations can involve no overlap where the economy is effectively completely private, complete overlap where all production occurs as a public good, and partial overlap where the economy is characterized by a public good and two private goods. We show that agents will choose the efficient location in equilibrium.

The equilibrium of the model is summarized by an overlap decision ξ and two effort allocation functions. It is straightforward to verify that the concavity of the production function implies that the equilibrium allocation of effort over each agent’s private land will be uniform³ and that combined effort over the “commons” will also be uniform. We can therefore without loss of generality consider only equilibria in which each agent chooses a simple function to allocate effort. That is, each agent chooses one level of effort for his “private land” and one level of effort over the commons. Thus, we can represent the action choices in the effort allocation phase as e_p^k for agent k ’s effort level over his private area, $\forall k \in \{i, j\}$, and e_c^k for agent k ’s effort over the commons area.

We first consider the final stage of the game, where each agent simultaneously chooses his effort allocation, taking his total available effort as given. Agent i solves in this effort allocation stage

$$\max_{e_p^i, e_c^i} 2(\rho - \xi)(e_p^i)^\alpha + \xi(e_c^i + e_c^j)^\alpha$$

³Technically, all statements about the effort allocation function are statements about how the function behaves except on a set of measure zero; we ignore the multiplicity of equilibria generated by admitting measure-zero differences in actions.

$$\text{s.t. } 2(\rho - \xi)e_p^i + 2\xi e_c^i.$$

This problem has first order conditions (where λ_i is the Lagrange multiplier)

$$2(\rho - \xi)\alpha(e_p^i)^{\alpha-1} - \lambda_i 2(\rho - \xi) = 0 \quad (3.1)$$

$$\xi\alpha(e_c^i + e_c^j)^{\alpha-1} - \lambda_i 2\xi = 0 \quad (3.2)$$

$$2(\rho - \xi)e_p^i + 2\xi e_c^i = E_i. \quad (3.3)$$

Since in any equilibrium $E_i = E_j \equiv E$ (because asymmetric actions in the prisoner's dilemma stage cannot occur in equilibrium), the equilibrium effort functions of both agents must be identical. This gives the following expression for the effort functions:

$$e_p = \frac{2^{\frac{1}{1-\alpha}} E}{(\rho - \xi) 2^{\frac{2-\alpha}{1-\alpha}} + \xi} \quad (3.4)$$

$$e_c = \frac{E}{2 \left[(\rho - \xi) 2^{\frac{2-\alpha}{1-\alpha}} + \xi \right]} \quad (3.5)$$

$$e_c^T = \frac{E}{(\rho - \xi) 2^{\frac{2-\alpha}{1-\alpha}} + \xi} \quad (3.6)$$

where e_c^T represents the combined effort of the two agents at each point in the commons. Thus, defining $a = 2^{\frac{1}{1-\alpha}}$, $b = 2^{\frac{2-\alpha}{1-\alpha}}$, the value of the game to one agent conditional on symmetric actions in the prisoner's dilemma and a fixed overlap can be given as

$$V(E, \xi) = 2(\rho - \xi) \left(\frac{aE}{(\rho - \xi)b + \xi} \right)^\alpha + \xi \left(\frac{E}{(\rho - \xi)b + \xi} \right)^\alpha$$

or

$$V(E, \xi) = \left(\frac{E}{b\rho + \xi(1-b)} \right)^\alpha (a\rho + \xi(1-a)). \quad (3.7)$$

It will become clear later that the payoffs along any equilibrium path can be represented in this form, with ξ and E appropriately chosen. It will also be convenient to refer to the following functions, which give the value accruing to an agent following a particular overlap decision and a particular set of actions in the prisoner's dilemma stage, assuming equilibrium play in the effort allocation stage:

- Let $V(C, C, \xi)$ be the value accruing to one agent following overlap of ξ and mutual cooperation in the prisoner's dilemma.

- Let $V_D(D, C, \xi)$ be the value accruing to an agent who plays D against an agent who plays C , given overlap ξ .
- Define $V(D, D, \xi)$ and $V_C(D, C, \xi)$ analogously.

In general, it will be the case that $V_D(D, C, \xi)$ is decreasing in ξ , while $V(C, C, \xi)$ is initially decreasing but may then increase for ξ close to ρ . $V_C(D, C, \xi)$ will, of course, increase in ξ . The behavior of these functions is best understood by in terms of the costs of crowding.

3.3.1 Costs of Crowding

These costs consist of two separate components: crowding together reduces the total area available for use, which is costly since the production function is concave in effort. I refer to this cost as the area-under-cultivation (AuC) loss. Demange [2005] identifies a similar tradoff between increasing returns to scale and a desire for variety in the context of cooperative club formation.

The second cost associated with crowding is the distortion away from efficient effort allocation that results from introducing public goods into the economy. I shall refer to this cost as the tragedy-of-the-commons (ToC) loss. The decomposition is useful since the behavior of the two inefficiencies exhibit important differences. To fully address the first problem, agents must locate on disjoint plots, while to solve the second problem agents can either locate on disjoint plots or locate on top of each other. More generally, the AuC-inefficiency is monotonically increasing in ξ , while the ToC-inefficiency reaches a global maximum at some $\xi \in (0, \rho)$.⁴

To measure the AuC and ToC effects, use as a benchmark the socially optimal value of the game for one agent when $\xi = 0$ and the cooperation decisions in the prisoner's dilemma are fixed. Then, as ξ increases from zero, the change in the (egalitarian) socially optimal payoff for an agent represents the AuC effect, while the difference between the new social optimum at any ξ and the equilibrium payoff (assuming the fixed cooperation decision) is the loss from the ToC. The AuC effect measures the change in the technological feasibility constraint when agents overlap, while the ToC effect measures the costs from the creation of a public good and the resulting externalities on effort. Appendix 3.8.9 demonstrates this decomposition for the parametric case

⁴This property follows almost immediately from the continuity of the ToC.

considered in this section. The relative importance of these two costs will play an important role in determining when cooperation will be maintained.

3.3.2 Maintaining Cooperation

Two approaches are necessary to determine when agents can maintain cooperation. First, taking ξ as given, note that if the solution to the optimal effort function (the final subgame) for both agents is interior in the sense that $e_c^i > 0$ for $i \in \{1, 2\}$ regardless of the actions taken in the prisoner's dilemma stage, then cooperation will always be sustained. This follows from the fact that the marginal product of investment in the private pasture must be equal for both agents in order for both agents to choose to invest in the commons; otherwise, one agent could improve his profit by reallocating effort either to or away from the commons.

To find the cutoff beyond which an agent who is “cheated” in the prisoners dilemma will refuse to invest in the commons, assume without loss of generality that agent i has deviated from the cooperative strategy. Now, assume that $e_c^j = 0$. Then, by the first order conditions of the maximization problem (equations 3.1 and 3.2),

$$e_c^i = \frac{1 + \eta}{2((\rho - \xi)a + \xi)}.$$

Now, note that the marginal product for an investment in the commons is

$$\xi \alpha (e_c^T)^{\alpha-1}.$$

So, in order for $e_c^j = 0$ to hold, it must be the case that the marginal product for investing in the private plot of the cheated agent must exceed the marginal product of investing in the commons. Thus, the requirement is that

$$\xi \alpha \left(\frac{1 + \eta}{2((\rho - \xi)a + \xi)} \right)^{\alpha-1} < 2(\rho - \xi) \alpha \left(\frac{1 + \theta}{2(\rho - \xi)} \right)^{\alpha-1}.$$

This reduces to the requirement that

$$\frac{1 + \eta}{1 + \theta} > \left(\frac{\xi}{\rho - \xi} \right)^{\frac{1}{1-\alpha}} \left(1 + \frac{1}{a} \left(\frac{\xi}{\rho - \xi} \right) \right).$$

This says, as would be expected, that a sufficiently large reward for deviating from

cooperation or a sufficiently large penalty for continuing to cooperate when the other agent deviates will result in the cheated agent finding it optimal to invest solely in his own plot. Note that setting $\theta = -1$, which implies that the cheated agent ends up with no effort to allocate at all, guarantees that he will not invest in the commons. The continuity of the right-hand side in θ then guarantees that, for any ρ and $\eta, \xi < \rho$, there exists some $\bar{\theta} > -1$ such that the cheated agent will choose not to invest in the commons. Finally, a large enough overlap will always guarantee an interior solution (and therefore cooperation) when $\theta > -1$.

The concavity of the production function also plays an important role here. As production becomes linear ($\alpha \rightarrow 1$), $\frac{1}{a} \rightarrow 0$ and $\left(\frac{\xi}{\rho-\xi}\right)^{\frac{1}{1-\alpha}} \rightarrow 0$. Thus the greater the degree of diminishing returns to effort (that is, the smaller the α), the easier it is to maintain cooperation by having non-trivial investment in the commons by both agents. Intuitively, this result follows from the fact that an agent who ends up having less effort to apply in total will tend to want to free ride off the investment in the commons of his wealthier neighbor. But, if the returns to investment in his own pasture diminish very quickly, he will not find it optimal to completely ignore the commons when the other agent is spreading his effort over his own private pasture and the commons. If, however, returns are roughly constant, the impoverished agent will not exhaust the returns in his own pasture before running out of effort.

The importance of the previous discussion follows from the fact that, if the optimal solution following a deviation from cooperation in the prisoner's dilemma is interior for all agents, then cooperation will always be sustained. This conclusion follows immediately from the fact that, when the game has an interior solution, all agents must have the same marginal utility of private land; the symmetry of the private plots then guarantees an egalitarian solution. And, since the total amount of effort available is higher under cooperation, it is clear that $V(C, C, \xi) > V_D(D, C, \xi) = V_C(D, C, \xi)$. That is, an agent is better off continuing to play C in the prisoner's dilemma stage instead of deviating to D .

The more interesting cases are those in which the condition for a corner solution holds. Again, such a solution can be guaranteed for all ξ by choosing $\theta = -1$, so the following discussion proceeds without considering the possibility of an interior solution following cheating.

Under this assumption, we find the value of the game to the agent who deviates

from cooperative play. Recall again that $e_c^j = 0 \Rightarrow$

$$e_c^i = \frac{1 + \eta}{2(a(\rho - \xi) + \xi)} \quad (3.8)$$

$$e_p^i = 2^{\frac{\alpha}{1-\alpha}} \frac{1 + \eta}{a(\rho - \xi) + \xi} \quad (3.9)$$

which then gives the value of the game to the deviator i as

$$V_D(D, C, \xi) = \left(\frac{1 + \eta}{2} \right)^\alpha [a\rho + \xi(1 - a)]^{1-\alpha}.$$

We know that

$$V(C, C, \xi) = \left(\frac{1 + \beta}{b\rho + \xi(1 - b)} \right)^\alpha [a\rho + \xi(1 - a)]$$

and that, in order to sustain cooperation in the prisoner's dilemma phase, it is necessary for $V(C, C, \xi) \geq V_D(D, C, \xi)$. This implies a cutoff for cooperation in terms of overlap of

$$\xi > \frac{b\rho \left(\frac{1+\eta}{1+\beta} - 1 \right)}{2 - \frac{1+\eta}{1+\beta} + b \left(\frac{1+\eta}{1+\beta} - 1 \right)}.$$

Since the expression on the right hand side is strictly positive (since $b \in (4, \infty)$), this implies, of course, that no overlap leads to uncooperative play. Furthermore, the right hand side will always be less than one, so $\xi = \rho$ implies that cooperation will always prevail when agents locate on top of each other. The threshold for ξ is increasing in η and decreasing in β , as would be expected. Defining ξ^* as the threshold for cooperation, observe that $\xi^* = \frac{b\rho}{\frac{1}{z} + b - 1}$ so $\frac{\partial \xi^*}{\partial z} = \frac{b\rho}{(1+z(b-1))^2}$, which is positive. Thus, ξ^* is increasing in z and therefore increasing in η and decreasing in β since $z > 0$ for all $\eta > \beta$.

The above threshold has all of the properties that one would expect; the stronger the prisoner's dilemma and the less concave the production function, the harder it is to sustain cooperation. To discuss this cutoff in a more concrete fashion, consider the parameterization (that is without loss of substantial generality) where $\rho = \frac{1}{2}$ and $\beta = 1$. Then, figure 3.3 shows the minimal value of ξ that can sustain cooperation when θ is small enough to guarantee a corner solution. Finally, we can note that the threshold is increasing in b , and therefore increasing in α , the curvature parameter.

This discussion establishes how play will proceed in any subgame following the location decision. Now, this information can be used to determine the optimal location. Again, the optimal location from the social perspective will coincide with the location

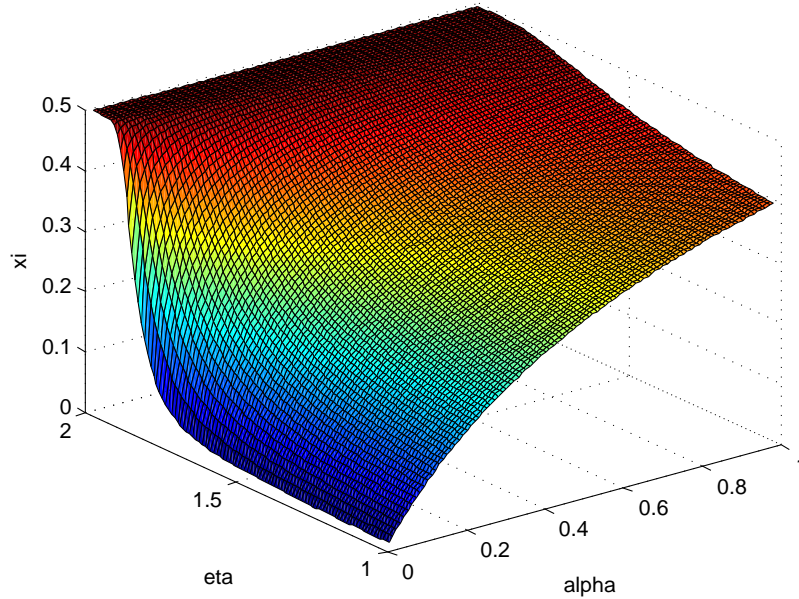


Figure 3.3: Minimal ξ to sustain cooperation

that the agents choose. The value of the game for each possible location appears in figure 3.4 (normalizing $\rho = 1$ and $\beta = 1$ and assuming $\alpha = \frac{1}{2}$ and $\eta = 1.5$). Thus, in this case, the unique Pareto-efficient equilibrium of the game will be to locate at the ξ threshold and cooperate in the prisoner's dilemma.⁵ Furthermore, it should be clear that there are only three possible ξ 's that can be candidates for efficient equilibrium for any set of parameter values. Agents will either locate at the threshold and enjoy the fruits of cooperation while minimizing the necessary but inefficient overlap, or they will locate on disjoint plots, thus guaranteeing that they will fail to cooperate in the prisoner's dilemma but eliminating the inefficiencies from the tragedy of the commons and the failure to maximize the area under cultivation, or they may locate exactly on top of each other, guaranteeing cooperation and eliminating the tragedy of the commons, but minimizing the area under cultivation.

Figure 3.5 presents two parameterizations that result in the first and the third situation: The intuition behind this result is as follows. If the agents are confronted with a relatively unimportant opportunity to cooperate (that is, $\beta \approx 0$) but where the gains from deviating while the other agent cooperates are large, then the overlap necessary to induce cooperation will be too costly to be justified. Agents will then choose

⁵Other equilibria with an interior overlap are possible; they all involve cooperation in the prisoner's dilemma, and an inefficiently large overlap. They are all eliminated by forward induction.

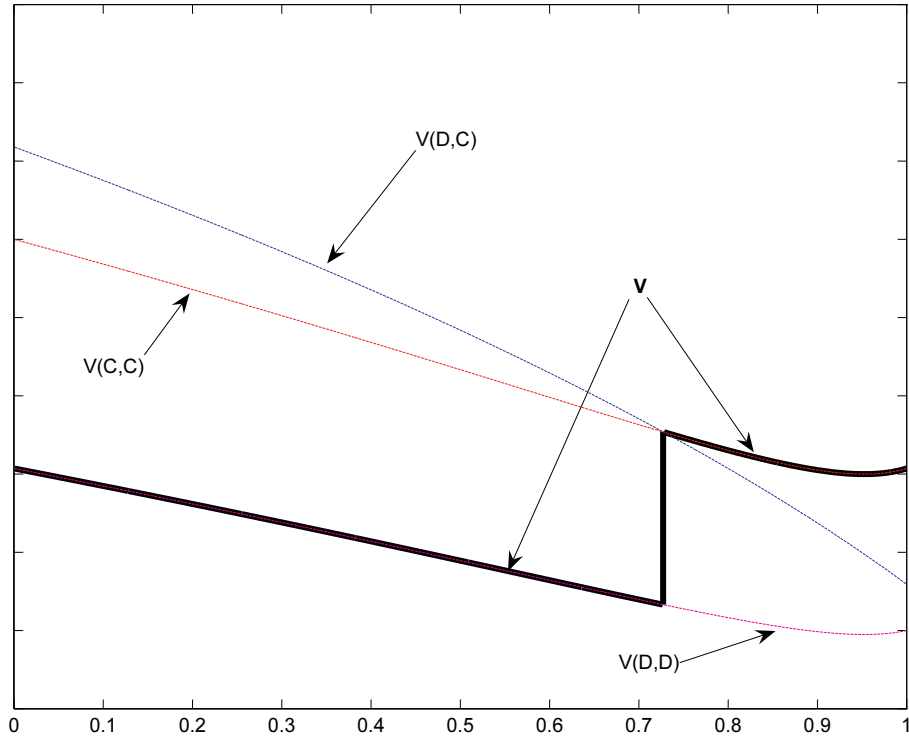


Figure 3.4: Value of the game as a function of ξ

zero overlap because, taking uncooperative behavior as given, the value of the game is decreasing in overlap in the relevant region due to the concavity of the production function and the tragedy of the commons. Adjusting the concavity of the production function changes what should be considered a small β and a large η . If, on the other hand, sustaining cooperation is worthwhile, the agents will often prefer the minimal degree of overlap in order to avoid the inefficiencies associated with reducing the area under cultivation and the tragedy of the commons. The inefficiencies associated with reducing the area under cultivation are clearly monotonically increasing in ξ for given behavior in the prisoner's dilemma. However, the tragedy of the commons is not monotonic, and indeed reaches a maximum at some interior point of $(0, \rho)$.⁶ Therefore, under certain circumstances, once agents have determined that sustaining cooperation is worthwhile they may choose to “go all the way” and just locate on top of each other. In particular, as $v(\cdot)$ becomes increasingly linear, the agents will tend to distort effort away from the commons even for very large common plots. And, the costs of moving

⁶This follows immediately from the continuity of the tragedy of the commons and the fact that any measure of this inefficiency would have to be zero at both $\xi = 0$ and $\xi = \rho$

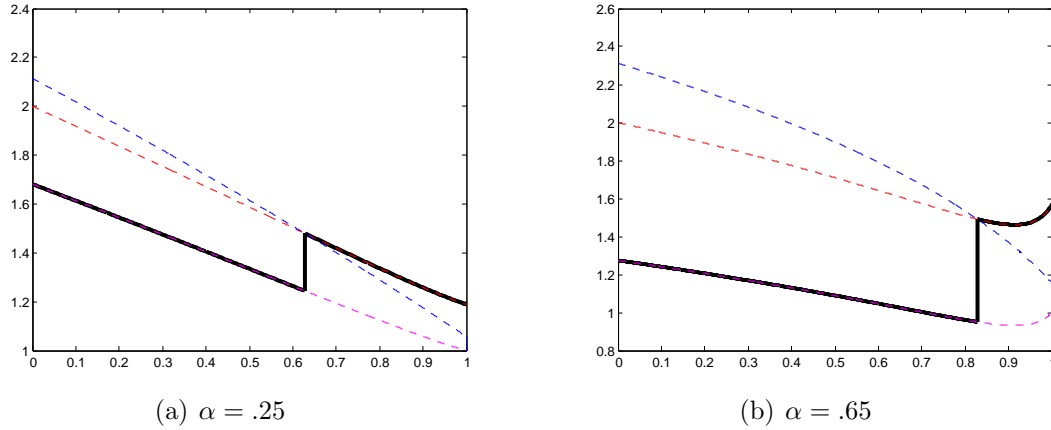


Figure 3.5: Corner Solutions for Overlap

more together to create a total overlap are small since the AuC loss is small when $v(\cdot)$ is almost a linear function. Thus, in addition to the role α plays in the minimal overlap necessary to sustain cooperation, α is also an important determinant of the qualitative characteristics of the equilibrium.

Note two potentially important features of this setup. First, multiple efficient equilibria are in fact possible, up to a maximum of three. However, generically it is impossible to have multiple equilibria; the value functions do not contain any regions over which the set of critical points have measure greater than zero. Second, the assumption that total labor effort is inelastic is essential for the argument that the tragedy of the commons is non-monotonic in ξ . If the model were to include a utility specification that includes leisure, the tragedy of the commons would persist even when agents locate on top of each other. For example, Schmitz [2002] documents that before the privatization of the land settlers at Jamestown were found bowling in the midst of a famine rather than planting crops on the land from which each settler received an equal portion of the crop. While the model here does not directly address this situation, it should be clear that the main result will still hold; the only important difference is that agents will now never find it optimal to locate on top of each other. To see that there could still be an interior equilibrium note that, faced with a choice between locating on disjoint plots or locating on top of each other, agents would still choose the latter for a sufficiently valuable prisoner’s dilemma. Thus, there still must be a non-trivial set of games for which equilibrium ξ is interior, even when agents can receive utility from leisure.

3.4 General Characteristics of Equilibria

This section considers the general model, where agents locate in \mathbb{R}^n and the production function $v(\cdot)$ is any strictly increasing, continuous, concave function. We confirm that the qualitative characteristics of the equilibrium described in the example hold for the general case, including generic uniqueness under forward induction and efficiency of the location choice. We also show that the sequential arrival game is equivalent to a simultaneous arrival game, indicating that location choice will be stable.

We first formally define certain objects necessary to consider the general model:

- Define the *aggregate plot* as $\overline{B}(x_i) \cup \overline{B}(x_j)$. That is, the aggregate plot is the total area used by either agent.
- Define the *commons* as $\overline{B}(x_i) \cap \overline{B}(x_j)$, the area where both agents produce and consume.
- Define the *private plot* of agent k as $\overline{B}(x_k) \cap (\overline{B}(x_{-k}))^c$, the area used by only agent k .

We also define the following efficiency concepts:

- A sequence of actions generates a *first-best outcome* if it maximizes total production over location decisions, play in the prisoner's dilemma phase, and effort allocation.
- A location decision generates a *second-best outcome* if it maximizes total production conditional on equilibrium play in the induced subgame.

We now collect some useful results that are immediate consequences of the concavity of $v(\cdot)$. The proofs are omitted to save space, and all statements should be understood to apply except on a set of measure zero:

- Proposition 15.** • *A first-best allocation requires uniform effort over the aggregate plot.*
- *Any equilibrium allocation will take the form of a simple function, where the steps correspond to the private plot and the commons.*

The above proposition applies also to any game with more than two agents where effort is allocated over local commons shared between subsets of agents. We therefore exploit this proposition to prove existence of equilibrium in the subgame for an arbitrary (finite) number of agents N :

Lemma 2. *Every effort allocation subgame has a pure strategy Nash equilibrium for $N \geq 1$.*

Since action choices in the first and second stage are finite, this is sufficient to prove the existence of an equilibrium for the general game. Furthermore, we can conclude that no equilibrium allocation will provide a first-best outcome, since first-best will require uniform effort allocation, cooperation in the prisoner's dilemma stage, and measure zero overlap.

We will now return to the case where $N = 2$ and establish the existence of a second best equilibrium, in the sense that the location decision will maximize productivity under the constraint that play in the second and third stages (the cooperation game and the effort allocation game) is in equilibrium. Furthermore, this efficient equilibrium is the only equilibrium to survive an intuitive forward induction refinement. Some simplification and little loss of generality is achieved by assuming, as we will for the remainder of the paper, that $\theta = -\omega$. That is, an agent who plays C against a D loses all of his effort endowment.

We establish the second-best efficiency result by first ruling out the existence of equilibria with asymmetric payoffs. Then, we show that this guarantees that the second-best efficient location decision, which must be either null overlap, complete overlap, or the minimal overlap necessary to sustain cooperation, must be an equilibrium of the game. We then characterize the inefficient equilibria and show that they are eliminated by forward induction since the location decision is a credible signal of which equilibrium is expected in the subgame starting in the cooperation stage.

As in the example, we define $V(C, C, \xi)$ as the per agent value of cooperation in the prisoner's dilemma following an overlap decision ξ , where effort allocation is in equilibrium in the subgame and $V(D, D, \xi)$ analogously. Similarly, $V_C(D, C, \xi)$ is again the value that accrues to the agent who plays C when play in the prisoner's dilemma is D, C , and $V_D(D, C, \xi)$ is the value accruing to the agent who plays D . V, V_C , and V_D will henceforth be referred to collectively as the value functions for the first stage.

The following single crossing condition, which is a consequence of the concavity of $v(\cdot)$, will be useful for establishing existence and uniqueness of the equilibrium. Two

value functions for the first stage are said to cross at some ξ if the values are equal at that ξ ; for example, V crosses V_D at ξ' if $V(C, C, \xi') - V_D(C, C, \xi') = 0$:

Proposition 16. $V(C, C, \cdot)$ and $V_D(D, C, \cdot)$ cross in ξ at most once.

An almost immediate corollary of proposition 16 is

Lemma 3.

$$\{\xi | V_D(D, C, \xi) \geq V(C, C, \xi)\} = \{\xi | \xi \leq \xi^*\}$$

for some $\xi^* \in [0, 1]$.

Now, it is useful to demonstrate that asymmetric play will never occur along the equilibrium path. The following lemma and proposition show this:

Lemma 4. $V_C(D, C, \xi)$ is increasing in ξ .

This lemma simply states the obvious that if an agent loses all of his endowment in the prisoner's dilemma phase, he is better off overlapping as much as possible with his newly wealthy "partner" since that agent will, by virtue of his larger endowment and higher wealth, provide more effort to the commons.

Proposition 17. For all η , \nexists an equilibrium in which D, C occurs along the equilibrium path.

The intuition for this is also straightforward. If an agent were to play C expecting his opponent to play D , he could increase his own endowment by more than he decreased his opponents by deviating to D and replicating the original allocation, with additional effort to spare.

Note that the above proposition does not rule out the possibility that D, C will be called for in some subgames. In fact, there exist games where such an outcome can be supported for a small region of ξ . This is somewhat surprising since in most cases an overlap small enough to make an agent prefer to refuse to cooperate against a cooperator will be too small to prevent a cooperator from wanting to deviate against a non-cooperator.

The preceding proposition is most useful for establishing the following result:

Proposition 18. The sequential arrival game is equivalent to the simultaneous arrival game for two agents when agents play pure strategies in the location phase.

Proof. Since play in the prisoner’s dilemma is always symmetric and the plots are symmetric, all payoffs will be symmetric. So, in a sequential arrival game, the optimal choice by the second agent will maximize the welfare of both players, subject to strategic constraints. Therefore, even if the first agent to arrive were given an opportunity to move, he would have no incentive to do so. Conclude that the sequential and simultaneous arrival games are payoff, cooperation, and location equivalent. \square

This proposition is crucial to establishing firm predictions about location in the model. Since sequential arrival is equivalent to simultaneous arrival, location decisions are stable in the sense that, following the arrival of the second agent, the first agent will not have an incentive to move before the start of the rest of the game. This finding does not hold for larger games, indicating that a simultaneous arrival game and a sequential arrival game will not have equivalent outcomes. As we will demonstrate by example in a later section, the efficient network of local public goods in a analogous game with more than two agents will not always be stable. This fact implies that agents will not necessarily reach a second best location configuration.

Proposition 19. *A pure strategy subgame perfect equilibrium exists.*

Following similar logic, we can see that the second-best location decision *will* be an equilibrium of the game. That is, if a social planner can control location decisions but is unable to enforce cooperation or effort allocations, then the social planner cannot improve upon decentralized location decisions. Thus, we have the following welfare proposition:

Proposition 20. *For any location choice, there exists a pure strategy subgame perfect equilibrium which provides a weak pareto-improvement over any fixed location.*

This result is immediate from the fact that players can choose the location that maximizes total productivity, and payoffs are symmetric in this case. Any asymmetric equilibrium of the exogenous location game will have to occur in the region where $V(D, C, \xi) > V(C, C, \xi)$, and thus the payoff for the agent playing D is worse than the payoff to each player at $V(C, C, \xi^*)$, where ξ^* is the threshold for cooperation.

3.4.1 Uniqueness

While the above proposition establishes that agents *can* in equilibrium reach the second-best optimal allocation, it does not guarantee that this equilibrium is unique,

or even that all equilibria are payoff equivalent. Indeed, inefficient equilibria do exist. We first characterize these equilibria and present a forward induction argument that selects only the efficient equilibria. These will be essentially unique in that, generically, the refined equilibrium identifies a single equilibrium overlap; this overlap can occur only at no overlap, complete overlap, or the minimal overlap necessary to sustain cooperation.

These inefficient equilibria fall into two categories; those that are inefficient because cooperation fails when it could have been sustained at reasonable cost, and those that are inefficient because the overlap is unnecessarily large. We consider these two types of equilibria in that order.

Proposition 21. • (a) Any equilibrium in which D, D is played must have $\xi = 0$.

- (b) Such an equilibrium exists if and only if $V(D, D, 0) \geq \max\{V(C, C, \rho), V(C, C, \xi^*)\}$, where ξ^* is such that $V(D, D, \xi^*) = V_C(D, C, \xi^*)$.

Proof. Part (a) just states that if agents will not cooperate, one agent will always have an incentive to move away so as to avoid the inefficiencies associated with crowding. Part (b) points out that an agent in a game that is being played uncooperatively can always force cooperation by deviating to complete overlap or overlap that makes C dominate D assuming equilibrium play in the final subgame.

(a) $V(D, D, 0) > V(D, D, \xi)$ for all $\xi \neq 0$, and $V_D(D, C, 0) > V(D, D, 0)$, so if equilibrium calls for (D, D, \cdot) and $\xi \neq 0$, $\xi = 0$ and (D, D) is a profitable deviation.

(b) If $V(D, D, 0) \geq \max\{V(C, C, \rho), V(C, C, \xi^*)\}$, assume equilibrium calls for D if $\xi < \xi^*$ and C otherwise. Then, all deviations from $(D, D), \xi = 0$ to $\xi < \xi^*$ cannot be profitable. But, deviations to ρ or ξ^* are not profitable by assumption, and since $V(C, C, \xi)$ does not have a local maximum, $\xi \in (\xi^*, \rho)$ is not profitable.

If $V(D, D, 0) < \max\{V(C, C, \rho), V(C, C, \xi^*)\}$, then a deviation to either ρ or $\xi^* + \varepsilon$ and C is profitable, since the other agent must play C when $\xi \in (\xi^*, \rho]$. \square

There are three types of equilibria in which C, C is played: those with $\xi = \rho$ (complete overlap), those with $\xi^* = \{\xi | V_D(D, C, \xi) = V(C, C, \xi)\}$, and those with

$$\xi' \in \{\{\xi | V(C, C, \xi) > V_D(D, C, \xi)\} \cap \{\xi | V_C(D, C, \xi) < V(D, D, \xi)\} \cap \{\xi | V(C, C, \rho) \leq V(C, C, \xi)\}\}.$$

The last set is the set of always inefficient equilibria with cooperation. They are sustained by the (subgame perfect) threat of playing D when ξ is too low, even if ξ could

sustain cooperation. The existence of these equilibria follows from the fact that, even when cooperation could be sustained, D is still a best response to D except possibly for very large ξ . In most cases, this set will not be empty unless $V(D, D, 0) \geq V(C, C, \xi^*)$ or $V(C, C, \rho) \geq V(C, C, \xi^*)$ (that is, unless the efficient equilibrium involves a corner solution for ξ). However, a simple forward induction refinement can eliminate these equilibria. Specifically, if an agent plays $\xi \in [\xi^*, \xi')$ as a deviation, he will receive, at most, $V(D, D, \xi) < V(D, D, 0)$. So, it must be the case that such an action is associated with the belief that the other agent will play C . Given that the other agent must be expected to play C , the agent's own best response is clearly C .⁷ Thus, by forward induction, only the most efficient interior equilibrium will survive. Following this refinement, the choice for location can be reduced to $\xi \in \{0, \xi^*, \rho\}$. The same forward induction argument eliminates $\xi = 0$ when $V(C, C, \xi^*) > V(D, D, 0)$, so the prediction of the model is unique as long as $V(C, C, \xi^*) \neq V(D, D, 0)$, $V(C, C, \xi) \neq V(C, C, \rho)$, and $V(C, C, \rho) \neq V(D, D, 0)$. But, clearly, this will not occur for generic parameter values (since $V(D, D, 0) = V(C, C, \xi^*)$, etc., would occur only for very specific values of parameters regardless of the functional form of $v(\cdot)$), so under the forward induction equilibrium refinement the equilibrium path is unique. In the absence of the forward induction argument, a continuum of Pareto-ranked interior equilibria may exist. To summarize, the choice of ξ determines whether the second stage game will be a prisoner's dilemma, a coordination game, or a game with a unique, dominance solvable equilibrium that is efficient. When the game is a coordination game, agents can credibly signal their intentions, and thus coordinate on the efficient equilibrium.

We will now return to the decomposition of the costs of crowding into the tragedy-of-the-commons effect (ToC) and the area-under-cultivation effect (AuC). Considering these effects separately allows us to show that agents will locate at the minimum threshold necessary to sustain cooperation if they choose any cooperation at all.

We first argue that the ToC effect single peaked. That is, there is no tragedy of the commons when agents are completely separate or completely overlapped and the ToC loss has a unique local maximum for $\xi \in (0, \rho)$. This also implies that there is no local minimum for the ToC effect. This result, combined with the monotonicity of of the AuC effect, guarantees that the optimal cooperative equilibrium will involve either minimal necessary overlap or total overlap. As such, comparative statics on minimal overlap are equivalent to comparative statics on optimal location, with one

⁷Note that while C is not a best response to C in the regular prisoner's dilemma, it may be a best response in this case because of how play will occur in equilibrium in the final subgame.

only needing to verify that the optimal location does not jump to complete or null overlap.

Proposition 22. *The rate of change from the ToC loss is monotonic after reaching its local maximum.*

This proposition and the following lemma are sufficient to prove that the symmetric payoff functions $V(C, C, \xi)$ and $V(D, D, \xi)$ do not have local maxima.

Lemma 5. *The loss from the AuC effect is increasing and convex in ξ .*

Lemma 6. *$V(C, C, \xi)$ and $V(D, D, \xi)$ do not have local maxima for $\xi \in (0, \rho)$.*

Proof. Recall that the loss from overlap can be decomposed into the AuC and the ToC effect. The AuC effect is increasing and convex. The ToC effect is, by construction, zero at $\xi = 0$ and zero at $\xi = \rho$. It is also strictly positive in the interior. Since $[0, \rho]$ is compact and the ToC effect is continuous, the loss from ToC achieves a maximum somewhere in $[0, \rho]$; in fact, we can conclude that the ToC reaches a maximum in the interior of this set since the ToC clearly is positive for all overlap where $\xi \in (0, \rho)$. At any critical point, we know that the rate of change in the AuC and the rate of change in the ToC must be equal in absolute value, and the loss from the ToC must be decreasing. Now, as ξ shrinks from the critical point, the rate of change in the ToC loss must fall relatively fast since it reaches 0 in the interior of $[0, \rho]$. The AuC loss, however, does not reach zero until $\xi = 0$ at the earliest, since it is increasing and convex. Thus, unless the rate of change of the ToC loss is non-monotonic between the critical point and the local maximum, there can be at most one critical point. Since we know $V(C, C, 0) > V(C, C, \rho)$, this critical point, if it exists, must be a local minimum. \square

The above assumption and lemma guarantee that the values from symmetric play are initially decreasing (this is trivial to prove but omitted for brevity) and may achieve only one local minimum. This guarantees that, if an interior overlap is optimal, the minimal interior overlap is optimal since, when the minimal overlap is designated by ξ^* , we can conclude $V(C, C, \xi) < \max\{V(C, C, \xi^*), V(C, C, \rho)\}$ for all $\xi \in (\xi^*, \rho)$.

We can now introduce some simple comparative statics results that confirm the intuition about the tradeoff between overlap and cooperation. Define the minimal overlap necessary to sustain cooperation as ξ^* :

Proposition 23.

$$\frac{d\xi^*}{d\eta} > 0.$$

and

$$\frac{d\xi^*}{d\beta} < 0.$$

These follow from the fact that $V(C, C, \xi)$ is invariant in η while $V_D(D, C, \xi)$ is strictly increasing in η , and vice versa for β . Thus, since ξ^* is unique, the crossing point must move in the stated direction as the effort payoffs change. For small changes in η or β , this implies that the equilibrium overlap selected by forward induction will also have these characteristics, with the caveat that, for some η and some β the equilibrium will jump between the interior overlap and complete overlap.

3.5 Applications and Implications

3.5.1 Financial Services

The fundamental prediction of the model is that when agents have a strong reason to need to cooperate with each other and when locating close together is not excessively costly in terms of distortions associated with economic overlap and inefficient coverage of the relevant space, they will congregate close together. This intuition can be applied to consider why certain industries cluster together while others that perhaps seem closely related tend to be dispersed. One notable example of this contrast is in the financial services industry. We observe extreme concentration for certain types of financial services and dispersion for others. Specifically, firms that primarily provide investment banking services are extremely concentrated. This geographical concentration can be seen to create economic overlap in a variety of ways. Training of employees becomes more of a public good when costs of job switching are low, as will occur when firm headquarters are located within a few blocks of each other. Proprietary information, which may be costly to generate, will also be far less private among concentrated firms. Finally, investment banking firms overlap in an even more significant economic space in that much of their most profitable activity is undertaken by syndicates of firms. Firms contribute effort to a joint project (say, an IPO), and then share revenues. Such an arrangement will clearly lead to distortions from efficient effort on joint projects, with firms preferring to expend their resources on projects that are not part of the syndicate and that therefore do not have the public element to them. However, this

industrial structure can lead to cooperation on other joint endeavors. If one investment bank must rely on the contributions of other members of a syndicate, said bank is less likely to exploit an opportunity to gain profit for itself at the cost of seriously weakening its future partner. Such opportunities for cooperation pervade the investment banking industry, including such things as maintaining non-competitive pricing of services (Chen and Ritter [2000] and Lowery [2008]) and providing emergency liquidity following shocks.

Firms that specialize in identifying underpriced securities for investment opportunities, on the other hand, tend to locate in a far more dispersed fashion. Since it has been documented that such firms tend to generate a significant part of any excess returns from investment in securities associated with firms located nearby (Coval and Moskowitz [2001]), this lack of concentration can be thought of as associated with a high area-under-cultivation cost. Failing to cover all possible geographic areas leaves some mispricing unexploited. Furthermore, such investment businesses will have few valuable opportunities to cooperate as they are effectively engaged in a fixed-sum game. Any profits accruing to one firm for identifying a mispricing will be unavailable to any other firm. This situation contrasts sharply with the investment banking firms, where cooperation to maintain monopoly pricing would prove valuable.

3.5.2 Formal Institutions and Social Capital

One implication of this model is that social capital in the form of economic overlap and formal enforcement institutions are substitutes. We would thus predict that, as formal institutions develop that allow agents to commit to cooperate on joint projects, economic overlap will disappear. In economies without formal institutions, we expect to see social norms that enforce sharing of wages, housing, and other economic gains in order to foster cooperation amongst members of a group. Such sharing will of course undermine incentives to engage in productive activity and individuals will expend disproportionate effort on activities that generate benefits that cannot be shared. Evidence of such arrangements can be found in, for example, Venkatesh [2006], who examines in detail the economic and social arrangements of individuals living outside the formal economy. As formal contracting makes cooperation feasible between individuals without links, the gains from breaking away from the inefficient overlap will lead to a breakup of such cooperative arrangements. This interpretation implies that the absence of formal institutions will undermine economic growth and development

not by directly undermining the ability of individuals to cooperate, but because the social and economic arrangements necessary to maintain cooperation are inefficient compared to those that will arise in the presence of formal institutions.

3.6 Three Agent Sequential Arrival

Certain of the elements of the model extend in an obvious way to the case with $N > 2$ agents. The location decision now determines, for each agent, how he will overlap with all of the other $N - 1$ agents, although the location decision no longer maps as directly into an overlap choice. Production is shared evenly among all agents who overlap on a given area. Thus, one agent may share a bilateral overlap with some other agent but a trilateral overlap with different agents, for example. The exact specifications of the dimension of the location space and the shape of each plot will generate restrictions on the feasible overlap configurations. It is less clear how the second-stage cooperation opportunity should scale up. Each agent could have a bilateral cooperation opportunity with every other agent, or all agents together could play some form of N -agent cooperation game. Furthermore, the behavior of the maximum cooperation bonus as a function of N would have an influence on the optimal and equilibrium networks. A thorough analysis of such an expanded model are left to future research. This section, however, considers an example with three agents. Here, I demonstrate that the efficiency result for two agents does not immediately extend to multilateral interactions.

To explore the structures that arise in a larger network, consider a sequential arrival game with three agents. For expositional purposes, we will first assume that agents commit to a location without taking into consideration the arrival decisions of future agents. Thus, the second agent to arrive will locate as if he were playing a two-agent game. Also, for simplicity, we will initially assume that agents cannot create trilateral commons. Assume that parameters are such that the second agent will choose a partial overlap with the first. Now, a third agent must choose how to locate. Under the assumption that a trilateral commons is prohibited, this reduces to a decision over how much of a commons to create with one of the agents. Without loss of generality, assume that the third agent arrives from the left, and designate this agent l . Agent l intrudes on the private pasture of agent m , and agent r continues to overlap only with agent m . An example configuration appears in figure 3.6. Consistent with the two agent game, ξ will parameterize half of the overlap between m and r ; however, ε

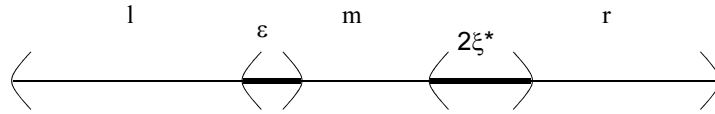


Figure 3.6: Three Agent Chain

will parameterize the *total* overlap between l and m . Consider now an example where $\rho = 1$, $\beta = 3$, $\eta = 3.1$, $\theta = -1$, and $\alpha = 0.15$. The second agent will enter with $\xi_2^* = 0.1039$, thus guaranteeing that cooperation between m and r can be sustained. When the third agent arrives, the cooperation configuration may change. If the third agent should choose a small but positive overlap with one of the current agents, then cooperation will completely break down. This follows from the fact that agents m and r were located such that they would just manage to sustain cooperation. Upon the intrusion of the third agent, agent m now has a smaller, and therefore less valuable, private plot. This will tend to make the nonnegativity constraint for contributions by l bind, breaking any potential link between l and m . The argument for the collapse of cooperation between m and r is more subtle. After the intrusion, the value of the game for a fixed cooperation configuration falls for the middle agent, both if he adheres and if he deviates against the right agent. But, the value of deviating falls slightly more since the benefit of having the extra effort to distribute over the private plot decreases as the size of the private plot decreases, whereas the value to the original commons remains relatively stable.

But, a larger overlap by the entering agent can serve to restore cooperation. In particular, for certain parameter values, there is an interior overlap by the entering agent that restores cooperation between m and r and also generates cooperation between l and r . To see this, observe that, as ε increases, the value of adhering or deviating from said strategy declines for the middle agent. But, adhering remains better for all values of ε except ones that almost eliminate the middle agent's private plot. Note the difference between this scenario and the previous scenario in which only m and r were considering cooperating. Now, the right agent is appreciably wealthier, and

consequently even after a deviation by the middle agent will tend to desire a positive contribution to the commons. Indeed, his wealth will be $1 + \theta + \beta = 3$, which is not appreciably smaller than the wealth of the middle agent after a deviation, $1 + \eta = 4.1$. But, as ε becomes large, the nonnegativity constraint of the right agent will begin to bind as the middle agent desires an increasing level of the right commons, and around this point it will become better for the middle agent to deviate. The value to the left agent follows a similar pattern. However, the right agent will, for small ε , prefer to deviate, as he is the only agent that will contribute to the right commons after all agents adhere to the proposed cooperation profile. This follows from the fact that he is relatively wealthy, as the cooperation calls for him to engage in two profitable prisoner's dilemmas instead of one. As epsilon grows, however, the situation begins to reverse as the "outside option" of the middle agent gets worse as his private plot shrinks. There will, therefore, be a small region of ε over which all contributions to the commons are interior. Over this region, cooperation *may* be feasible. It is still possible that cooperation will break down if the bonus for deviating is large enough, but for the parameterization discussed here cooperation is maintained over a small region. The left agent must therefore choose between isolation and this partial overlap that sustains cooperation in two prisoners' dilemmas. In this example, the partial overlap is appealing, and will therefore be selected. As a consequence, the right agent becomes much better off, while the middle agent, due to the reduction in his private plot, is worse off.

This raises the obvious question of how the second agent would have optimally located were he aware that a third agent would arrive. This question is best resolved by inspecting figure 3.7, which shows the value to each agent as a function of the overlap between the first two agents, assuming that the third agent locates in his own best interest. Over the region of small ξ , agent l 's value is flat because he will choose isolation when ξ is too small to facilitate cooperation; after the increase associated with a ξ that permits cooperation (and thus induces an interior ε , agent r 's value declines linearly in ξ , while the value to l and m decline in a nonlinear fashion; the vertical line, representing the myopic ξ , shows that all agents prefer the smaller, rational ξ_3^* to the myopic ξ_2^* .

From this, it is clear that the agent that ends up on the right prefers to locate so as to induce the third agent to choose an interior overlap. The agent who ends up being in the middle, however, prefers to choose ξ such that cooperation breaks down completely. Assuming that agents are risk neutral and that the third agent enters

with equal probability from either side, the second agent will risk ending up in the middle. This location will not be the same as the optimal location in the two agent game. Decreasing ξ , which would destroy cooperation in the two agent game, is now possible; the third agent will locate so as to restore cooperation, and all agents end up better off, in expectation, than in the myopic game. In fact, with forward looking actions by agent 2, the eventual network formed is second best optimal as agents are all appropriately incentivized to maximize the area under cultivation subject to the incentive compatibility constraints of the chosen cooperation condition. But, we can see that this network will not be stable in the sense of Jackson and Wolinsky [1996]; the middle agent, if permitted, will prefer to deviate to isolation after the third agent joins.⁸ But, then one of the two remaining agents will prefer to form a link with the agent who has just left the network, restoring bilateral cooperation with that agent. Thus, the network will cycle under pairwise stability.

This three agent example shows that the incentives to choose efficient overlap continue to play a role in games with more than two agents, but the increased complexity associated with the introduction of indirect links and the need to anticipate future arrival decisions breaks the immediate link between efficient and equilibrium outcomes. Thus, in games with more than two players, the ability of agents to successfully exploit cooperation opportunities at minimal cost will depend on the specific elements of the environment under consideration. This disconnect introduces the possibility that a social planner could have a useful role in enforcing location decisions even if he cannot directly enforce cooperation or control effort allocation decisions.

3.7 Conclusion

This paper demonstrates the role that overlapping economic activity can play in facilitating cooperation. In contrast with the standard Hotelling location problem, the structure of this model can generate partial overlaps and can therefore rationalize a variety of geographical and economic arrangements that are observed in superficially similar settings. A relatively parsimonious explanation for cooperation in finite prisoner's dilemmas is presented, with predictions for cooperative behavior that do not depend on patience or repetition.

In the context of political economy, this model is complementary to the Tiebout

⁸In this sense, eliminating overlap is treated as equivalent to breaking a link.

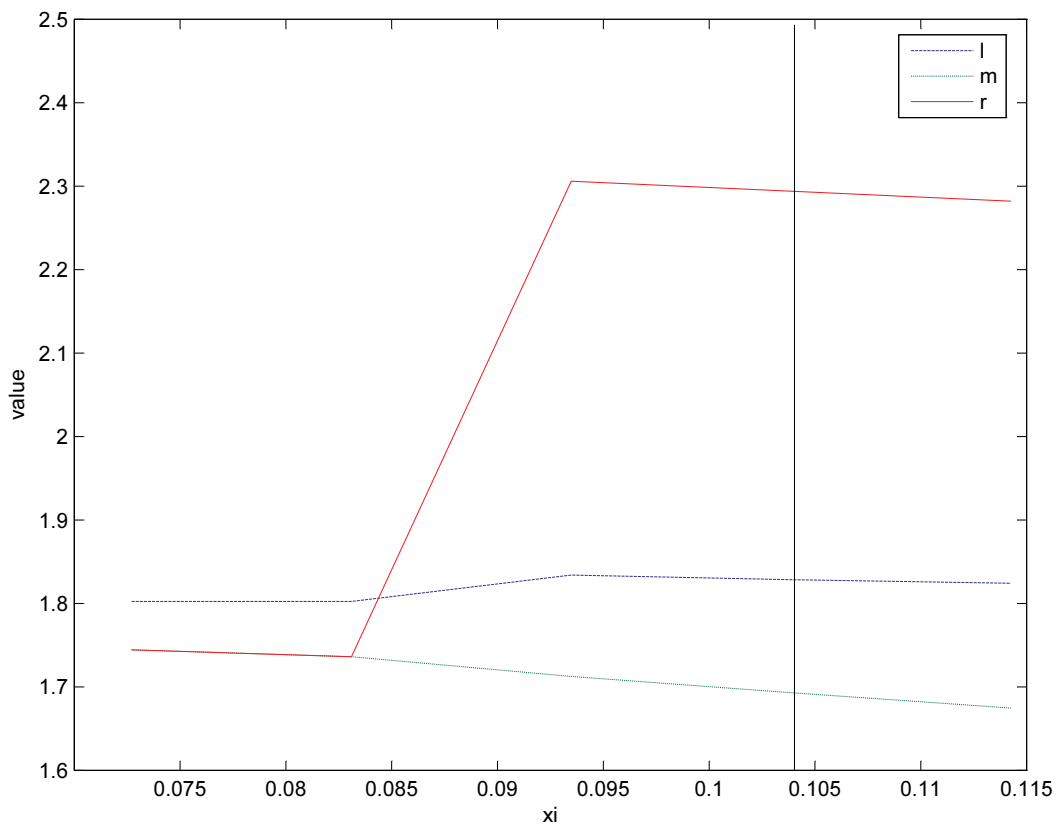


Figure 3.7: Value for given ξ

approach to location decisions. Here, agents choose the type of economic unit to develop based on the underlying tradeoff between the value of and inefficiencies from public projects, but they lack access to political institutions that can facilitate cooperation through locally involuntary taxation.⁹ Since political institutions themselves are a form of cooperation and presumably develop at a later stage than society itself, a model that can describe agglomeration without reliance on these institutions is helpful.

3.8 Appendix

3.8.1 proof of lemma 2

Proof. The strategy space of the effort allocation subgame is all non-negative function $e_i()$ such that $\int_{\Omega} e_i d\mu \leq E_i$. But, as shown above, the concavity of v guarantees that optimal allocations must be uniform over any plot of a particular public good configuration. That is, if an area is an overlap between n agents, then that entire area must have a uniform effort allocation or one agent could make himself (and all the other agents involved in the plot) better off by reallocating effort. And, we can restrict attention to equilibria where agents provide uniform effort over each plot in which they are involved because, for any public plot, the best response to uniform effort by the other contributors is uniform effort. So, we can reduce the strategy space to a finite vector, since the number of potential configurations is finite. (For example, with two agents in \mathbb{R}^1 , there will be at most a private plot and one commons for each agent, with three agents a private plot, an overlap with one agent, an overlap with the other agent, and an overlap with both, etc.). It is now clear that the effort budget constraint is a simplex of the same dimension as the number of configurations. This implies that each strategy space is a nonempty convex compact subset of a Euclidean space. Finally, payoffs are continuous and quasi-concave in the strategy space since v is, by assumption, continuous and concave, and the payoff function can be written as

$$\sum_{j=1}^G \mu(g_j) v(e_{ij}),$$

⁹In the two agent case the location decision is second best when play in the P-D cannot be controlled. The multilateral case remains to be analyzed.

where G is the number of configurations, g_j is the j^{th} plot in which i is involved (arbitrarily ordered), and e_{ij} is the level of effort agent i puts into each point in plot g_j . This payoff function is then a multiple of a convex combination of univariate concave functions, which is itself concave. So, by Glicksberg [1952], a pure strategy Nash equilibrium exists. \square

3.8.2 Proof of lemma 3

Proof. $V_D(D, C, 0) = 2\rho v\left(\frac{1+\eta}{2\rho}\right) > 2\rho v\left(\frac{1+\beta}{2\rho}\right) = V(C, C, 0)$ and $V(C, C, \rho) = \rho v\left(\frac{2+2\beta}{2\rho}\right) > \rho v\left(\frac{1+\eta}{2\rho}\right)$, so assumption 2 guarantees that the value functions cross exactly once. \square

3.8.3 Proof of lemma 4

Proof.

$$V_D(\xi) = (2\rho - 2\xi)v(e_p(\xi)) + \xi v(e_c(\xi)).$$

Now, let $\xi' < \xi$, and assume agent D does not alter the effort allocation from ξ . He then receives

$$V'_D = (2\rho - 2\xi)v(e_p(\xi)) + 2(\xi - \xi')v(e_c(\xi)) + \xi'v(e_c(\xi)),$$

where the first term is the value of the old private plot, the second term the value of the new private plot, and the third the value of the new commons. Now, since $\xi' < \xi$, if the value of the smaller commons exceeds the value of the original, larger commons, it must be that effort is reallocated to the commons. In order for that to be optimal, we need

$$\frac{\xi'v'(e_c(\xi))}{2\xi'} > \frac{(2\rho - 2\xi)v'(e_p(\xi))}{2\rho - 2\xi} \Rightarrow \frac{v'(e_c(\xi))}{2} > v'(e_p(\xi))$$

which is false by the original optimality conditions, or

$$\frac{\xi'v'(e_c(\xi))}{2\xi'} > \frac{2(\xi - \xi')v'(e_c(\xi))}{2(\xi - \xi')} \Rightarrow \frac{v'(e_c(\xi))}{2} > v'(e_c(\xi)),$$

which is false since v is increasing. So, effort is reallocated away from the now smaller commons (and put into the previously common area), making the value of the commons fall. This implies that V_c must fall as the commons shrinks since the only value that accrues to the cooperating agent is that from the commons. \square

3.8.4 Proof of proposition 17

The following proposition is needed:

Proposition 24. *If $\eta + \theta < 0$ (that is, the total value of the cooperation phase is less than zero when play is asymmetric), \nexists an equilibrium where D, C is played in any subgame.*

Proof. Assume that there is an equilibrium with D, C . Any allocation that was feasible to D before the allocation is feasible to C after the deviation since $\eta < 1$. That is, agent C is gaining more effort than agent D is losing; this is achieved by having both agents solve agent D 's optimization problem (for D 's welfare, *not* for C 's welfare after a deviation to D). Clearly, agent D will not have an incentive to deviate from this allocation since it was reached by optimizing his welfare given technological constraints. Thus, $V(D, D, \xi) \geq V_C(D, C, \xi)$ for all ξ . But, from this allocation, if C marginally and uniformly reduces the effort he puts into the commons and reallocates that effort to his private plot, the value accruing to him must increase since the marginal value of his private plot is at a maximum (since it currently receives zero effort) and he does not have to divide this with the other agent. D 's best response to this action is to *increase* his contribution to the commons, so C cannot be worse off. \square

Proof. Assume D, C is played along the equilibrium path. Then

$$V_D(D, C, \xi) \geq V(C, C, \xi) \quad (3.10)$$

$$V_D(D, C, \xi) \geq V(C, C, \rho) \quad (3.11)$$

$$V_C(D, C, \xi) \geq V(D, D, 0) \quad (3.12)$$

$$V_C(D, C, \xi) \geq V(D, D, \xi) \quad (3.13)$$

$$V_C(D, C, \xi) \geq V(C, C, \rho). \quad (3.14)$$

These inequalities guarantee, in order, that adhering to the proposed $\{\xi, (D, C)\}$ equilibrium satisfies: D will not deviate to playing C at the proposed location; D will not deviate to total overlap, which brings about C, C play (and a local maximum for $V(C, C, \xi)$); C will not deviate to no overlap, which will always imply D, D play and a global maximum of $V(D, D, \xi)$; C will not deviate to play D at the current location; and C will not deviate to total overlap, which guarantees C, C .

But, $V(D, D, 0) \geq V(D, D, \xi)$ for all ξ , so inequality 3 \Rightarrow 4. And, $V_D(D, C, \xi) - V_C(D, C, \xi) - (2\rho - 2\xi)v(e_p) > 0$, so $V_D(D, C, \xi) > V_C(D, C, \xi)$, so 5 \rightarrow 2. Now, recall

that inequality 1 will hold for sufficiently small ξ ; that is, inequality 3 takes the form of a threshold, by lemma 3.8.2. But, by lemma 4, $V_C(D, C, \xi)$ is strictly increasing in ξ . Therefore, it is sufficient to show that

$$\{\xi | V_D(D, C, \xi) = V(C, C, \xi), V_C(D, C, \xi) \geq V(D, D, 0), V_C(D, C, \xi) \geq V(C, C, \rho)\} = \emptyset$$

in order to establish that no equilibrium where D, C is played exists, since if the inequalities fail at the threshold, they must fail for all ξ below the threshold.

But, $V(D, D, 0) = (2\rho)v\left(\frac{1}{2\rho}\right)$ and $V(C, C, \rho) = \rho v\left(\frac{1+\beta}{\rho}\right)$. So, only one of the two inequality conditions can be operable in any (generic) problem. Specifically, inequality 3 is operable if and only if

$$2v\left(\frac{1}{2\rho}\right) > v\left(\frac{1+\beta}{\rho}\right);$$

otherwise, inequality 5 is operable. Which inequality matters will depend on the payoff to cooperating, the size of the plot, and the curvature of v .

For example, in the parametric CES case, the condition for 3 to be the relevant constraint becomes

$$\begin{aligned} 2\left(\frac{1}{2\rho}\right)^\alpha &> \left(\frac{1+\beta}{\rho}\right)^\alpha \\ 2^{\frac{1}{\alpha}}\frac{1}{2\rho} &> \frac{1+\beta}{\rho} \\ 2^{\frac{1-\alpha}{\alpha}} &> 1+\beta \\ \beta &< 2^{\frac{1-\alpha}{\alpha}} - 1 \\ \log(1+\beta) &< \frac{1-\alpha}{\alpha}\log(2) \\ \log_2(1+\beta) &< \frac{1-\alpha}{\alpha} \\ \alpha &< \frac{1}{\log_2(1+\beta) + 1}. \end{aligned}$$

So, when v exhibits near constant returns to scale, inequality 5 is operable, while for extreme diminishing returns to scale, inequality 3 is. As $\beta \rightarrow \infty$, the cutoff falls to zero, while as $\beta \rightarrow 0$, it increases toward 1.

Now, designate ξ^* as the ξ at which inequality 1 holds with equality and assume $V_C(C, C, \rho) > V_D(D, D, 0)$ and D, C can occur. But, $V_C(D, C, \xi)$ is strictly increasing in ξ and $V_C(D, C, \xi) \leq V(D, C, \xi)$ and equal only at $\xi = \rho$. So, $V(D, C, \xi) \leq$

$V_D(D, C, \rho) = \rho v\left(\frac{1+\eta}{2\rho}\right)$. So,

$$\begin{aligned} V_C(D, C, \xi^*) &< \rho v\left(\frac{1+\beta}{2\rho}\right) \\ V_C(D, C, \xi^*) &> \rho v\left(\frac{2+2\beta}{2\rho}\right). \end{aligned}$$

But, $2+2\beta > 1+\eta$, so this is a contradiction.

If instead $V_C(C, C, \rho) < V_D(D, D, 0)$ and D, C occurs,

$$\begin{aligned} V_D(D, C, \rho) &< V(C, C, \rho) \\ V(C, C, \rho) &< V(D, D, 0) \\ V_D(D, C, \rho) &< V(D, D, 0) \end{aligned}$$

and since $V_C(D, C, \xi) < V_D(D, C, \rho)$, then $V_C(D, C, \xi) < V(D, D, 0)$, which is a contradiction. \square

3.8.5 Proof of lemma 5

Proof. Where $E^T = E_1 + E_2$,

$$AuC^{loss}(\xi) = 4\rho v\left(\frac{E^T}{4\rho - 2\xi}\right) - (4\rho - 2\xi)v\left(\frac{E^T}{4\rho - 2\xi}\right)$$

so

$$\begin{aligned} \frac{\partial AuC^{loss}(\xi)}{\partial \xi} &= -\left[(4\rho - 2\xi)v'\left(\frac{E^T}{4\rho - 2\xi}\right)\left(\frac{E^T}{(4\rho - 2\xi)^2}\right)(2) - 2v\left(\frac{E^T}{4\rho - 2\xi}\right)\right] \\ &= 2\left[v\left(\frac{E^T}{4\rho - 2\xi}\right) - \left(\frac{E^T}{4\rho - 2\xi}\right)v'\left(\frac{E^T}{4\rho - 2\xi}\right)\right] \end{aligned}$$

But, for all $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that f is increasing, concave and $f(0) = 0$, $f(x) - xf'(x) > 0$, so, letting $f = v$ and $x = \frac{E^T}{4\rho - 2\xi}$, $AuC^{loss} > 0$.

Now,

$$\begin{aligned} \frac{\partial^2}{\partial \xi^2} &= 2\left[v'(x)\frac{\partial x}{\partial \xi} - \left(v''(x)\frac{\partial x}{\partial \xi}x + \frac{\partial x}{\partial \xi}v'(x)\right)\right] \\ &= 2\left[-v''(x)\frac{\partial x}{\partial \xi}\right]. \end{aligned}$$

But,

$$\frac{\partial x}{\partial \xi} = \frac{2E^T}{(4\rho - 2\xi)^2} > 0,$$

and

$$v''(x) < 0,$$

so AuC^{loss} is increasing and convex in ξ . □

3.8.6 Proof of proposition 16

Proof. Assume $\exists \xi^*$ such that $V(C, C, \xi^*) = V_D(D, C, \xi^*)$. This implies that, at ξ^* , the benefit to one agent of obtaining extra effort by playing D in the prisoner's dilemma phase is exactly offset by the costs to this agent associated with the reduction in the other agent's contribution to the commons. As ξ increases above ξ^* , the benefit of having additional effort to allocated over ones private plot declines since the private plot is smaller and $v(\cdot)$ is concave. At the same time, the costs associated with depriving the other agent of resources, some of which he would employ in the commons, increases since that agent's private plot is also smaller for the larger value of ξ , while the common plot is larger. Thus, for $\xi > \xi^*$, $V(C, C, \xi) > V_D(D, C, \xi)$. Since $V_D(D, C, 0) > V(C, C, 0)$, this completes the proof. □

3.8.7 Proof of proposition 19

Proof. Each agent must choose a location on \mathbb{R} . Take one agent's location choice as given and consider the payoff function for the location choice of the other agent. All points that lead to no overlap have the same value to that agent since he will obviously play D in the prisoner's dilemma and spread his effort uniformly over his plot. We can then consider only the payoff function for $\xi \in [0, \rho]$, where ξ is again the overlap. Since each location decision generates a subgame with an equilibrium, we know that the payoff function is well defined for every ξ over the (compact) interval $[0, \rho]$. But, the payoff is bounded from above since it could never exceed (for example) $2\rho v\left(\frac{2+2\beta}{2\rho}\right)$, so the payoff function in terms of ξ must achieve it's maximum on $[0, \rho]$. Now, observe that, whenever play in the prisoner's dilemma is symmetric, the effort allocation problem is symmetric and, by the concavity of v , we know that the outcome will be symmetric. Proposition 17 demonstrates that, in equilibrium, play will be symmetric. So, whatever optimal choice is made in the location game by the agent we

are considering will be optimal also for the other agent. Thus, at this location, neither has an incentive to deviate. \square

3.8.8 Proof of proposition 22

Proof. Initially, as ξ increases away from zero, more effort per area is put into the private plot. But, as ξ grows the returns on the private plot diminish more rapidly, implying that more effort should be invested in the (growing) commons. These two effects must exactly offset each other at one point since $ToC(\xi = 0) = ToC(\xi = \rho) = 0$ (by the mean value theorem). But, as ξ grows beyond the point at which $\frac{dToC}{d\xi} = 0$, the concavity of $v(\cdot)$ guarantees that the returns to total investment in the private plot diminish even more rapidly. As the commons is also increasing in size as ξ increases, the total effect must be a continued decline in the ToC. \square

3.8.9 Tragedy of the Commons and Area under Cultivation: Parametric Example

Calculating the value of the game to a single agent with socially optimal effort functions is a simple matter. The production technology is uniform across the area under cultivation (since the feature that distinguishes the commons from the private plots is the distributional technology), so a socially optimal effort function will spread effort uniformly over the AuC. In particular, effort at any location is given by

$$e = \frac{2E}{4\rho - 2\xi} = \frac{E}{2\rho - \xi},$$

where $E = E_i = E_j$ (where E_i is again agent i 's effort endowment and $E_i = E_j$ because the focus is on potentially equilibrium path allocations). So,

$$V^{so}(\xi, E) = \frac{1}{2} \left(\frac{E}{2\rho - \xi} \right)^\alpha (4\rho - 2\xi)$$

or

$$V^{so}(\xi, E) = E^\alpha (2\rho - \xi)^{1-\alpha}.$$

This gives, in the benchmark $\xi = 0$ case,

$$V^{so}(E, 0) = E^\alpha (2\rho)^{1-\alpha}.$$

Note that the above expression can also be derived from the general value function in the competitive game since $V^{so} = V^{CE}$ when $\xi = 0$ (and therefore labor markets are effectively complete).

This now gives an expression for the total inefficiency resulting from a commons

$$F^T(\xi, E) = E^\alpha (2\rho)^{1-\alpha} - \left(\frac{E}{b\rho + \xi(1-b)} \right)^\alpha [a\rho + \xi(1-a)],$$

the inefficiency from lower land usage

$$F^{AuC}(\xi, E) = E^\alpha (2\rho)^{1-\alpha} - E^\alpha (2\rho - \xi)^{1-\alpha}$$

and the inefficiency from the tragedy of the commons

$$\begin{aligned} F^{ToC}(\xi, E) &= F^T(\xi, E) - F^{AuC}(\xi, E) \\ &= E^\alpha (2\rho - \xi)^{1-\alpha} - \left(\frac{E}{b\rho + \xi(1-b)} \right)^\alpha [a\rho + \xi(1-a)]. \end{aligned}$$

To simplify the following, consider a normalization where $E = 1$ and $\rho = \frac{1}{2}$. Then

$$F^{AuC}(\xi) = 1 - (1 - \xi)^{1-\alpha} \quad (3.15)$$

$$F^{ToC}(\xi) = (1 - \xi)^{1-\alpha} - \left(\frac{1}{a + \xi(1-b)} \right)^\alpha [2^{\frac{\alpha}{1-\alpha}} + \xi(1-a)] \quad (3.16)$$

Note that

$$\frac{\partial F^{AuC}(\xi)}{\partial \xi} = \frac{1 - \alpha}{(1 - \xi)^\alpha} > 0,$$

indicating, as expected, that the inefficiencies from reducing the area under cultivation are strictly increasing in ξ , while

$$\frac{\partial F^{ToC}(\xi)}{\partial \xi} = -\frac{1 - \alpha}{(1 - \xi)^\alpha} + (a - 1) \left(\frac{1}{a + \xi(1-b)} \right)^\alpha + (1 - b) \alpha \left(2^{\frac{\alpha}{1-\alpha}} + \xi(1-a) \right) \left(\frac{1}{a + \xi(1-b)} \right)^{1+\alpha},$$

which is increasing initially and then decreasing.

Combining these two derivatives, it is clear that the inefficiencies associated with generating a commons are increasing up to some threshold and then ambiguous after that. The existence of cases both where the total inefficiency increases monotonically and where the total inefficiency reaches a maximum at some interior point has already been demonstrated by example and will depend on the concavity of the production

function. Note also that, when the strategies in the prisoner's dilemma are taken as a given, it can never be the case that $\xi = \rho$ is a global minimum for the total inefficiency since both elements of the inefficiency are minimized at $\xi = 0$, while only the ToC is minimized at $\xi = \rho$.

To understand the root of this potentially non-monotonic behavior, consider the effect of increasing the area of the commons. The larger commons now implies that a greater area is subject to the tragedy of the commons; however, as agents focus more intensely on their own private plots the returns decline. Not only are they applying more total effort to the private plot, the size of that plot is shrinking as the commons increases. At some point, the returns to the private plot can diminish to the point that it actually becomes more desirable to reallocate effort *to* the commons as the commons grows. Since effort in the commons is always sub-optimal, this results in an increase in value. Note, however, that this increase will not always swamp the loss from the change in AuC, so not all parameters will induce the uptick for large ξ .

Bibliography

- D Abreu, D Pearce, and E Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5):1041–1063, September 1990.
- D. Abreu, P. Milgrom, and D. Pearce. Information and timing in repeated partnerships. *Econometrica*, 59:1713–33, 1991.
- Dilip Abreu, David Pearce, and Ennio Stacchetti. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251 – 269, 1986.
- Viral Acharya and Tanju Yorulmazer. Cash-in-the-market pricing and optimal resolution of bank failures. *Review of Financial Studies*, 21:2705–2742, 2008.
- Reena Aggarwal. Stabilization activities by underwriters after initial public offerings. *The Journal of Finance*, 55(3):1075–1103, 2000.
- Masaki Aoyagi. Bid rotation and collusion in repeated auctions. *Journal of Economic Theory*, 112(1):79–105, September 2003.
- Masaki Aoyagi. Efficient collusion in repeated auctions with communication. *Journal of Economic Theory*, 134(1):61 – 92, 2007.
- S Athey, K Bagwell, and C Sanchirico. Collusion and price rigidity. *Review of Economic Studies*, 71(2):317–349, 2004.
- Susan Athey and Kyle Bagwell. Optimal collusion with private information. *The RAND Journal of Economics*, 32(3):428–465, 2001.
- Michael R. Baye and Dennis W. Jansen. Repeated games with stochastic discounting. *Economica*, 63(252):531–541, November 1996.
- J. Benoit and V. Krishna. Finitely repeated games. *Econometrica*, 53:905–922, 1985.
- Ted Bergstrom, Larry Blume, and Hal R. Varian. On the private provision of public goods. *Journal of Public Economics*, 29:25–49, 1986.
- B. Douglas Bernheim and Aniruddha Dasgupta. Repeated games with asymptotically finite horizons. *Journal of Economic Theory*, 67(1):129–152, 1995.

- Sreedhar T. Bharath and Amy K. Dittmar. Why do firms use private equity to opt out of public markets. Working paper, March 2007.
- James Booth and Lena Chua. Ownership dispersion, costly information, and IPO underpricing. *Journal of Financial Economics*, 41(2):291–310, 1999.
- Yann Bramoulle and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.
- Markus K. Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. June 2007.
- Bruce Carlin, Miguel Lobo, and S Viswanathan. Episodic liquidity crises: Cooperative and predatory trading. *The Journal of Finance*, 62:2235–2274, 2008.
- Archishman Chakraborty, Simon Gervais, and Bilge Yilmaz. Security design in initial public offerings. October 2008.
- Thomas Chemmanur and Paolo Fulghieri. A theory of the going-public decision. *Review of Financial Studies*, 12(2):249–279, Summer 1999.
- Hsuan-Chi Chen. Competition and collusion in the IPO market. September 2001.
- Hsuan-Chi Chen and J.R. Ritter. The seven percent solution. *The Journal of Finance*, 55(3):1105–1131, June 2000.
- Joshua D. Coval and Tobias J. Moskowitz. The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, 109(4):811–841, 2001.
- Pedro Dal Bo. Tacit collusion under interest rate fluctuations. *RAND Journal of Economics*, 38(2):1–8, Summer 2007.
- Gabrielle Demange. Group formation: The interaction of increasing returns and preference diversity. In Gabrielle Demange and Myrna Wooders, editors, *Group Formation in Economics*. Cambridge University Press, 2005.
- Peter M. DeMarzo, Ilan Kremer, and Andrzej Skrzypacz. Bidding with securities: Auctions and security designs. *American Economic Review*, 95(4):936–959, September 2005.
- Avinash Dixit. On modes of economic governance. *Econometrica*, 72(2):449–481, March 2003.
- Avinash Dixit. *Lawlessness and Economics*. Princeton UP, 2004.
- Darrell Duffie, Nicolae Garleanu, and Lasse Heje Pedersen. Over-the-counter markets. *Econometrica*, 73(6):1815–1847, November 2005.

- Dennis Epple and Richard Romano. Public provision of private goods. *Journal of Political Economy*, 104(1):57–84, February 1996.
- Dennis Epple, Thomas Romer, and Holger Sieg. Interjurisdictional sorting and majority rule: An empirical analysis. *Econometrica*, 69(6):1437–1465, Nov 2001.
- A. Fernandes and C. Phelan. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, 91:223–247, 2000.
- Roland Fryer. A model of social interactions and endogenous poverty traps. NBER Working Paper 12364, July 2006.
- Edward Glaser, David Laibson, and Bruce Sacerdote. An economic approach to social capital. *The Economic Journal*, 112:F437–F458, November 2002.
- I.L. Glicksberg. A further generalization of the Kakutani fixed point theorem with applications to nash equilibrium points. *Proceedings of the American Mathematical Society*, 38:170–174, 1952.
- Matthew Goldstein and David Henry. Bear bets wrong. *Businessweek*, October 2007.
- Sanjeev Goyal. Strong and weak links. *Journal of the European Economic Association*, 3:608–616, April-May 2005.
- Avner Greif. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge UP, 2006.
- Avner Greif. Contract enforcement and economic institutions in early trade: The maghribi traders’ coalition. *The American Economic Review*, 83(3):525–548, June 1993.
- Avner Greif, Paul Milgrom, and Barry R. Weingast. Coordination, commitment, and enforcement: the case of the merchant guild. *Journal of Political Economy*, 102(4):745–776, 1994.
- Galina Hale and Joao Santos. Evidence on the costs and benefits of bond ipos. 2006.
- A Hameed, W Kang, and S Viswanathan. Asymmetric comovement in liquidity. mimeo, Duke University, 2005.
- Makoto Hanazono and Huanxing Yang. Collusion, fluctuating demand, and price rigidity. *International Economic Review*, 48(2):483–515, 2007.
- Robert S. Hansen. Do investment banks compete in IPOs?: The advent of the ‘7% plus contract’. *Journal of Financial Economics*, 59(3):313–346, March 2001.

- Harrison Hong and Jeremy C. Stein. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6):2143–2184, December 1999.
- Harrison Hong, Terrence Lim, and Jeremy C. Stein. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, 55(1):265–295, February 2000.
- Ibbotson. *Cost of Capital Quarterly, 1999 Yearbook*. Ibbotson Associates, 2nd edition, 1976.
- Matthew Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.
- Cathleen Johnson and Robert P Gilles. Spatial social networks. *Review of Economic Design*, 5:273–299, 2000.
- Michael A. Jones. Cones of cooperation, Perron-Fröbenius Theory and the indefinitely repeated Prisoners’ Dilemma. *Journal of Mathematical Economics*, 30:187–206, 1998.
- Michael A. Jones. The effect of punishment duration of trigger strategies and quasifinite continuation probabilities for Prisoners’ Dilemmas. *International Journal of Game Theory*, 28(4):533–546, 1999.
- K.L. Judd, S. Yeltekin, and J. Conklin. Computing supergame equilibria. *Econometrica*, 71(4):1239–1254, 2003.
- Michihiro Kandori and Hitoshi Matsushima. Private observation, communication and collusion. *Econometrica*, 66(3):627–652, May 1998.
- Laurie Kirgman, Wayne H. Shaw, and Kent L. Womack. The persistence of IPO mispricing and the predictive power of flipping. *The Journal of Finance*, 54(3):1015–1044, 1999.
- Stephen Knack and Phillip Keefer. Does social capital have an economic payoff? a cross-country investigation. *Quarterly Journal of Economics*, 112:1251–1288, 1997.
- Laurie Krigman, Wayne H. Shaw, and Kent L. Womack. Why do firms switch underwriters? *Journal of Financial Economics*, 60:245–284, 2001.
- Tim Loughran and Jay Ritter. Why has IPO underpricing changed over time? *Financial Management*, 33(3):5–37, 2004.
- Michelle Lowery and Susan Shu. Litigation risk and IPO underpricing. *Journal of Financial Economics*, 65(3):309–355, 2002.
- Richard Lowery. Imperfect monitoring and fixed spreads in the market for ipos. June 2008.

- Richard Lowery and Bryan Routledge. Trust, leverage, and liquidity. 2008.
- George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford UP, 2006.
- Marco Pagano, Fabio Panetta, and Luigi Zingales. Why do companies go public? an empirical analysis. *The Journal of Finance*, 53(1):27–64, 1998.
- Nagpurnanand R. Prabhala and Manju Puri. December 1998.
- Susan Pulliam and Serena Ng. Default fears unnerve markets: Partners in credit deals face big write-downs as bond issuer teeters. *The Wall Street Journal*, January 2008.
- Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster, New York, 2000.
- Jay Ritter. Investment banking and securities issuance. In G. Constantinides, M. Harris, and R. Stultz, editors, *Handbook of the Economics of Finance*. Elsevier, 2003.
- Jay Ritter. The costs of going public. *Journal of Financial Economics*, 19(2):269–281, December 1987.
- J Rotemberg and G Saloner. A supergame-theoretical model of price wars during booms. *American Economic Review*, 76:390–407, 1986.
- Bryan Routledge and J von Amsberg. Social capital and growth. *Journal of Monetary Economics*, 50(1):167–193, January 2003.
- Bryan R. Routledge and Stanley E. Zin. Model uncertainty and liquidity. working paper, Carnegie Mellon Univeristy, June 2004.
- David Schmitz. The institution of property. In David Schmitz and Elizabeth Willott, editors, *Environmental Ethics: What Really Matters. What Really Works*. Oxford UP, 2002.
- SEC. SEC charges CSFB with abusive IPO allocation practices; CSFB will pay 100 million to settle SEC and NASD actions: Millions in IPO profits extracted from customers in exchange for allocations in “hot” deals. press release, January 2004. Available from: <http://www.sec.gov/news/headlines/csfbipo.htm>.
- Stephem Spear and Sanjay Srivastava. On repeated moral hazard with discounting. *Review of Economic Studies*, 55(4):599–617, 1987.
- Charles M. Tiebout. A pure theory of local expenditures. *The Journal of Political Economy*, 64(5):416–424, Oct 1956.
- Sami Torstila. The clustering of ipo gross spreads: International evidence. *Journal of Financial and Quantitative Analysis*, 38(3):673–694, September 2003.

Frans van Dijk and Frans van Winden. Dynamics of social ties and local public good provision. *Journal of Public Economics*, 64:323–341, 1997.

Sudhir Alladi Venkatesh. *Off the Books: The Underground Economy of the Urban Poor*. Harvard University Press, 2006.

Peter G Warr. The private provision of a public good is independent of the distribution of income. *Economics Letters*, 13(2-3):207–11, 1983.

Noah Williams. Persistent private information. January 2008a.

Noah Williams. On dynamic principal-agent problems in continuous time. September 2008b.